Developing novel enzymes for the laundry industry

Ziyuan Liu

A thesis submitted for the degree of Doctor of Philosophy

December 2010

Institute for Cell and Molecular Biosciences
Newcastle University

Acknowledgements

I would like to thank Harry who gave me this PhD opportunity and all of his guidance, support and patience. It is greatly appreciated from the bottom of my heart. Also, special thanks must go to my supervision team, Dave Bolam, Dianne Ford and James Flint, for their useful advice and cares in my final year. I also want to thank all the members of our laboratory, both past and present, for their friendship and help in my project: Art Rogowski, Louise, Nic Smith, Motern Nelson, Mark Proctor, Tibor, Hongjun Zheng, Yanping Zhu, Alan Cartmell, Lauren Mckee, Liz and Joanna Norman. I really appreciate the support the excellent technical assistance from Carl Morland.

I want to thank Unilever for offering my tuition fee and living cost during the past three years. I would also like to thank Paul Stevenson for organising my trip to Port Sunlight and for all of his help during my stay in the lab at Unilever.

Finally, I wish to thank all my friends and my family, especially auntie, mum and dad. Mum, I could not have done this without your love.

Abstract

Grass stains containing chlorophyll are the most difficult stains to remove from cotton fabrics in laundry processes. The grass cell walls are highly complex networks of polysaccharides interlocking with each other, which makes it recalcitrant to enzyme degradation. In nature, glycoside hydrolases that target plant cell walls often contain non-catalytic carbohydrate binding modules (CBMs), which promote enzymes access to insoluble substrates and thus potentiate catalysis. This project aims to use xylan-specific CBMs to target enzymes in washing powder onto grass stains to improve the removal of these stains.

The binding capacity of CBMs to grass stains arrayed on nitrocellulose was initially assessed using immunological methodologies that focused on the detection of the His_6 tag of the CBMs. The data showed that CBM15 and CBM2b1,2 bound to grass weakly, in the presence of Marvel. When the surfactant Sodium Lauryl Ethoxy Sulphate (SLES) was added, both CBMs showed strong binding to grass stains arrayed on the membranes in the presence of Marvel. Different methods to detect the two CBMs binding to grass stains displayed on cotton were explored in the absence of Marvel. The most appropriate method was direct labelling of CBM2b and CBM15 using fluorescent dye (Alexa 488) followed by confocal and fluorescent microscopy. The confocal data showed that both CBMs bound to grass stains and cotton, in the absence of SLES. The confocal data also showed that SLES strongly increased CBM2b1,2 targeting to grass stains. To quantify the CBM adsorption to grasses and cotton, standard fluorescence microscopy was used to provide a quick and efficient image formation system. Statistical analysis showed that neither CBMs significantly bound to grass stains, in the absence of SLES. There was significant increased targeting of both CBMs to grass stains in the presence of SLES. They both showed significant binding to non-stained cotton, either with or without SLES. CBM2b1,2 and CBM15 were fused to the xylanases *C. japonicus* Xyn10A and *N. patriciarum* Xyn11A respectively; the CBM-enzymes were purified and assessed in detergent assays. Statistical analysis showed that CBM2b1,2-Xyn10A efficiently released grass stains the most and showed a significant difference to the control with no enzymes, in surfactant formulation containing a high percentage of SLES (high SLES blend). Xyn10A, Xyn11A, CBM15-Xyn10A, CBM15-Xyn11A and CBM2b1,2-Xyn11A were not significantly different to the control, but they showed higher averages of the amount of grass dye released than the control, in high SLES blend.

This thesis also reports the biochemical properties and crystal structure of *v*CBM60, a homolog of CBM60 in *C. japonicus*, which uniquely displays broad ligand specificity, targeting xylans, galactans and cellulose. The crystal structure of *v*CBM60 displays a β-sandwich with the ligand binding site comprising a broad cleft formed by the loops connecting the two β-sheets. Ligand recognition at site 1 is, exclusively, through hydrophobic interactions, while binding at site 2 is conferred by polar interactions between a protein-bound calcium and the *O2* and *O3* of the sugar. The observation that ligand recognition at site 2 requires only a β-linked sugar that contains equatorial hydroxyls at *C2* and *C3*, explains the broad ligand specificity displayed by *v*CBM60. Another CBM from family 35 in *C. thermocellum* has been studied in this project. Site directed mutagenesis of ten residues around its binding site were carried out. The binding of wild type and mutants of *Ct*CBM35 to galactomannan was determined by Isothermal Titration Calorimetry (ITC). The data showed that all ten mutants abolished binding to galactomannan. Within a wider context for the CBM35 family, where the ligand binding sites of galactose-specific *Ct*CBM35 and uronate-specific CBM35s are similar structures, but display divergent ligand specificity, the signature residues in their ligand binding sites are compared and discussed in this study.

List of Figures and Tables

Abbreviations

A_x	Absorbance at x nm
AGE	Affinity Gel Electrophoresis
Amp	Ampicillin
APS	Ammonium Persulphate
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
BX	Birchwood xylan
C2	Cellobiose
C3	Cellotriose
C4	Cellotetraose
C5	Cellopentaose
C6	Cellohexaose
Carob	Azo-carob galactomannan
CAZy	Carbohydrate Active Enzyme server
CBM	Carbohydrate Binding Module
CD	Catalytic Domain
CFE	Cell-Free Extract
Cf	*Cellulomonas fimi*
*Cf*CBM2b1,2	Family 2b Carbohydrate-Binding Modules from *Cellulomonas fimi*
Cj	*Cellvibrio japonicus*
*Cj*CBM15	Family 15 Carbohydrate-Binding Modules from *Cellvibrio japonicus*
*Cj*CBM60	Family 60 Carbohydrate-Binding Modules from *Cellvibrio japonicus*
*Cj*Gal53A	Galactanase 53A from *Cellvibrio japonicus*
*Cj*Xyn10A	Xylanase 10A from *Cellvibrio japonicus*
Ct	*Clostridium thermocellum*
*Ct*CBM22	Family 22 Carbohydrate-Binding Modules from *Clostridium thermocellum*
*Ct*CBM35	Family 35 Carbohydrate-Binding Modules from *Clostridium thermocellum*
3D	Three-Dimensional
dNTP	Deoxynucleotidetriphosphate
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribonucleic Acid
2,4-DNP	2,4-Dinitrophenol
DNSA	3,5-Dinitrosalicyclic Acid
FITC	Fluorescein isothiocyanate
Gal	D-Galactose
G1	Galactose
G2	Galactobiose
G3	Galactotriose
G4	Galactotetraose

GAX	Galactoarabinoxylan
GFP	Green Fluorescent Protein
gfp	Gene encoding GFP
GH	Glycoside Hydrolase
HG	Homogalacturonan
His_6-tag	Polyhistidine tag
HMW	High molecular weight marker
HRP	Horse Radish Peroxidase
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl-β-D-Thiogalactopyranoside
IX	Insoluble oat spelt xylan
ITC	Isothermal Titration Calorimetry
Kan	Kanamycin
kb	kilo base
kDa	kilo Dalta
LB	Luria-Bertani Broth
LMW	Low molelular weight marker
M1	Mannose
M2	Mannobiose
M3	Mannotriose
M4	Mannotetraose
M5	Mannopentaose
M6	Mannohexaose
MQ H_2O	MilliQ ultra pure double distilled 18.2 MΩ water
MALDI-MS	Matrix-Assisted Laser Desorption Ionisation Mass Spectrometry
Mwt	Molecular weight
MP	Marvel powder
Np	*Neocallimastrix patriciarum*
*Np*Xyn11A	Xylanase 11A from *Neocallimastrix patriciarum*
NMR	Nucleic Magnetic Resonance Spectroscopy
NSB	Non-specific binding
OSX	Oat Spelt Xylan
p	Plasmid
P	Primary cell wall
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
PCW	Plant Cell Wall
PD	Pull Down
PEG	Polyethelene Glycol
PNP-X_2	1,4-Nitrophenyl-β-D-Xylobiose
PT linker	Proline-Threonine linker
RAX	Rye arabinoxylan

RG	Rhamnogalacturonan
RGI	Rhamnogalacturonan I
RGII	Rhamnogalacturonan II
S	Secondary cell wall
SDM	Site-Directed Mutagnesis
SDS	Sodium Dodecylsulphate
SLES	Sodium Lauryl Ethoxy Sulphate
sp	species
SX	Soluble oat spelt xylan
TEMED	N,N,N,N'-Tetra Methyl Ethylene Diamine
TLC	Thin Layer Chromatography
WAX	Wheat arabinoxylan
WT	Wild type
X2	Xylobiose
X3	Xylotriose
X4	Xylotetraose
X5	Xylopentaose
X6	Xylohexaose
XG	Xyloglucan

Content

Chapter 1. Introduction

Laundry detergents are becoming increasingly important nowadays, as they can be loaded and used during the washing cycles, impart softness, antistaticness, resiliency to fabrics, mild to eyes and skin and show good dispersibility in water. As the washing powder is continuously and commonly consumed, the production and sale has become an important industry. There are many different brands of laundry detergent sold in the market and many of them claim special qualities as selling points, such as, targeting some specific stains, maintaining bright colour of the fabrics. A laundry detergent composition is a formulated mixture of raw materials that can be classified into different types based on their properties and functions in the final product. These are surfactants, builders, bleaching agents which remove dirt, stain, and soil from surfaces or textiles and give them pleasant feel and odour. There are non-bio or biological washing powders, depending on whether the washing powder contains enzymes. It is traditionally in a powdered solid form, but the use of liquid laundry detergents has gradually increased over the years. Environment-friendly, with a high stain removing efficiency and high solubility in concentrated form are the properties required in a good laundry powder.

1.1 Structure of the plant cell wall

The plant cell wall constitutes a complex extracytoplasmic structure that has many biological roles including physical support and protection, energy storage, intracellular communication, water movement, cells development and microbial defence (Brett and Waldren, 1996). The cell wall consists of several layers (Fig 1.11). The first layer to be deposited, soon after cell division, is the middle lamella, which is formed at the cell plate as two cells divide. After cell division, and during cell expansion, the primary cell wall is formed and deposited over the middle lamella. After expansion, cell that display specialised functions such as transport of water (xylem) or structure (woody or fibrous tissues), synthesise an impervious secondary cell wall (S), deposited between the primary cell wall and the plasma membrane. The secondary cell wall can be deposited in more than one layer, S1, S2 or S3 (Beguin and Aubert, 1994).

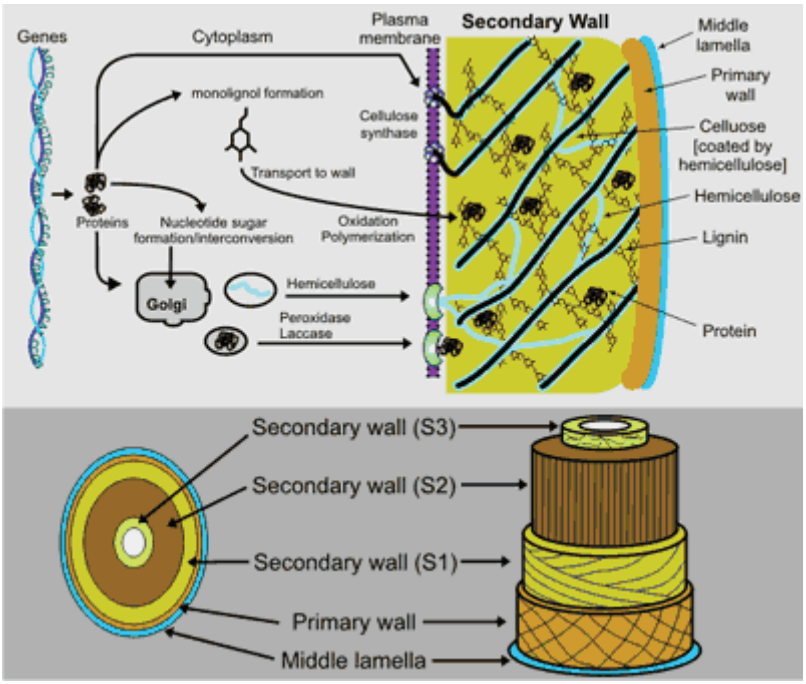


Fig 1.11 Structure of the plant cell wall

Schematics of the plant cell wall. S1, S2, S3 indicate different layers of the secondary cell wall (www.ccrc.uga.edu).

Primary cell walls consist of a complex network of cellulose, hemicellulose and pectin (Brett and Waldren, 1996; Cosgrove, 2005), which cross-link with each other *via* mainly non-covalent interactions. Secondary cell walls are similar to primary cell walls but, in addition, contain varying amounts of phenolic compounds such as lignin, providing additional rigidity and compressive strength, as well as rendering the walls hydrophobic and water impermeable (Whetten and Sederoff, 1995). While in primary cell walls the cellulosic microfibrils are ranged in a random order (Muller, et al., 2002), in the S layer they appear to be arranged in parallel (McNeil, et al., 1984).

The physical properties of the plant cell wall depend upon the interactions between its different components, which include hydrogen bonding between cellulose and hemicellulose (Tomme, et al., 1995), covalent ester or ether linkages between lignin and hemicellulose (Brett and Waldron, 1990; Grabber, et al., 1998), and covalent cross-linking of cell wall glycoproteins (Fry, 1982). Covalent cross-linkages between ferulic acid and other phenolic compounds are also implicated in the extreme recalcitrance of some plant cell walls to biodegradation (Weimer, et al., 2000).

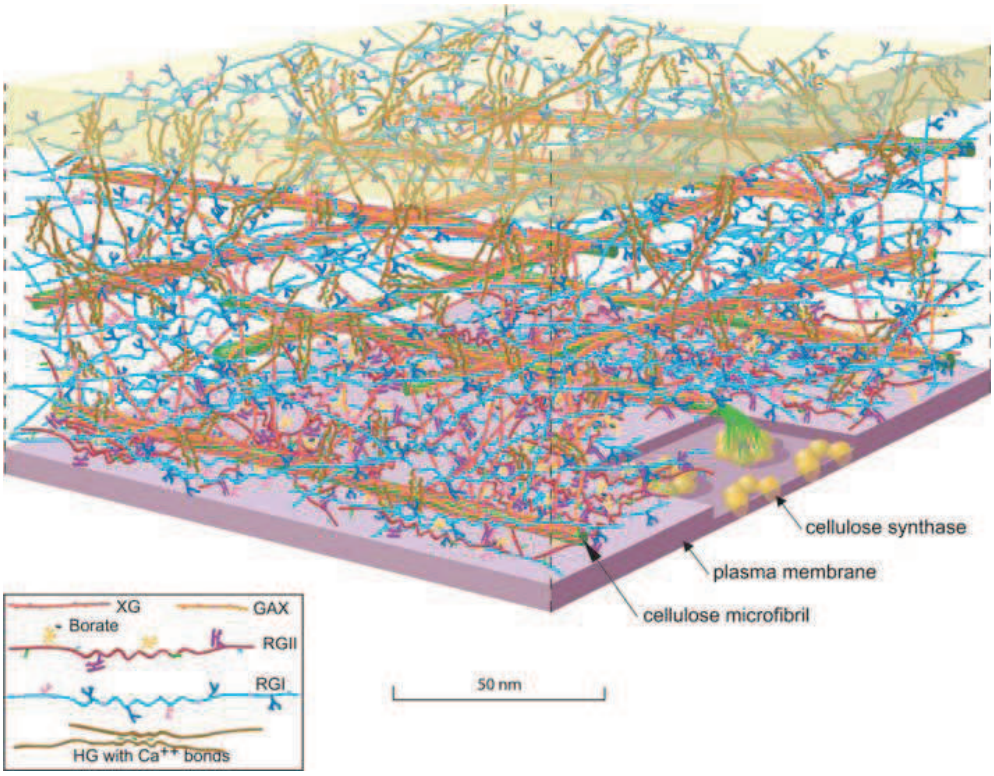


Fig 1.12 Model of primary cell wall

Scale model of the polysaccharides in an *Arabidopsis* leaf cell. Abbreviations: XG, xyloglucan; GAX: glucuronoarabinoxyaln; RGI: rhamnogalacturonan I; RGII: rhamnogalacturonan II; HG: homogalacturonan (Somerville, et al., 2004).

A cross-section model illustrating the complexity of the *Arabidopsis* leaf primary cell wall is shown in Fig 1.2. The plant cell wall is structurally complex and the composition between species, developmental stage and tissues can also be very different (Bacic, et al., 1988; Brett and Waldren, 1996; Popper and Fry, 2003). Polysaccharides make up 90 % of the plant cell wall and can be divided into three groups: cellulose, hemicellulose and pectin (McNeil, et al., 1984).

1.1.1 Cellulose

Cellulose is the most abundant polysaccharide within the plant cell wall (Fig 1.3). It is the one of most abundant polysaccharides on Earth. It consists of a chain of β 1-4 linked glucose (Glc) residues which are at least 300 molecules long. The repeating unit in cellulose is the disaccharide cellobiose with the second residue glc rotated 180 º compared to the first glc (Carpita, 1997). Cellulose chains found in primary cell walls

have a degree of polymerisation of about 300–15000 residues (Brown, 1999), and controlled cellulose biosynthesis allows arrangement of 30 - 100 chains which can be aligned side-by-side, through an intra and inter-chain hydrogen bonding network, to form microfibrils of 3-5 nm diameter (Beguin and Aubert, 1994; Lehtio, et al., 2003). The cellulose microfibrils are generally crystalline, although these ordered structures are interspersed with amorphous regions (Teeri, 1997), particularly at the surface (Beguin and Aubert, 1994). The structure of cellulose varies from species to species; in cotton, cellulose is about 70 % crystalline (Wood, 1988), whilst in *Valonia macrophysa* it is almost 100 % crystalline; this algae has very large microfibrils of several hundred glucan chains (Lehtio, et al., 2003; Nishiyama, et al., 2003). Thus although the glucan chain is chemically simple, cellulose has complex suprastructures, which make it indigestible to all organisms, except the few that produce a consortia of cellulolytic enzymes. The β1-4 linkage of cellulose gives it its linear structure, which confers high tensile strength, lending itself to a purely structural role in the cell wall, in contrast to the α-linked glc polymers starch, which forms an easily accessible helix ideal for carbon storage (Beguin and Aubert, 1994). The glucan chains in cellulose crystallize to form stable microfibrils in six different crystalline polymorphs. Cellulose I is the main form and it appears to have a semi-crystalline morphology that is interrupted by a random arrangement of amorphous regions (Atalla, et. al., 1993), particularly at the surface (Beguin and Aubert, 1994). In an expanding cell wall, cellulose accounts for 20 – 30 % (w/w) of dry weight.

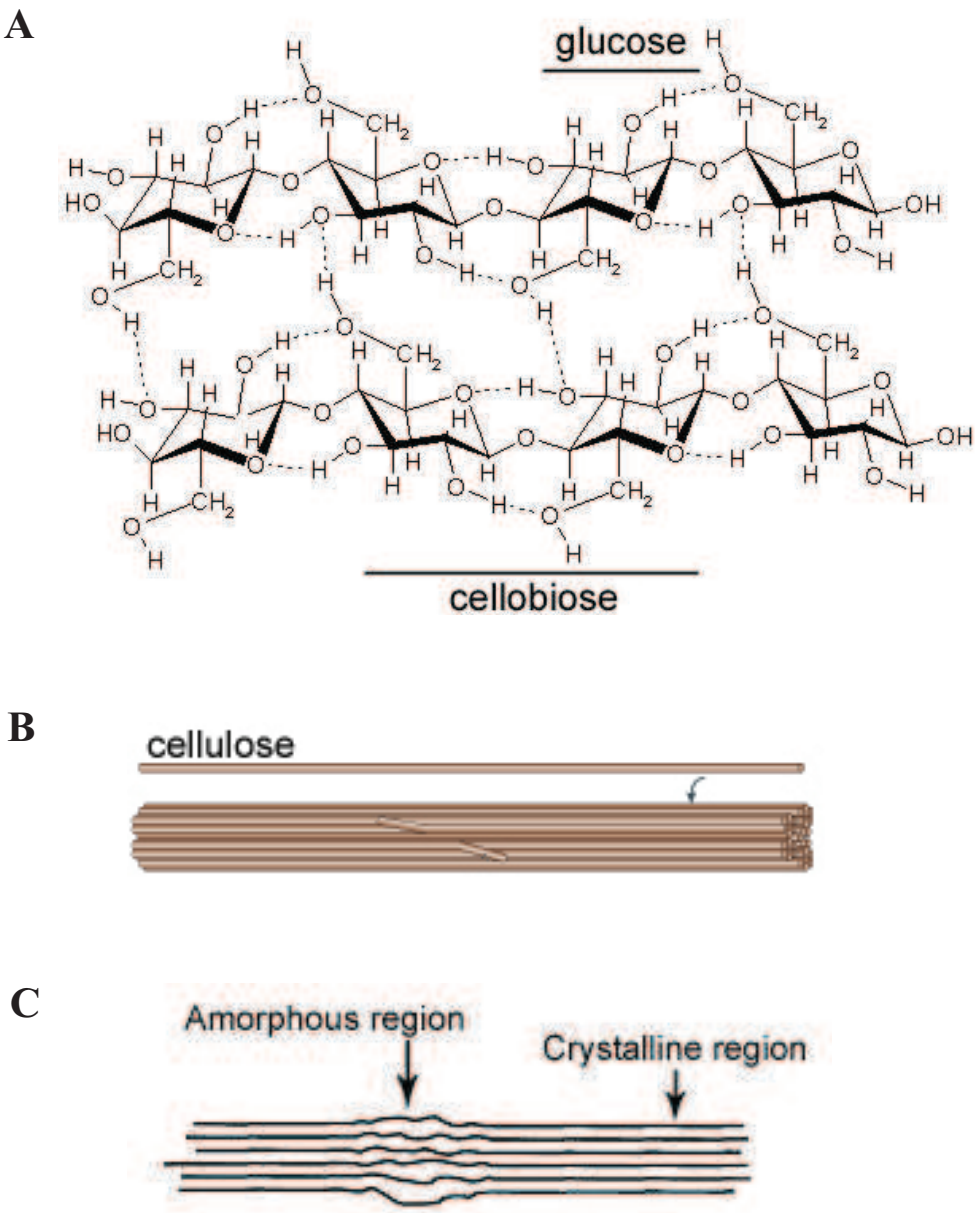


Fig 1.13 Structure of cellulose

Panel A displays the glucose linkages in cellulose, and the intra and inter-chain hydrogen bonds between β-glucan chains, showing repeating disaccharide units. Panel B: cellulose microfibrils. Panel C: crystalline and amorphous regions within cellulose microfibrils (adapted from Cosgrove, 2005 and Béguin and Aubert, 1996).

1.1.2 Hemicelluloses

The cell walls of most higher plant cells are composed of large amounts of cellulose fibres embedded in a continuous non-crystalline gel-like matrix (space between cellulose microfibrils), of which the major components are the hemicelluloses. The matrix is highly variable in composition dependent upon the plant species, stage of development and type of cell, and it can also contain pectins, glycoprotein and lignin (Brett and Waldron, 1990). These molecules are all, physically interlocked and interact by van der Waal's forces and hydrogen bonds (Brett and Waldron, 1996). The four predominant forms of hemicellulose polysaccharides found in the plant cell wall are xylan, xyloglucan, mannan and glucomannan (Brett and Waldron, 1990).

1.1.2.1 Xylans

The heteropolymer xylan can account for as much as 30 % (w/w) of the dry matter of many land plants and, in Chlorophyceae (one of the classes of green algae) and Rhodophycae (one of the classes of red algae), they substitute entirely for cellulose (Joseleau, et al., 1992). Thus, xylans play an important structural role in the plant cell wall. These polymers have a β-1,4 linked xylopyranosyl (Xyl) backbone comprising > 200 residues (Fig 1.4), decorated with side-chains (Liab, et al., 2000) that may include acetyl groups, arabinosyl (α-1,2 and α-1,3 linked) and α-1,2 4-O-methylglucuronic acid (Gilbert and Hazlewood, 1993). They are amorphous in structure and do not form fibres. Xylans, including arabinoxylans, glucuronoxylans and glucuronoarabinoxylans, are minor components of the primary cell walls of dicotyledons (or dicot; the name for a group of flowering plants whose seed typically contain two embryonic leaves) and non-graminaceous monocotyledons (Darvill, et al., 1980). By contrast, the secondary cell wall of hardwoods and the primary cell wall of the bran of grasses (Graminiae), are the principal source of xylans (Ebringerova and Hromadkova, 2000). In hardwoods, the polymers are decorated with α-1,2 4-O-methylglucuronic acid, and are also extensively acetylated at C2 and/or C3 (Darvill, et al., 1985; Puls and Schuseil, 1993; Sunna and Antranikian, 1997).

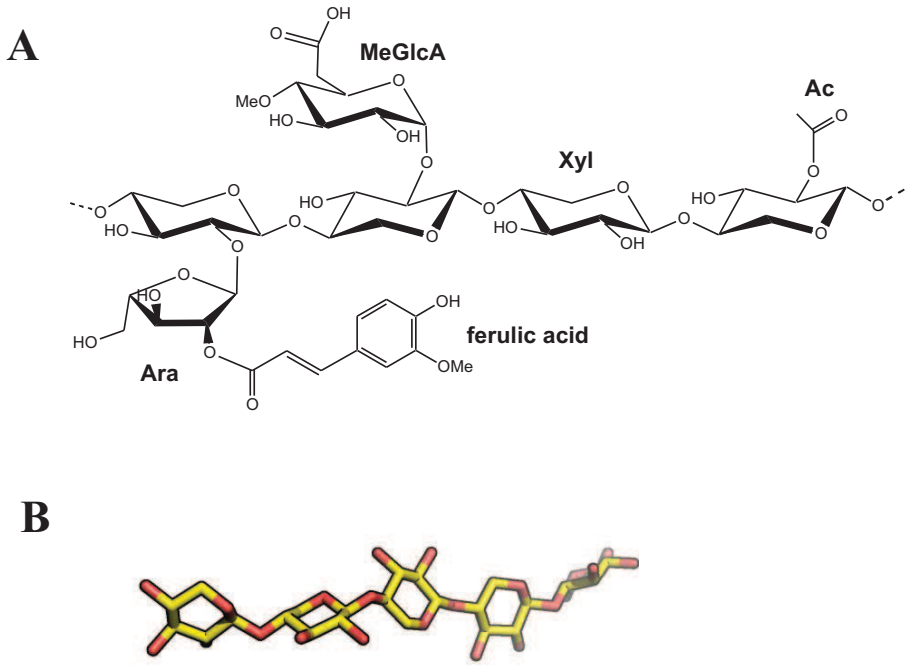


Fig 1.14 Structure and major types of substituents of xylan

Panel A – structure of a hypothetical xylan chain showing the major types of possible substituents. Abbreviations are: Ac, acetyl group; Ara, arabinofuranose; MeGlcA, 4-O-methylglucuronic acid; Xyl, xylopyranose. Panel B – Stick representation of xylopentaose of a three-fold helix conformation, which is from the crystal structure of the oligosaccharide in complex with Xyn10C-CBM15 from *Cellvibrio japonicus* (PDB code 1GNY, (Xie, et al., 2001a)).

By contrast, xylan found in the bran of grasses and grain, such as, wheat or rye, are predominantly substituted with α-1,2 or α-1,3-linked arabinofuranosyl residues (Ara), with low levels of 4-O-methylglucuronic acid substitution. In these arabinoxylans the Ara themselves can be substituted with ferulic acid at C5, which can dimerise and cross-link arabinoxylans to other hemicelluloses, pectins and lignin (Brett and Waldron, 1996; Joseleau, et al., 1992). The backbone of soluble non-feruloylated xylans is believed to be twisted into a three-fold helical structure (Atkins, 1992), which is consistent with the observed crystal structure of a family 15 carbohydrate binding module in complex with xylopentaose (Pires, et al., 2004; Xie, et al., 2001a).

1.1.2.2 Xyloglucans

Xyloglucan is the main hemicellulose found in the primary cell wall of dicotyledonous plants where it is thought to cross-link cellulose microfibrils (Pauly, et al., 1999a), being important for the flexibility of the cell wall (Whitney, et al., 1999). Xyloglucan consists of a backbone of β-1,4-Glc linked residues, that are decorated at regular intervals with α-1,6-Xyl units (Fig 1.5). Other commonly found side-chains include (1→6)-α-Xyl-(1→2)-β-Gal, (1→6)-α-Xyl-(1→2)-β-Gal-(1→2)-β-Fuc, (1→6)-α-Xyl-(1→2)-β-Gal-(1→2)-β-Ara and (1→6)-α-Xyl-(1→2)-β-Ara. In addition, Gal residues may be O2-acetylated (Aspinall, 1980; Brett and Waldron, 1996; Carpita, 1997). Xyloglucan chains can form hydrogen bonds with cellulose chains, usually in amorphous phases, or at the edges of cellulose microfibrils (Hayashi, et al., 1994; Ogawa, et al., 1990), thus xyloglucan polymers cross-link cellulose fibrils, creating a rigid structure (McCann, et al., 1990).

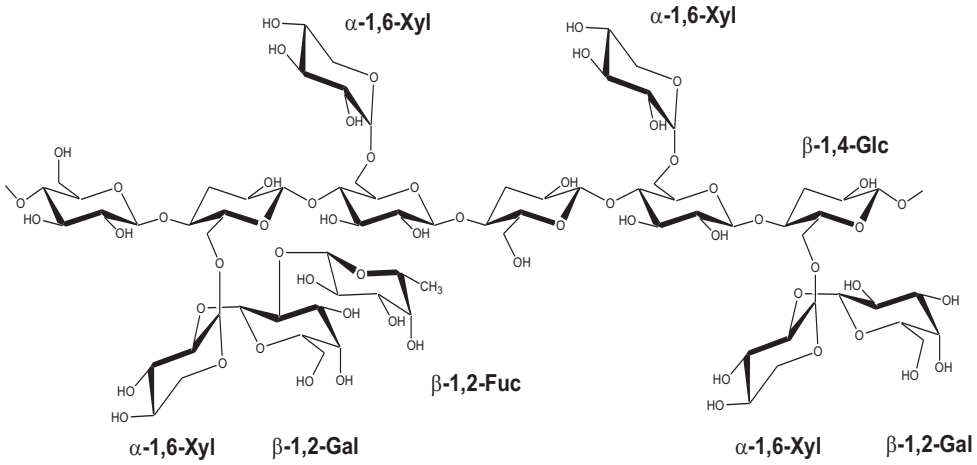


Fig 1.15 Structure of xyloglucan

Structure of a hypothetical xyloglucan chain showing the major types of possible substituents. Abbreviations are: Xyl, xylopyranose; Glc, glucopyranose; Gal, galactopyranose; Fuc, fucopyranose (Gloster, et al., 2007).

1.1.2.3 Mannans

Mannans include the following five polysaccharides: undecorated mannan, galactomannan, glucomannan, galactoglucomannan and glucuronomannan (Fig 1.6), (Brett and Waldren, 1996). These polysaccharides are the primary hemicellulose in gymnosperms (Puls and Schuseil, 1993), and are also abundant in the cell walls of various seed endosperms such as carob and coffee beans, and are the principal storage

carbohydrate of palm (Aspinall, 1980). As well as their storage function, mannans may also have a structural role in the cross-linking of cellulose microfibrils (Whitney, et al., 1998). Undecorated mannans form similar microfibrils to cellulose and consist of β-1,4 linked mannopyranosyl (Man) units in chains of varying length. In galactomannan the Man backbone is decorated with α-1,6-linked Gal side-chains with a ratio of Man to Gal residues, ranging from 1:1 to 1:5 (Aspinall, 1980; Brett and Waldren, 1996). This soluble polysaccharide forms part of the secondary cell wall thickening and is involved in the imbibing and taking up of water by seeds (Brett and Waldren, 1996). Glucomannan is the major hemicellulose component of the secondary cell wall of the gymnosperms, accounting for 25 % of the dry weight of wood (Puls and Schuseil, 1993). The backbone of this polysaccharide is a heteropolymer of β-1,4 linked Man and Glc residues that can be decorated on the mannosyl moieties with α-1,6 linked Gal to form galactoglucomannan (Brett and Waldren, 1996). Glucuronomannan is found in small amounts in many cell walls, it consists of a backbone of alternating α-1,4-linked Man and β-1,2-linked glucuronic acid residues with side-chains of Gal, and Ara (Whitney, et al., 1998). Mannans can be acetylated in different ratios at both C2 and C3 (Andersson, et al., 2007; Brett and Waldren, 1996; Puls and Schuseil, 1993)

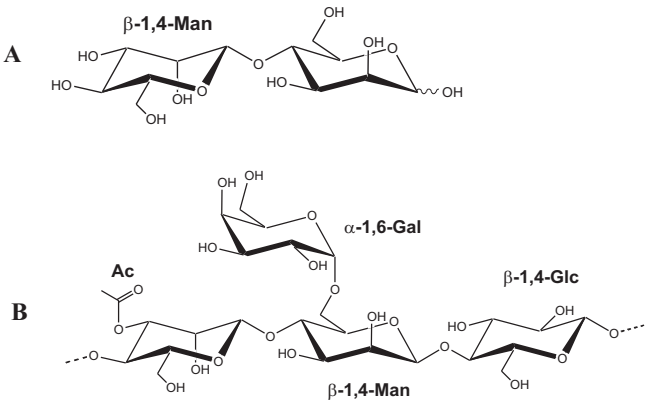


Fig 1.16 Structure of mannans

Panel A displays the structure of mannobiose. The repeating disaccharide in β-mannan and the backbone of galactomannan. Panel B displays galacto/glucomannan which has a backbone of Man and Glc residues randomly distributed. It also contains α-1,6-linked Gal side-chains, and the O-2 and O-3 of the mannose units can be substituted with acetate groups. Galactomannan is similar to galactoglucomannan except the backbone is exclusively Man. Abbreviations are: Man, mannopyranose; Glc, glucopyranose; Gal, galactopyranose (Shallom and Shoham et al., 2003).

1.1.3 Pectins

Homogalactoronic acid (HGA) is a homopolymer of 100-200 residues composed of α-(1,4)-linked galacturonic acid (GalA) units that are often highly methyl-esterified at C-6 and sometimes acetyl-esterified at C-2 or C-3 (Ralet, et al., 2001). Xylogalacturonic acid (XGA) is similar to HGA, except that it is substituted with single β-1,3-Xyl or β-1,4-Xyl residues (Schols, et al., 1995). Rhamnogalacturonan II (RGII) is a very complex polysaccharide but its structure appears to be remarkably conserved in all vascular plant. Its backbone is a short stretch of HGA (≈ 9 α-(1,4)-GalA) substituted with four different side-chains. RGII consists of at least 12 different monosaccharides in more than 20 different linkages. The residues in RGII include some monosaccharides that are rarely found in other polysaccharides, such as apiose, aceric acid, 3-keto-3-deoxy-manno-octulosonic acid (Kdo) and 3-deoxy-lyxo-2-heptulosaric acid (Dha) (Ishii and Matsunaga, 2001; Matsunaga, et al., 2004; O'Neill, et al., 2004). The backbone of RGI consists of the repeating disaccharide unit: (1→2)-α-Rha-(1→4)-α-GalA, which is predominantly substituted at C-4 of the Rha residues with neutral sugar side-chains (Willats, et al., 2001). Rhamnogalacturonan I (RGI) side-chains generally consist of arabinan, galactan and arabinogalactan (Schols and Voragen, 1994). The schematic structure of pectin was shown in Fig 1.7.

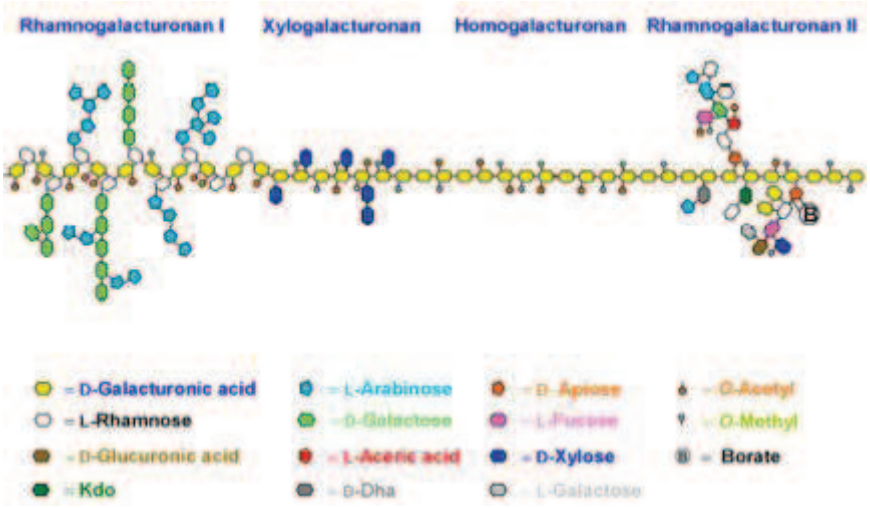


Fig 1.17 Schematic structure of pectin
Abbreviations: Kdo, 3-keto-3-deoxy-manno-octulosonic acid; Dha, 3-deoxy-lyxo-2-heptulosaric acid (www.jbei.org/feedstocks/cwb.shtml).

1.1.4 Neutral pectic polymers (Pectic galactan)

Side-chains of RGI occur in blocks described as “hairy regions” alternating with smooth undecorated structures (Brett and Waldron, 1996). Arabinan has a backbone of

α-1,5 linked Ara residues with single Ara substitutions at O2 and O3, and is linked to the Rha residues of the backbone of RGI (Lerouge, et al., 1993). Galactans are mostly linear chains of β-1,4 linked Gal residues while in arabinogalactan I this backbone is decorated with Ara oligosaccharides, Gal and ferulic acid (Carpita and Gibeaut, 1993). Highly branched arabinogalactan II are more common in plant cell wall proteoglycans, playing an important role in plant development (Rumyantseva, 2005), but may be found in gymnosperm and especially larch cell walls (Brett and Waldron, 1996). The structure comprises highly branched β-1,3 and β-1,6 linked Gal polymers decorated with Ara and small amounts of glucuronic acid.

1.1.5 Aromatic residues in plant cell wall polysaccharides

Aromatic compounds are thought to play an important role in the structure and function of the plant cell wall. Ferulic acid can be linked to both hemicellulose (Smith and Hartley, 1983) and pectin (Rombouts and Thibault, 1986), and is able to cross-link these polysaccharides to each other, as well as to the heterogeneous aromatic polymer, lignin (Ishii, 1997; Lam, et al., 1994). This cross-linked structure confers rigidity upon the cell wall. An increase in ferulic acid cross-links during ageing of the plant cell suggests a function in limiting cell growth (Fry, 1979; Wakabayashi, et al., 1997).

1.1.6 Proteins and minerals of the plant cell wall

In addition to the plant cell wall polysaccharides, 10 % of the cell wall dry weight are composed of proteins, which are almost all glycosylated. Plant O-glycans are usually linked to serine, threonine or hydroxyproline residues of the polypeptide backbone, and are commonly designated “hydroxyproline-rich glycoproteins” or HGRPs. One the three main categories of HGRPs are extensins, which are heavily glycosylated by short O-linked glycans, consisting of one to four Ara residues β-linked to hydroxyproline, and one or two Gal residues α-linked to serine. These molecules form a left-handed helix with three residues per turn, giving an overall rigid structure, which is maintained by the glycan side-chains, and play a structural role within the wall. These glycoproteins form a network that is orientated perpendicular to the polysaccharide networks, thus potentially strengthening the cell wall and playing a role in defence from pathogens (Roberts and Shirsat, 2006; Shanmugam, 2005).

Proteins that belong to the superfamily known as expansins, are also present in the plant cell wall. Expansins mediate cell wall disassembly by reversible disruption of hydrogen bonds at cellulose-hemicellulose interfaces, in response to stress produced by the internal pressure of the cell (Cosgrove, 2000). These proteins are involved in many processes where cell wall loosening is crucial, such as plant growth or organogenesis. Expansins have two domains, one that is structurally related to family 45 glycoside hydrolases (GH45), the other to grass pollen allergen domain 2. The crystal structure of expansins reveal aromatic residues along one plane in both domains which could be involved in polysaccharide recognition (Yennawar, et al., 2006).

Enzymes are also evident in the plant cell wall. They catalyse the degradation of structural polysaccharides and are classified in two major groups: exopolysaccharidases and endopolysaccharidases (Cosgrove, 2005; Minic and Jouanin, 2006).

1.2 Microbial plant cell wall glycoside hydrolases

The plant cell wall consists of a complex network of polysaccharides and is thus highly inaccessible to enzyme attack (Brett and Waldron, 1996). The vast majority of the enzymes which degrade plant cell wall are synthesised by anaerobic or aerobic mesophilic and thermophilic, bacteria and fungi, present in a variety of habitats including the rumen of herbivore guts and in top soil (Warren, 1996). Polysaccharides have remarkable diversity, a consequence of the wide variety of naturally occurring monosaccharides and the stereochemistry and location of the very stable glycosidic bonds that link the sugars. An example of the stability of cellulose is that it has an estimated half-life of five million years, whereas DNA has a half-life of "only" 140,000 years (Wolfenden, et al., 1998). The sugars that are components of polysaccharides are valuable sources of carbon and energy for microorganisms, and must be released from complex polymers prior to utilization by the host organisms. The major enzymes that catalyse plant cell wall polysaccharides hydrolysis are as glycoside hydrolases.

1.2.1 Family classification and nomenclature

Due to the great number of glycoside hydrolases in Nature, a classification based on sequence and fold similarity, displaying hydrophobic cluster analysis, was proposed by Henrissat and colleagues. (Henrissat, 1991; Henrissat and Bairoch, 1993; 1996). This classification, today, includes 118 glycoside hydrolases families in the CAZy database (Coutinho and Henrissat, 1999). In 1998, a scheme for naming glycoside hydrolases was proposed that took into account the activity and family location of the enzyme (Henrissat, et al., 1998). For example, Abf62 is an arabinofuranosidase located in glycoside hydrolases family 62. A capital letter is added to the end of the name to distinguish iso enzymes from the same organism. Furthermore, to distinguish enzymes from different organisms, two letters from the organism are added to the beginning of the name. For example, Abf62A from *Cellvibrio japonicus* is named *Cj*Abf62A.

Some glycoside hydrolases have also been grouped into clans (labelled from A to N, Fig 1.18) based on their structural fold and conservation the catalytic apparatus and mechanism (Henrissat, et al., 1995; Henrissat and Davies, 1997). The largest clan is GH-A, which includes enzymes from families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79 and 86 (http://www.cazy.org/fam/acc_GH.html). The members of this clan share the $(\beta/\alpha)_8$ - TIM barrel fold; the proteins adopting this structure contain eight parallel β-strands forming the inner barrel surrounded, on the outside, by eight α-helical cylinders, which are originally described for the chicken muscle triose phosphate isomerase, TIM (Banner, et al., 1975). The mechanisms of glycoside hydrolases are described in Section 1.2.3.

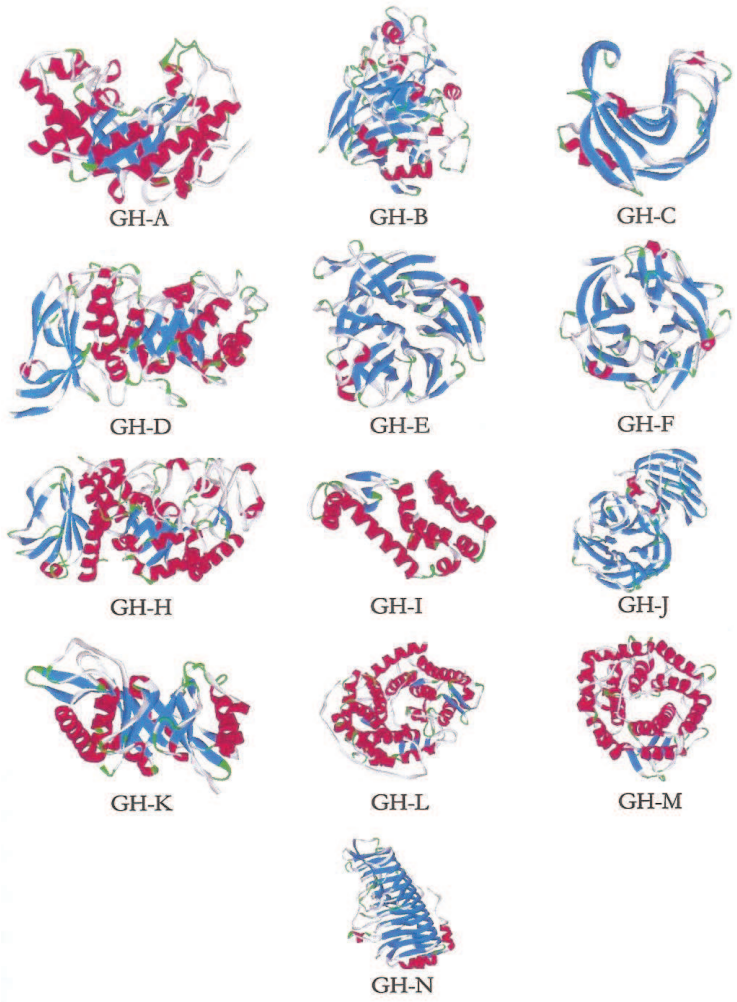


Fig 1.18 Secondary structure of glycoside hydrolases for each clan

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-glucanse; GH-N, *A. aculeatus* rhamnogalacturonase A (GH28) (reproduced from Pell, et al., 2004).

1.2.2 Endo/Exo acting mode of glycoside hydrolases

The GHs display either an endo- or exo-acting mode of action (Sinnott, 1990). The endo-acting enzymes often contain an "open" active site cleft while the majority of the exo-acting enzymes display a pocket shaped active site region (Davies and Henrissat, 1995). The cellobiohydrolases present the third mode of action, which appears to have evolved from the endo-acting cleft with loops covering the opening to create a tunnel like active site. The loops that form the "ceiling" of the tunnel appear to open (Armand, et al., 1997) allowing cellulose to bind to the cleft. The loops then close to reform the tunnel and successive cellobiose molecules are released from the active site as the enzyme "processes" along the bound cellulose (Rouvinen, et al., 1990).

1.2.3 Catalytic mechanism

There are two major mechanism by which glycoside hydrolases cleave glycosidic bonds that are often named after the stereochemical outcome of the reactions: inverting (α to β or *vice versa*) and retaining (α to α or β to β) (Koshland and Clarke, 1953; Withers, 2001). The inverting mechanism (Fig 1.19a) proceeds *via* a single step with one residue acting as the general base removing a proton from a water molecule, which then attack the anomeric carbon of the glycone sugar, and the other residue acting as the general acid donating a proton to the leaving group (Sinnott, 1990).

The retaining mechanism usually requires two catalytic carboxylates, a nucleophile and an acid/base (McCarter and Withers, 1994; Sinnott, 1990). In clan GH-A the catalytic residues are reported to be glutamates. The double-displacement reaction mechanism occurs in two separate reaction steps, glycosylation and deglycosylation (Fig 1.19b). In the glycosylation step, a carboxylate acts as the nucleophile and directly attacks the anomeric carbon of the glycone sugar, while the general catalytic acid-base acts as an acid and donates a proton to the glycosidic oxygen, aiding leaving group departure. As a result, a covalent bond is formed between the nucleophile and the anomeric carbon of the sugar. In the second step the charged general acid base activates an incoming water molecule by abstracting a proton, which then cleaves the covalent bond between the catalytic nucleophile and the glycone sugar. This reaction results in an overall retention of the anomeric carbon. However, it is also possible for an incoming sugar to participate in the deglycosylation step in which case a transglycosilation reaction will

occur (Sinnott, 1990). Such a reaction results in the formation of a new glycosidic bond between the glycone and the incoming sugar.

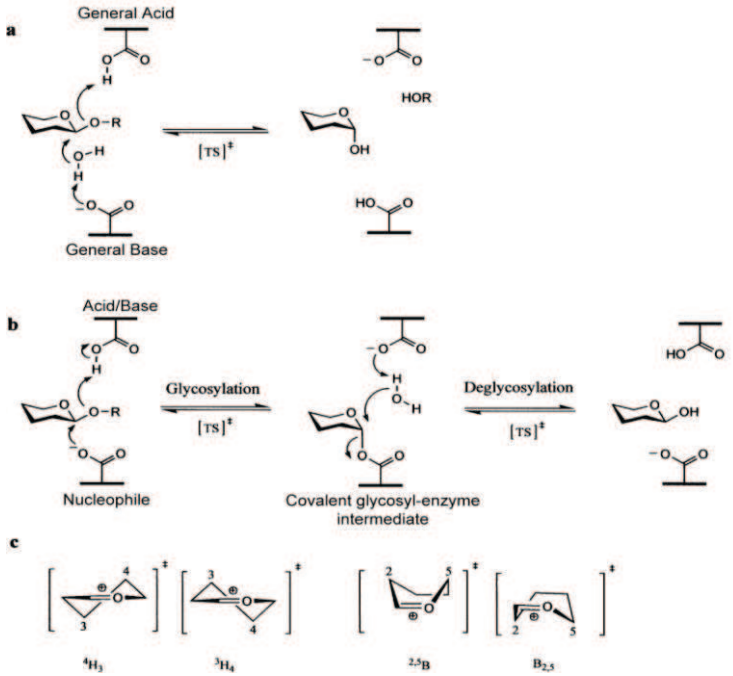


Fig 1.19 General mechanisms for inverting and retaining glycosidases.

The inverting (a) and retaining (b) reaction mechanisms of glycoside hydrolases. Both pathways pass through transition states ([TS]) possessing strong oxocarbenium-ion-like character. Such transition states (c) are now thought to display one of four possible conformations, 4H3 and 3H4 half chairs and 2,5B and B2,5 boats. Increasing evidence suggests that different enzyme families harness different transition state conformations (adapted from Davies, et al., 2003).

Binding and positioning of the substrate in the active site is dependent on complex protein-carbohydrate interactions. The amino acid(s) that participate in the binding of one sugar monomer are grouped into a subsite. It was proposed that the subsites should be numbered by positive and negative integers (Davies, et al., 1997). The negative numbered subsites are positioned in the glycone and the positive in the aglycone region of the active site (Fig 1.20). Glycosidic bond hydrolysis occurs between the -1 and +1 subsites (Davies, et al., 1997).

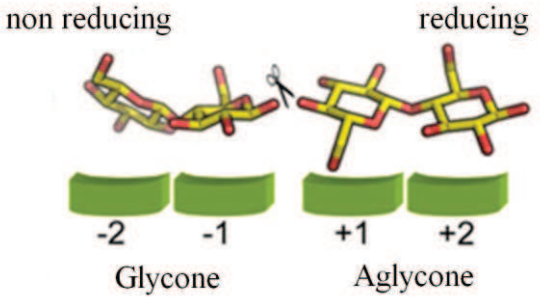


Fig 1.20 Schematic representation of GH subsites.

The hydrolysis occurs between the -1 and +1 subsites, symbolised by the scissor. Negative numbered subsites are located in the glycone and positive numbered in the aglycone. In this figure, the reducing end is positioned in the +2 subsite.

Some microorganisms utilise the entire plant cell wall as a source of carbon and energy, producing the full range of requisite GHs to saccharify these structures. Many microorganisms, however, express enzymes that hydrolyse limited regions of the plant cell wall (Tomme, et al., 1995). Microorganisms with partial hydrolytic ability typically include plant pathogens, which in some instances only require limited cellulolytic capacity to gain access to the plant tissues (Tomme, et al., 1995), and bacteria that live in environments such as the rumen, where they function as part of a consortium acting synergistically to completely hydrolyse plant biomass (Beguin and Aubert, 1994; Williams and Withers, 1985). The following is a description of the main plant cell wall glycoside hydrolases.

1.2.4 Main plant cell wall glycoside hydrolyses

1.2.4.1 Cellulases

Because of their great economic potential in the bioconversion of cellulosic waste into biofuel and feed stocks, cellulases are by far the most extensively studied glycoside hydrolases (Antoni, et al., 2007; Hahn-Hagerdal, et al., 2006). There are four major classes of cellulases; endoglucanases, which hydrolyse cellulose to glucooligosaccharides; cellobiohydrolases release cellobiose from crystalline cellulose; β-glucosidases degrade cellobiose and small oligosaccharides to glucose, and exoglucanases release glucose from cellulose and cellooligosaccharides. The distinction

between exoglucanases and cellobiohydrolases is not always clear due to the differences in the methods used to study these enzymes (de Vries and Visser, 2001).

Endo-acting cellulases (GHs from families 5, 6, 7, 8, 9, 12, 26, 44, 45, 48, 51, 61 and 74), termed endo-β-1,4-glucanases, catalyse the random cleavage of the internal glycosidic bonds within cellulose, and possess an "open-cleft" active site. These enzymes are active on cellulosic substrates that are accessible, generally, non-crystalline soluble glucans, cellulosic derivatives such as carboxymethylcellulose (CMC), and the amorphous regions of cellulose microfibril (Gilbert and Hazlewood, 1993).

Endo-β-1,4-glucanases and β-glucosidases are also able to degrade the backbone of xyloglucan. An endo-β-1,4-glucanase from *Aspergillus aculeatus* has been purified and was specific for the substituted xyloglucan backbone (Pauly, et al., 1999b). This enzyme was not able to hydrolyse cellulose and treatment of plant cell walls with the enzyme liberated only xyloglucan oligosaccharides. A *Bacillus licheniformis* GH12 enzyme demonstrated broad *endo*-glucanase activity and only tolerated backbone branching, a *Paenibacillus pabuli* GH5 enzyme explicitly required chain α(1 → 6)-xylosylation and showed increased activity on substrates where branching units were further extended with Gal-β(1 → 2) residues (Gloster et al. 2007). An atomic resolution structure of a GH44 *endo*-(xylo)glucanase has been published and the reaction stereochemistry for this family has been established as retaining (Kitago, et al., 2007). Notably, the latter work provides a structural representative for the GH family currently known to contain *endo*-(xylo)glucanases.

Cellobiohydrolases release cellobiose from cellulose chains in a sequentially manner. A processive mode of action appears to be due to cellobiose diffusion out of the active site while cellulose chains remain bound to the enzyme. This processive model has been proposed to be important for promoting the hydrolysis of crystalline cellulose. In constrast, the rigid crystalline structure of the cellulose microfibril does not allow physical access to endo-acting enzymes. In general, cellobiohydrolases have been shown to attack the cellulose chains from the non-reducing end. However, some cellobiohydrolases, such as GH7 enzymes, exemplified by CBH I from *Trichoderma reesei*, hydrolyse cellulose from the reducing end (Vranska and Beily, 1992). In

addition, it is believed that same cellobiohydrolases can also exhibit some endo-activity (Armand, et al., 1997; Boisset, et al., 2000; Meinke, et al., 1995). This is due to the flexibility of the extended loops that form the roof of the "tunnel" active site which allow the enzymes to bind to internal regions of the cellulose chain (Armand, et al., 1997; Davies and Henrissat, 1995; Varrot, et al., 1999). This is supported by the crystal structure of the D416A mutant of the *Humicola insolens* cellobiohydrolase Cel6A (GH6) in complex with a non-hydrolysable oligosaccharide, which displays the loops that form the roof in the active site in an "open" conformation (Varrot, et al., 2002), and also show how a cellulose chain is accommodated in the "tunnel" active site (Divne, et al., 1994; Divne, et al., 1998). Some cellobiohydrolases, however, display an exclusively exo-activity.

The final class of cellulases, β-glucosidases, hydrolyse cellobiose and short cellooligosaccharides into glucose (GHs from families 1 and 3). They display a "pocket" active cleft which accommodates the cellobiose substrate. These enzymes have been shown to act synergistically with cellobiohydrolases by relieving end product inhibition of the cellobiohydrolases (Hazlewood and Gilbert, 1998).

1.2.4.2 Xylanases

The β-1,4-linked Xyl backbone of xylans is randomly hydrolysed by xylanases, referred to as endo-β-1,4-xylanases (Gilbert, et al., 1992; Harris, et al., 1994). They hydrolyse internal bonds linking adjacent xylosyl residues by a double displacement acid-base assisted catalytic mechanism leading to retention of the anomeric configuration (Davies and Henrissat, 1995), and the generation of a range of xylooligosaccharides (Pell, et al., 2004a). Xylanases are generally found in GH families 10 and 11 (CAZy server), where GH 10 enzymes adopt a $(\beta/\alpha)_8$ - TIM barrel fold (Harris, et al., 1994) typical of clan GH-A, while the GH11 biocatalysts display a β jelly-roll fold, typical of clan GH-C (Torronen, et al., 1994). The active sites of both GH10 and GH11 xylanases consist of an "open-cleft", consistent with their endo mode of action (Biely, et al., 1981; Charnock, et al., 1998). The catalytic acid-base and nucleophile residues of GH10 enzymes are located at the end of β-barrels 4 and 7, respectively (Harris, et al., 1994; Lo Leggio, et al., 1999). It has been reported that some of these GH10 and GH11 xylanases can accommodate, and even interact with side-chains of substituted xylans,

enhancing their catalytic efficiency(Dumon, et al., 2008; Fujimoto, et al., 2004; Pell, et al., 2004a; Pell, et al., 2004b; Vardakou, et al., 2005).

The side chains on xylans are often removed first. The Ara*f* side chains are removed by arabinofuranosidases from GH43, GH51, GH54, and GH62 (Gilbert, 2008). Two arabinofuranosidases from family GH43 remove the *O3* side chain from Xyl residues that are decorated at both *O2* and *O3* with Ara*f* (Sorensen, et al., 2006). The 4-*O*-methyl-D-GlcUA appended to the Xyl positioned at the –2 subsite was hydrolysed by GH5 xylanase (Vrsanska, et al., 2007). The uronic side chains are released from the nonreducing end of xylooligosaccharides by GH67 α-glucuronidases (Nurizzo et al., 2002), although recent data showed that GH115 α-glucuronidases remove the uronic acid decorations from the internal regions of xylan (Ryabova et al., 2009). After all these side chains on xylan were degraded, the xylan backbone is more exposed and further hydrolysed by xylanases efficiently.

1.2.4.2 Mannanases

Mannanases, termed endo-β-1,4-mannanases are found mainly in GH5 and 26 (Dhawan and Kaur, 2007), with one example in GH family 44 (Gibbs, et al., 1992). These enzymes display a $(\beta/\alpha)_8$ - TIM barrel fold, catalysing bond cleavage *via* the double-displacement retaining mechanism. These glycoside hydrolases catalyse the random hydrolysis of glycosidic bonds within the backbone of mannan and heteromannans such as galactomannan, glucomannan or galactoglucomannan (Le Nours, et al., 2005). GH26 mannanases are mainly prokaryotic whereas GH5s are also produced by fungi, plants and some animals (Dhawan and Kaur, 2007).

A major discrimination between two types of mannanase is whether they contain cellulose binding domain, thus the cellular location of these enzymes are directed to are different. The mannan-degrading system of *C.japonicus* utilises three GH26 and four GH5 mannannase, all contain cellulose binding domain, directing these enzymes to insoluble plant cell walls and carry out the initial hydrolysis of mannans and glucomannans that are integral to these composite structures. The 'solubilised' mannans are then further hydrolyzed by *Cj*Man5A, mannobiohydrolases *Cj*Man26A, *Cj*Man26B and *Cj*Man26C by their attachment to membrane lipids (Cartmell, et al.,

2008). The *Cj*Man26A and *Cj*Man26C, although they may display complementarity or *endo-exo* synergy in which (random) internal glycosidic bond cleavage by the *endo*-mannanase creates new reducing ends that are targeted by the *exo*-mannanase. The major oligosaccharides generated by these two enzymes, mannobiose and, to a lesser extent, mannotriose, will then be hydrolyzed by the *exo*-acting mannosidase, releasing mannose from both the disaccharide and trisaccharide (Cartmell, et al., 2008). Hence, due to the heterogeneity and complex chemical nature of plant mannans, its complete breakdown into simple sugars requires the synergistic action of endo-β-1,4 mannanases and exo-acting β-mannosidases (Dhawan and Kaur, 2007). Additional enzymes, such as β-glucosidases, α-galactosidases and acetyl mannan esterases (Tenkanen, 1998) are also required to remove side-chain sugars and acetyl groups that are attached to the mannan backbone.

1.2.4.3 Pectate lyases

Pectate lyases (or polygalacturonate lyases) constitute a family of enzymes that share 29-91 % sequence similarity (Harrissat and Heffron, et al., 1995). These enzymes cleave α-1,4-linked GalA units at C4 by a β-elimination mechanism that generates a Δ 4,5 unsaturated galacturonic acid at the non-reducing end of oligomeric products (Albersheim and Neukom, 1960). Pectate lyases are produced by many bacteria and some pathogenic fungi, with endo-pectate lyases being more abundant than exo-pectate lyases. The CAZy database (Coutinho and Henrissat, 1999) has classified polysaccharide lyases into 18 families, with pectate lyases located in families 1, 2, 3, 9 and 10. Several crystal structures of pectate lyases revealed a common parallel β-helical topology (Akita, et al., 2001; Lietzke, et al., 1994; Pickersgill, et al., 1994; Yoder and Jurnak, 1995), however, a notable exception is provided by family 10 pectate lyases which display a $(\alpha/\alpha)_6$ toroid fold (Charnock, et al., 2002b; Novoa De Armas, et al., 2004) or family 2 from *Yersinia enterocolitica* which adopts a rare $(\alpha/\alpha)_7$ barrel fold (Abbott and Boraston, 2007). Superimposition of bound oligosaccharide substrates within the catalytic centre of these structurally unrelated pectate lyases show that there is a remarkable similarity in the architecture of the catalytic apparatus, highlighting the conserved β-elimination mechanism (Charnock, et al., 2002b). Pectate lyases have an absolute requirement for Ca^{2+} ions (Margo, et al., 1994), and crystal structures of both apo and pectate lyases substrate complexes revealed that the metal ion binds directly to

the enzymes, linking the substrate to the proteins (Herron, et al., 2003). However, a recent study (Abbott and Boraston, 2007) reported utilisation of Mn^{2+} or Ni^{2+} instead of Ca^{2+} in metal-assisted β-elimination. The mechanistic contribution of transition metals in the β-elimination reaction remains to be determined.

A consortium of numerous enzymes is required to facilitate the degradation of pectins. These include pectin methyl esterases (CE family 8, CE8), pectin acetyl esterases (CE12) and rhamnogalacturonan acetylesterases (CE12), which remove the methyl and acetyl groups from pectin. Pectin lyases (PL1), exopolygalacturonate lyases (PL9), polygalacturonases (GH28) and exopolygalacturonases (GH28) degrade homogalacturonan. Rhamnogalacturonan lyases (PL4 and PL11) and rhamnogalacturonases (GH28) degrade rhamnogalacturonan. The branches of rhamnogalacturonan, composed of galactan, arabinan and arabinogalactan, are cleaved by several enzymes including galactanases (GH53) and arabinases (GH43) (Brown, et al., 2001).

1.3 Carbohydrate-binding modules

Most of the enzymes that degrade the plant cell wall are microbial glycoside hydrolases (Section 1.2) that display a modular architecture comprising catalytic and non-catalytic modules, which are joined by a flexible linker rich in hydroxyl amino acids (Fig 1.11).

Fig 1.21 Molecular architecture of *Cellulomonas. fimi* Xyn11A

Domain structure of *Cellulomonas. fimi* Xyn11A, showing a catalytic modules and two family 2b carbohydrate-binding modules. N-terminal signal peptide (cheques), GH11 catalytic module (vertical stripes), glycine linker (white), CBM2b-1 (diagonal stripes), Proline-Threonine (PT) linker (black), CE4 bifunctional xylanase/esterase (horizontal stripes), CBM2b-2 (vertical dashes), numbered with amino acid sequence. Residues 230-245 and 542–557 have sequences characteristic of linkers (Bolam, et al., 2000).

Many of these non-catalytic modules contain from 30 to about 200 amino acids and are present as single or multiple copies. They often bind to specific oligo- and polysaccharides derived from the plant cell wall and are thus defined as carbohydrate-binding modules hereafter referred to as CBMs (Boraston, et al., 1999). CBMs have been found in several polysaccharide-degrading enzymes including cellulases, xylanases, endomannanases, acetylxylan esterases, isomaltodextranases, arabinofuranosidases, pectate lyases, β-agarases, dextranases and β-glucosidases (Brown, et al., 2001; Finnegan, et al., 2005; Hatada, et al., 2004; Lymar, et al., 1995; Margolles-Clark, et al., 1996; Miyanaga, et al., 2004; Ohta, et al., 2004; Pires, et al., 2004; Stalbrand, et al., 1995). A study reported that a CBM appended to cytochrome b_{562}, which may play a role in the localization of the redox protein on the surface of cellulose or on the fungal sheath (Yoshida, et al. 2005). Two CBMs were described as independent modules, binding to β-1,3-glucans and polygalacturonic acid, respectively (Barral et al. 2005) (Abbott, et al. 2007). Some CBMs are also present in non-catalytic proteins and are part of a scaffolding protein that organizes the catalytic subunits into a multienzyme complex known as the cellulosome. Expansins, which are believed to play a role in non-hydrolytic cell wall expansion, are homologues to CBMs and possess cellulose binding capabilities *in vitro* (Cosgrove, 2000).

1.3.1 CBM nomenclature

CBMs are defined using the same nomenclature system used for GHs (Henrissat, et al., 1998). For example, the CBM from *Clostridium thermocellum* Xyn10B, which belongs to family 22, is called *Ct*CBM22 or, to be more exact, *Ct*Xyn10BCBM22. This GH contains several repeating CBMs of the same family, each CBM is numbered according to their relative position to the *N*-terminus of the enzyme. Thus, a triplet of family 22 CBMs present in a *Clostridium thermocellum* xylanase is called *Ct*CBM22-1, the second *Ct*CBM22-2 and the third *Ct*CBM22-3. There is at present 60 families of CBMs, based on amino acid sequence similarities, found in the continuously updated carbohydrate-active enzyme data base CAZy (http://www.cazy.org/). In addition to known CBMs, there are a large number of other modules found in several GH which have unknown function and lack homology with characterized proteins, and as such have been classified as X-modules. Some of them have subsequently been shown to be CBM and have been reclassified as such. It is likely that some of the uncharacterized X-modules will also display a carbohydrate binding function.

1.3.2 Carbohydrate binding

CBMs modulate GH activity through three different mechanisms; they have a proximity effect, a targeting function and contribute to non-hydrolytic substrate disruption. By binding to the substrate of the enzyme, CBMs bring the appended catalytic module into intimate and prolonged contact with its substrate, which is thought to increase the rate of polysaccharide degradation (Bolam, et al., 1998). Several studies have shown that this proximity effect occurs against insoluble polysaccharides but not soluble substrates. Indeed, proteolytic excision or genetic removal of CBMs from the catalytic modules decreases, dramatically, the activity of the enzymes against insoluble, but not soluble polysaccharides (Bolam, et al., 1998; Hall, et al., 1995; Tomme, et al., 1988).

The potential enzyme-targeting function of CBMs in the context of intact primary and secondary cell wall deconstruction was evaluated (Herve, et al., 2010). The capacity of *C. japonicus* pectate lyase Pel10A to degrade pectic homogalacturonan in primary cell walls was potentiated by cellulose binding modules from family 3a and 2a, but not by xylan-directed modules from CBM family 15 and 2b. Conversely, the capacity of arabinofuranosidase mediating removal of side chains from arabinoxylan in xylan-rich and cellulose-poor wheat grain endosperm cell walls was increased by a xylan-binding CBM15 and CBM2b but less so by a cellulose-specific module CBM2a. The capacity of xylanases to degrade xylan in secondary cell walls was potentiated by both xylan- and cellulose-directed CBMs. This demonstrates that CBMs can potentiate the action of a cognate catalytic module toward polysaccharides in intact cell walls through the recognition of nonsubstrate polysaccharides. The cellulose binding CBMs therefore have strong proximity effects within cell wall structures, bringing the catalytic module to the cellulose surface of plant cell wall and increase the access of the catalytic modules to their substrates in plant cell walls. This explains why many noncellulase cell wall hydrolases have evolved to contain cellulose-directed CBMs.

CBMs also fulfil a targeting function. Numerous GH contain CBMs that bind to the surface of crystalline polysaccharides, referred to as type A modules. By contrast, type B CBMs bind to single polysaccharide chains that act as substrates for the appended catalytic module. For example, type B CBMs, which bind to xylan, mannan and laminarin, are present in xylanases, mannanases and laminarases,

respectively(Simpson, et al., 2002; Tunnicliffe, et al., 2005; van Bueren, et al., 2005). This targeting function directs enzymes to specific polysaccharides within complete macromolecular structures such as the plant cell wall. The target of the CBMs can be more specific. Indeed, the family 9 CBM from 10A xylanase of *Thermotoga maritima* binds specifically to the reducing ends of polysaccharides, suggesting that the module targets damaged regions of the plant cell wall (Boraston, et al., 2001a; Notenboom, et al., 2001b), enabling enzymes to be recruited to structures that are accessible to GH attack. Recent data demonstrates that four distinct family 35 CBMs all display conserved specificity for Δ4,5-anhydrogalacturonic acid, which is a signature molecule of active pectin degradation taking place in plant cell walls. The architectures of the ligand binding sites are also conserved. However, they were appended to three different hydrolases and a glucosaminidase, which contributes to the detoxification process in the bacteria. Hence, these observations reveal that four CBM35s display divergent biological functions not only dictated by their carbohydrate binding specificity, but also the context of the target ligands.

Furthermore, family 1, family 2a and family 3 CBMs, appended to the same catalytic module, display different capacities to degrade crystalline cellulose (Carrard, et al., 2000). This implies that different regions of this chemically invariant polysaccharide are recognized by these CBMs. In addition, the ability of an enzyme to degrade non-crystalline cellulose can be influenced by the capacity of CBMs to recognize different regions of this polysaccharide, as demonstrated by competition binding studies of CBMs from families 17 and 28 (Boraston, et al., 2003a).

The third role of CBMs is to disrupt the structure of insoluble polysaccharides. This function was first documented for the *N*-terminal family 2a CBM of Cel6A from *Cellulomonas fimi* (Din, et al., 1994). More specifically, the CBM was able to disrupt the structure of cellulose fibres, resulting in the release of small particles without any detectable hydrolytic activity. In addition, this CBM was able to prevent the flocculation of microcrystalline bacterial cellulose (Gilkes, et al., 1993). Similar phenomena were also reported for other CBMs (Banka, et al., 1998; Gao, et al., 2001; Levy, et al., 2002; Xiao, et al., 2001). However, the potentiation of this disruption on cellulose activity was very limited, questioning the biological significance of these observations. In addition, it was reported that the two ligand binding sites of family 20

starch-binding CBMs are required to disrupt the structure of amylose (Giardina, et al., 2001), exemplified by the observation that these modules potentiate amylase activity, albeit at a modest level, when not physically linked to the enzyme. Recently, it has been shown that efficient chitin degradation depends on the action of a small non-catalytic family 33 CBM, CBP21, which binds to the insoluble crystalline substrate, leading to structural changes in the substrate which makes it more accessible to its target enzymes. CBP21 mediated a substantial potentiation in the hydrolysis of crystalline β-chitin by chitinases A and C *in trans*, while it is essential for complete degradation of chitin by chitinase B (Vaaje-Kolstad, et al., 2005). Another interesting observation was reported when a CBM was applied to dental plaque polysaccharides (biofilm containing mainly fructan and glucan), which resulted in its dispersion, thereby removing and preventing plaque formation (Fuglsang and Tsuchiya, 2001).

CBMs are not only appended to GHs; some of these modules can be linked to other enzymes or even exist as independent entities (not accessory module within a larger catalytic protein), exemplified by the CBM33 protein CBP21. Human laforin is one example where a CBM is connected to a phosphatase, where the function of the module is to direct the enzyme to glycogen. Mutations in this CBM led to mis-targeting of the phosphatase, causing the so called Lafora disease which is a type of epilepsy (Wang, et al., 2002). The highly allergenic olive pollen protein Ole e 10, is an example of an independent CBM not linked to another polypeptide module. This CBM is suggested to interact with β-1,3-glucans and regulate enzymatic activity of the cell wall synthesis/degradation pathways during pollen germination, as part of a multi-protein complex (Barral, et al., 2005). Recently, a fourth role has been proposed for CBM32 from *Yersinia enterocolitica*, another example of independent CBM. The protein displayed maximum affinity for highly polymerized forms of galacturonic acid. It was hypothesised that the role of *Ye*CBM32 may be to retain polygalacturonic acid in the periplasm of the bacterium and thus is integrated to polygalacturonic acid transport during pectin degradation (Abbott, et al., 2007).

1.3.3 CBM folds

Among the 59 families of CBMs in CAZy, the structures of these modules were classified into seven folds (Fig 1.12). The dominant fold is the β-sandwich (Richardson, 1981), being found in families 2, 3, 4, 6, 9, 11, 15, 17, 22, 27, 28, 29, 32, 34, 35, 36, 40, 41, 42, 44 and 47. This fold consists of two β-sheets of three to six antiparallel β-strands which adopt a jelly-roll topology. The β-sandwich CBMs bind at least one metal atom which, in general, appears to be structural. The second most prevalent 3D structure is the β-trefoil fold (fold 2) displayed by family 13 CBMs (Murzin, et al., 1992). This fold contains a 12 stranded β-sheet that form six hairpin turns. The other folds are the "cellulose binding" and the oligonucleotide/oligosaccharide binding folds (fold 3-5) and hevein-like fold (fold 6-7).

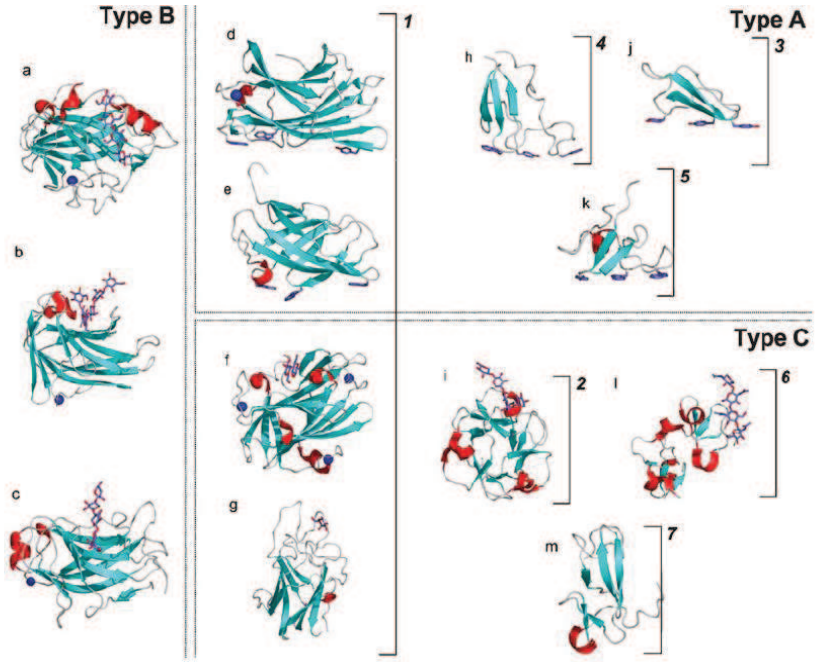


Fig 1.22 Fold relationships among CBMs
Dotted boxes surround examples of CBMs belonging to the functional types A, B, and C. Brackets with numbers indicate examples of CBMs belonging to fold 1–7. CBMs shown are as follows: (a) family 17 CBM, *Cc*CBM17, from *Clostridium cellulovorans* in complex with cellotetraose ; (b) family 4 CBM, *Tm*CBM4-2, from *Thermotoga maritima* in complex with laminariohexaose; (c) family 15 CBM, *Cj*CBM15, from *Cellvibrio japonicus* in complex with xylopentaose; (d) family 3 CBM, *Ct*CBM3, from *Clostridium thermocellum*; (e) family 2 CBM, *Cf*CBM2, from *Cellulomonas fimi*; (f) family 9 CBM, *Tm*CBM9-2, from *Thermotoga maritima* in complex with cellobiose; (g) family 32 CBM, *Mv*CBM32, from *Micromonospora viridifaciens* in complex with galactose; (h) family 5 CBM, *Ec*CBM5, from *Erwinia chrysanthemi*; (i) family 13 CBM, *Sl*CBM13, from *S. lividans* in complex with xylopentaose; (j) family 1 CBM, *Tr*CBM1, from *Trichoderma reesi*; (k) family 10 CBM, *Cj*CBM10, from *Cellvibrio japonicus*; (l) family 18 CBM from *Urtica dioca* in complex with chitotriose; (m) family 14 CBM, tachychitin, from *Tachypleus tridentatus*. Bound ligands are shown as 'liquorice' representations, while bound metal ions are shown as a blue spheres (reproduced from Boraston, et al., 2004).

1.3.4 Structure function relationships of CBMs

CBMs have been grouped into three types to reflect the gross structure of the ligands recognized by these protein modules, and the topology of their binding sites is consistent with the nature of their target polysaccharides. Thus, type A CBMs, which recognize crystalline cellulose and/or chitin, contain a binding site that comprises a planar platform composed of aromatic amino-acids, which are exposed to solvent (Fig 1.13A). This planar architecture is thought to be complementary to the flat surfaces of its ligand (McLean, et al., 2000; Tormo, et al., 1996). The specificity of type A modules appear to be invariant, although they are found in a range of enzymes including cellulases, mannanases, xylanases, pectinases and esterases. It includes members of CBM families 1, 2a, 3, 5 and 10 (Boraston, et al., 2004).

The ligand recognition sites of "glycan-chain-binding" CBMs, referred to as type B modules, are described as grooves or clefts, and comprise several subsites that accommodate the individual sugar units of the polymeric ligand (Fig 1.13B). The interaction site is often located on the concave surface of the proteins whose topology complements the conformation of the ligand (Notenboom, et al., 2001b). However, it is now recognized that several CBMs display a binding site on the edge of the β-fold, in the loops connecting the β-strands (Boraston, et al., 2006; Henshaw, et al., 2006; Jamal-Talabani, et al., 2004; van Bueren, et al., 2005). In addition, some CBMs contain multiple distinct ligand binding sites. This is exemplified by the crystal structure of CBM6-2 from *Cellvibrio mixtus* in complex with β-1,3-1,4-linked glucan tetrasaccharides, which binds β-1,4-1,3 glucan at the concave binding site and glucose at the edge of the β-fold (Pires, et al., 2004). Biochemical studies demonstrate that the affinity of type B CBMs for their target ligands increases up to the hexasaccharide, but is negligible for oligosaccharides with a degree of polymerization less than three. In addition, the crystal structures of CBMs in complex with their ligands revealed that the depth of the groove varies in these "chain-binding" CBMs, from very shallow, in *Cc*CBM17 (Notenboom, et al., 2001a), to sufficiently deep to accommodate the entire width of a pyranose ring (Boraston, et al., 2002b). Contrary to type A modules, the specificity of type B CBMs is much more diverse; target ligands include amorphous cellulose, xylan, β-1,3-glucans, β-1,3-1,4- mixed linkage glucans, galactomannan,

glucomannan and starch. Hydrogen bonds play an important role in this diversity, while in Type A CBMs hydrophilic interactions hominate ligand recognition.

The crystal structure of a xylan-binding CBM (CBM15) reveals a β-jelly roll structure of two β-sheets displaying an extended groove that runs along the face of the protein. Within this deep-walled canyon, two tryptophan residues form hydrophobic stacking interactions with two xylopyranose, reflecting the approximate 3-fold helical conformation of bound xylan. The observation that the 2' band 3' hydroxyls of the sugars that comprise the polysaccharide are solvent exposed provides an explanation for the specificity of the CBM (Xie, et al., 2001a), as the backbone saccharide polymer is often decorated with an array of different sugars and acetate moieties. The structures of CBM27 (Boraston, et al., 2003b) and CBM6 (Boraston, et al., 2003a) bound to 6^3,6^4-α-D-galactosylmannopentaose and xylooligosaccharide, respectively, also revealed information on the extent to which the side-chain of xylans can be accommodated in the binding site of these proteins. Type B CBMs include examples from families 2b, 4, 6, 15, 17, 20, 22, 27, 28, 29, 34, 35 and 36 (Boraston, et al., 2004).

The last class of CBMs, is the 'small-sugar-binding' type C modules (Fig 1.13C). These CBMs are lectin-like and contain smaller binding sites than type B modules, displaying affinity for mono-, di- or tri-saccharides.

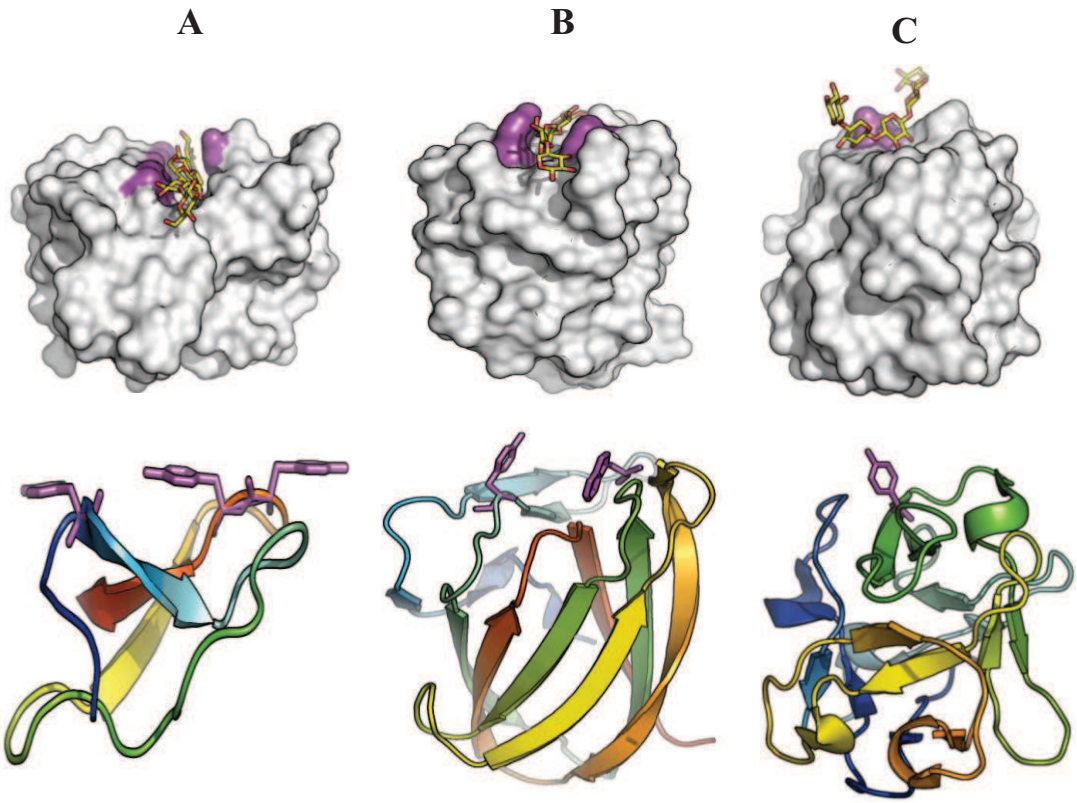


Fig 1.13 Binding site topology of type A, B and C CBM

Panel A displays Family 1 CBM, *Tr*CBM1 from *Trichoderma reesi*. Upper panel displays a surface representation of the CBM showing the planar platform of the binding site. Aromatic residues are colored in magenta. Lower panel displays a cartoon representation of the same CBM, showing the aromatic residues of the binding site (PDB code ICBH, (Kraulis, et al., 1989)). Panel B displays Family 6 CBM, *Ct*CBM6 from *Clostridium thermocellum* (PDB code 1UXX, (Pires, et al., 2004)). Upper panel displays a surface representation of the CBM showing the cleft in complex with xylopentaose. Aromatic residues are colored in magenta and polysaccharide in red and yellow sticks. Lower panel displays a cartoon representation of the same CBM, showing the aromatic residues of the binding site, situated at the edge of the β-sandwich. Panel C displays Family 13 CBM, *Sl*CBM13 from *Streptomyces lividans*. Upper panel displays a surface representation of the CBM showing the shallow binding site in complex with xylopentaose. Aromatic residues are colored in magenta. The two structures are shown in identical orientations. Lower panel displays a cartoon representation of the same CBM, showing the aromatic residue of the binding site (PDB code 1MC9, (Notenboom, et al., 2002)).

Structures of the CBMs typically reveal a preformed binding site with little conformational change between the free and the ligand bound states. The type C CBMs

include examples from families 6, 9, 13, 14, 18, 32 and 35 (Boraston, et al., 2004). Identification and characterisation of type C CBMs is lagging behind type A and B modules, probably due to their lower abundance in plant cell wall GH. CBMs from families 13 and 32 seem to be more prevalent in bacterial toxins or enzymes that attack eukaryotic cell surfaces or matrix glycans.

1.3.5 Structural determinants of polysaccharide binding

In CBMs, the aromatic side-chains of tyrosine, tryptophan and, less commonly, phenylalanine make hydrophobic interactions with the sugar rings of the target ligand (Pell, et al., 2003). The side-chains can be planar, twisted or form a sandwich. The planar conformation of the aromatic amino acid side-chains is characteristic of type A CBMs (Boraston, et al., 2004). In type B binding sites where the apolar platform is twisted the planes of two to three aromatic amino acid side-chains are rotated relative to one another (Bolam, et al. 2001). Where the binding site of type B CBMs adopts a "sandwich" topology, the aromatic amino acid side-chains stack against the β and α face of a single sugar residue at the heart of the binding site, while adjacent binding sites may induce a twist in the bound polysaccharide or oligosaccharide (Boraston, et al., 2004).Replacement of such amino acids at the binding site of a CBM has been shown to reduce or completely abolish ligand recognition (Nagy, et al., 1998; Pell, et al., 2003; Ponyi, et al., 2000), suggesting that aromatic residues play a pivotal role in carbohydrate binding. The orientation of aromatic side-chains is critical for both binding affinity and specificity of the CBM. Thus, the xylan-binding Xyn11ACBM2b from *Cellulomonas fimi* was converted into a cellulose binding module by introducing a mutation that resulted in the reorientation of the side-chain of a tryptophan residue (Simpson, et al., 2000). Consequently, the tryptophans in the mutated CBM formed a planar hydrophobic platform, which is typical of type A CBMs.

Although not as important as the aromatic residues, polar amino acids in the binding sites of CBMs have been shown to play a significant role in ligand recognition. Lectin and other sugar-binding proteins bind to carbohydrate through extensive networks of direct and indirect hydrogen bonds (Garcia-Hernandez, 1999). This is due to the amphipatic characteristic of the carbohydrates, which contain hydroxyl groups that are able to form hydrogen-bonds with polar residues in the binding-site. In contrast, the

binding-site of CBMs presents a relative paucity of hydrogen bonds with ligand. The reason is unknown, but may reflect the necessity to accommodate highly decorated polysaccharides that are interacting with other carbohydrate polymers within the plant cell wall (Boraston, et al., 2004). Hydrogen bonds contribute to the locking of the carbohydrate ligand into the binding cleft, explaining why both the binding enthalpy and entropy are affected by the replacement of direct bonding residues (Pell, et al., 2003). Mutation to alanine of polar residues involved in hydrogen bonding had only a minor effect on the affinity of type A CBMs for crystallising cellulose(McLean, et al., 2000), whereas large decreases in ligand affinity (Notenboom, et al., 2001a; Pell, et al., 2003) and even the complete loss in binding (Xie, et al., 2001b) have been observed in type B CBMs. Hydrogen bonds present in deep binding clefts thus appear to have a larger impact on ligand binding CBMs than those present on more shallow solvent-exposed binding sites. The exact position of the polar amino acids in the binding site make a significant contribution to ligand specificity since these residues participate in defining the complementary shape of the binding site to the target carbohydrate.

Previous studies have shown that CBMs from families 4, 6, 9, 22, 32, 35 and 47 contain one or more calcium ions located at sites remote from the ligand binding cleft, suggesting a structural role for the metal (Abbott, et al., 2007; Boraston, et al., 2002b; Boraston, et al., 2006; Charnock, et al., 2000; Czjzek, et al., 2001; Notenboom, et al., 2001b). Recent studies, however, have shown that calcium can play a direct role in CBM-ligand interactions. Family CBM35 provides one of the first examples of calcium playing a direct role in the binding of a CBM to its target ligand (Montanier, 2009) Indeed, resolution of the crystal structure of Xyn43ACBM36 from *Paenibacillus polymyxa* in complex with xylotriose demonstrated that calcium mediated ligand recognition through direct interaction between the bivalent metal and a xylose residue of the ligand (Jamal-Talabani, et al., 2004). CBM6 from *Saccharophagus degradans* agarase 16B also required calcium to bind to its ligand, however, the metal ion did not play a direct role in ligand recognition but contributed to the conformation adopted by Tyr 40, which was essential to ligand recognition (Henshaw, et al., 2006). This suggests that the involvement of calcium in the association of CBMs with their target ligands may be a more common phenomenon than previously believed.

1.4 Laundry powder formulations

Laundry detergent, or washing powder, is a mixture of a number of chemical ingredients. They are mainly surfactants, builders, bleach and its activator TAED, the antiredeposition agent carboxyl methyl cellulose (CMC), fragrances and the most importantly: enzymes. Enzyme-containing washing powder is also called 'biological'. The enzymes are required to work with equal efficiency at a range of temperatures, normally from 40 to 90 °C. A good detergent primarily relates to its cleaning performance, but also includes its biodegradability, it energy and water saving capacity and its minimal impact on aquatic system.

The surfactant is the most important part in laundry powder as it reduces the surface tension of the fabric and helps other detergent ingredients penetrate the water barrier more easily to act on the stains. Laundry detergents may contain more than one type of surfactants to aid the wash action. Different surfactants display various abilities to remove stains and responses to water hardness. The builders include phosphates, carbonates, silicates and citrates. Their role is to combine with mineral salts in the water which otherwise would interfere with the cleaning. Phosphates also stop the dirt from settling back on clothes. Some other ingredients are also present in the washing powder, such as fluorescers, which increases the visual whitening effect.

Different enzymes which target specific stains are added to detergents. The commonly used ones are protease, lipases, cellulase and amylase. They, respectively, break down proteinaceous stains such as blood, grass and egg; fatty and greasy stains such as lipstick and butter stains; starchy substances such as chocolate pudding. Enzymes enhance the removal of stains remarkably and have been used in detergents for the past 25 years.

During the wash cycles the fabric is made 'wetter' by the surfactants in the rinse step, the enzymes then penetrate the water barrier and reach the surface. Stains are broken down by enzymes and lifted away from the surface of the clothes, being prevented from redepositing on fabrics by antideposition agents. Washing powder or liquid is used in almost every household worldwide, and there are numerous brands of washing powders on the market, with varied instructions and recommended dosage. Hence, the

production and sale of these reagents is of industrial significance. Innovative ideas aimed at improving the performance of washing powders on difficult stains that are hard to remove, such as grass and blood are being explored.

1.4.1 Biosurfactant structure and roles

It is also very important to study the interactions between the different components in the laundry detergent. Surfactant is the term for surface active agents that compromise heterogeneous and long-chain molecules containing both hydrophilic and hydrophobic moieties. These surfactants are defined as cationic, anionic, nonionic and amphoteric, depending on the overall charge. By varying either the hydrophobic or hydrophilic component of the surfactant, a number of factors can be modified, e.g. wetting, emulsifying, foaming ability and dispersive abilities. Surfactant is known to have the tendency to adsorb to surfaces forming an interface between fabrics and water, which can synergistically assist other ingredients to penetrate the barrier between water and stains more easily. Surfactants produced by microorganisms are called 'biosurfactants'. For example, Sodium Lauryl Ethoxy Sulphate (SLES) is the main biosurfactant used for studying the effect of CBMs in this project. The chemical structure of SLES is shown in Fig 1.14 below. The surfactant contains a hydrophilic head, pointing to the water molecules; and long hydrophobic fatty acid chains pointing away from water that is attracted to grease and stains. It is soluble and Ca^{2+} tolerant. It reduces the surface tension of the fabric; loosens and removes soil on fabrics; and solubilizes or suspends solid in the wash solutions (Bajpai and Tyagi, 2007).

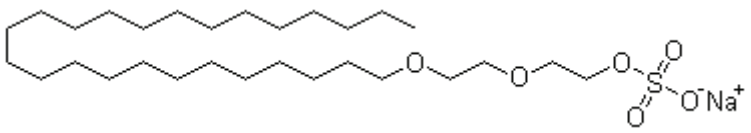


Fig 1.14 The structure of linear surfactant Sodium Lauryl Ethoxy Sulphate (SLES) Molecular formular: $C_{12}H_{25}O.(C_2H_4O)_2.SO_3.Na$ (R&D Department, Unilever)

1.4.2 Application of CBMs in laundry industry

The wide diversity in specificity and affinities of CBMs creates considerable scope for the applications of these modules in bioseparations and bioprocessing (Bayer, et al., 1994; Greenwood, et al., 1992; Levy and Shoseyov, 2002; Shoseyov, et al., 2006; Tomme, et al., 1998; Volkov, et al., 2004). Most commonly, CBMs are fused to other biologically active proteins and used as tags, such as affinity tags, immobilisation tags or targeting tags. Furthermore, these proteins have found uses as analytical tools in research and as diagnostics. Three examples of the practical applications of CBMs are described.

Since type A CBMs have high affinity for cheap and safe matrixes such as cellulose, they represent suitable tags for affinity protein purification (Boraston, et al., 2001b; Rodriguez, et al., 2004). Screening tools such as phage-display and protein microarray are used for the simultaneous analysis of the binding properties of many proteins. Both of these methods are under constant development and the use of fusion proteins with CBMs have been shown to overcome some of the technical obstacles faced by these systems (Berdichevsky, et al., 1999; Ofir, et al., 2005).

CBMs also represent valuable probes to study plant cell wall structure. For example, thermophilic bacterial CBMs, in which the ligand-binding aromatic residues stack against the two faces of the same xylose sugar, bind only to highly exposed xylan chains (McCartney, et al., 2004), while mesophilic CBMs derived from families 2b and 15, in which the ligand-binding site interacts with single faces of the backbone sugars, are able to bind to xylans that are in close association with other structural polysaccharides in plant cell walls (McCartney, et al. 2004). CBMs are also the only highly specific cellulose probes, comparing to antibodies that recognises cello-oligosaccharides and a range of β-glucans (Blake, et al., 2006). Recent studies have revealed modules that bind specifically to hemicellulose side chains. Thus CBM4-2 from a GH54 arabinofuranosidase binds specifically to the arabinose side chains of arabinoxylan (Miyanaga, et al., 2004), while Gilbert and Fontes have revealed two novel CBMs that recognise the galactose residues present in mannans and XGs, and recent studies on a xylan-specific CBM35 has shown that it binds to glucuronic acid side chains, but not the 4-methyl derivative of the uronic acid, which is more

commonly associated with decorated xylans (Harry Gilbert, unpublished data). By using CBMs as probes to plant cell walls, Knox's lab (Centre for plant sciences, University of Leeds) carried out elegant experiments by direct labelling CBMs with fluorescent dyes to gain information in the context of plant cell walls (Willats, et al., 2001; McCartney, et al. 2004; Blake, et.al. 2006).

Biosorption is an emerging technology that has been shown to be effective in removing very low levels of heavy metal from waste water. A recent study reported the cloning and expression of a recombinant protein composed of a *Clostridium*-derived cellulose-binding CBM fused to a synthetic phytochelatin. The immobilized sorbent was shown to be highly effective in removing cadmium at the level of parts per million (Xu, et al., 2002).

A possible textile-associated CBM application is exploited and researched in the detergent industry in this project. These modules could be deployed in the fabric targeting of recombinant enzymes to specific stains. This could be achieved by fusion of CBMs with the desired enzyme (von der Osten, et al., 2000a; b). There are potential benefits of appending a CBM to a catalytic enzyme. They promote enzyme access to insoluble polysaccharides and this potentiates catalysis by bringing the enzyme and the insoluble highly recalcitrant substrate into close proximity (Boraston, et al., 2004), which results in enhanced catalytic activity of enzymes. The enzymes will be in high concentration near the substrates and hence increase the rate of breakdown (Bolam, et al., 1998). Appropriate CBMs can contribute to laundry technologies and act as a linker attaching the stain at one end, and the corresponding enzyme molecule on the other end.

There are different types of polysaccharides present on the surface of stains. Hence, fusing CBMs that target specific molecules within stains present on clothes, to enzymes used in laundry detergent, may improve the cleaning process. These CBMs can act as molecular probes targeting stain areas.

Therefore, the main objectives of this study are as follows:

1. Using an immunological approach, to select xylan-specific CBMs that target grass stains arrayed on nitrocellulose membranes.
2. To develop appropriate labelling methods which allow CBM binding to grass stains arrayed on cotton to be detected, to explore the effect of surfactant SLES on CBM targeting to grass stains.
3. To determine the most effective xylanase families to remove grass stains. Fuse the most promising enzymes to grass binding CBMs and evaluate their capacity to remove grass stains from cotton.

Chapter 2. Materials and methods

2.1 Molecular Biology

High quality MQ H_2O (18.2 MΩ) produced by a Millipore Milli-RO 10 Plus Water Purification System was used for all experiments as the solvent in all solutions, unless otherwise stated.

2.1.1 Bacterial strains and plasmids

The *Escherichia coli (E. coli)* strains and plasmids used in this research are listed in Table 2.1 and Table 2.2

Strain *Escherichia coli*	**Description**	**Use**	**Reference**
BL21(DE3)	F⁻ *dcm ompT hsdS*(r_B- m_B-) gal (DE3)	Protein Expression	Studier. et al. (1986)
Tuner™ (DE3)	F⁻ *ompT hsdS*B (r_B^- m_B^-) *gal dcm lacY1* (DE3)	Protein Expression	Novagen
Origami B™ pLysS	F⁻ *ompT hsdS*B (r_B^- m_B^-) *gal dcm lacY1 ahpC gor522* ::Tn*10* (Tc^r) *trxB*::*kan* (DE3) pLysS (Cm^r)	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*ΔM15 Tn10(Tet^r)]	DNA Replication	Novagen
One Shot™ TOP10	F′*mcrA (mrrCB-hsdRMS-mrr) ø*80*lacZ*ΔM15 *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

Plasmids	Size (kb)	Phenotype/ Genotype	Reference
pCR®-Blunt	3.5	Kan^r, Zn^r, *ccdB*	Invitrogen Corp.
pET 16b	5.7	Amp^r, T7, *lac*, *lac*I^q	Novagen
pET 22b	5.5	Amp^r, T7, *lac*, *lac*I^q	Novagen
pET28a	5.4	Kan^r, T7, *lac*, *lac*I^q	Novagen

Table 2.2 Plasmids used in this study. The multiple cloning regions (MCRs) of these plasmids are given in Appendix A.

2.1.2 Media and growth conditions for propagation of bacteria

Escherichia coli strains were grown in liquid media at 37 °C (unless otherwise stated), in Luria-Bertani broth (LB) (1 % (w/v) Bacto ©tryptone, 1 % (w/v) NaCl and 0.5 % (w/v) yeast extract, pH 7.4). Conical baffle flasks used were at least two times the volume of the media to provide sufficient aeration by rotary shaking at 180rpm. For expression of seleno-methione proteins, a specific medium was prepared, using a kit from Athena Enzyme Systems™ comprising a selenoMet Medium Base™, a selenoMet™ Nutrient Mix and a SelenoMethionnine Solution following the manufacturer's instructions. Growth media were solidified by the addition of 2 % (w/v) Bacteriological agar No.1 (Oxoid), before autoclaving. The agar plate surface was dried by placing open face down at 65 °C for 15 min and allowed to cool at room temperature before storage at 4 °C for up to three to four weeks. The media were sterilised by autoclaving, after which appropriate antibiotics were added as required.

2.1.3 Selective media

For antibiotic selection, appropriate volumes of stock solutions (Table 2.3) were added to media that had been cooled to approximately 50 °C. Isopropylthio-ß-D-galactoside (IPTG) was added to strains containing *lacI*^{q}, either on plasmids or in the genome for induction of transcription of recombinant genes controlled by *lacO*. IPTG was made at a stock concentration of 1 M in MQ H_2O and added to liquid media to a final concentration of 1 mM for BL21 and B834 *E. coli* strains. For Tuner *E. coli* strain IPTG was most commonly added to media to a final concentration of 0.2 mM.

Antibiotic	Stock concentration	Working concentration	Storage
Ampicillin	50 mg/ml in water	50 µg/ml	4 °C for less than 3~4 days
Kanamycin	10 mg/ml in water	50 µg/ml	-20 °C
Chloramphenicol	34 mg/ml in ethanol	34 µg/ml	-20 °C

Table 2.3 Antibiotics used in this study.

2.1.4 Storage of DNA and bacteria

Glycerol stocks of each strain (25 % v/v glycerol) were kept at -80 °C in cryovials and restreaked every 30 days onto appropriate agar media containing antibiotics as required. Bacterial colonies on agar plates were stored at 4 °C for up to two weeks. DNA plasmids were stored at -20 °C in EB buffer (10 mM Tris/HCl buffer, pH 8.5).

2.1.5 Sterilisation

Unless otherwise stated, all solutions and media were sterilized by autoclaving using either an Astell Hearson 2000 series Autoclave or a Prestige Medical Series 2100 Clinical Autoclave at 121 °C, 32 lb/inch^{-2} for 30 min.

2.1.6 Chemicals, Enzymes and media

Chemicals, enzymes and media used in this study, which are not specified in the text, were from Sigma Chemical Company.

2.1.7 Centrifugation

Harvesting of bacterial cells from 1000 ml cultures was carried out by centrifugation at 5,000 g for 10 min in 500 ml centrifuge pots (Nalgene), using a Beckman J2-21 centrifuge with a JA-10 rotor. Bacterial cells from 10 ml cultures were harvested in 25 ml sterilin universal containers by centrifugation at 5,000 g in a MSE Mistral 3000i bench centrifuge with a swing out rotor. Eppendorf tubes (1.5 ml) were centrifuged up to 13,000 g using a Heraeus Instruments Biofuge *pico* benchtop centrifuge.

2.1.8 Plating bacteria

A glass spreader was sterilised by immersion in ethanol 100 % (v/v) for 1 min, followed by removal of excess ethanol by passing through a Bunsen burner flame and allowing the ethanol to burn off. Typically, a 100 µl bacterial suspersion was spread with a sterilised spreader evenly over the agar surface. Bacterial plates were incubated inverted in a 37 °C incubator (Laboratory Thermal Equipment Ltd) for approximately 16 h.

2.1.9 Chemically competent *E. coli*

E.coli strains were made competent for taking up plamid DNA using $CaCl_2$ using various methods. A 1 ml aliquot of a 10 ml overnight LB culture of *E. coli* was used to inoculate a one litre non-baffled flask containing 100 ml of LB. The culture without antibiotic was incubated at 37 °C at 180 rpm until log phase was reached ($A_{600}\approx0.4$). After 10 min on ice, the cells were harvested by centrifugation at 500 g at 4 °C for 5 min. The medium was decanted and the cells gently re-suspended in 8 ml of ice cold 100 mM $MgCl_2$. The previous step was repeated and the cell pellet re-suspended in 4 ml of ice cold 100 mM $CaCl_2$. After 2 h on ice, the cells were competent for transformation with plasmid DNA. Aliquots of 100 µl were stored at -80 °C with 25 % (v/v) glycerol in 1.5 ml Eppendorf tubes.

2.1.10 Transformation of competent *E. coli*

A 100 µl aliquot of competent cells were gently thawed on ice for 10 min.Typically, a 2 µg of plasmid DNA were added, mixed by gentle swirling and stored on ice for 30 min.

The suspension was then heat shocked by incubation in a Techne Dri-Block DB-2A at 42 °C for 2 min and returned immediately to ice for a further 2 min. Transformation with plasmid encoding amplicillin or kanamycin resistance were plated directly onto agar containing 50 µg/ml ampicillin or kanamycin before being inverted and incubated at 37 ºC for 16 h.

2.1.11 Rapid small scale plasmid preparation

For plasmid preparation, DNA was transformed into *E. coli* XL1-Blue or One Shot TOP10. Overnight 5 ml bacterial cultures were centrifuged at 5,000 g for 5 min and the supernatant removed. All traces of residual media were carefully removed by centrifuging the pellet for a further 2 min and decanted. Purification of plasmid DNA was carried out using a Qiagen QIAspin Prep kit as described in the manufacturer's instructions.

2.1.12 Measuring the concentration of DNA solutions

DNA concentration was determined either by estimating the size/intensity of the DNA band by comparison to quantified DNA ladders on agarose gel by DNA eletrophoresis (Section 2.1.13), or by spectrophometry. The absorption of appropriately diluted DNA (in MQ H_2O) between 230 nm and 340 nm was measured in quartz cuvettes on a Pharmacia Ultrospec 4000 spectrophotometer. Double-stranded DNA at 50 µg/ml or single-stranded DNA gives an A_{260} of 1.0, and oligonucleotide at 33 µg/ml.

2.1.13 Agarose gel electrophoresis of DNA

The separation and determination of the sizes of linear DNA molecules were carried out by electrophoresis through submerged horizontal gels. The composition of all buffers and reagents are given in Appendix C. Mini-gels used for rapid analysis were prepared with 0.4-1 g (0.8-2 %) of agarose in 50 ml 1 x TBE buffer and microwaved until dissolved. After cooling to approximate 50 °C, 0.5 µg/ml (final concentration) of ethidium bromide was added to make DNA visible, before gels were caste in mini-gel trays (Applied Biosystems). Once the gel had set, it was submerged in 50 ml of 1 x TBE buffer. DNA was electrohporesed at approximate 60 mA with a constant current

for 1 h (LKB Bromma 2197 Power Supply). The samples were prepared by the addition of loading buffer composition and 10 µl was pipetted into the wells of the gel. Standard DNA samples (HyperLadder I and IV from Bioline) comprising 10 µl (5 µg) were also electrophoresed to determine the size of DNA fragments.

2.1.14 Visualisation of DNA and photography of agarose gel

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

2.1.15 Determination of DNA fragment size by agarose electrophoresis

The size of linear double stranded DNA molecules can be determined as the rate of migration through agarose gels is inversely proportional to the Log 10 of the size of the nucleic acid. Hence, the sizes of DNA fragments were determined by comparing their eletrophoretic mobility with that of the DNA standards of known size.

2.1.16 Purification of DNA fragments

2.1.16.1 Restriction enzyme digestion of DNA

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

2.1.16.2 Purification of insert and vector DNA

Restriction digested vector and insert DNA fragments were separated by agarose gel electrophoresis using high purity Seachem Gold™ Agarose and TBE buffer. DNA was isolated from agarose gels by excision of the required band from the gel with a clean scalpel blade. DNA purification was carried out using a Perfectprep Gel Cleanup Sample Kit (Eppendorf) as described in the manufacturer's instructions.

2.1.16.3 Purification of PCR products

DNA PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) as described in the manufacturer's instructions.

2.1.17 Ligation of insert and vector DNA

In order to enhance the efficiency of the digestion of the PCR products by restriction enzymes, blunt PCR products were cloned into pCR-Blunt plasmid using a Zero Blunt PCR Cloning Kit from Invitrogen. Insert and pCR-Blunt plasmid were mixed at a molar ratio of at least 10:1. A typical ligation reaction of blunt PCR products and pCR-Blunt was made up in 1.5 ml Eppendorf tubes as follows:

1 µl	pCR-Blunt (25 ng), in TE buffer (100 mM Tris/HCl and 10 mM EDTA), pH 8.0
4 µl	Blunt PCR product (≈ 250 ng)
1 µl	10 × Ligation Buffer (60 mM Tris/HCl, pH 7.5, 60 mM $MgCl_2$, 50 mM NaCl, 1 mg/ml bovine serum albumin, 70 mM β-mercaptoethanol, 1 mM ATP, 20 mM dithiothreitol, 10 mM spermidine)
3 µl	Sterile water
1 µl	Invitrogen T4 DNA Ligase (4 U/µl)
10 µl	Total Volume

The ligation reactions were incubated at 16 °C for 1 h and One Shot® TOP10 supercompetent cells were transformed as described (Section 2.1.10). Insert and expression vector DNA molecules with compatible cohesive ends were mixed at a molar ratio of 3:1. A typical ligation was made up in 1.5 ml Eppendorf tubes as follows:

100 ng	Vector DNA
300 ng	Insert DNA
3 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 to 14 µl	Sterile water
1 µl	InvitrogenT4 DNA Ligase (4 U/µl)
15 µl	Total Volume

Each ligation reaction was mix gently and centrifuges briefly to bring the contents to the bottom of the tube and incubated at 16 °C for 2 h.

2.1.18 Polymerase chain reaction

The polymerase chain reaction (PCR) developed by Mullis and Faloona (Mullis and Faloona, 1987) was used in this study for gene amplification. 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 before despatch from MWG.

2.1.18.1 Standard PCR protocol

This method requires two oligonucleotide primers, one complementary to each strand of the DNA molecule, which flank the region of DNA to be amplified. A thermostable DNA polymerase catalyses the synthesis of the complementary DNA strand in the presence of dNTPs. Oligonucleotide primers were designed such that complementary sequences were approximately 25 bases in length having a G/C content of at least 40 %. The primers were longer where low G/C content was present. The annealing temperature (T_M) of the primers were > 45 ºC and within 5 ºC for the each primer pair. T_M were calculated using the following equation (http://www.basic.northwestern.edu/biotools/oligocalc.html):

$$T_M=65+41 \times (aG+bC-16.4)/(xA+yTaG+zC)$$

Where a,b,x,y are the number of the bases A, T, G and C in the DNA sequence, respectively. If possible, primers were also designed with G or C as the nucleotide at both 3'-end and 5'-end. The required restriction site sequences were added to the 5'-end of primers with the addition of the sequence CTC as at the extreme 5'-end to enable digestion of the PCR product for ligation into vectors. PCRs were made up in sterile 0.2 ml Eppendorf tubes as follows:

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)
3 µl	25 mM $MgCl_2$
5 µl	dNTPs (2 mM each)
1 µl	50 µM oligonucleotide primer #1
1 µl	50 µM oligonucleotide primer #2
≈ 60 ng	Template DNA
up to 49 µl	PCR Grade Water
1 µl	Novagen KOD DNA Polymerase (2.5 U/µl)
50 µl	Total Volume

Control reactions lacking target DNA were always carried out. A 30 µl layer of mineral oil (BDH) was added, prior to PCR, to the surface of each reaction mixture to prevent evaporation where heated lid apparatus was unavailable. PCR reactions were performed with a Robocycler (Stratagene). The standard thermocycle was as follows:

No. of cycle	Temp	Time
1	95 °C	3 min
	95 °C	30 sec
25	5 °C below the T_M of the primer pair	1 min
	72 °C for 1 or 2 min/kb fragment size	
1	72 °C	10 min
	5 °C	24 hr

After the PCR was completed, a 5 µl aliquot of each reaction was analysed by electrophoresis on a mini-gel.

2.1.18.2 Quickchange site-directed mutagenesis

Single amino acid mutagenesis was carried out with a Quickchange Site-directed Mutagenesis Kit from Stratagene, following the manufacturer's instruction. This method utilized an appropriate double-stranded recombinant plasmid DNA and two synthetic oligonucleotide primers (MWG-Biotech AG, Germany) containing the required mutation flanked by 10-15 nucleotides that fully complemented the DNA template. The oligonucleotide primers are extended during temperature cycling by

using the high fidelity proof reading DNA polymerase *Pfu*Turbo. Reactions were made up in sterile 0.5 ml Eppendorf tubes as follows:

20 ng	DNA template
5 µl	10 x 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)
1 µl	dNTPs mix
125 ng	50 µM oligonucleotide primer 1
125 ng	50 µM oligonucleotide primer 2
Up to 49 µl	PCR Grade Water
1 µl	*Pfu*Turbo DNA polymerase
50 µl	Total Volume

Each reaction tube was transferred to the Thermal cycler (Techne PHC-3) and cycling reactions were carried out as follows:

No. of cycle	Temp	Time
1	95 °C	3 min
1	95 °C	30 sec
18	55 °C	1 min
	68 °C for 1 or 2 min/kb of plasmid length	

After thermocycling the reaction tubes were kept on ice. The DNA was then transformed into *E. coli* One Shot TOP10 supercompetent cells (Section 2.1.10).

2.1.19 Automated DNA sequencing

DNA sequencing employed the Value Read service from MWG Biotech AG, Ebersberg, Munich, Germany using ABI 3700 sequencers. Plasmid DNA (1.5 µg prepared, as described in section 2.1.11) was dried by vacuum lysophilization at room temperature in 1.5 ml tubes. Plasmids were then sequenced with appropriate primers listed in Table 2.4 below. An example chromatogram is given in Fig 2.1 below.

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

Table 2.4 Plasmid sequencing primers

Typical sequencing data for a gene inserted into a pET vector is shown in Fig 2.1.

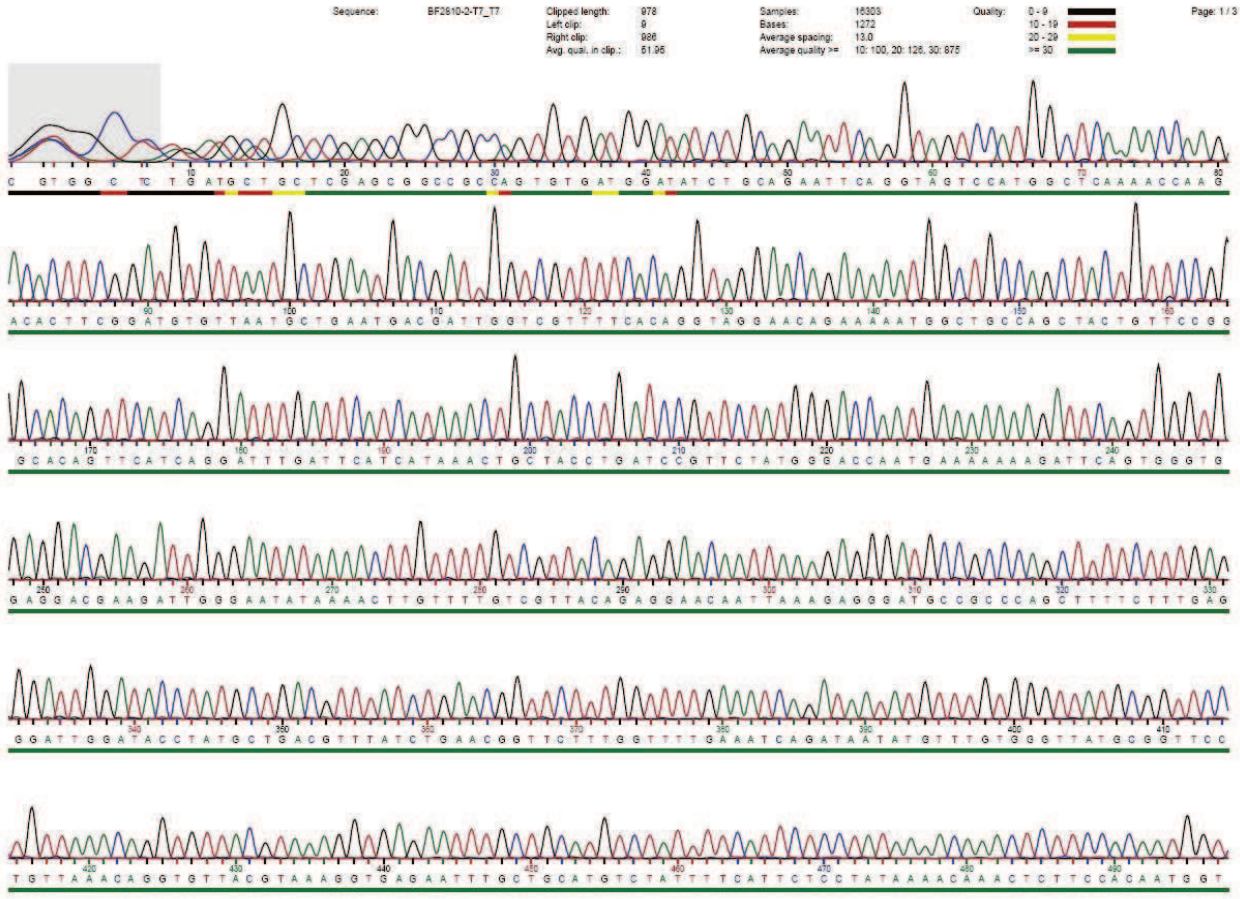


Fig 2.5 An example chromatogram of automated sequencing of the gene encoding *Cj*CBM15 using the standard T7 forward sequencing primer.

2.1.20 Sequence alignments and analysis

Amino Acid sequence alignments, translation and inversion-complementation of nucleic sequences were carried out using tools available on the web server InfoBiogen (http://bioinfo.hku.hk/services/menuserv.html).

2.1.21 Over-expression of recombinant proteins in *E. coli*

Unless otherwise stated, recombinant proteins were over-expressed in *E. coli* as follow. A streak of *E. coli* colonies harbouring an appropriate plasmid were inoculated into 5 ml LB in the presence of appropriate antibiotics and grown with aeration at 180 rpm at 37 °C for 16 h. The cultures were then used as the innoculum for a 2 litre baffled flask

containing 1 litre of LB supplemented with appropriate antibiotics. The cultures were incubated with aeration at 180 rpm at 37 °C until the absorbance at 600 nm reached 0.6. Depending upon strain and the proteins over-expressed, the culture was cooled down to 30 °C or 16 °C, or maintained at 37 °C, and IPTG was added at a final concentration of between 0.2 and 1 mM, and the cells were maintained at this temperature for the duration of recombinant protein expression (4 h or 16 h). Cells were harvested by centrifugation at 5,000 g for 10 min. The supernatant was decanted and the cell pellet re-suspended in 10 ml of Talon buffer (20 mM Tris/HCl, pH 8.0 containing 300 mM NaCl) per litre of original culture. The cell suspension was stored in a 25 ml Sterilin tube for 16 h at -20 °C. The cell suspension was then thawed and cell-free extract (CFE) was prepared as described below (Section 2.1.22).

2.1.22 Preparation of cell-free extract and pellet fractions from *E. coli*

Aliquots (10 ml) of thawed cell suspensions (Section 2.1.21) were sonicated for 3 min (on ice), using a sonicator (B. Braun Labsonic U) set at low intensity ≈45 Watts and 0.5 second cycling. Lysed cell suspension was transferred to a 50 ml centrifuge tube (Nalgene) and cell debris pelleted at 30,000 g for 30 min at 4 °C. The supernatant comprising the CFE was retained for further use. The cell debris-containing pellet (insoluble fraction) was resuspended in 10 ml of Talon buffer containing 8 M urea and incubated with occasional shaking at 37 °C until essentially transparent.

2.1.23 Sodium dodecyl sulphate-polyacrylamide gel elctrophoresis (SDS-PAGE)

Analysis of proteins was carried out using SDS-PAGE as described by Laemmli (Laemmli, 1970) to analyse the size, relative purity and quantity of protein. Gels consisting of 10-15 % polyacrylamide (Acrylogel 3; BDH Electran) were typically used dependent on size of protein of interest. SDS-PAGE was carried out using AE-6450 apparatus supplied by ATTO Corporation. A pair of glass plates (12 cm x 10 cm and 12 cm x 9 cm) were sprayed with ethanol and rinsed with distilled water. The plates were clipped together with a rubber gasket in between. The resolving gel (Table 2.5) was poured into the plates covered with a layer of water and allow to set.

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 2.5 12.5 % Resolving gel composition for SDS-PAGE

The water was then removed and the stacking gel (Table 2.6) was poured on top of the resolving gel and allowed to polymerise with a comb (12 wells) in place.

Volume	Component
3.75 ml	0.25 M Tris/HCl buffer, pH 8.8 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 2.6 Stacking gel composition for SDS-PAGE

The comb and the rubber seal around the glass plates were removed after the gel was polymerised. Protein samples were prepared by taking 10 µl of diluted proteins, 10 µl of loading buffer (Table 2.7) and boiling for 3 min.

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	ß-mercaptoethanol
0.1 %	Bromophenol blue dye

Table 2.7 Loading buffer for SDS-PAGE

The plates were placed in the gel tank that was filled with running buffer (Table 2.8). Samples (20 µl) were loaded into the wells and were electrophoresed at a current of 35A per gel.

Concentration	Component
32 mM	Tris / 190 mM glycine, pH 8.3
0.1 %	SDS

Table 2.8 Running buffer for SDS-PAGE

After electrophoresis, the gel was soaked in Instant Blue (containing Coomassie dye, ethanol, phosphoric acid and solubilising agents in water) for 20 min with gentle shaking. The gel was then soaked in distilled water for 1 h until all of the protein bands were visible and recorded using a digital camera (Canon). The Mwt of proteins separated by SDS-PAGE were estimated by comparing their electrophoretic mobility with protein standards of known molecular weight (See Appendix C).

2.1.24 Protein purification

2.1.24.1 Immobilised metal affinity chromatography (IMAC)

Immobilised metal affinity chromatography (IMAC) was used to purify recombinant proteins containing a stretch of 6 Histidine residues at the N- or C-terminus (His_6 tag). The electron-rich histidine residues interact with an electropositive transition metal (Cobalt or Nickel) bound to a solid chromatographic support and this interaction can be dissociated with imidazole. The metal affinity matrix used was Talon Fast Flow (Clontech Laboratories Inc), in which the His_6 tags interact with Cobalt, or a Chelating Sepharose Fast Flow matrix (Ap Biotech) in which the His_6 tags interact with Nickel.

2.1.24.2 IMAC using an isocratic flow by gravity

A 2 ml bed volume Talon column was made in a Fast flow column (Clontech Laboratories Inc). The column was then equilibrated with at least 2 x 10 column volumes of Talon buffer. A 30 ml of CFE (Section 2.1.22) was filtered (0.45 µm Acrodisc Gelman), applied to the column and allowed to flow through by gravity. The column with bound His-tagged protein was then washed with 2 x 20 ml of Talon buffer, before a step wise elution gradient of 5 ml Talon buffer containing 10 (3 x 3 ml) and

100 (3 x 4 ml) mM imidazole. Each fraction was analysed by SDS-PAGE (Section 2.1.23) and those containing large amounts of the target proteins of sufficient purity were dialysed against 2 x 200 volumes of a suitable buffer.

2.1.24.3 Ion-exchange chromatography (IEC)

A Bio-Rad BioLogic DuoFlow System with a flow of 2ml/min connected to a UNO Q12 anion exchange column (Bio-Rad), and using a Bio-Rad Biofrac fraction collector, was used to carry out a secondary purification step when required, after IMAC purification. Protein fractions were dialysed into 10 mM Tris/HCl, pH 8.0 and loaded onto the column using a 50 ml super-loop. The column was then washed with 60 ml of the equilibrating buffer to remove unbound proteins, before elution with a 120 ml 0-500 mM NaCl gradient in 10 mM Tris/HCl, pH 8.0. Protein elution and detection was as described previously.

2.1.24.4 Gel filtration fast performance liquid chromatography (FPLC)

Using the same FPLC system as above, connected to a Pharmacia XK 16 HiLoad 16/60 Superdex 75 prep grade gel filtration column (Fig 2.2), a final round of purification was carried out. The protein was concentrated to 1.1 ml (Section 2.1.26), loaded onto the column using a 1 ml static loop and eluted from the column using 10 mM Tris/HCl, pH 8.0 containing 150 mM NaCl. Fractions (1.5 ml) were collected and analysed by SDS-PAGE (Section 2.1.23). Pure fractions were pooled and concentrated from further analysis.

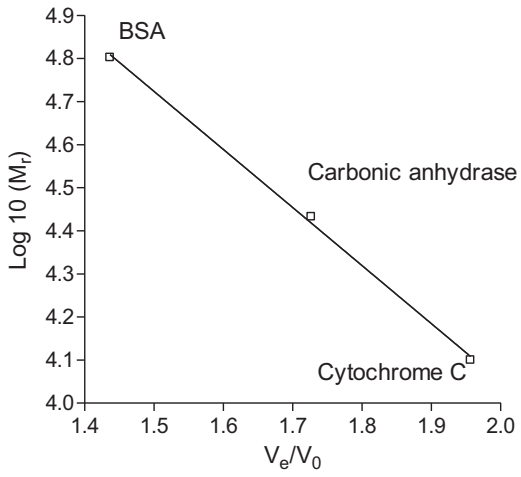


Fig 2.6 Calibration of the Pharmacia XK16 HiLoad™ 16/60 Superdex™ 75 column

Calibration of the column using proteins of known M_r (Kit for molecular weights 12 kDa - 200 kDa from Sigma, using Bovine Serum Albumin (66 kDa), Carbonic anhydrase (29 kDa) and Cytochrome C (12.4 kDa); V_e = void elution volume of the protein, V_0 = volume of the column (44.55 ml). The equation of the resulting line is: y = -1.4047x + 6.8535.

2.1.25 Quantification of purified protein

Pure protein concentration was determined by the method of Gill and von Hippel (Gill and von Hippel, 1989) and Pace et al. (Pace et al, 1995). Protein was suitably diluted in an appropriate buffer and the absorbance at 280nm was determined by scanning from 230-360 nm (Pharmacia Biotech Ultrospec 4000, UV/vis spectrophotometer). The protein concentration was determined using the formula:

$$A=\varepsilon\, C\, l\, D$$

Where $A=A_{280}$

ε =molar extinction coefficient at 280 nm

l=length of light path

D=dilution factor

C=molar concentration of sample

The molar extinction coefficient was calculated from the primary structure using the ProtParam calculator on the Expasy molecular biology server (http://us.expasy.org/tools/protparam.html)

2.1.26 Concentration proteins

Protein solutions were concentrated after filtering (0.2 µm) using 20 ml or 2 ml Vivaspin centrifugal concentrators (VivaScience) with 5, 10, or 30 kDa molecular weight cut off filters as appropriate. Centrifugation was performed at 3,500 g using a MSE Mistral 3000i bench centrifuge with a swing out rotor.

2.2 Crystallography

In order to solve the crystal structure of the CBM60 in complex with different ligands, molecular replacement was used. For the last technique, both versions of CBM60 (with and without His_6 tag) were prepared and purified as follows.

2.2.1 Protein sample preparation

2.2.1.1 Wild type with His_6 tag

Proteins desired for crystallization were required to be highly homogenous; >95 % homogeneity as assessed by SDS-PAGE. Protein samples were buffer exchanged into 10 mM Tris/HCl buffer, 5 mM Ca^{2+}, pH 8.0; and further purified by IEC (Section 2.1.24.3). Proteins were concentrated and extensively dialysed in sterile MQ H_2O. Protein was then diluted and ligands added to the required concentration for screening. Protein or protein/ligand mixes were spun down at 13,000 g for 3 min to remove aggregated protein.

2.2.1.2 Wild type without His_6 tag version

In order to express CBM60 without tag version, *E. coli* strain was transformed with plasmid DNA encoding CBM60 with a stop codon prior to the sequence encoding the His_6 tag, and this protein was over-expressed using the same condition used to produce the His-tagged version of the protein.

An aliquot of 12.6 ml of CFE of *E. coli* strains expressing CBM60 with no His_6 tag was mixed with 1 g of finely-sonicated insoluble oat spelt xylan, and incubated in an orbital shaker at 4 ºC overnight. After washing of the polysaccharide three times with 20 ml of 10 mM Tris/HCl buffer, pH 8.0, it was mixed with 5 ml of 100 % glycol ethylene for 1 hour at 4 ºC to elute bound protein. The xylan was removed by centrifugation at 13,000 rpm for 3 mins. The glycol ethylene containing bound protein was dialysed against 3 x 1000 vol of 10 mM Tris/HCl buffer, pH 8.0, prior to anion exchange chromatography. Protein eluted from the polysacharides was subjected to SDS-PAGE using a 10 % (w/v) polysacrylamide gel prior to functional assays.

2.2.2 Protein crystallization screen

Commercially available and other recommended crystallization screens were used to obtain preliminary crystallization conditions for proteins. JCSG+ Suite and PACT Suite screens were from Qiagen. A screen based on JCSG+ Suite screen was obtained from Dr Rick Lewis (conditions listed Appendix B). Crystal screens were carried out

using a protein concentration of 10 mg/ml with a varying ratio of protein to crystallography reagent (2 µl or 1 µl protein: 1 µl reagent) and in the presence of appropriate ligand typically at 10 mM (co-crystallisation). Conditions that produced the best possible crystals (suitable size and form) were further optimised. Typically, the concentration of salt and precipitant were varied against a range of pH, around the condition determined from the initial screening.

2.2.3 Growth of crystals

For the most part, crystallography of proteins was carried out using the hanging drop vapour diffusion technique by hand. However, in order to screen more conditions, sitting drop vapour diffusion was used with a robot.

2.2.3.1 Hanging drop vapour diffusion

The hanging drop technique for crystal growth requires that a small aliquot of protein or protein/ligand mix (0.7-2 µl) be mixed with an amount of crystallizing mother liquor (0.7-2 µl) on a siliconized cover slip. Cover slips (18 x 18 mm, Scientific laboratory supplies) were pretreated with Aqua SilTM, dried according to manufacturer's instructions and polished with a silk scarf prior to dispensing protein or crystallization liquor. Cellstar, 25-well tissue culture, polystryrene, non-pyrogenic, DNase and RNase free plates (Greiner, Bio-one)were prepared by applying High Vacuum Grease (dow Corning) to the rim of each well using a 10 ml syringe with a 200 µl Gilson tip, trimmed to allow flow of the grease. Typically 0.5 ml of mother liquor was pipette into each well, although different well volumes could be used for alteration of diffusion rate. Once cover slips were setup, as previously described, they were immediately inverted and sealed above the corresponding mother liquor by firmly pressing down the cover slip against the vacuum grease as in Fig 2.3. Crystallization trays were incubated at 20 °C in a synyo MIR-153 incubator. Drops were viewed, using a Leica MZ-6 crystallization microscope, immediately following completion of the tray, the following morning and every day for the next week to look for appearance of crystals or conditions that might lead to crystal. After the first week trays were assessed periodically until the mother liquor in the wells appeared to have dried up.

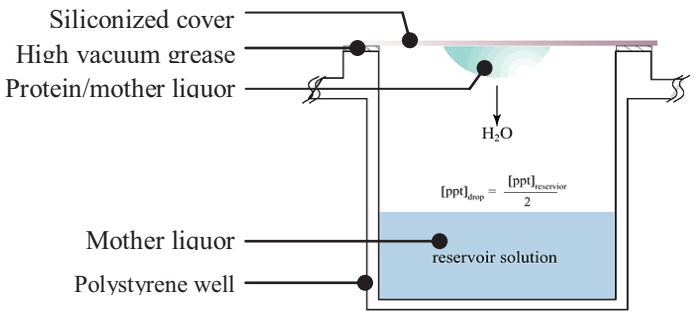


Fig 2.7 Set-up of hanging drop vapour diffusion

Diffusion due to loss of water from hanging drop in vapour phase causes equilibrium to occur between drop and reservoir. Super saturation of protein occurs as water diffuses away from drop (www.hamptonresearch.com).

2.2.3.2 Sitting drop vapour diffusion

Crystallization was conducted using a mosquito™ (TTP Labtech) nanolitre pipetting robot with 96-well crystallization plates (Greiner). Crystallization mother liquor (100 µl) was dispensed into each well of the plates, and 7 µl of protein was dispensed into each of the 4 sample reservoirs before 100 nl was was aliquoted by the robot onto each of the crystallization shelves (Fig 2.4). Mother liquor (100 µl) was dispensed onto each crystallization shelf. The tray was then sealed using sealing film with contact adhesive and crystal conditions were then viewed as described before. Drops were viewed using a Leica MZ-6 crystallization microscope.

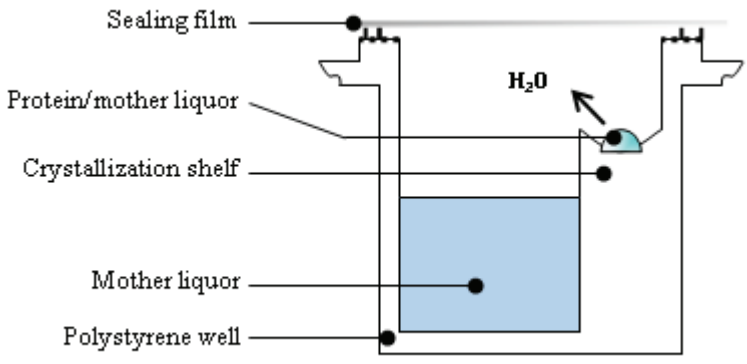


Fig 2.8 Set-up of sitting drop vapour diffusion

Similar in principle to hanging drop vapour diffusion but droplet is mounted on shelf within crystallization compartment rather than on cover slip above compartment (www.hamptonresearch.com).

2.2.4 Visualisation of structures

Analysis of crystals structures with ligands were carried out using PyMol (Delano Scientific), which was also used to prepare figures of molecular structures.

2.3 Biochemistry

2.3.1 Fluorescence Spectroscopy

Fluorescence spectroscopy was carried out using a Cary Eclipse Varian spectrofluorimeter. The excitation wavelength was 280 nm and a slit width of 5 nm was used for excitation and emission, while emission spectra were recorded from 290 nm to 450 nm. The temperature was set to 20 °C and the measurements were carried out in 5 mm x 5 mm quartz cuvettes. The stock concentration of protein was 1 mM in 50 mM Hepes buffer, pH 8.0, containing 5 mM $CaCl_2$. Protein (250 µl) and different concentration of ligands (250 µl) were mixed, five scans were averaged for each concentration of ligands. Samples were corrected by subtraction from a buffer blank. When a shift of the wavelength of maximum emission Emax was observed, the varycentric mean wavelength of the integral between 300 and 430 nm was determined according to the following equation:

$$\lambda m = [\sum F(\lambda) x \lambda] / [\sum F(\lambda)]$$

Where λm=barycentric mean wavelength, F (λ) is the fluorescence at wavelength λ. Measurements were taken every 2 nm (Lakey et al., 1989). Titrations were carried out with increasing concentrations of ligand, until a plateau was reached. The barycentric mean wavelength at the plateau was considered as 100 % binding, and the binding ratio at the other ligand concentrations was determined. The dissociation constant Kd was extracted from the curve according to the one site binding (hyperbola) equation:

$$Y = Bmax \times X / (k_D + X)$$

Bmax is the maximal binding and k_D is the concentration of ligand required to reach half-maximal binding (GraphPad Prism 4.02). Finally, the association constant k_A was calcultated as $k_A = 1/k_D$. X is the ligand concentration in nM or pM and Y is the total binding in sites/cell.

2.3.2 Isothermal Titration Calorimetry (ITC)

The thermodynamic and binding affinities of carbohydrate protein interaction were investigated by ITC using a MicroCal™ VP-Isothermal Titration Calorimeter. ITC measurements were made at 25 ºC following standard procedures. The proteins were dialyzed extensively against buffer and the ligands were dissolved in the dialysis buffer to minimize heats of dilution. Degassed protein solution (~1.5 ml) at a high concentration of 80-200 μM was equilibrated in a reaction cell maintained at 25 ºC. To this solution, 28 aliquots (10 μl) of ligands at an appropriate concentration were automatically injected, with rapid stirring (307 rpm), at 300 s intervals. Typically, oligosaccharides ligands were titrated at ≈ 5 mM and soluble polysaccharide ligands at 1- 10 mg/ml. Binding was monitored by measuring heat released (exothermic binding) or absorbed (endothermic binding) for each aliquot of ligand. During the titration, the difference in electrical power required to maintain the temperature of the reaction cell versus the temperature of the reference cell was recorded, and from these differences the heat change on binding calculated.

The molar concentration of binding sites present in polysaccharide ligands was determined following the method of Szabo, et al., (2001). After determining that the test protein contained a single binding site, by titrations with oligosaccharides of known molarities, the molar concentration of polysaccharide ligands was fitted in an iterative fashion until the n-value was as close as possible to 1. Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression using a single site-binding module (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:

$$-RT\ln k_A = \Delta G = \Delta G - T\Delta S$$

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

2.3.3 Affinity Gel Eletrophoresis (AGE)

The affinity eletrophoresis method using native polyacrylamide gels containing soluble polysaccharide ligands was based on the method of Tomme et al (1996). A continuous gel system was used with the same gel apparatus (ATTO Corporation) as used for SDS-PAGE (Section 2.1.23). Gels contained 7.5 % (w/v) acrylamide (Acrylogel 3; BDH Electran™) in 25 mM Tris, 250 mM glycine buffer, pH 8.3, which comprised the Running and Sample buffers. For ligand-containing gels, appropriate polysaccharides were added at 0.001-0.1 % final concentration prior to polymerisation. Pure proteins (10 μg) in 7 μl of loading buffer, containing 5 % (v/v) glycerol and 0.0025 % Bromophenol Blue final concentration, were electrophoresed at 10 mA/gel at room temperature for 2.5 h. Proteins were then stained in 0.4 % (w/v) Coomassie Blue, 40 % (v/v) methanol, 10 % glacial acetic acid for ≈ 1 h and then destained in 40 % (v/v) methanol, 10 % glacial acetic acid. Gels with and without ligands were run in the same gel box with identical samples loaded on each. BSA (15 μg) was used as a negative, non-interacting control. By determining the relative mobility of the protein in the presence and absence of ligand versus BSA, over a range of ligand concentrations, the dissociation constant k_D of the protein for its ligand could be calculated according to the method of Takeo et al (1984). A double reciprocal plot of relative mobility (1/R-r) against ligand concentration 1/[L] was plotted, and the intercept of the linear regression analysis line of the data, with the x-axis, yields 1/ k_D. R= distance migrated by test protein divided by the distance moved by the control non-binding protein （BSA）in the absence of ligand; r=distance moved by the test protein divided by the distance moved by the control non-binding protein in the presence of ligand.

2.3.4 Thin Layer Chromatography (TLC)

TLC plates (Silicagel 60, 20 x 20 cm^2, Merck) were cut to the desired size (the minimum height was 10 cm). Between 2 and 4 μl of digested potato pectic galactan samples (Chapter 5) were spotted on the plate, separated by 7 mm. Solvent (50 ml) comprising freshly made1-butanol/acetic acid/water (2:1:1, v/v) was poured into a glass chromatography tank (23 x 23 x 7.5 cm^3) and covered tightly. Vapours were allowed to equilibrate for at least 2 h before use. The TLC plate was placed into the tank and samples allowed to migrate until the running buffer reached ≈ 1 cm to the top of the

plate. The plate was dried carefully using a hairdryer and put back in the tank for another 1 h. The plate was dried again and immersed for a few seconds in an orcinol sulphuric acid reagent [sulphuric acid/ethanol/water 3:70:20 v/v (orcinol, 1 ‰)], dried carefully and heated until sugars were revealed, at 120 ºC (5-10 min). Plates were recorded using a digital camera (Canon PowerShoot A75). To estimate the size of the sugars, different size marker standards were run in parallel, depending on the nature of the samples. 'Mannose' standards consist of mannose, mannobiose, mannotriose mannotetrose, mannopentose and mannohexahose (Sigma, Megazyme) were at a final concentration of 10 mM and 1.5 µl was spotted on the TLC plate. In order to detect the presence of sugar in a sample quickly, 2 µl were spotted on a section of TLC plate and immersed for a few second in orcinol sulphuric acid reagent, dried carefully and heated until sugars were revealed, as previously described.

2.3.5 Microarray assays on nitrocellulose sheets

Test polysaccharides of 10 mg/ml were applied as 2 µl aliquots to untreated nitrocellulose sheets (Scheicher & Schuell, Germany) in a 5-fold dilution series. Sheets were left to dry at room temperature for at least 30 min prior to blocking for 1 h with 5 % (w/v) milk protein (MP, "Marvel", Premier Beverage, UK) in Tris-buffered saline (TBS; 10 mM Tris/HCl, pH 7.4 containing 150 mM NaCl). The nitrocellulose sheets were incubated with 1.25 µg/ml of the appropriate carbohydrate-binding module (CBM) in 5 % (v/v) MP/TBS for at least 1 h. The nitrocellulose sheets were then washed extensively under running water for at least 2 min, incubated in water for 5 min while rocking, and washed for a further 2 min under running water, prior to incubation with a 5 ml of 1000-fold dilution of anti-his horseradish peroxidase (anti-his HRP) conjugate (Sigma, UK) in MP/TBS for 1 h. After washing in water as above, sheets were incubated in freshly prepared HRP substrate to detect CBM binding (25 ml of deionised water, 5 ml MeOH containing 10 mg/ml 4-chloro-1-naphthol and 20 µl 30 % (v/v) H_2O_2). The reaction was stopped by washing the nitrocellulose sheets with water and pictures were taken.

2.3.6 Anti His-Horse Radish Peroxidase (HRP) assay

Grass (G) had been boiled for 20 min. Separate fractions of the boiled grass were soaked with 100 % acetone (GA) and 0.5 % biosurfactant SLES solution overnight (GB), respectively. They were washed with 10 mM Tris/HCl buffer, pH 8.0, three times before being arrayed on nitrocellulose membranes. CBM at 0.3 µg/ml final concentration was applied to the nitrocellulose membrane, as it was determined as the optimal concentration to use in the presence of Marvel. The grass and pure ligands arrayed on the membrane were incubated with CBM for 1h at 4 ºC in the dark room. The membrane was washed with 10 mM Tris/HCl buffer, pH 8.0 to take away unbound proteins. Anti-His HRP was used at 10^{-4} dilution of the stock and added to the membrane for 1 h at 4 ºC in the dark room. The membrane was washed again, with 10 mM Tris/HCl buffer, pH 8.0, before the substrate for HRP from the ECL+ kit was applied. The membrane and a piece of X-ray film were overlaid between two X-rays cassettes for 3~10 min. The signals were displayed on the X-ray film exposed to the nitrocellulose membrane. It identified where CBM binding occurred.

2.3.7 Fluorescence analysis of the binding of protein to stained cotton

Equal molar concentration of free dye 488 and CBM-488 (as the positive control in this case) were applied to grass stains on cotton in 10 mM Tris/HCl buffer, pH 8.0, either in the presence and or absence of 1 g/L SLES; 0.2 mM $CaCl_2$ and 5 mM NaCl. The samples were incubated for 1 h on an orbital shaker in the dark room at 4 °C. The samples were then washed using 10 mM Tris/HCl buffer, pH 8.0 and assessed microscopically. With confocal fluorescent microscopy, the fluorescent signals emitted in regions from 500 to 550 nm and from 660 to 710 nm were detected simultaneously, as the chlorophyll in grass cells emits in the range 500 to 550 nm and the CBMs studied here were labelled by Alexa dye 488 (488) which emit in the region from 660 to 710 nm. With fluorescence microscope, images of CBM binding to cotton were captured on an Olympus DP50 fluorescence microscope as Alexafluor-488 fluorescence emission at x 40 magnification and using the FITC filter. Fluorescence intensity (representing CBM binding to cotton) was analysed using analysis® 3.1 software. After conversion to greyscale images, the software assigned values to each pixel (8 bits per pixel) of between 0 (black) and 255 (white) and calculated total fluorescence intensity as the integral of all values in the field.

2.3.8 Enzyme assays

Unless otherwise stated, all enzyme assays were performed at, and all reagents and cuvettes pre-warmed to, 30 ºC. All assays were repeated at least three times. The same set of Gilson pipettes was used throughout each assay. Graphs were plotted in GraphPad Prism 4.02 which was also used to calculate slopes, gradients and standard errors. The non-linear 'one-site' binding model was used to fit kinetic data to estimate k_M and k_{cat}.

2.3.8.1 4-Nitrophenyl-β-D-xylobiose (PNP-X_2) assay

A 500 µM PNP-X_2 (Invitrogen) stock was freshly prepared in water and kept on ice until used. In a 1.5 ml eppendorf tube, the enzymatic assays were carried out in 50 mM sodium phosphate buffer pH 7.0 as follows:

PNP-X_2 (final concentration at 50 µM)	50 µl
Buffer	400 µl

Solutions were warmed up to 30 ºC in a water bath, before the addition of xylanases was added as follows:

Enzymes (stock concentration at 10, 5, 1, 0.5 and 0.2 µM)	50 µl
Total volume	500µl

Eppendorf tubes were vortexed and 450 µl of each reaction volumes (prior to the enzyme addition) were transferred to glass curvettes in a Pharmacia Biotech Ultrospec 4000, UV/vis spectrophotometer at 30 ºC. The reactions were incubated for 1 to 2 min, before 50 µl of the enzymes at appropriate concentration were added to each curvette to initiate the reaction. The release of 4-nitrophenyl (PNP) from the substrate was measured at an absorbance of 420 nm, and the rate of absorbance change was calculated. The reaction was stopped after 15~20 min. A non-enzyme control and a control only containing substrate were also set up in each round. The extinction coefficient of PNP is 15,245 $M^{-1}cm^{-1}$).

2.3.8.2 3, 5-Dinitrosalicyclic acid assay (DNSA).

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

2.3.9 Polysaccharide complete acid hydrolysis

To obtain galactobiose, galactotriose for crystallography studies of CBM60 with different ligands, potato pectic galactan sample (10 mg/ml in water) was incubated with 10 M HCl (final concentration) at 37 ºC and 52 ºC for 8 ~12 h and neutralised by addition of 10 M sodium hydroxide. Samples were centrifuged for 3 min (13,000 g) and a 2 µl sample was spotted on TLC (Section 2.3.4). The constituent sugars were determined by comparison with standard manno-sugars and galactose.

2.3.10 Enzymatic digestion of potato pectic galactan

Potato pectic galactan (Sigma) was dissolved at 10 mg/ml and solubilised in cell free extract containing over-expressed galactanase A (Section 2.1.21) at room temperature and 37 ºC. To follow the reaction and to ensure the galactan was digested to completion, aliquots of 50 µl were removed at 0 min, 30 min, 60 min, 1 h, 2 h, 4 h, 6 h and 8 h and were boiled for 5 min and centrifuged 3 min at 13,000 g to remove any protein aggregates or undigested galactan. Aliquots of 2 µl sample were spotted on TLC (section 2.3.4). The constituent sugars were determined by comparison with standard manno-sugars and galactose.

2.3.11 Oligosaccharide size exclusion chromatography

Appropriately digested pectic galactan products (Section 2.3.10) were separated by size exclusion chromatography using P2 Bio-Gel (Bio-Rad) matrix packed in 3 Glass

Econo-Column™ (2.5 cm x 80.0 cm) with a flow adaptor (Bio-Rad) at 0.7 ml/min in degassed MQ H_2O. Fully digested and freeze dried galactan (Section 2.3.13) was dissolved in 8 ml MQ H_2O to a final concentration of 0.72 g/ml and loaded directly onto the first columns, which was then run with MQ H_2O as mobile phase at 0.5 ml/min using a peristaltic pump(LKB Bromma 2132 microperpex ™). The 5ml fractions were collected continuously between 20 h and 39 h after loading using a Bio-Rad model 2110 fration collector. A 2 µl aliquot of every three fraction was analysed by TLC (Section 2.3.4) identifying fractions of interest to be pooled and freeze dried for further analysis.

2.3.12 Preparation of polysaccharide substrates and ligands

2.3.12.1 Soluble oat-spelt xylan

The soluble fraction of oat-spelt xylan was prepared as follow. Oat-spelt xylan (Sigma, 40 g) was resuspended in 400 ml of MQ H_2O, the pH of this suspension was adjusted to 10.0 by addition of 1 M sodium hydroxide (≈ 1 ml) and stirred at room temperature for 1 h. The mixture was then neutralised by addition 1 M acetic acid and centrifuged at 10,000 g for 10 min to pellet the insoluble xylan. The supernatant was removed and centrifugation was repeated to remove residual insoluble material, before being poured into shallow trays and frozen at -80 ºC (Ghangas, et al., 1989). This solution was then lyophilised (Section 2.3.13) and the resulting solid comprised the soluble oat-spelt xylan which was stored desiccated at room temperature.

2.3.12.2 Other polysaccharides

All the monosaccharides, oligosaccharides and polysaccharides used in this study were dissolved at various concentrations in water or in ITC buffer, except when digested or used for analysis. The sugars were purchased from Sigma, Megazyme or Seikagaku Corporation.

2.3.13 Concentrating purified sugars or protein by freeze drying

Purified oligosaccharide and polysaccharide sugars or proteins were frozen to -80 ºC and then lyophilised in a Christ Alpha 1-2 Freeze Drier at -60 ºC.

Chapter 3. Developing novel recombinant enzymes for the laundry industry

3.1 Introduction

3.1.1 Glycoside hydrolases (GH), Carbohydrate binding modules (CBM) and their applications in the laundry industry

Polysaccharide-degrading glycoside hydrolases are generally modular enzymes. They often contain one, or multiple, non-catalytic carbohydrate binding modules (CBMs) appended to the catalytic domain by a flexible linker sequence. CBMs promote enzyme access to insoluble polysaccharides by bringing glycoside hydrolases and the insoluble highly recalcitrant substrate into close proximity (Bolam, et al., 1998), which results in enhanced catalysis.

This project focuses on the use of novel CBM-containing enzymes to remove grass stains arrayed on cotton fabrics. Grass cell walls are highly complex macromolecular structures that consist of repertoires of polysaccharides that interact with each other through extensive networks (O'Neill, et al., 2003). The intimate association of polysaccharides within cell walls limits the access of hydrolytic enzymes that attack these composite structures, and thus the degradative process is relatively slow (Hall, et al., 1995). Xylan, which accounts for up to 30 % of plant biomass is one of the most abundant polysaccharides embedded in plant cell walls. Hence, to reduce the accessibility problem, and thus increase the capacity of all the enzymes to degrade grass stains in the laundry industry, xylan was selected as the target ligand for the CBMs. Although CBMs are grouped into sequence-based families and classified into 'types' A, B and C based on the topology of the binding site (Section 1.3.2), one of the intriguing features of CBMs is the diversity of families that contain members which apparently display the same ligand specificity against purified polysaccharides. CBMs from families 2b, 4, 6, 15 and 22 all contain protein modules that recognize the xylan backbone (Boraston, et al., 2004). Hence, CBMs located in three of these families (Table 3.1), CBM2b, 15 and 22, which display specificity for xylans, either purified or

embedded in the plant cell walls, as described in McCartney, et al., (2006) were selected as potential targeting modules for grass stains that contain xylan. It has been hypothesized that by fusing these modules to enzymes and the catalytic module can be targeted to its substrate in grass stains, leading to more efficient removal of the stain.

3.1.2 Cell wall nature of cotton fibers

Cotton fibers consist almost exclusively of cellulose, a β-1,4-linked polymer of glucose (Fig 3.1). Cotton textiles are manufactured in a process involved the following stages: cultivating and harvesting, preparatory process, spinning, weaving and finishing. They are made of cotton fibers, which is a cellulose based material. There are two types: pure cellulose and modified cellulose such as cellulose acetates. In this study, pure cellulose-based cotton was used to array the grass stains. The fiber is hollow and, under a microscope looks like a twisted ribbon. This chapter assesses the capacity of selected xylan targeting CBMs to bind to grass stains arrayed on different surfaces, with cotton being the most important within the context of the laundry industry.

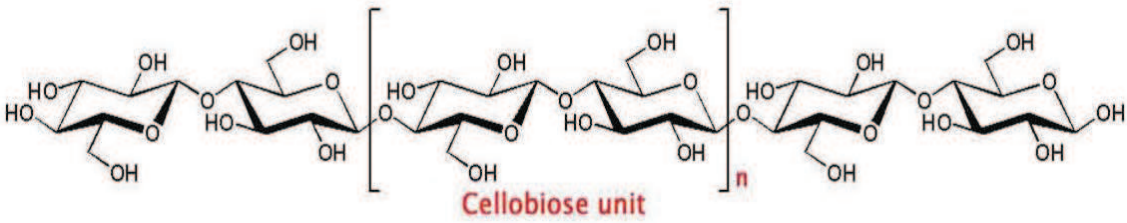


Fig 3.1 Structure of cellulose in cotton fibers (www.indiantextilejournal.com)

3.2 Results

Protein over-expression and solubility

The *Cj*CBM15, *Cf*CBM2b1,2 and *Ct*CBM22 are derived from *Cellvibrio japonicus* xylanase 10C, *Cellulomonas fimi* xynlanase 11A and *Clostridium thermocellum* xylanase 10B. The DNA encoding sequences were encoded in pET22b, pET22b, and pET21a vectors, respectively. In the following result chapters, *Cj*CBM15, *Cf*CBM2b1,2 and *Ct*CBM22 are referred as CBM15, CBM2b1,2 and CBM22.

CBM15 and CBM22 were expressed in the *E. coli* strain BL21 as described in Section 2.1.21. The recombinant strain was cultured at 37 °C and expression was induced by the addition of 1 mM IPTG and incubation for 4 h at 37 °C. CBM2b1,2 was expressed in the *E. coli* strain Tuner as described in Section 2.1.21. The recombinant strain was cultured at 37 °C and expression was induced by the addition of 0.2 mM IPTG and incubation overnight at 16 °C. As each CBM contains a *C*-terminal His_6 tag supplied by the vector pET, the proteins were purified by immobilized metal affinity chromatography (IMAC) as described in Section 2.1.24. Recombinant forms of CBM15, CBM22 and CBM2b1,2 were expressed at high yield after induction with the lactose analogue IPTG. CBM15 and CBM22 were produced in soluble form but CBM2b1,2 formed insoluble inclusion bodies.

3.2.1 Protein purification of soluble CBMs

IMAC of the soluble CBMs was assessed by SDS-PAGE (Fig 3.2). During the purification a proportion of the CBMs did not bind to the Talon columns as the recombinant proteins were evident in the flow through fractions. This may be due to the matrix being saturated with His_6-tagged target proteins or; more likely, a proportion of the CBMs were misfold, and the His tag was not available to bind to the column. It is also possible that the His tag had been cleaved from some of the CBM modules and hence these products would not bind to the Talon column. However, there was not a great amount of loss of target CBM in the flow through. Generally, IMAC generated proteins with a high degree of purity. There were two bands appeared in CBM15 purification gel, this could be due to that there were some degradation products of CBM15 during the purification process.

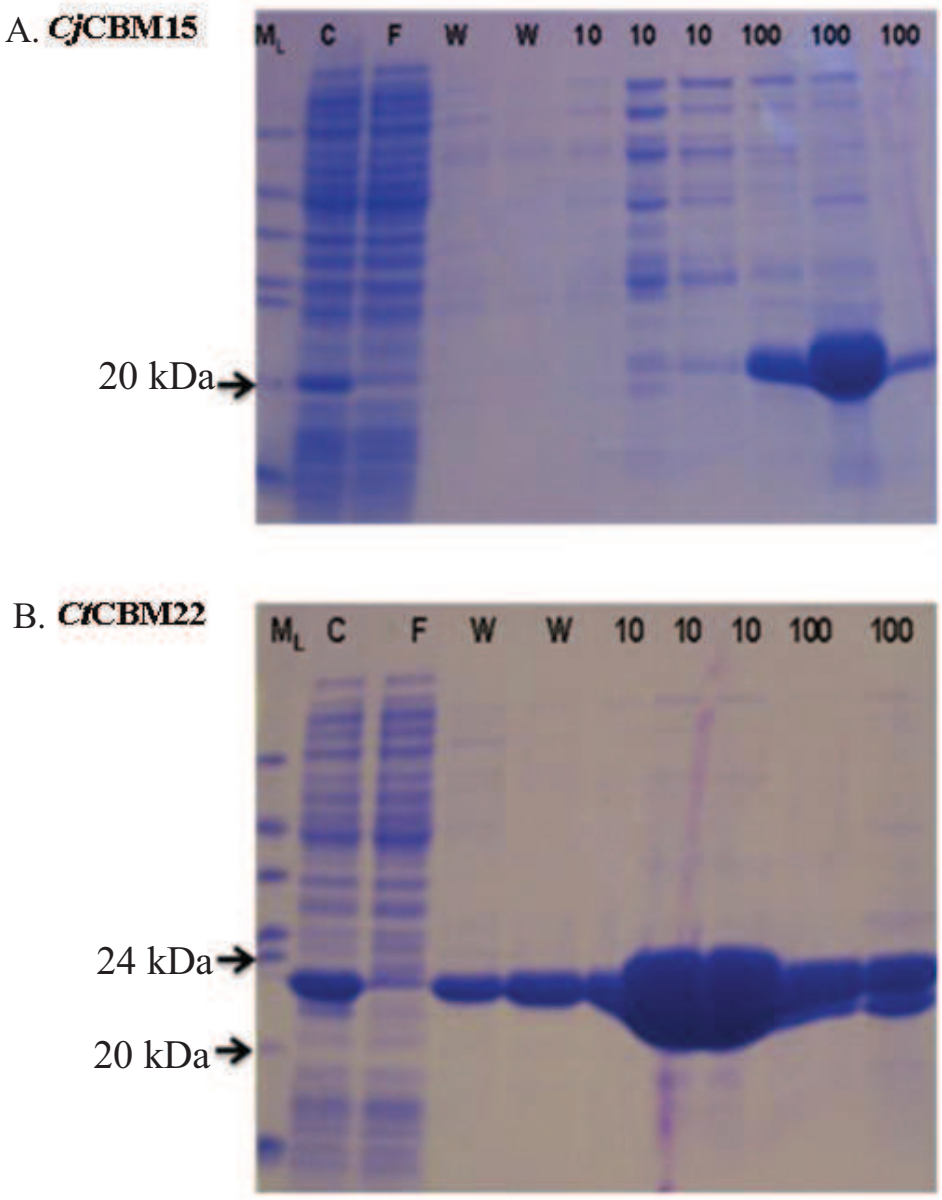


Fig 3.2 SDS-PAGE of CBM15 and CBM22 purified by IMAC

For both panels, SDS-PAGE was carried out using a 12.5 % polyacrylamide gel. Panel A and B display the purification of CBM15 and CBM22, respectively. The lanes contained 5 µl aliquot of cell free extract and flow-through (Lanes C and F respectively), 10 µl of wash with 20 mM Tris/HCl pH 8.0, containing 300 mM NaCl (W), 10 µl of elution with 10 mM imidazole (10) and 100 mM imidazole (100). Lanes M_L contained Sigma low molecular weight standards.

3.2.2 Protein purification of insoluble CBM2b1,2

The insoluble protein fraction containing of CBM2b1,2 was resuspended in 1x Talon buffer (Section 2.1.24) with 8 M urea and incubated at 37 °C overnight. Most of the inclusion body was dissolved by this procedure. Redissolved pellet was applied to the Talon resin, and all buffers contained 8 M urea, to retain CBM2b1,2 in its soluble form. CBM2b1,2 was eluted in Talon elution buffer with 100 mM imidazole. Although poorly expressed, a reasonable amount of the CBM, however, was purified from the inclusion body (Fig 3.3).

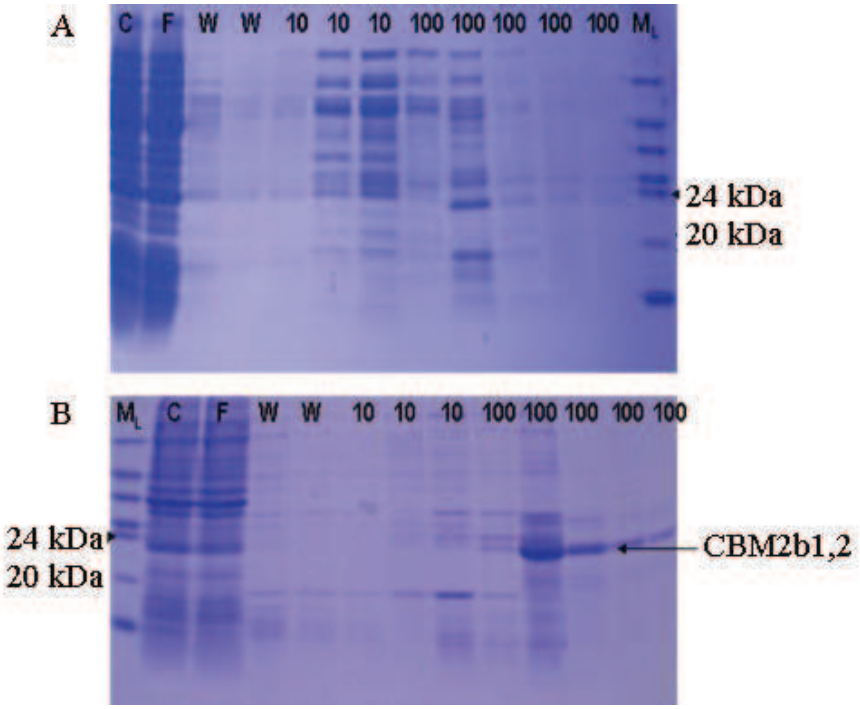


Fig 3.3 SDS-PAGE of CBM2b1,2 purified by IMAC in 8 M urea

SDS-PAGE was carried out using a 12.5 % polyacrylamide gel, as described in Section 2.1.24. Panels A and B display the purification of *Cf*CBM2b1,2 in soluble form and from insoluble inclusion bodies, respectively. The lanes contained 5 µl aliquot of cell free extract and flow-through (Lanes C and F respectively), 10 µl of wash with 20 mM Tris/HCl pH 8.0, containing 300 mM NaCl (W), 10 µl of elution with 10 mM imidazole (10) and 100 mM imidazole (100). Lanes M_L contained Sigma low molecular weight standards.

CBM2b1,2 was refolded by decreasing the urea concentration from 8 M to 4 M, 2 M, 1 M, 0.5 M, 0.25 M, 0.125 M to 0 M step-wise. Minor precipitation occurred during the refolding process. The refolded CBM2b1,2 was subjected to subsequent binding assays to confirm that it was functional. There was no significant precipitation of the other CBMs during dialysis.

3.2.3 Functional assay and binding specificity studies

3.2.3.1 Results from affinity gel electrophoresis (AGE)

AGE was carried out to see if the CBMs produced were functional. CBM15 and CBM22 were previously shown to bind to soluble oat spelt xylan (OSX) (Charnock, et. al., 2000; Szabo, et. al., 2001), therefore a non-denaturing gel containing 0.1 % soluble OSX was loaded with 10 μg of each CBM, in parallel with a gel containing no ligand. BSA was also loaded onto the gel to act as a negative control, as it does not bind to the polysaccharide ligands. Electrophoretic migration of the BSA derived bands were the same on gels containing ligand or no ligand. In contrast, the CBM15 and CBM22 modules migrated much slower on native gels containing OSX compared to gels containing no polysaccharide. These data demonstrate that both CBMs bound to xylan (Fig 3.4). Since the pH of the running buffer in AGE is 8.3, which is very close to the isoelectric point (pI) of CBM2b1,2 (pI at 8.4), the protein will have no net charge and thus fail to migrate in AGE experiments. To test if CBM2b1,2 is functional, pull down experiments deploying insoluble xylan was used instead of AGE.

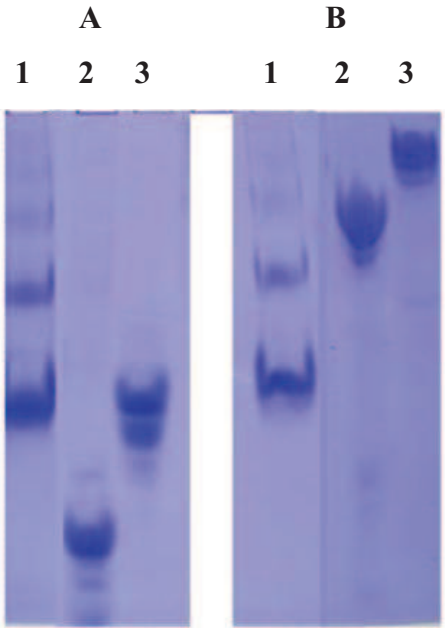
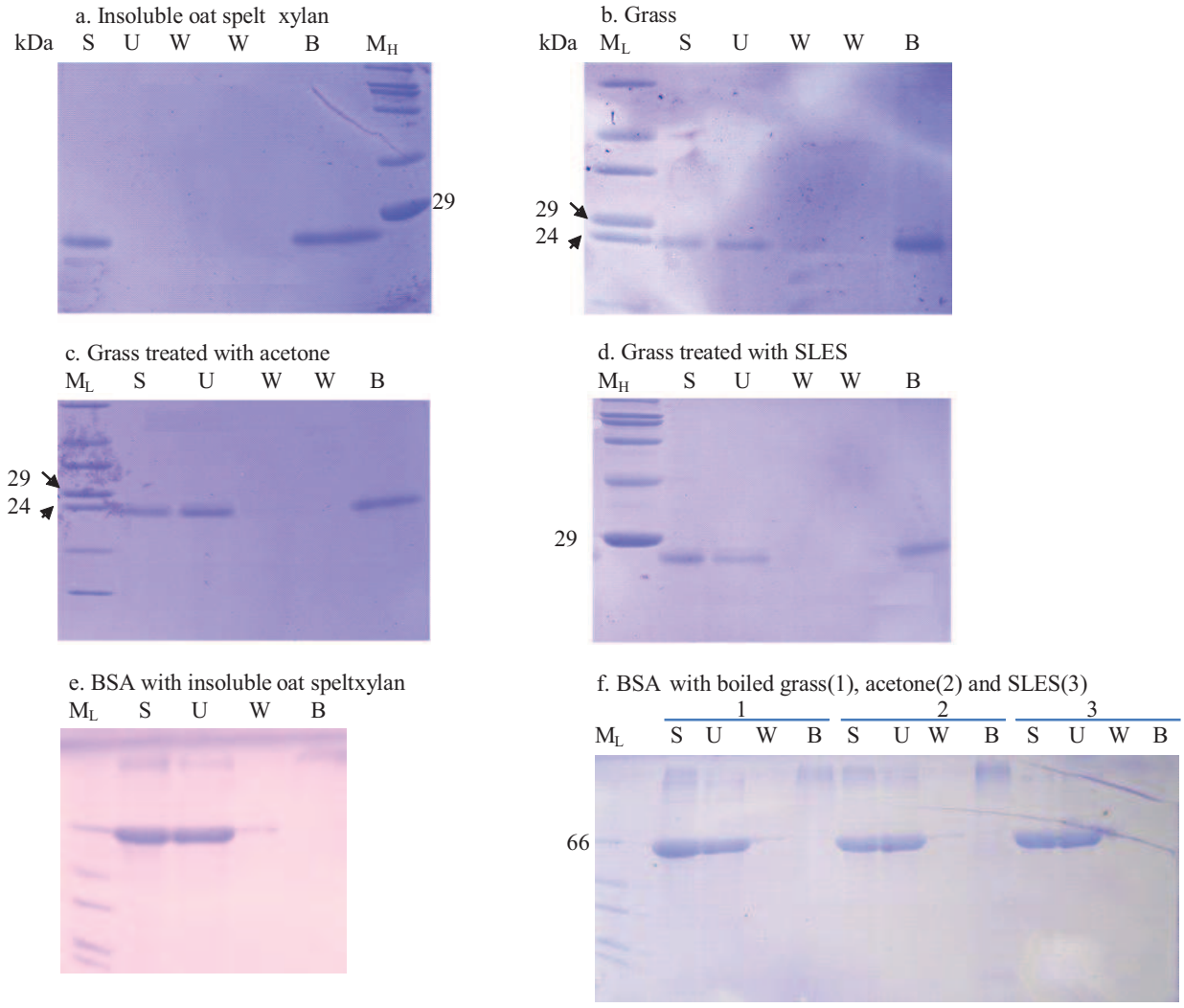


Fig 3.4 Non-denaturing affinity gel electrophoresis of CBM15 and CBM22

Non-denaturing affinity gel electrophoresis (AGE) was carried out, using 12.5 % (w/v) polyacrylamide gels in a Tris/glycine buffer system, pH 8.3. The gels were run for 1.5 h at 20 mA, and were loaded with BSA (lane 1), which was the negative control, CBM15 (lane 2) and CBM22 (lane 3). Electrophoresis was in the presence (Panel B) and absence (Panel A) of 0.1 % oat spelt xylan. A 10 μg aliquot of protein was loaded into each well.

3.2.3.2 Pull down assay

Ligand (~ 2.5 mg insoluble oat spelt xylan or boiled grass) was mixed with CBM2b1,2 (~ 40 μM) and material in the bound and unbound fractions were analyzed by SDS-PAGE. The data (Fig 3.5) showed that almost all of the CBM2b1, 2 bound to insoluble OSX. A significant proportion of the CBM also bound to grass suggesting that the hemicellulosic polysaccharide within the plant cell wall is accessible to the protein. Binding appeared to be specific as the BSA (negative control) did not interact with the insoluble polysaccharides, and there was no precipitation of the CBM, which would have given a false positive result.



The data also showed that CBM2b1,2 bound to grass dewaxed with acetone or the surfactant SLES, although binding was not as strong as observed with OSX. Thus, the biosurfactant did not adversely affect the activity of CBM2b1,2, although it did not appear to enhance access of the protein modules to the hemicellulose embedded in the cell wall.

3.2.4 Binding assessment using immulogical methodologies

After the xylan specificity of all the CBMs was confirmed, the binding capacity of CBMs to grass stains arrayed on nitrocellulose and cotton was initially assessed using immunological methodologies that focused on the detection of the His_6 tag of the CBMs. Anti-His antibody conjugated to horse radish peroxidase (HRP) was used, and the substrate (present in the ECL+ kit) for the peroxidase generated a chemoluminescent product which could be detected at high sensitivity on X-ray films. (Section 2.3.6).

3.2.4.1 Assessment of boiling grass

Grass, which was initially arrayed on nitrocellulose membranes, reacted with the substrate from the ECL+ kit, even when the CBM and the anti-His HRP antibody were both absent; a dark spot was observed (Fig 3.6B). This suggested that the grass itself contained peroxidases that reacted with the ECL substrate. This view was supported by the observation that boiled grass did not give any signal with the ECL substrate in the absence of both CBM and antibody (Fig 3.6A). Thus boiling the grass inactivated the endogenous grass peroxidases.

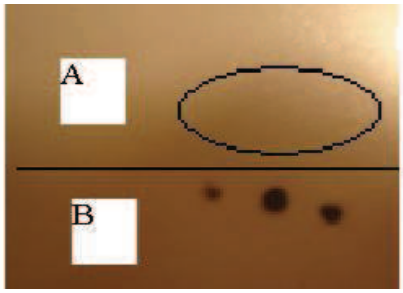


Fig 3.6 Assay of peroxidase activity in grass arrayed on nitrocellulose membranes

Boiled grass and untreated grass were arrayed on membranes in triplicate, after which the chemoluminescent substrate (ECL+ kit) was added and the membrane subjected to autoradiography. In Panel A, boiled grass displays no autoradiographic signal. Panel B displays an autoradiographic signal generated by grass that had not been boiled.

3.2.4.2 Assessment of Marvel milk powder as a blocking reagent

Marvel prevents non-specific binding of the antibody as the milk powder prevents general background of antibody to the membrane. When marvel was absent, there was high background due to non-specific binding of antibody to the surface; the fluorescent signal was therefore always observed (Fig 3.7A). When 5 % Marvel milk powder (MP) was added to the Tris buffer, the background was almost eliminated (Fig 3.7B); Thus, Marvel-Tris buffer containing 5 % (w/v) skim milk powder were used to throughout the following experiments.

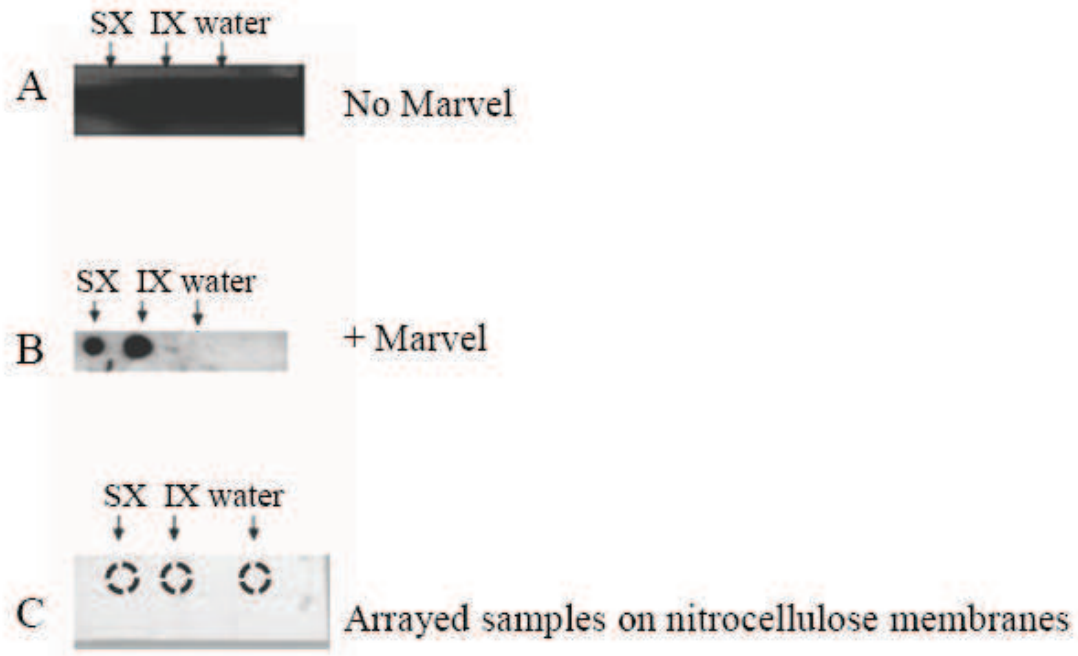


Fig 3.7 Assay of CBM2b1,2 binding to xylan arrayed on nitrocellulose membranes, in the presence and absence of Marvel

The experiment was carried out at 4 ºC, as described in Section 2.3.6. CBM was applied at 0.3 µg/ml, which was determined as the optimal final concentration to use on nitrocellulose filters. Anti-His HRP was used at 10^{-4} dilution of the stock, before the substrate from the ECL+ kit was applied. The membrane and X-ray film were placed between two X-rays cassettes for 10 min, and the signals identified where CBM binding occurred. Panel A and Panel B display the X-ray film exposed to the nitrocellulose filters in the absence and presence of Marvel, respectively. Panel C is the nitrocellulose membrane containing soluble OSX (SX), insoluble OSX (IX) and water arrayed on nitrocellulose membranes.

After these experiments have resolved the main problems associated with the method used (elimination of the high background and removal of endogenous grass peroxidase activity), the binding of CBMs to boiled grass arrayed on nitrocellulose was assessed (Fig 3.8). CBM2b1,2 concentration was used at 0.3 μg/ml, as this was the optimal concentration to use among 0.03 μg/ml (no signal produced), 0.3 μg/ml and 3 μg/ml (signal produced all over the sample), which were tried in a series of dilution of the protein stock on nitrocellulose membranes using this assay. It was observed that no obvious signal was evident in Panel B, in which no antibody was added after CBM incubation, demonstrating that boiling the grass removed endogenous grass peroxidases. A pale background, however, was observed in the negative control where antibody only was added. Although MP was used and significantly blocked the non-specific binding of antibody to the surface, it was not possible to remove the non-specific binding completely. It was also found that even when CBM was not added, the antibody alone produced a background. However, the background was minimized which likely enables an assessment of CBM binding to potential ligands. The data displayed in Fig 3.8 showed that CBM2b1,2 bound strongly to insoluble and soluble OSX and weakly to boiled grass. There was a faint signal developed on the spot where boiled grass (G) was arrayed. However, this conclusion needed to be confirmed as the observed binding was weak.

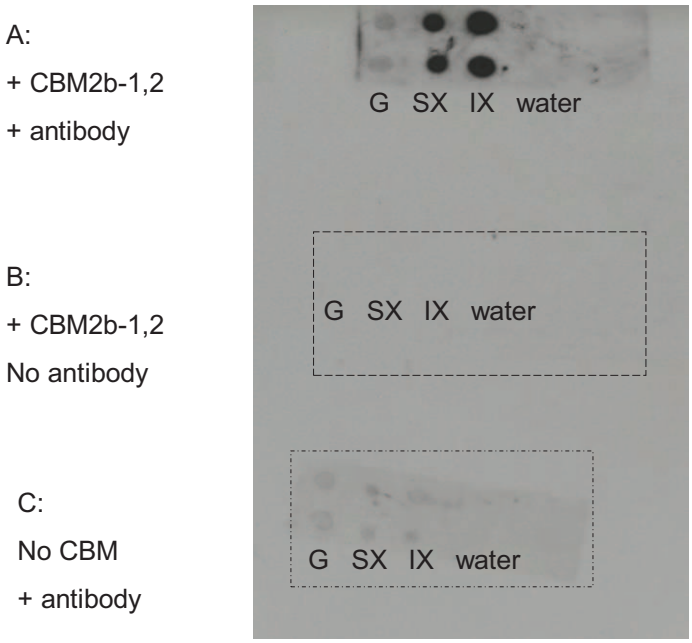
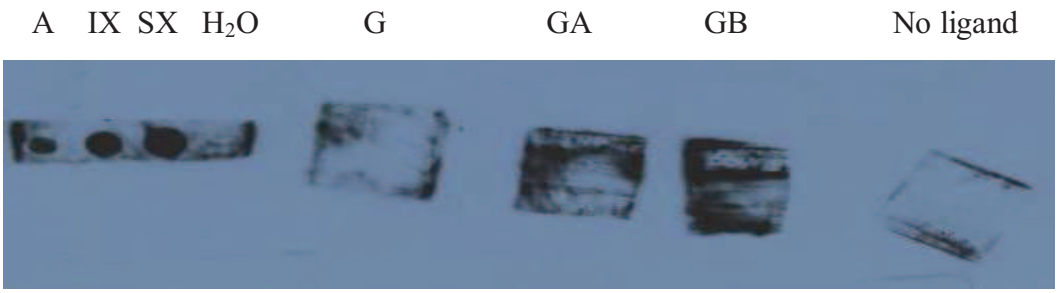


Fig 3.8 Assay of CBM2b1,2 binding to boiled grass arrayed on nitrocellulose membranes
The experiment was carried out in the presence of Marvel at 4 ºC, as described in Section 2.3.6. CBM was applied at 0.3 µg/ml, which was determined as the optimal final concentration to use on nitrocellulose filters. Anti-His HRP was used at 10^{-4} dilution of the stock, before the substrate from the ECL+ kit was applied. The membrane and X-ray film were placed between two X-rays cassettes for 10 min, and the signals identified where CBM binding occurred. Panel A displays the binding data with both CBM and antibody added in the reaction. Panel B and C are negative control reactions, with no CBM and no antibody present, respectively. Abbreviations: G, boiled grass; SX, soluble oat spelt xylan; IX, insoluble oat spelt xylan.

It was believed that the cuticle waxy layer above the grass cell walls restricted CBM2b1,2 binding to xylan molecules within the grass cell walls. Therefore, grass was pretreated with acetone or the commercial biosurfactant SLES to remove the waxy layer. After this treatment, binding assay using anti-His HRP conjugate were carried out to access the capacity of CBM2b1,2, to bind to boiled grass and boiled dewaxed grass arrayed on nitrocellulose membranes (Fig 3.9).



A: Avicel (cellulose)
IX: Insoluble oat spelt xylan
SX: Soluble oat spelt xylan
G: Boiled grass
GA: Boiled grass pretreated with acetone
GB: Boiled grass pretreated with biosurfactant SLES

Fig 3.9 Assay of CBM2b1,2 binding to dewaxed grass arrayed on nitrocellulose membranes

The experiment was carried out in the presence of Marvel at 4 ºC, as described in Section 2.3.6. CBM was applied at 0.3 µg/ml, which was determined as the optimal final concentration to use on nitrocellulose filters. The CBM which contains His_6 tag, was detected via anti-His HRP at 10^{-4} dilution of the stock, before the substrate for HRP was applied. The membrane and X-ray film were placed between two X-rays cassettes for 10 min, and the signals identified where CBM binding occurred.

CBM2b1,2 bound strongly to insoluble and soluble xylan. It also bound to Avicel, which is crystalline cellulose. The region containing no ligand displayed no signal. The control data validated the detection method. When tested against grass stains arrayed on nitrocellulose CBM2b1,2 displayed very weak binding, as these was only a faint shadow in the region containing the grass. However, after the grass had been subjected to dewax-treatment by acetone, or the biosurfactant SLES, there was a increase in binding of CBM2b1,2 to the plant material. It was believed that the acetone or biosurfactant SLES removed the waxy layer above the grass cell wall, exposing the xylan on the cell surface to CBM2b1, 2.

3.2.4.3 Screening and selection of grass binding modules

Once the assay was validated using CBM2b1,2, the capacity of CBM22, CBM15 to bind to dewaxed grass was assessed. The binding results are shown in Fig 3.10. The data showed that CBM15 and CBM2b1,2 bound to dewaxed grass treated by biosurfactant more strongly than CBM22 . CBM15 appeared to bind better to grass than CBM2b1,2. Thus, CBM15 and CBM2b1,2 were the best candidates for targeting enzymes to the grass stains, and were therefore employed in the subsequent research programme. It should be mentioned that, in the binding results presented in Fig 3.10 using CBM2b1,2 generated a weaker signal than observed in the previous experiment in Fig 3.9. This was due to the exposure time to X-ray film, which was 3 minutes, while in the previous experiment it was 10 minutes.

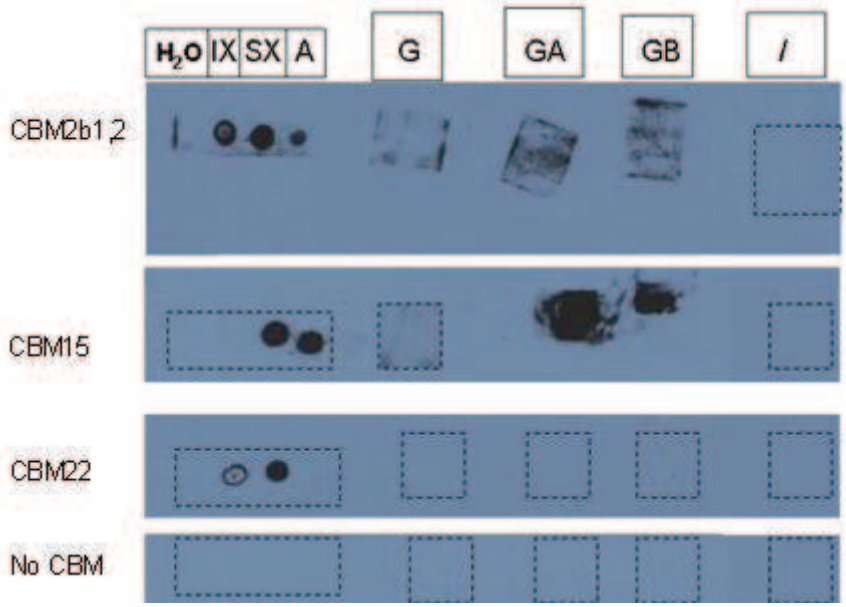


Fig 3.10 Assay of CBM2b1,2 (repeat), CBM15 and CBM22 binding to dewaxed grass arrayed on nitrocellulose membranes The experiment was carried out in the presence of Marvel at 4 ºC, as described in Section 2.3.6. All CBMs were applied at 0.3 µg/ml, which was determined as the optimal final concentration to use on nitrocellulose filters. The CBMs, which contain His_6 tag, were detected via anti-His HRP at 10^{-4} dilution of the stock, before the substrate for HRP was applied. The membrane and X-ray film were placed between two X-rays cassettes for 3 minutes, and the signals identified where CBM binding occurred. The panels display the binding profile of CBM2b1,2, CBM15 and CBM22, respectively, to boiled grass untreated(G), treated with acetone (GA) and treated with biosurfactant SLES (GB), arrayed on nitrocellulose membrane. In the negative control, no CBM was added. A piece of membrane (/) with no ligands arrayed was included. Abbreviations: A, Avicel; IX, insoluble oat spelt xylan; SX, soluble oat spelt xylan.

Isothermal titration calorimetry data has shown that CBM15, 22 and 2b1,2 bind to different xylans (Bolam, et. al, 2001; Szabo *et. a*l, 2001; Charnock *et. al* 2001)(Table 3.1). These data, however, can not predict how well these individual CBM target grass cell walls as they are complex systems of interlocking xylan chains decorated with different side chains. For example, CBM15 by ITC data shows 39.6 times lower affinity to oat spelt xylan (OSX), Rye arabinoxylan (RAX) and Wheat arabinoxylan (WAX) than CBM2b1,2, although they both target grass, based on the peroxidase assay. Indeed, CBM15 appeared to bind to grass cell walls more extensively than CBM2b1,2 (Fig 3.10). While CBM22 bound to arrayed soluble and insoluble OSX, the module did not bind to grass in either form. Therefore, CBM22 does not target grass although it binds tightly to xylan.

Protein	OSX	RAX	MGX	GAX	Xylohexaose	Cellohexaose
*Cf*CBM2b1,2	51.6	11.4	5.7	–	0.6	0.01
*Cj*CBM15‡	1.3	0.5	0.8	–	2.0	0.2
*Ct*CBM22§	7.6	11.0	–	–	15.0	0.08

Table 3.1 Binding of CBMs to purified ligands as determined by ITC

Values are association constants (M^{-1} x 10^4). –, Not determined; NB, no binding detected. Data are from Bolam, et. al, 2001; Szabo L et. al, 2001‡; and Charnock S.J et. al 2001§.

3.2.4.5 CBM binding studies on cotton

The binding experiments on arrayed grass initially deployed nitrocellulose and Marvel to eliminate non-specific binding. These two experimental factors are essential to give reliable binding data without producing false positive. But for industrial applications, one more aspect should be considered. The experimental condition should mimic the actual industrial situation, i.e., the laundry cycle. Thus the binding of CBM2b1,2 and CBM15 to grass stains arrayed on cotton fabrics, in the absence of Marvel, was explored in the next stage of the project.

Grass stains on cotton were manufactured by IGER stain company and ready to use (www.igergru.ibers.aber.ac.uk). The binding of the CBMs arrayed on cotton was then assessed. This requires the optimal concentration of the secondary antibody to be determined. Based on the following criteria: i) a significant difference between incubation containing CBM and antibody and incubation containing only antibody; ii)

The signal at the grass stain area should be more evident than in regions consisting only of cotton. This is because the CBM used should display selectivity for grass, compared to cotton. From the experiment displayed in Fig 3.11, a 10^{-4} dilution of anti-His HRP was optimum for CBM detection on cotton. A 10^{-5} or 10^{-6} dilution of anti-His HRP was too dilute either to detect the CBM15 binding on the samples, or too dilute to react the subsequent substrate for HRP from the ECL+ kit to produce a signal.

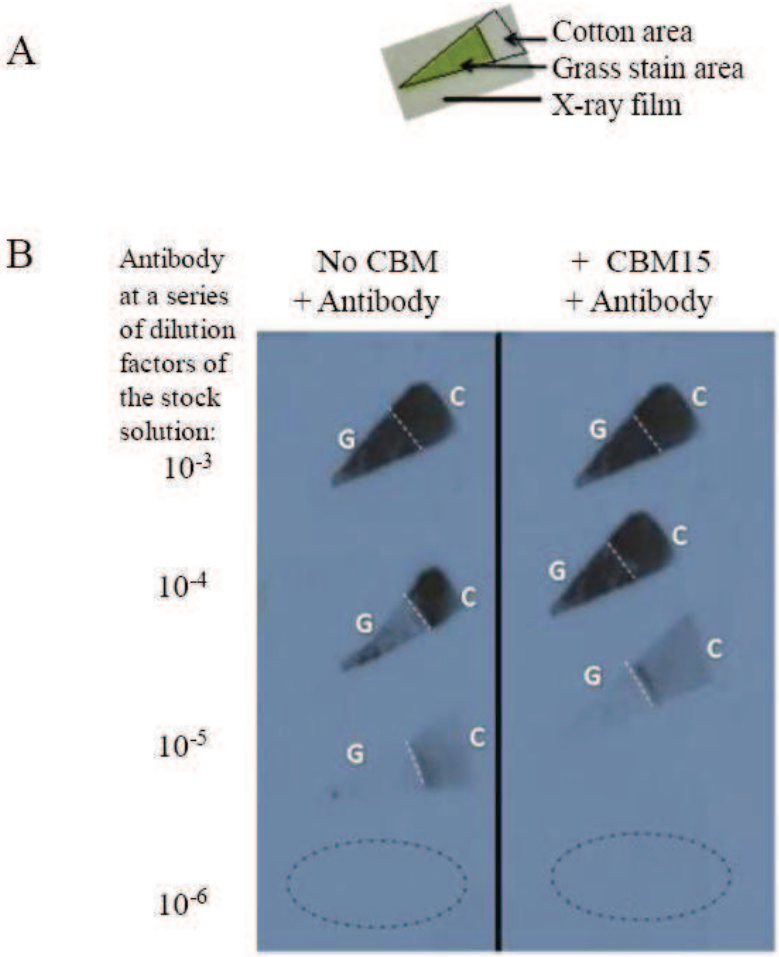
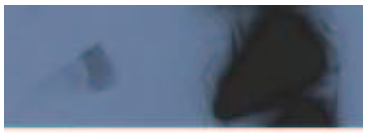


Fig 3.11 Determining the optimal concentration of antibody to use in binding assays of CBM to grass arrayed on cotton

The experiment was carried out on cotton textiles, in the presence of Marvel at 4 ºC, as described in Section 2.3.6. The grass stains arrayed on cotton were incubated with CBM15 at 0.3 µg/ml (optimal concentration determined previously on nitrocellulose) for 1 h at 4 ºC, and the samples were washed with 10 mM Tris/HCl buffer, pH 8.0. They were then incubated with the secondary antibody anti-His HRP, at a series of antibody dilution factors 10^{-3}, 10^{-4}, 10^{-5} and 10^{-6} of the stock solution, for 1 h at 4 ºC, before the substrate for HRP was applied. Panel A displays the actual sample, grass stains arrayed on cotton, with cotton area and grass stains area indicated next to it. Panel B displays the autographic data on X-ray film at 10 min exposure time. Grass areas and cotton areas are indicated as 'G and C'.

The assays were also tested in the presence and absence of Marvel, without CBM added. A very strong signal appears all over the cotton without Marvel, preventing the detection of CBM bound to the grass (Fig 3.12). This indicated the secondary antibody binds to cotton matrix non-specifically. Although different approaches have been tried, such as increasing the washing period, decreasing the exposure time, the non-specific binding issue caused by the anti-His HRP was difficult to solve when using cotton as the matrix. Due to the massive non-specific binding signal produced by anti-His HRP in the absence of Marvel, an alternative method was explored in which CBMs were directly labelled with small fluorescent molecules were explored (Section 3.2.6). Moreover, the binding assay protocols only involve a single binding-wash step, and the binding signal can be detected directly. Hence, this method would simplify the protocol and therefore are likely to eliminate the non-specific binding issues. For the direct labelling methods, fluorescent microscopy was used to obtain binding signals.



+ Marvel **no Marvel**

Fig 3.12 Assay of antibody binding to grass stains arrayed on cotton

The experiment was carried out as described in Section 2.3.6, in the presence and absence of Marvel. The grass stains arrayed on cotton were incubated with the antibody at 10^{-4} dilution of the stock solution, in the absence of CBMs, for 1 h at 4 ºC. The samples were washed using 10 mM Tris/HCl buffer, before the substrate from the ECL+ kit was applied. This figure displays the autographic data on X-ray film at 10 min exposure time.

3.2.5 GFP-CBM fusion protein assay

3.2.5.1 Trial experiment using GFP-CBM29

Since a vector encoding CBM29 fused to green fluorescent protein (GFP) was available in our lab, and mannan was a known ligand for the protein, a trial experiment was carried out to develop the use of GFP for CBM detection using the CBM29-GFP fusion protein. It was seen under UV light that the GFP-CBM29 fusion protein bound to its ligand (mannan) arrayed on nitrocellulose membrane, whereas the GFP protein alone did not bind to any of the purified ligands immobilized on the membrane (Fig 3.13). There was no binding when no protein was added to the negative control reaction. Therefore, GFP fusion appeared to be a viable method to detect CBMs. Ultimately the CBM-GFP strategy was unsuccessful and thus the binding analysis of CBM15-GFP and CBM2b1,2-GFP will be described in brief. Various recombinant plasmids were generated, which encode CBM2b1,2 and CBM15 fused by a 15 residue Proline-Threoning linker (PT linker) to GFP. Details of the construction and verification of the plasmids are present in the Appendix D.

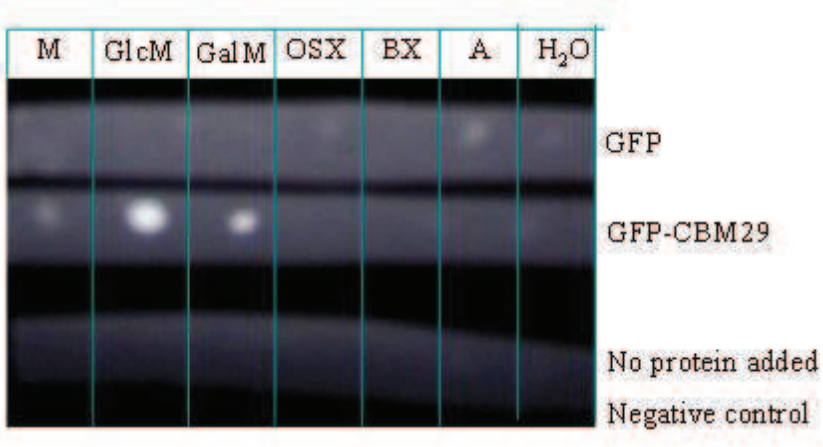


Fig 3.13 Assay of GFP-CBM29 binding to mannan arrayed on nitrocellulose membranes

The data displays the binding of proteins to seven different ligands arrayed on nitrocellulose membranes, in the presence of Marvel. The experiment was carried out in 10 mM Tris/HCl buffer at pH 8.0 at 4 °C. GFP and GFP-CBM29 fusion protein were used at 1 mg/ml. A negative control reaction (with no protein added) was included. Ligands arrayed: M, mannan; GlcM, glucomannan; GalM, galactomannan; OSX, soluble oat spelt xylan, BX, birchwood xylan; A, arabinan, and H_2O.

3.2.5.2 ITC of CBM-GFP fusions to oat spelt xylan

ITC was used to evaluate the binding to xylan of CBM2b1,2, CBM15, CBM15-GFP and CBM2b1,2-GFP. ITC is a quantitative analytical method that characterizes protein-ligand interaction. The binding constant, reaction enthalpy and entropy can be accurately measured where a sigmoid binding curve has occurred.

Comparing the best fit curve in Fig 3.14 Panel A and B, CBM2b1,2 and CBM15 display a higher affinity for soluble oat spelt xylan than the modules fused to GFP. The reduced affinity of GFP-CBM fusion proteins could be due to that GFP may cause a steric clash with the binding site in CBM2b1,2 and CBM15 modules. Hence, GFP could possibly prevent xylan chains from coming into the ligand binding site, leading to a weaker binding. These quantitative data explain why CBM15-GFP and CBM2b1,2-GFP did not bind efficiently to pure ligands or grass stains arrayed on nitrocellulose, discussed in the following paragraph.

From the data presented in Fig 3.14 and Table 3.2, GFP, and GFP fused to either CBM2b1,2 or CBM15, all bound to grass that had been treated with surfactant and arrayed on nitrocellulose. As the reporter molecule (GFP) binds to grass stains, it can not be used to explore CBM recognition of these arrayed plant cell walls. It was decided, therefore, to exploit direct labelling of CBM15 and CBM2b1,2 using fluorescent dye molecules (Section 3.2.6).

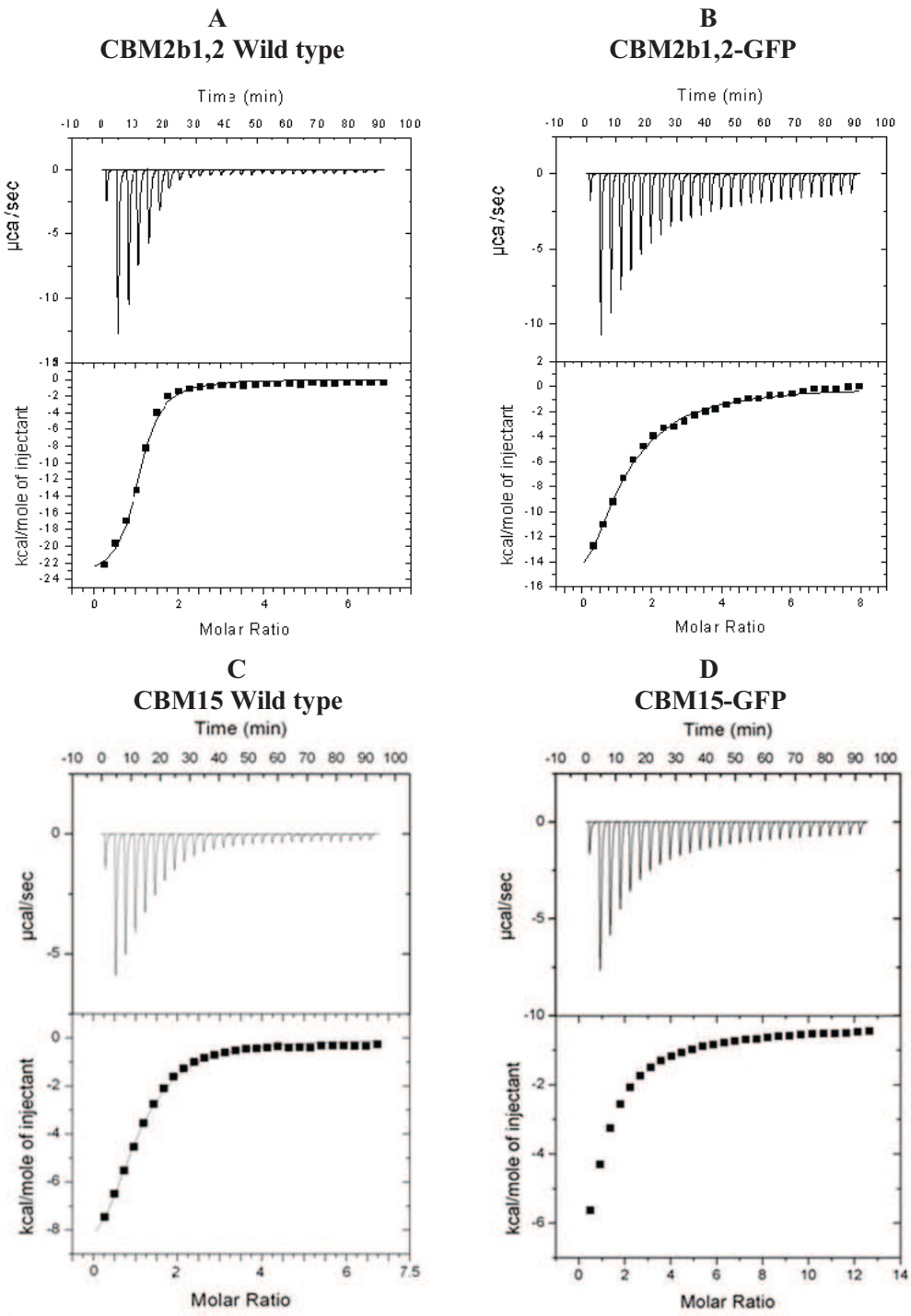


Fig 3.14 ITC data of wild type CBMs from family 15 and 2b1,2 and the protein fused to GFP fusion module binding to oat spelt xylan

ITC was carried out as described in Section 2.3.2, in 10 mM Tris/HCl buffer pH 8.0, at 25 ºC. Panel A and B display titration of CBM2b1,2 and CBM2b1,2-GFP, respectively, at 70 µM against 1 % oat spelt xylan. Panel C displays titration of CBM15 at 100 µM against 1 % oat spelt xylan and Panel D displays titration of CBM15-GFP at 100 µM against 2 % (w/v) oat spelt xylan.

Protein	$K_A \times 10^4$ (M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n[a]
CBM2b1,2	18.5±0.25	-18.14	-24.1± 6.10	-5.96	1.02± 0.02
CBM2b1,2-GFP	6.56±0.16	-12.53	-19.69± 3.20	-7.16	1.02 ±0.02
CBM15	1.45±0.09	-14.02	-21.8±3.00	-7.78	1.04 ±0.03
CBM15-GFP	0.519±0.15	-4.99	17.33±9.64	-12.34	1.01 ±0.50

Table 3.2 Table of the binding affinity and thermodynamic parameters of wild type CBMs from family 15 and 2b1,2 and the protein fused to GFP fusion module binding to oat spelt xylan

The ITC data were fitted to a single site binding model. For polysaccharide ligands in which the molar concentration of binding sites is unknown, the n-value was iteratively fitted to as close as possible to one, by adjusting the molar concentration of the ligand, the rational for this approach is described Section 2.3.2. ITC was carried out as described in Section 2.3.2, in 10 mM Tris/HCl buffer pH 8.0, at 25 °C. Protein concentrations were at 70~100 μM. Oat spelt xylan was used at 10 mg/ml.

3.2.5.3 CBM15-GFP binding to grass on nitrocellulose membranes in Marvel

E. coli strains BL21 containing p15PT-GFP with GFP cloned at C terminus, were induced with 1 mM IPTG for four hours to induce of the expression of the recombinant fusion protein. The CBM15-GFP fusion proteins are His-tagged, and were purified by IMAC using 10 and 100 mM imidazole to elute the recombinant proteins from the Talon columns (Fig D3.4 in Appendix D). The binding of the CBM15-GFP fusion to xylan arrayed on nitrocellulose was assessed (Fig 3.15). The protocol was established in the previous trial experiment, and the CBM15-GFP was used at 1 mg/ml determined as the minimal concentration to produce a signal on nitrocellulose membranes, meanwhile GFP used as a negative control was employed at the same concentration.

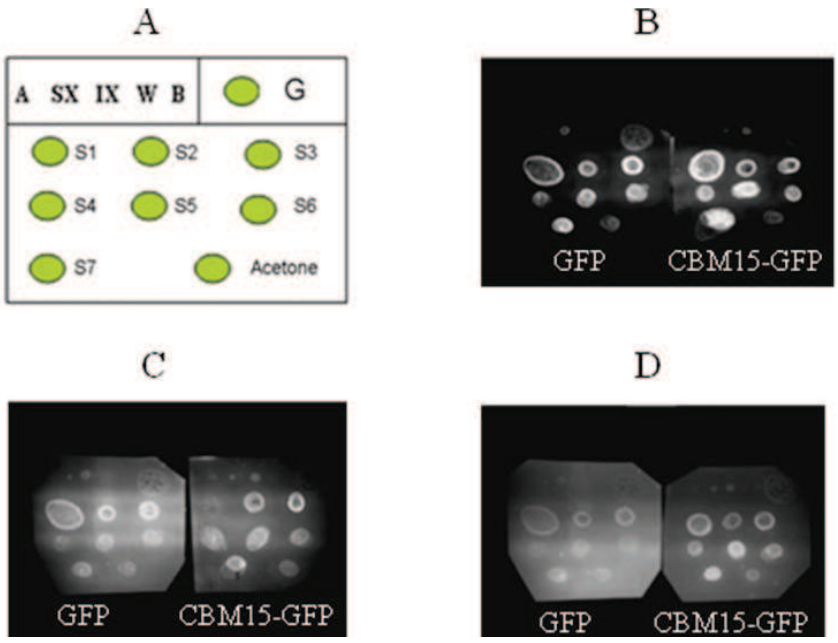


Fig 3.15 Assay of CBM15-GFP binding to grass arrayed on nitrocellulose membranes

The binding experiment was carried out at 4 °C for 1 h, in 10 mM Tris/HCl buffer, pH 8.0, in the presence of Marvel. Panel A displays the array pattern of the following ligands: intact grass stain and grass pre-incubated with different biosurfactants and acetone. Panel B, C and D display the binding results, with protein concentration at 1 mg/ml (as the minimal concentration to produce a signal on nitrocellulose membranes). The same control was used and shown in Panel B, C and D. Grass was ground in liquid nitrogen and respended in 10 mM Tris/HCl buffer, pH 8.0. Ligands arrayed: A, Avicel; SX, soluble xylan; IX, insoluble xylan; W, water; B, buffer; G, grass; S1-S7: grass pretreated by surfactant SLES (coded as S1), LAS (S2), Neodal (S3), and four biosurfactants (S4-S7) received from University of Ulster in July 2007, which were of similar properties to surfactant SLES, but produced in different strains of bacteria.

3.2.5.4 CBM2b1,2-GFP binding to grass on nitrocellulose membranes

E. coli strains Tuner containing p2b1,2PT-GFP with GFP cloned at C terminus, were induced with 0.2 mM IPTG for overnight to induce of the expression of the recombinant fusion protein. The CBM2b1,2-GFP fusion proteins are His-tagged, and were purified by IMAC using 10 and 100 mM imidazole to elute the recombinant proteins from the Talon columns (Fig D3.4 in Appendix D). The binding of the CBM2b1,2-GFP fusion to xylan arrayed on nitrocellulose was assessed (Fig 3.16). The protocol was established in the previous trial experiment, and the CBM2b1,2-GFP was used at 1 mg/ml determined as the minimal concentration to produce a signal on nitrocellulose membranes, meanwhile GFP used as a negative control was employed at the same concentration.

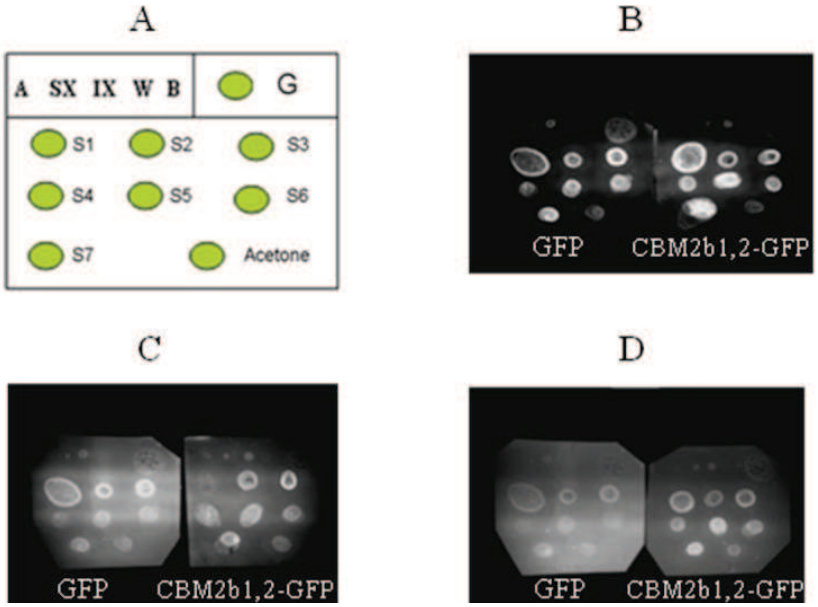


Fig 3.16 Assay of CBM2b1,2-GFP binding to grass arrayed on nitrocellulose membranes
The binding experiment was carried out at 4 °C for 1 h, in 10 mM Tris/HCl buffer, pH 8.0, in the presence of Marvel. Panel A displays the array pattern of the following ligands: intact grass stain and grass pre-incubated with different biosurfactants and acetone. Panel B, C and D display the binding results, with protein concentration at 1 mg/ml (as the minimal concentration to produce a signal on nitrocellulose membranes). The same control was used and shown in Panel B, C and D. Grass was ground in liquid nitrogen and respended in 10 mM Tris/HCl buffer, pH 8.0. Ligands arrayed: A, Avicel; SX, soluble xylan; IX, insoluble xylan; W, water; B, buffer; G, grass; S1-S7: grass pretreated by surfactant SLES (coded as S1), LAS (S2), Neodal (S3), and four biosurfactants (S4-S7) received from University of Ulster in July 2007, which were of similar properties to surfactant SLES, but produced in different strains of bacteria.

3.2.6 Binding of CBM15 and CBM2b1,2 to grass stains on cotton, using direct fluorescent labeling

3.2.6.1 Labelling of CBM with fluorescent dye and calculation of the concentration and degree of labeling (D.O.L) of dye-CBM conjugant after labelling

There are various reagents and detection methods to label a protein for in vitro analysis. Beside indirect detections using immunological methodologies such as peroxidase assays, direct labeling either using a protein reporter, for example, GFP, or using a small molecule marker, such as radioactive elements, or fluorophores. For grass stains arrayed on cotton, direct labeling has more advantages than indirect methods, as cotton is a matrix which has protein absorbing properties and likely bind to detection molecules.

In this section, fluorescent dyes were explored as potential labeling molecules. Among the fluorescent dyes available in the current market, Alexa Fluor dyes are superior fluorophores with fluorescence emissions that span the visible spectrum. Alexa fluor conjugates exhibit brighter fluorescence and greater photostability than the conjugates of other spectrally similar fluorophores. It also has characteristics such as water solubility and pH insensitivity from pH 4.0-10.0, which make these molecules highly robust tools. Succinimidyl esters provide an efficient and convenient way to selectively link the dye to primary amines ($R=NH_2$) on peptides or proteins. Unlike other reactive moieties, succinimidyle esters demonstrate very low reactivity with aromatic nitrogens and –OH groups, hence they are unlikely to cross react with the cotton matrix.

The chlorophyll in grass cells emits in the range 500 to 550 nm. However, the CBMs studied here were labelled by Alexa dye 488 (488) which emit in the region from 660 to 710 nm. Hence, the two emission regions are discrete from each other. Alexa dye 488 has a formula weight of 643 and its extinction coefficient is 71000. Approximately 10 mg of CBM2b1,2, CBM15 and BSA were dissolved in 1 ml of 0.1 M sodium bicarbonate buffer at pH 8.3-9.0, respectively. A 50 µl aliquot of the dye at 10 mg/ml dissolved in DMSO was slowly added to the protein solution while stirring. The reaction was incubated for 1 h at room temperature with continuous stirring with foil covered on the top of the tube. The protein conjugates (CBM15-488, CBM2b1,2-488 and BSA-488) were separated from unreacted free dye (called ‘Free Alexa 488’) by

size exclusion using a PD10 column, and stored at 2-6 ºC. An aliquot of 10 times diluted stock protein was measured by spectrometry from 200 to 800 nm at room temperature. An example scan of the protein-dye conjugants is showed in Fig 3.17.

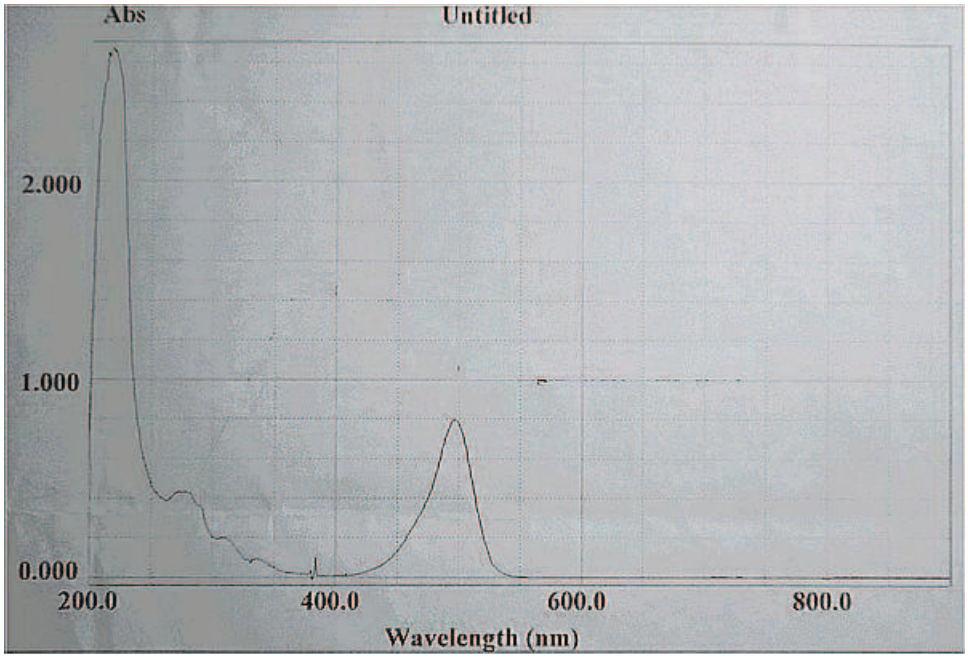


Fig 3.17 An example of CBM2b1,2-488 scan profile after labelling

From the scan profile, A280=0.459 x 10 = 4.59

A495=0.846 x 10 = 8.46

Ratio of dye absorbance at 280/495 is 0.11 (according to the instruction book),

Therefore if the absorbance of the protein-dye conjugate 495 is 8.46, then

The absorbance due to the dye at 280nm is 8.46 x 0.11 =0.9306 (w)

The total absorbance at 280 (A280) due to the protein–w=4.59-0.9306=3.66 (y), and thus protein concentration can be determined

Hence, concentration of CBM2b1,2= y/extinction coefficient of CBM2b1,2

=3.66/37830

=96.7 μM

The concentration of dye (x) can be determined from the value at 495 and the molecular extinction coefficient (according to the instruction book) at this wavelength = 8.46/71000=119 μM

Now x/y gives ratio of dye bound to protein, which is the degree of labeling =119/96.7=1.23, hence on average every CBM2b1,2 protein molecule is labeled by one dye molecule.

3.2.6.2 Fluorescent photographs of the dye Alexa 488 displays non-specific binding to cotton and grass

After the CBM was labeled with Alexa 488, referred as CBM2b1,2-488 or CBM15-488, it is essential to test if the dye absorbs to cotton non specifically. Hence, equal molar concentration of free dye 488 and CBM2b1,2-488 (as the positive control in this case) were applied to grass stains on cotton in the presence and absence of SLES. The binding results are shown in Fig 3.18; both grass areas and non-stained cotton areas were analyzed by fluorescent microscopy.

The results showed that dye 488 did not bind to grass stains (image a and b) or cotton (image e and f), in the presence or absence of SLES. In contrast, CBM2b1,2-488 bound to the grass stain (image c and d), and binding appeared to be more extensive when SLES was present. CBM2b1,2-488 also bound to cotton (image g and h). By contrast, CBM2b1,2-488 appeared to bind more tightly to cotton when SLES was absent (image g). From the above data, Alexa 488 does not bind to grass or cotton, and therefore was a useful tool for monitoring the binding of CBMs to grass stains arrayed on cotton.

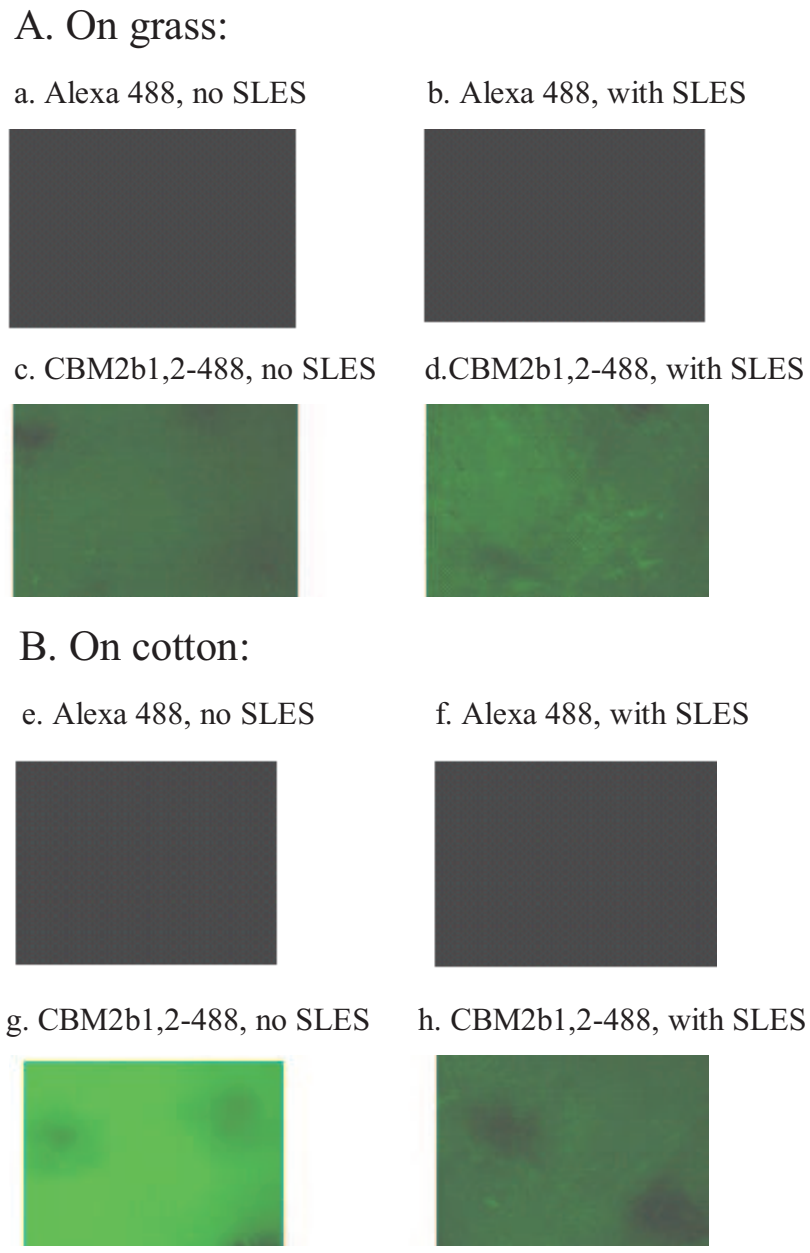


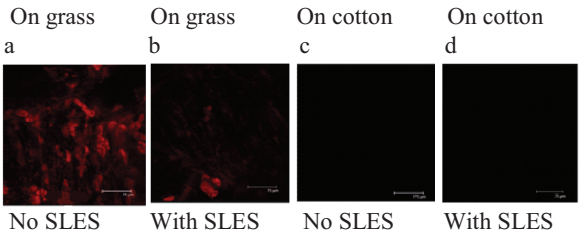
Fig 3.18 Photo-micrographs of fluorescence detection of Alexa dye 488 and CBM2b1,2-488 bound to grass stains and cotton

The experiment was carried out as described in Section 2.3.7, in the presence and absence of SLES. Panel A displays binding of the 488 and CBM2b1,2-488 on grass stains, while Panel B exhibits the binding of the 488 and CBM-488 conjugant to regions of cotton that do not contain grass stains. Panel A and B were assessed microscopically at x 40 magnification using fluorescence microscopy using an FITC filter that detects the green fluorescence emitted by the Alexa 488.

3.2.6.3 Optimal concentration of CBM-dye to use on cotton matrix

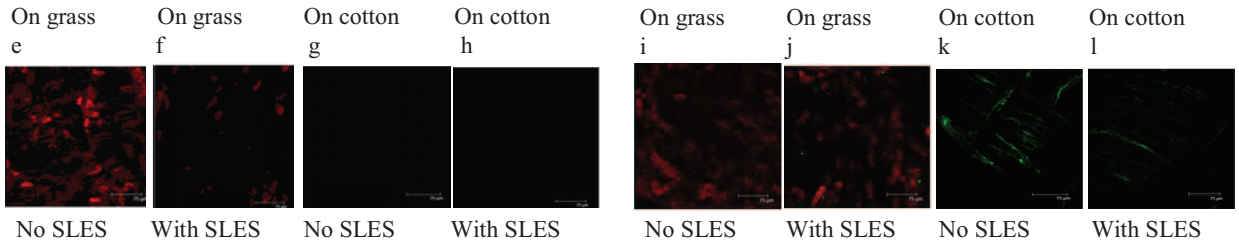
It is important to find out the optimal concentration of CBM that differentiates between the binding of the modules to grass and cotton. Hence, the binding of a series of dilutions of CBM2b1,2-488 (0 to 15 µM) to cotton arrayed with grass stains was carried out (Fig 3.19).

A. Control, CBM2b1,2-488=0:



B. CBM2b1,2-488 at 0.015 µM:

C. CBM2b1,2-488 at 0.15 µM:



D. CBM2b1,2-488 at 1.5 µM:

E. CBM2b1,2-488 at 15 µM:

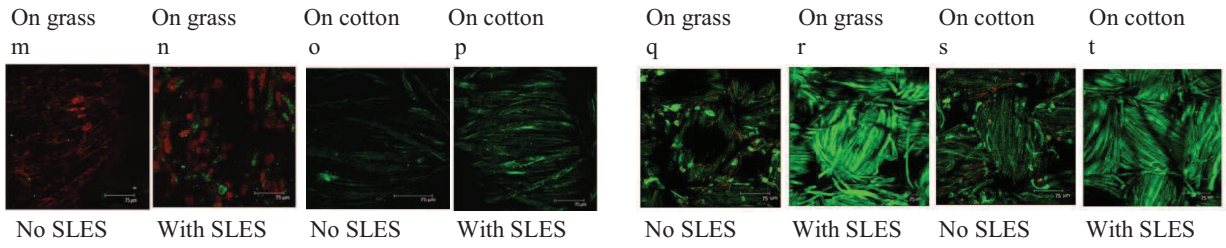


Fig 3.19 Photo-micrographs of CBM2b1,2-488 binding to grass stains and cotton at a series of concentrations The experiment was carried out as described in Section 2.3.7, in the presence and absence of SLES. The red and green fluorescent signals were simultaneously detected by confocal fluorescent microscopy in the emission range from 500-550 nm (red fluorescence from chlorophyll) and the emission range from 660-710 nm (green fluorescence from CBM2b1,2-488). Each panel displays confocal data of Panel A. CBM2b1,2-488 at 0; Panel B. CBM2b1,2-488 at 0.015 µM; Panel C. CBM2b1,2-488 at 0.15 µM; Panel D. CBM2b1,2-488 at 1.5 µM; Panel E. CBM2b1,2-488 at 15 µM. The bar in each image at right corner represents 75 µm, [----].

The data presented in Fig 3.19 revealed that as the concentration of the labeled CBM increased higher levels of the protein could be detected on the arrayed cotton, through the appearance of green fluorescence. Grass stains could be visualized through the red fluorescence emitted by the chlorophyll. SLES increased the binding of the CBM to cotton and grass, with increasing CBM2b1,2-488 concentration. It was apparent, however, that there was no preferential binding to grass arrayed on the cotton. Indeed, the protein appeared to bind more extensively to the untreated cotton.

3.2.6.4 Binding results of CBM2b1, CBM15, CBM60 to grass stains

Apart from CBM2b1,2, it is also of interest to know how CBM15, CBM2b1 and CBM60 target grass stains (Fig 3.20). As CBM2b1,2 is a dual module of CBM2b1 and CBM2b2, and its binding affinity to xylan and cellulose are increased 10 fold due to avidity effects, compared to CBM2b1 and CBM2b2 (Bolam et al., 2001). Hence, CBM2b1,2 binds to grass stain well, but also it gives non-specific bindings on cotton regions. It is possible that a single module, CBM2b1, would still target grass stains, but would display reduced binding to cotton.

The CBM60 module was also selected (Chapter 5) as it binds to a wide range of polysaccharides in plant cell walls, including xylans. Hence it may function as another possible grass targeting module. Thus, both xylan binding modules CBM60 and CBM15 were tested in this experiment together with CBM15, to see if their weaker affinity for cellulose (10 fold less than that of CBM2b1,2) would give less or no non-specific binding on cotton. All the CBM modules were labeled by Alexa 488 and used at 1.5 μM. The ratios of labeling of CBM2b1, CBM15 and CBM60 were 3.5, 3.4 and 1.4, respectively.

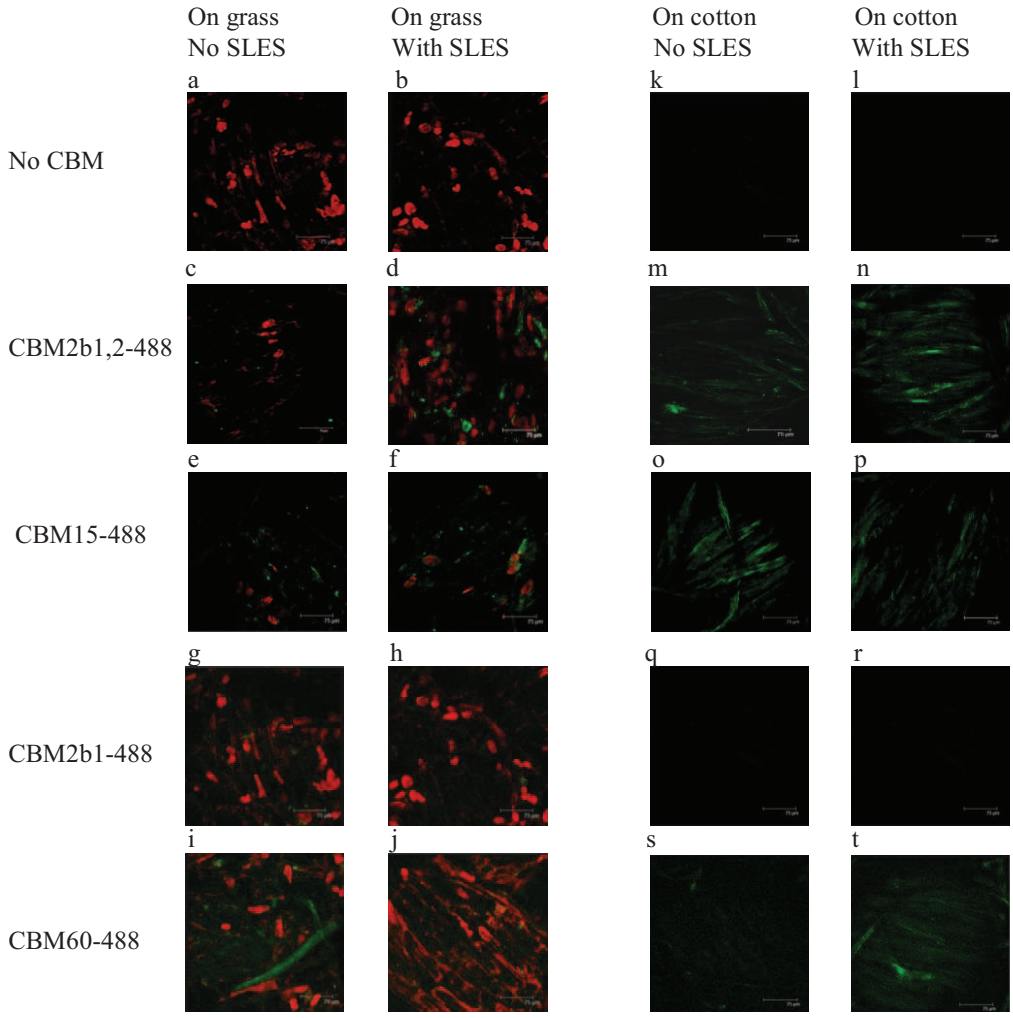


Fig 3.20 Photo-micrographs of four different labeled CBM modules binding to grass stains and cotton The experiment was carried out as described in Section 2.3.7, in the presence and absence of SLES. The red and green fluorescent signals were simultaneously detected by confocal fluorescent microscopy in the emission range from 500-550 nm (red fluorescence from chlorophyll) and the emission range from 660-710 nm (green fluorescence from CBM2b1,2-488). Images a-j and k-t display the confocal data of grass stains and cotton, respectively, incubated with CBM2b1,2-488, CBM15-488, CBM2b1-488 and CBM60-488 at 1.5 μM. CBM2b1,2-488 data were the same as those displayed in Fig 3.18. The bar in each image at right corner represents 75 μm, [----].

From the result presented in Fig 3.20, the negative controls (image a and b) revealed only red chlorophyll and no green CBM-488 molecules. The background of the untreated cotton displayed no signals (image k and l). The binding of CBM2b-1 to grass stains was too weak to be detected in the presence and absence of SLES (image g and h). A green fluorescence signal was evident in images i and j, in which the arrayed cotton was incubated with CBM60. However, the green fluorescence signal derived

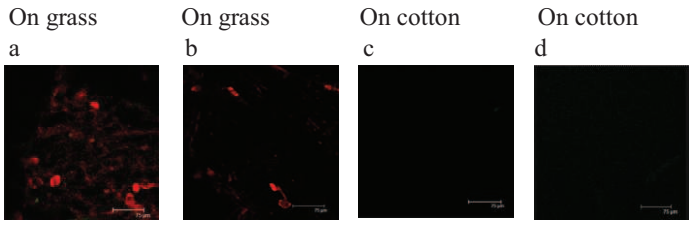
from CBM60, and the red fluorescence emitted by chlorophyll, did not overlap suggesting that the protein is binding to the cotton and not the grass stains (image i). Paradoxically CBM60 did not appear to bind to untreated cotton, particularly in the absence of SLES similar to CBM2b1 (image s and q). It appears that CBM60 binds to regions where grass has been arrayed but not to untreated cotton without SLES (image I and s). As the nature of the binding of CBM60 to the grass stains was unclear, this module, together with CBM2b1, which do not bind to grass stains, were not explored further.

CBM15, unlike CBM2b1 and CBM60, targeted grass stains to some degree (image e), which increased in the presence of SLES, which was shown in image f as quanlitative data. It also bound to cotton, which produced non-specific binding signals, both in the absence and presence of SLES (image o and p). Compared to CBM2b1,2, the binding of CBM15 on grass and cotton regions was weaker. To confirm this, CBM2b1,2-488 and CBM15-488 were used in the quantification experiments describe below.

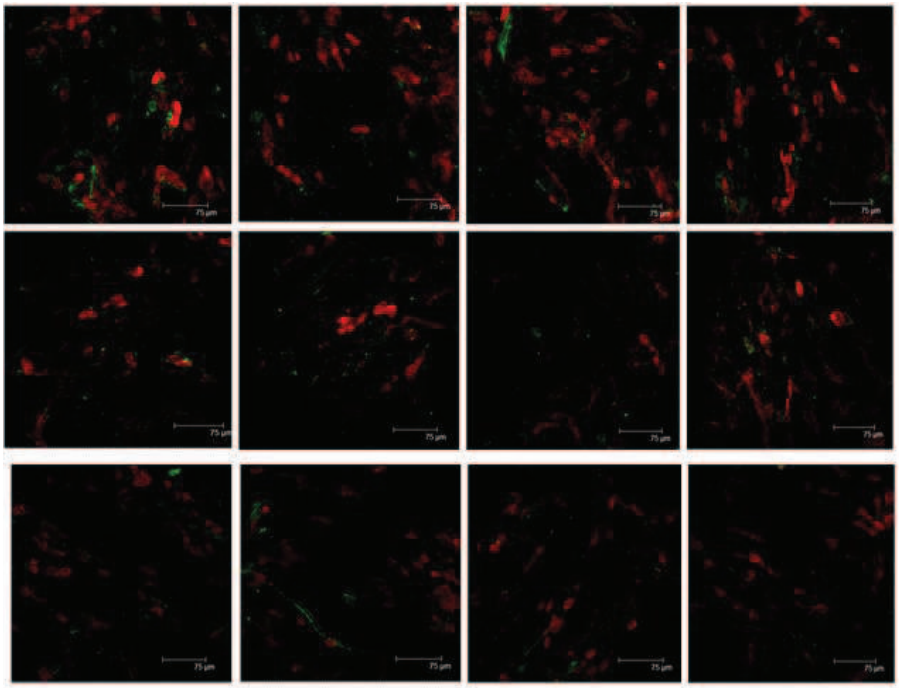
3.2.6.5 Effect of SLES made to CBM2b1,2 binding to grass stains

Since the preliminary data (Section 3.2.6.2 and 3.2.6.3) clearly showed that CBM2b1,2 displayed increasing targeting of grass stains in the presence of SLES, it was necessary to repeat the experiment to provide more extensive information. Since the increase targeting of grass stains of CBM2b1,2 in SLES was relatively more obvious than other protein modules, thus only CBM2b1,2 was tested at qualitative stage before carrying out quantification in a larger scale. The following experiment was carried out by doing triplicates for each condition, and sampling four discrete areas of each treatment. The following data show 12 images from each treatment, derived from three separate experiments. The results are shown in Fig 3.21.

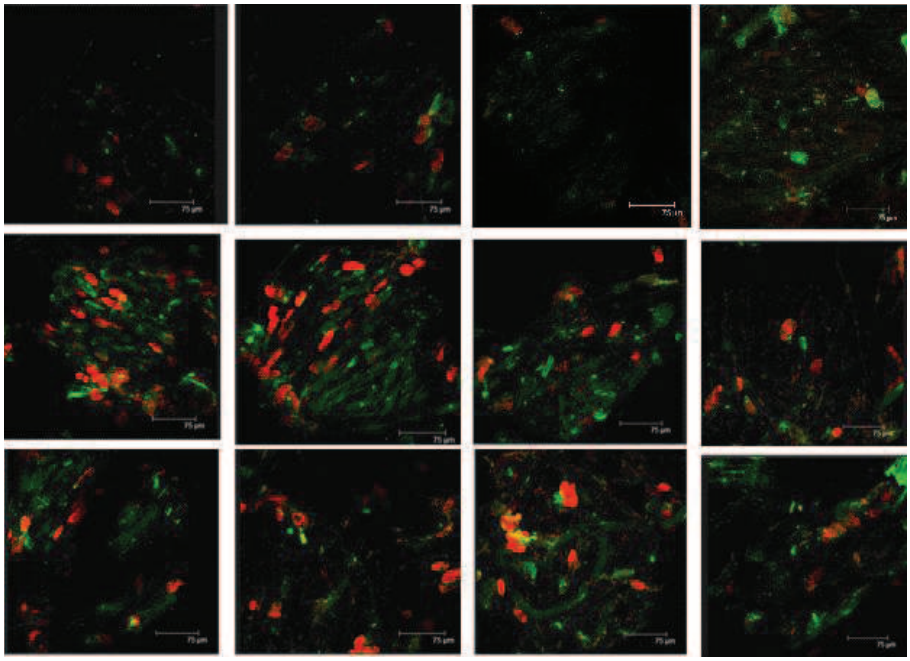
A. no CBM (controls):



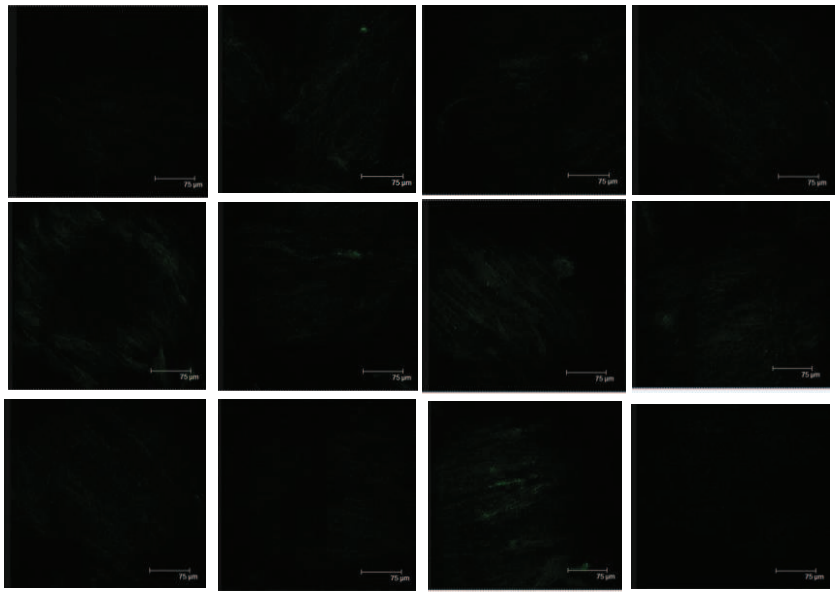
B. CBM2b1,2-488 binding to grass regions, in the absence of SLES:



C. CBM2b1,2-488 binding to grass regions, in the presence of SLES:



D. CBM2b1,2-488 binding to cotton regions, in the absence of SLES:



E. CBM2b1,2-488 binding to cotton regions, in the presence of SLES:

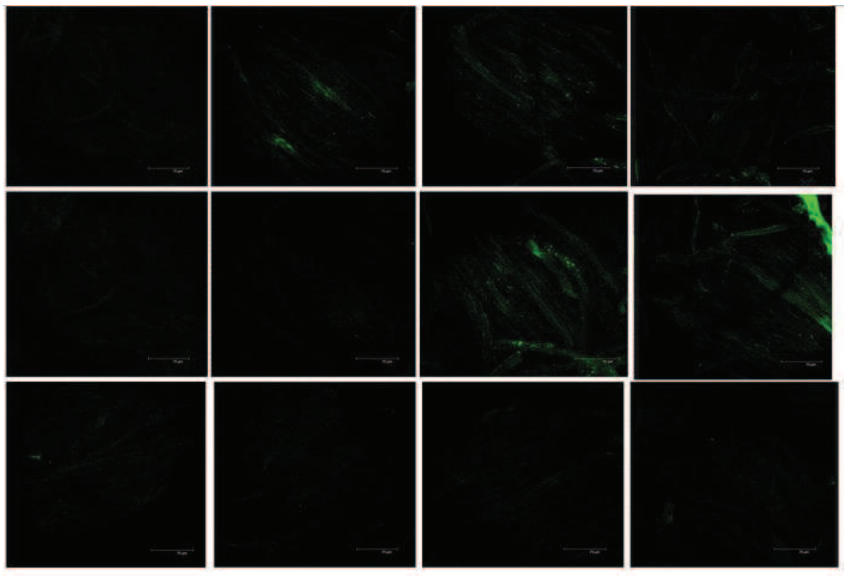


Fig 3.21 Photo-micrographs of CBM2b1,2-488 binding to grass stains and cotton (repeated at a large scale)

The experiment was carried out as described in Section 2.3.7, in the presence and absence of SLES. The red and green fluorescent signals were simultaneously detected by confocal fluorescent microscopy in the emission range from 500-550 nm (red fluorescence from chlorophyll) and the emission range from 660-710 nm (green fluorescence from CBM2b1,2-488). Panel A (image a to d) displayed the confocal data from controls. Panel B, C, D and E displayed the 12 confocal images for each condition of CBM2b1,2-488 (at 1.5 μM) on grass with no SLES, grass with SLES, cotton with no SLES and cotton with SLES, respectively. The bar in each image at right corner represents 75 μm, [----].

The negative control, without CBM-dye added, showed a clear background on cotton and only red chloroplast particles in the areas containing grass stain. There were no green particles (CBM-dye) around the surface of grass cells. For the negative controls, all the images taken from different regions on the samples looked very similarly clear. Hence, one representative image was displayed for each type of control.

On grass stains without SLES (Panel B), binding of CBM2b1,2 was observed, through the appearance of green fluorescent signals, although the level of binding was weak. On cotton without SLES, the non specific binding appeared also weak (Panel D). In the presence of SLES, the amount of CBM2b1,2-dye molecules bound to the surface of grass cells significantly increased, evident by the brighter and more extensive green fluorescence around the red chlorophyll (Panel C). On cotton, only 2 of the 12 images contained more green fluorescence compared to cotton without SLES (Panel E).

The above confocal data is not only consistent with previous results obtained with the peroxidase assay, but also provides evidence that CBM2b1,2 binds to grass cell wall more strongly in the presence of SLES. Furthermore, the data suggest that CBM2b1,2 bind more extensively to grass stains than arrayed cotton in the presence of SLES. By using direct labeling combined with confocal images analysis, we could visualize the binding of CBMs to grass stains arrayed on cotton. At a more detailed level, these promising qualitative data suggest quantifying CBM absorption onto cotton and grass area, with or without surfactant SLES, was warranted.

3.2.7 Quantification of CBM adsorption to grass stain and cotton

To quantify the CBM adsorption to grass stain and cotton as accurately as possible, a larger area of the arrayed textile needs to be examined, together with an increase in the number of replicates. This requires a large amount of image information, which requires high throughput microscopy. Hence, normal fluorescence microscope was used for quantifying CBM-dye adsorption on grass stain areas and cotton areas, as it covers a larger field (0.5 mm^2 using x 40 magnification) and it generates images quicker than confocal microscopy, which covers a smaller field (0.14 mm^2).

The four experimental conditions assessed were as follows: grass with no SLES, cotton with no SLES, grass with SLES and cotton with SLES. Each material was treated with CBM-dye, no CBM, free dye, BSA or no added molecule. In total, 6 replicates of each combination of arrayed material and added molecules were quantified. For each image, the total number of pixels (X1) emitted by the fluorescent CBM-dye conjugant, or free dye were, calculated by anlyaisSIS® 3.1 imaging software. In total, 5 images were analyzed for each combination and the average of these readings was taken as a single data point. The same experiment was repeated six times for each combination. From the values, the average and standard error were calculated. A chart was produced and all pairs of data were analyzed for statistical significance by stats software Instats in Prism (Appendix E). Each group of data was also named by a letter (group A to L) in the tables and charts, next to the key. More detailed data collected for the six replicates from which the average values were generated in provided in Appendix E.

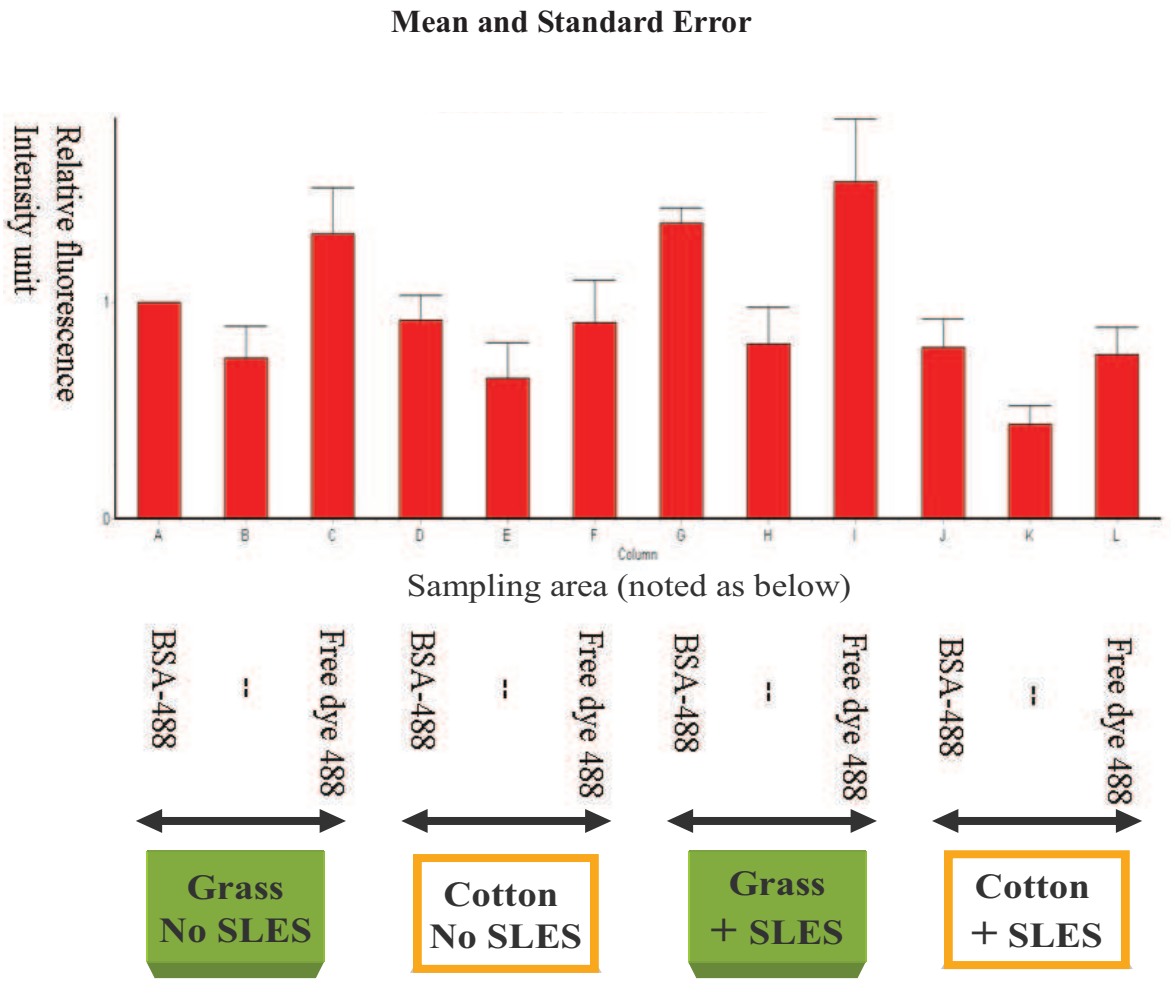


Fig 3.22 Relative fluorescence intensity emitted from dye (488) labelled BSA in grass and cotton regions, in the presence and absence of SLES

The experiment was carried out as described in Section 2.3.7. Negative controls (no protein, no dye incubations) and free dye 488 of equal molar to BSA-488 (at 1.5 µM) were included. Each group of data was named by an alphabetic letter A to L in the chart. The summarised data in group B to L were normalised to group A (CBM-488 binding to grass stains, in the absence of SLES). All pairs of data were analyzed for statistical significance by stats software Instats in Prism. Brackets with *, ** and *** above mean the difference between a pair of data were statistically significant, with $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. More detailed data, which were used to derive the figure, are presented in Table 3.3.

BSA-488	Grass, no SLES			Cotton, no SLES			Grass, +SLES			Cotton, +SLES		
Independent replicate	BSA A	-- B	Free dye C	BSA D	-- E	Free dye F	BSA G	-- H	Free dye I	BSA J	-- K	Free dye L
1	1	0.061	1.398	1.002	0.068	1.257	1.601	0.068	2.646	1.049	0.050	0.804
2	1	1.023	2.021	1.010	1.058	1.562	1.481	1.078	2.122	0.761	0.413	0.976
3	1	1.058	1.803	0.767	1.078	0.470	1.450	1.170	1.640	0.651	0.686	0.463
4	1	0.607	0.779	0.634	0.333	0.374	1.267	0.610	1.010	0.270	0.466	0.463
5	1	0.790	0.746	1.390	0.629	0.638	1.157	0.850	1.101	1.230	0.437	0.641
6	1	0.921	1.161	0.729	0.757	1.168	1.279	1.087	0.861	0.791	0.582	1.237
Average	**1**	**0.743**	**1.318**	**0.922**	**0.654**	**0.911**	**1.372**	**0.810**	**1.563**	**0.792**	**0.439**	**0.764**
Standard Error	0	0.152	0.214	0.112	0.163	0.197	0.067	0.170	0.289	0.136	0.088	0.125

Table 3.3 Relative fluorescence intensity emitted from dye (488) labelled BSA in grass and cotton regions, in the presence and absence of SLES The experiment was carried out as described in Section 2.3.7. Negative controls (no protein, no dye incubations) and free dye 488 of equal molar to BSA-488 (at 1.5 μM) were included. Each group of data was named by an alphabetic letter A to L in the table. The summarised data in group B to L were normalised to group A (CBM-488 binding to grass stains, in the absence of SLES). All pairs of data were analyzed for statistical significance by stats software Instats in Prism. More detailed data, which were used to derive the figure, are presented in Appendix E.

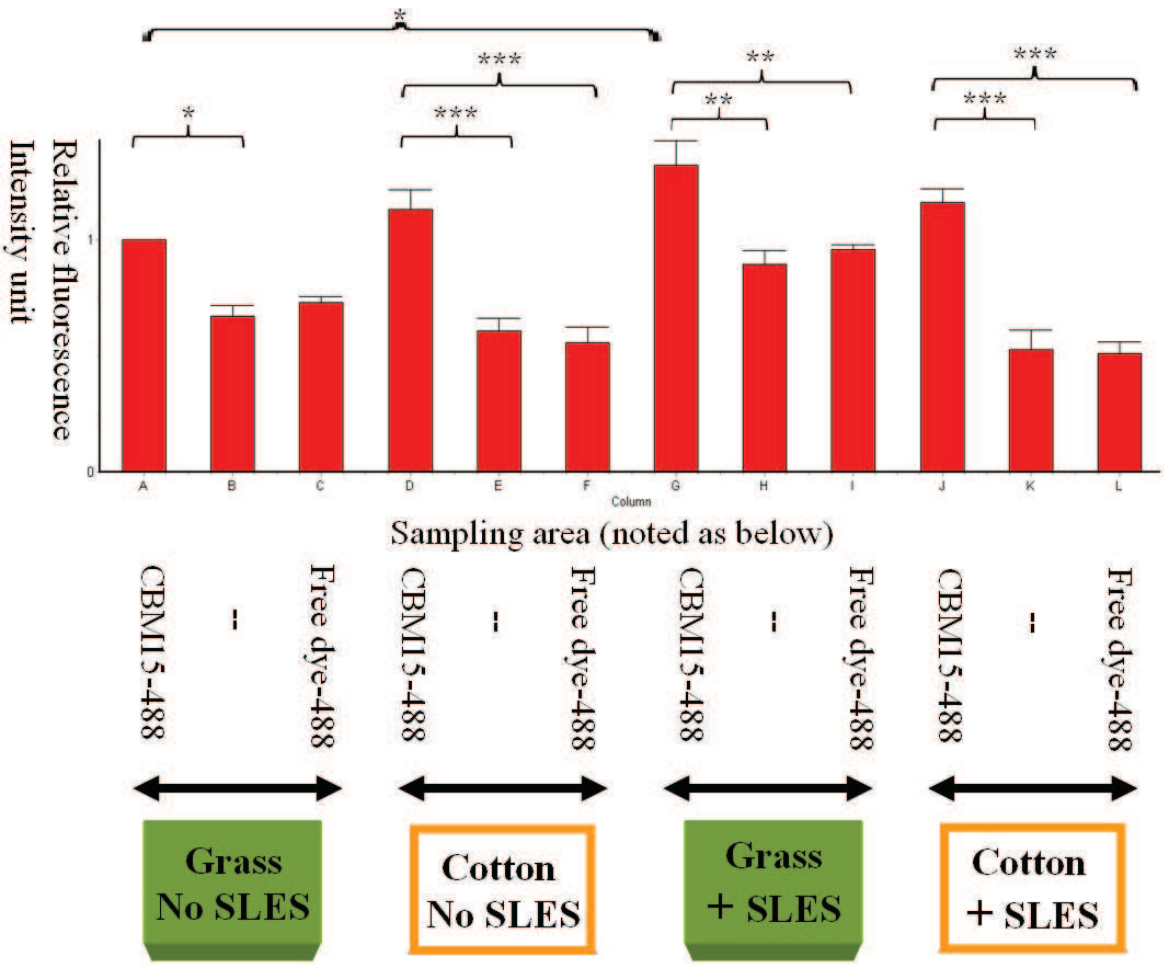


Fig 3.23 Relative fluorescence intensity emitted from dye (488) labelled CBM15 in grass and cotton regions, in the presence and absence of SLES

The experiment was carried out as described in Section 2.3.7. Negative controls (no protein, no dye incubations) and free dye 488 of equal molar to CBM15-488 (at 1.5 μM) were included. Each group of data was named by an alphabetic letter A to L in the chart. The summarised data in group B to L were normalised to group A (CBM-488 binding to grass stains, in the absence of SLES). Brackets with *, ** and *** above mean the difference between a pair of data were statistically significant, with $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. All pairs of data were analyzed for statistical significance by stats software Instats in Prism. More detailed data, which were used to derive the figure, are presented in Table 3.4.

CBM15-488	Grass, no SLES			Cotton, no SLES			Grass, +SLES			Cotton, +SLES		
Independent replicate	CBM A	-- B	Free dye C	CBM D	-- E	Free dye F	CBM G	-- H	Free dye I	CBM J	-- K	Free dye L
1	1	0.704	0.785	0.858	0.560	0.786	1.399	0.718	0.942	1.277	0.450	0.570
2	1	0.571	0.802	1.247	0.538	0.684	1.133	0.942	0.891	1.193	0.392	0.476
3	1	0.694	0.682	1.245	0.572	0.337	1.791	0.940	1.031	1.241	0.572	0.426
4	1	0.865	0.756	1.189	0.898	0.613	1.079	1.125	0.946	1.225	0.898	0.431
5	1	0.605	0.657	0.895	0.541	0.428	1.213	0.844	0.993	0.866	0.541	0.453
6	1	0.593	0.705	1.358	0.528	0.501	1.317	0.800	0.956	1.156	0.326	0.722
Average	**1**	**0.672**	**0.731**	**1.132**	**0.606**	**0.558**	**1.322**	**0.895**	**0.960**	**1.160**	**0.530**	**0.513**
Standard Error	0	0.045	0.024	0.084	0.059	0.068	0.105	0.058	0.020	0.061	0.083	0.047

Table 3.4 Relative fluorescence intensity emitted from dye (488) labelled CBM15 in grass and cotton regions, in the presence and absence of SLES The experiment was carried out as described in Section 2.3.7. Negative controls (no protein, no dye incubations) and free dye 488 of equal molar to CBM15-488 (at 1.5 μM) were included. Each group of data was named by an alphabetic letter A to L in the table. The summarised data in group B to L were normalised to group A (CBM-488 binding to grass stains, in the absence of SLES). All pairs of data were analyzed for statistical significance by stats software Instats in Prism. More detailed data, which were used to derive the figure, are presented in Appendix E.

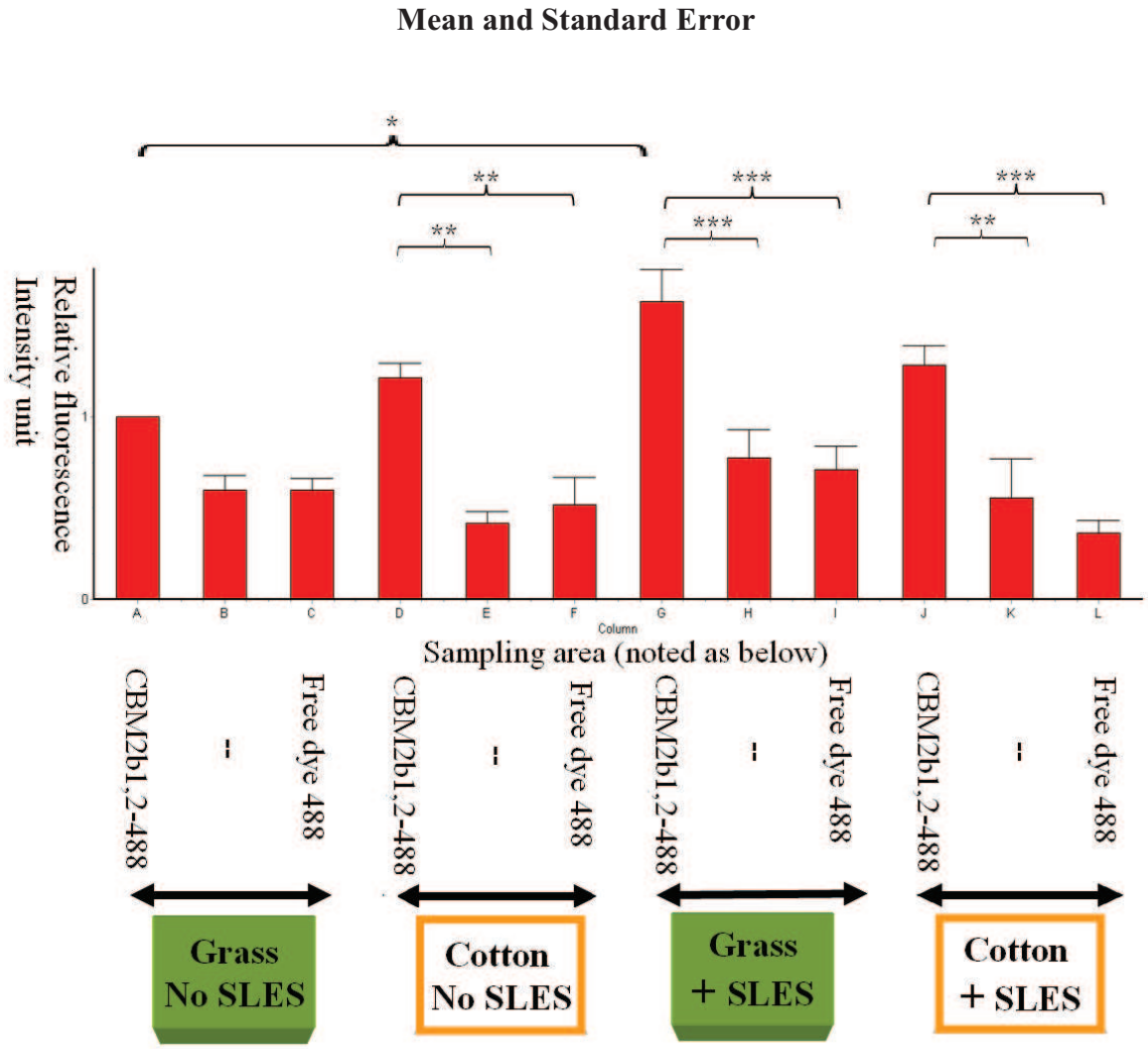


Fig 3.24 Relative fluorescence intensity emitted from dye (488) labelled CBM2b1,2 in grass and cotton regions, in the presence and absence of SLES

The experiment was carried out as described in Section 2.3.7. Negative controls (no protein, no dye incubations) and free dye 488 of equal molar to CBM2b1,2-488 (at 1.5 μM) were included. Each group of data was named by an alphabetic letter A to L in the chart. The summarised data in group B to L were normalised to group A (CBM-488 binding to grass stains, in the absence of SLES). Brackets with *, ** and *** above mean the difference between a pair of data were statistically significant, with $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. All pairs of data were analyzed for statistical significance by stats software Instats in Prism. More detailed data, which were used to derive the figure, are presented in Table 3.5.

CBM2b1,2-488	Grass, no SLES	Cotton, no SLES	Grass, +SLES	Cotton, +SLES

Independent replicate	CBM A	-- B	Free dye C	CBM D	-- E	Free dye F	CBM G	-- H	Free dye I	CBM J	-- K	Freedye L
1	1	0.956	0.860	1.214	0.513	1.267	1.604	1.111	1.219	1.372	0.650	0.534
2	1	0.546	0.625	1.442	0.476	0.378	2.462	1.366	0.983	1.716	1.552	0.619
3	1	0.690	0.680	1.344	0.638	0.403	1.562	0.494	0.445	0.919	0.502	0.245
4	1	0.468	0.503	1.072	0.380	0.377	1.549	0.442	0.592	1.344	0.130	0.303
5	1	0.491	0.531	1.307	0.207	0.363	1.418	0.577	0.604	1.130	0.256	0.277
6	1	0.441	0.383	0.925	0.291	0.335	1.196	0.664	0.427	1.214	0.261	0.196
Average	**1**	**0.599**	**0.597**	**1.217**	**0.418**	**0.520**	**1.632**	**0.776**	**0.712**	**1.282**	**0.559**	**0.362**
Standard Error	0	0.080	0.067	0.078	0.064	0.149	0.177	0.153	0.130	0.110	0.213	0.070

Table 3.5 Relative fluorescence intensity emitted from dye (488) labelled CBM2b1,2 in grass and cotton regions, in the presence and absence of SLES The experiment was carried out as described in Section 2.3.7. Negative controls (no protein, no dye incubations) and free dye 488 of equal molar to CBM2b1,2-488 (at 1.5 μM) were included. Each group of data was named by an alphabetic letter A to L in the table. The summarised data in group B to L were normalised to group A (CBM-488 binding to grass stains, in the absence of SLES). All pairs of data were analyzed for statistical significance by stats software Instats in Prism. More detailed data, which were used to derive the figure, are presented in Appendix E.

Conclusions are drawn from Fig 3.22 - Fig 3.24:

BSA-488 binding to grass and cotton

As the negative control, the amount of BSA on the cotton or arrayed grass was not significantly different to free dye. Therefore, any effect observed by the CBMs can be said to be due to a specific interaction with the matrix (Fig 3.22).

CBM15-488 binding to grass and cotton

In the absence of SLES, there is statistically significant binding of CBM15 to cotton, but not to arrayed grass. In the presence of SLES, CBM15 displays statistically significant binding to both cotton and grass (Fig 3.23).

CBM2b1,2-488 binding to grass and cotton

In the absence of SLES, there is significant binding of CBM2b1,2 to cotton, but not to arrayed grass. Although the binding of the CBM to grass in the absence of SLES is not significantly different to just the dye, there is trend for the protein to bind to grass more than just dye. CBM2b1,2 displays statistically significant binding to grass in the presence of SLES, and the module displays a statistically significant binding to cotton again in the presence of surfactant. These data are also consistent with the confocal results presented in Fig 3.24.

To conclude, despite the non-specific binding to cotton, due to it is presence in large excess compared to the arrayed material; both qualitative and quantitative data indicate that CBM2b1,2 and CBM15 target grass stains in the presence of SLES, which is consistent with the anti-His HRP assay and the confocal data. Adding SLES significantly increases the binding of both CBMs to grass stains. Therefore, the two CBMs were fused to xylanases to study how CBM-Xylanase conjugants performed in detergent assays (Chapter 4).

3.2.8 Generation of potential grass binding modules with increased grass binding affinity: CBM15-CBM15

It has been observed from ITC studies that the affinity of CBM2b1,2 for OSX is 18.5 x 10^4 M^{-1}, which is 12.8 times higher than that of CBM15 for the same ligand (Table 3.2). The enhanced affinity displayed by CBM2b1,2 is due to cooperative binding, or avidity; which occurs when two linked CBMs to the same polysaccharides, as the two modules work synergistically in binding to xylan chains (Bolam, et al., 2001). By contrast, individual CBM2b1 and CBM2b2 do not display high affinity for xylan. Thus, it is possible that linking two CBM15 modules in tandem, through a flexible linker, would increase the binding affinity of two duplicate module to xylan and grass stains compared to single copies of the CBM. A construct of CBM15-CBM15 in tandem was made by Dr. Louise Tailford (Fig 3.25).

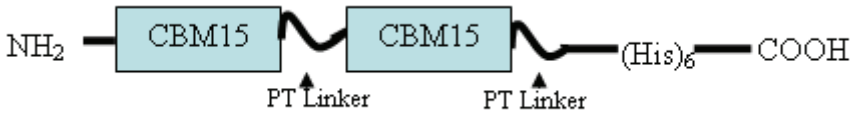


Fig 3.25 Schematic construct of CBM15-CBM15 in pET22b

The pET22b plasmid was engineered to produce CBM15-CBM15 constructs. A DNA fragment encoding a 15-amino acid proline and threonine rich linker (PT linker) was cloned between two genes encoding CBM15 modules. PT linkers are usually found in the modular xylanase, for example, in *Cellulomonas fimi* Xyn10A. The cloning work was carried out by Dr. Louise Tailford.

3.2.8.1 Protein expression and purification of CBM15-CBM15

The expression of CBM15-CBM15 was tested in *E. coli* strains BL21 and Tuner (Fig 3.26) (Section 2.1.21). The protein size observed was approximately 45 kDa, which is close to its theoretical size 44 kDa. The recombinant protein expression was induced in 1 L cultures at mid-exponential phase (A_{600}=0.6) by the addition of 1 mM IPTG (BL21) and incubation at 37 °C for 4 h and 0.2 mM IPTG (Tuner) and incubation at 16 °C overnight. The *C*-terminal His_6 tag of the recombinant protein was exploited in IMAC purification (Section 2.1.24). CBM15-CBM15 was eluted with Talon buffer containing 100 mM imidazole. CBM15-CBM15 was expressed at higher level in *E. coli* strain Tuner than BL21; hence this former strain was used to produce the protein.

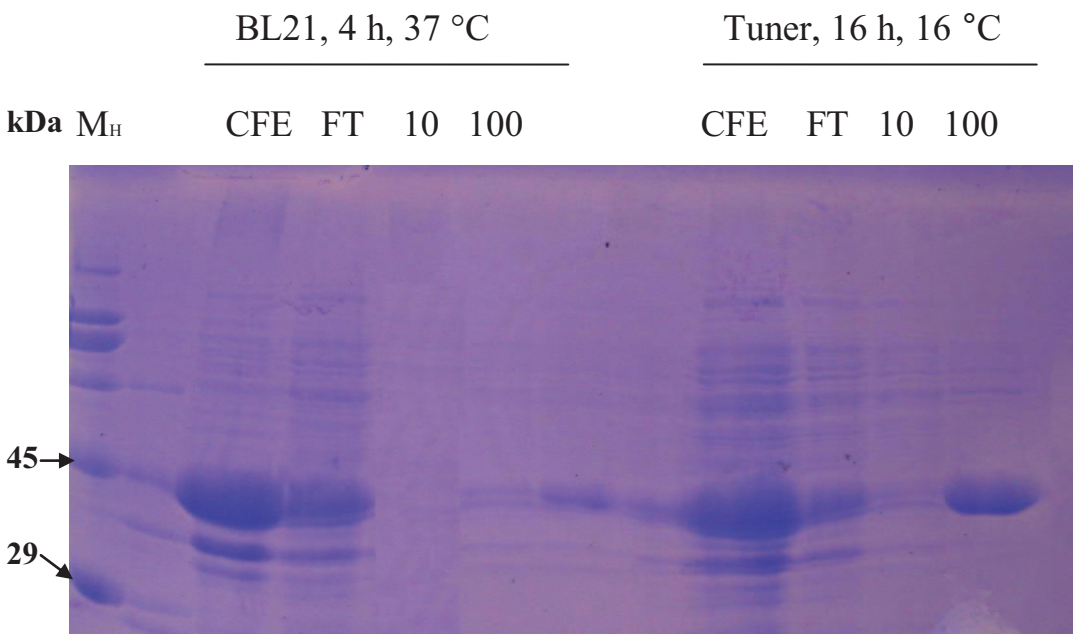


Fig 3.26 SDS-PAGE of CBM15-CBM15 purified by IMAC

Purification was carried out as described in Section 2.1.24. A 5 µl aliquot of cell free extract (CFE) and flow-through (FT) was loaded, then 10 µl of elution with 20 mM Tris/HCl, pH 8.0, containing 300 mM NaCl and 10 mM imidazole (10) and 10 µl of elution with 100 mM imidazole (100) . Lane M_H contained Sigma high molecular weight standards. Analysis was performed by SDS-PAGE, using a 12.5 % (w/v) polyacrylamide gel.

3.2.8.2 ITC studies of CBM15 and CBM15 tandem to oat spelt xylan

ITC was used to measure the affinity of the CBM15 and CBM15-CBM15 for oat spelt xylan (OSX). Analysis of the data confirm that the affinity of CBM15 for OSX was very similar (k_A =1.49 x 10^4 M^{-1} and k_A =1.5 x 10^4 M^{-1}) in the presence and absence of the surfactant SLES (Fig 3.27). The thermodynamics forces driving CBM15/OSX interaction were driven both enthalpy and entropy, the changes in enthalpy and entropy both had a positive impact on the binding event. When CBM15 was in tandem, the affinity for OSX increased by four fold higher (k_A=7.06 x10^4 M^{-1}), when SLES was added. The change in enthalpy became more negative, which favoured CBM15-CBM15/OSX interaction; whereas the change in entropy became more negative and thus gave a penalty to the reaction. A decrease of entropy resulted in a more positive ΔG (the total energy of the reaction), which makes the reaction less likely to happen.

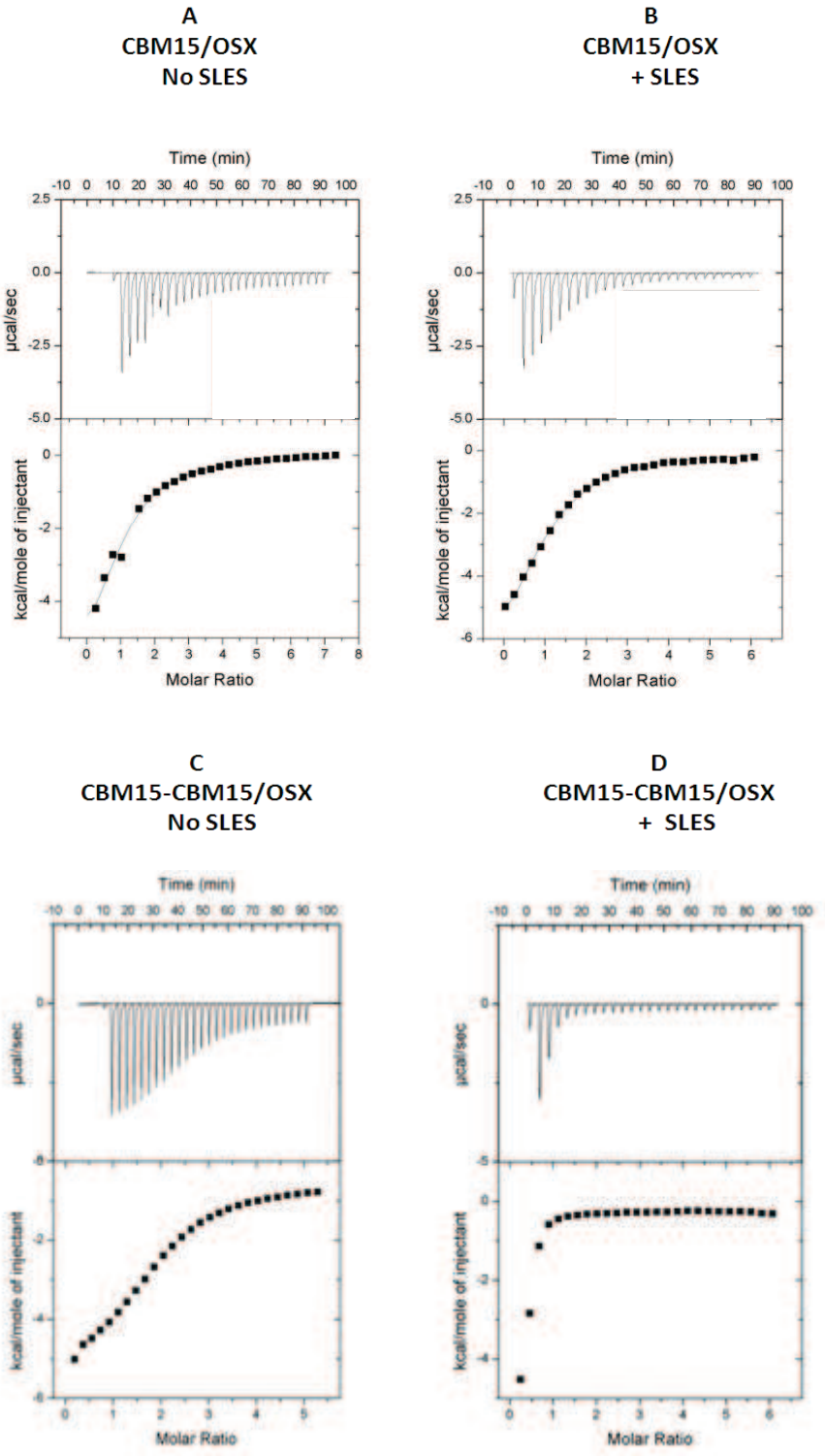


Fig 3.27 ITC data of CBM15 and CBM15-CBM15 binding to oat spelt xylan

The titrations were carried out in 10 mM Tris/HCl buffer, pH 8.0, at 25 ºC, as described in Section 2.3.2. Oat spelt xylan was at 10 mg/ml, and the protein concentration was at 100 µM. The binding affinity and thermodynamics of binding for each protein are summarized in Table 3.6.

Protein	Surfactant SLES at 1g/L	$k_A \times 10^4$ (M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n[a]
CBM15	-	0.626±0.49	-5.19	-2.18 ±0.24	3.01	1.04 ±0.03
CBM15	+	2.68±0.83	-6.05	-1.49 ±0.19	4.56	0.989 ±0.11
CBM15-CBM15	-	1.06 ± 0.11	-10.01	-7.60 ±0.41	-2.11	1.99 ±0.03
CBM15-CBM15	+	5.79 ±0.11	-6.48	-12.68 ±6.71	-6.20	*

Table 3.6 Table of the binding affinity and thermodynamic parameters of CBM15 and CBM15-CBM15 binding to oat spelt xylan

Oat spelt xylan was at 10 mg/ml, and the protein concentration was at 100 μM. The ITC data were fitted to a single site binding model. As the molar concentration of binding sites is unknown in xylan, the n-value was iteratively fitted to as close as possible to one, by adjusting the molar concentration of the ligand, the rational for this approach is described Section 2.3.2. ΔG and TΔS were calculated according to the equation, $\Delta G = -RT ln k_A = \Delta H - T\Delta S$.

It is noticed the k_A value of CBM15 to oat spelt xylan in Table 3.6 is half of that presented in Table 3.2. There can be some variations existed between the same protein/ligand reaction, which were carried out at two different times by ITC. It is possible that the protein prepared this time was not as active as that previously used in Table 3.2, due to protein aggregates or misfolding, or protein degraded to an extent during the protein purification process.

3.2.8.3 Grass pull-down assay

The capacity of CBM15, CBM15-CBM15, CBM2b1,2 (positive control) and BSA (negative control) to bind to grass was assessed by a pull down experiment. The data presented in Fig 3.28 showed that BSA, CBM15 and CBM15-CBM15 did not bind to grass in any of the three conditions tested. CBM2b1,2 did bind to grass in the presence and absence of SLES.

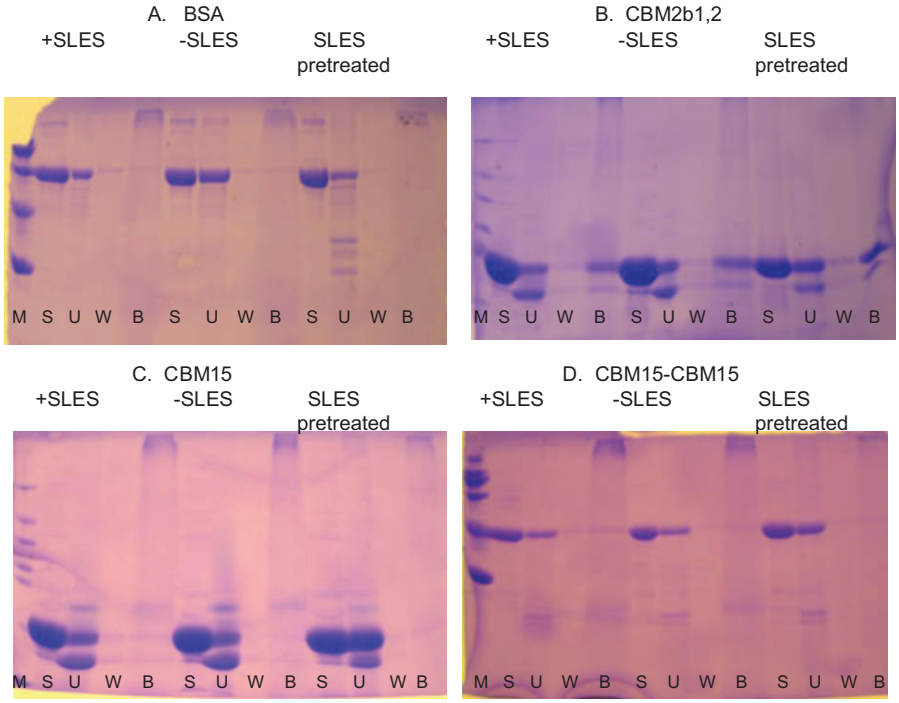


Fig 3.28 Pull down assay of CBM2b1,2, CBM15 and CBM15-CBM15 binding to grass

Purified BSA (Panel A, negative control), CBM2b1,2 (Panel B), CBM15 (Panel C) and CBM15-CBM15 (Panel D) at 100 μM was incubated with 10 mg of grass and incubated for 1 h on ice, in the presence of SLES at 1 g/L (+SLES), in the absence of SLES (-SLES) and with grass pretreated by SLES for 12 h (SLES pretreatment), respectively. Lane S, U, W and B contained the purified protein as starting materials, the supernatant after incubating with the ligand (unbound protein), the wash and the protein released from the polysaccharide by SDS. Samples in all the fractions were analysed by SDS-PAGE, using 10 % (w/v) gels. Sigma low molecular weight markers were on lanes M_L.

From the literature, CBM15 did not bind to insoluble xylan (Millward-Sadler, et al., 1995), nor to grass in this study. This could be due to the grass not containing much xylan both on the surface and under the waxy layer. For CBM2b1,2, it was expected that there would be an increase in binding in the presence of SLES, but this was not observed. This could be because the 1 h incubation at 4 °C was not carried out long enough for the detergent pretreatment process.

3.3 Discussion

There are seven families of CBMs that bind to xylan; which are family 2b, 15, 4, 6, 22, 35 and 60. Plant cell wall binding experiments using immunohistochemistry showed that CBM15 and CBM2b bind well to a variety of plant cell walls in sections,; particularly the xylan in secondary walls, in a range of species (McCartney, et al., 2006). In contrast, xylan-binding CBMs from families 4, 6, 22 and 35 display a more restricted capability to target secondary cell walls in plants. Therefore, CBM15 and CBM2b1,2 were selected as potential grass stain targeting modules. CBM22 was deployed as a negative control, as the module binds tightly to purified xylan but does not access the hemicellulose when it is embedded in the plant cell wall.

Approaches

The binding of selected CBMs to grass was first validated on nitrocellulose using Marvel to block protein non-specific binding to the matrix. CBM binding to grass arrayed on cotton without the addition of Marvel was then explored to better simulate the real situation. Initially, an HRP-antibody, which recognises a His tag, was used to detect the binding of CBMs to grass arrayed on nitrocellulose and cotton. However, the antibody bound extensively to the textile in the absence of Marvel. Furthermore, the grass had to be boiled to inactivate endogenous peroxidase enzymes contained in grass cells. Hence, the method was viewed as inappropriate for models that accurately

mimic the laundry industry and thus was not pursued further. To measure CBM binding on cotton without a blocking agent, direct labelling using GFP fused to CBMs was accessed. It simplified the protocol to a single binding-wash procedure. However, it was found that fusion of GFP (45 kDa) to the CBM led to a reduction of affinity of the protein for xylan. This could be due to that GFP may cause a steric clash with the binding site in CBM2b1,2 and CBM15 modules. It is possible that GFP prevents xylan coming into the ligand binding site, leading to a weaker binding. Thus, a much smaller fluorescent dye molecule (Alexa 488, Fwt at 680) was used instead, which not only simplified the protocol, but also provide a quick and easy labelling procedure. Moreover, the emission region of the dye is distinct from that of chlorophyll in grass, which allows simultaneous detection by confocal microscope and quantification by fluorescence microscopy.

There were three major reasons for switching to fluorescence from confocal microscopy. The most critical factor was that the fluorescence signal becomes weaker with time and this is problematic when quantifying the image as this is time consuming. Therefore quantification requires a minimum of five sampling points for each area, and for each condition six areas were analyzed; images were quantified for each experimental condition. For confocal systems, the time spent taking images is practically very long. The operation taking each image on a confocal microscope requires 1 min per image. However, using standard fluorescent microscope, an image can be recorded in 5 seconds. The time difference is because in fluorescence microscopy excitation and signal detection occur at the surface of each sample, whereas in image formation in confocal microscopy the software divides the sample into multiple micro-discs, penetrating the sample to a certain depth before taking the image. Certainly, using confocal microscopy to obtain qualitative information is ideal

since it provides clearer resolution of the images, but it was not optimal for quantification as the focusing process takes much longer than conventional fluorescence microscopy. Finally, the coverage of the field using confocal (~0.14 mm^2) is smaller than that of standard fluorescence microscopy (~0.48 mm^2). Hence, fluorescence microscopy was used for quantifying CBM binding.

With fluorescence microscope, images of CBM binding to cotton were captured on an Olympus DP50 fluorescence microscope as Alexafluor-488 fluorescence emission at 40x magnification and using the FITC filter. For each cotton sample, images of five random points were captured. For all samples analysed in a given experiment, image-gaining settings were optimised based on achieving minimal background using the negative control and remained constant for all data collection. Fluorescence intensity (representing CBM binding to cotton) was analysed using analysis® 3.1 software. After conversion to greyscale images, the software assigned values to each pixel (8 bits per pixel) of between 0 (black) and 255 (white) and calculated total fluorescence intensity as the integral of all values in the field. Data are expressed as mean ±SEM. For comparison across multiple experiments, all data were normalised to the mean value measured for binding of CBM-488 to grass stain in the absence of SLES.

Consistency of data

In the initial scoping experiment, deploying the HRP-antibody conjugate to detect CBM binding of CBM2b1,2 was evaluated twice and shown to bind to grass. However, CBM22 and CBM15 were tested only once by this method and it would have been prudent to have repeated these assays to provide confidence of the data generated. However, the HRP-antibody was only used in scoping experiment. It was decided to use a fluorescent labeling method to obtain more extensive qualitative and qualitative

data by repeating the experiment, doing triplicate for each condition, combined with a rigorous statistical analysis, before drawing conclusions on the targeting of CBMs to grass stains.

Despite the limited use of the HRP-antibody, using this method, CBM15 displayed a positive signal against Avicel. The data were consistent with confocal and fluorescence microscope data showing that CBM15 bound to cotton, which is essentially crystalline cellulose. It was also shown from previous studies that CBM15 binds weakly to Avicel and hydroxyethyl cellulose (Millward-Sadler, et al., 1995; Szabo, et al., 2001). It was also noted that CBM15 targeted grass arrayed on nitrocellulose, but did not bind to grass by the pull down assay. Given that the same source of grass harvested in Newcastle University was used in both assays, one explanation is that the sensitivity of the peroxidase assay is greater than using SDS-PAGE to detect CBM binding, which could detect protein down to 1 ~ 10 pg. Thus, the peroxidase assays involved secondary antibody detection of CBMs, which could be easily detected using the highly sensitive substrate used by secondary antibody-enzyme conjugates, while the detection limit of product by SDS-PAGE is ~10 ng. In this project, the protein concentration was at 0.3 μg/ml as the optimal concentration to use on nitrocellulose membranes in the presence of Marvel. Hence, if <10 ng of protein bound to the grass, while it could be detected by the peroxidase assay, SDS-PAGE analysis in the pull down assays would likely not be sufficiently sensitive to detect binding.

Compared to grass bound to nitrocellulose, CBM15 did not appear to display significant binding to grass arrayed on cotton. Possible explanation for this phenomenon is that the grass stain samples received from Unilever, which were

manufactured by a company, is from a different species of the grass to that used in the HRP-antibody experiments, which was harvested on-site from Newcastle University. Since two species may display structural variations in grass cell walls, this could result in the different targeting performance by CBM15. It is known that grass is a monocot, which belongs to the Poaeceae family under the taxonomic order Poaeles in monocotyledon families with the presence of ferulic acid ester-linked to their unlignified cell walls (Michelangeli, et al., 2003). It contains (1-3, 1-4)-β-D-glucans and a small proportion of pectic rhamnogalacturonans. The hemicellulose in grass cell walls is usually divided into four general groups: (i) xyloglycans (Xylans), (ii) mannoglycans (mannans), (iii) xyloglucans, and (iv) mixed-linkage ß-glucans (Michelangeli, et al., 2003). All of them display in many structural variations differing in side chain types, distribution, localization and/or types and distribution of glycoside linkages in the main macromolecular chain. For example, xylan in grass cell walls can be grouped into several structural subclasses: (i) homoxylans, (ii) glucuronoxylans, (iii) (arabino)glucuronoxylans, (iv) arabinoxylans, (v) (glucurono)arabinoxylans, and (vi) heteroxylans (Ebringerova, et al., 2006). Also, the proportion of the structural varieties of xylans in two different grass stains may occur, which may explain why CBM15 targeted grasses used in anti-His HRP assay better than the grass stain samples on cotton. For example, the content of L-Arabino-D-xylans (AX) in grass cell walls can vary from 0.15 % in rice to ~13 % in the whole grain from barley and rye, and up to 30 % in wheat bran (Michelangeli et al, 2003). The frequency of unsubstituted, mono- and di-substituted xylopyranose backbone residues with α-Araf (arabinofuranoside) side chains; the degree of esterification of some Araf residues, are all factors that could result in a large structural variation in the xylans in different species of grass (Ebringerova, et al., 2005), thus this could lead to a variation of the CBM15 targeting capacity to different grass stains. D-Glucurono-L-arabino-D-xylans (GAX), which are

the dominant hemicelluloses in the lignified tissues of grasses may also influence the variation in CBM binding. GAX isolated from various plant sources shows considerable variation of the MeGlcA (methyl-D-glucuronic acid) : Xyl : Ara ratios in the range 3–9:10:1–10 (Ebringerova, et al., 2006). With respect to generally utility, one has to consider that grass stains on clothes from customers, from different geographical locations, are unlikely to be the same type of grass provided by the stain company. Comparing the binding performance of CBM15 and CBM2b1,2, the latter seems more likely to interact with a wider range of grass cell walls, it is a more efficient as a stronger grass targeting module in both pull down assays and in the confocal experiment (Fig 3.21). Thus CBM2b1,2 was viewed as a more promising targeting module for the detergent industry.

The confocal images in Fig 3.19 showed that SLES increased the CBM binding on cotton, whereas Fig 3.18 (Section 3.2.6.2, dye check) showed the opposite. Since Fig 3.18 and Fig 3.19 only assessed one single point from each sample in early stage assays, the binding image in these two figures were preliminary. The more extensive random sampling and repeats carried out, the more reliable information of the CBM binding profiles is likely to be provided. This was exemplified by the experiment repeated in Fig 3.21, and the later quantification studies and data, are of more confidence, comparing to Fig 3.18 and 3.19. CBM15 binding to cotton is consistent with its capacity to interact with cellulose, albeit weakly. The preference for the matrix other the arrayed grass likely reflects the large amount of cellulose present in cotton, compared to the concentration of available xylan in grass cell walls. The confocal data in Fig 3.21, together with the images obtained using the fluorescence microscope (Section 3.2.7, Fig 3.24) showed a greater average fluorescence intensity of CBM2b1,2 bound to grass in the presence of SLES, compared to the CBM2b1,2

bound to cotton also in the presence of the surfactant. Hence, a trend was observed showing CBM2b1,2 displays a preference for grass stains over cotton in the presence of SLES. This is consistent with qualitative confocal data display in Fig 3.19 panel D, which showed CBM2b1,2 to bind more to grass stains than cotton in the presence of SLES, although it is inconsistent with Fig 3.18 panel D which was discussed before, only represents preliminary data that may well give an inaccurate picture of the binding profiles.

The structural basis for the observed binding

The topology of the binding site of CBMs determines the binding affinity and specificity. CBM2b1,2 has a very shallow binding cleft. Indeed, the topology of CBM2b1,2 places the module between type A CBM, a planer hydrophobic ligand binding site, and type B CBM, a deep cleft where a single sugar polymer chains are accommodated (Bolam, et al., 2001). CBM2b1,2 appears to be adapted to make contact with xylans chains which are in close association with other cell wall components, as the binding site is more exposed. By contrast CBM22 has a deeper cleft, which makes accessing xylan molecules that are in intimate contact with other polysaccharides more difficult, and therefore display limited recognition of secondary cell walls in grasses (Charnock, et al., 2000; Simpson, et al., 2000).

However, the theory that the depth of the binding cleft is related to the capacity of the CBM to bind to xylans embedded in the plant cell wall is less tenable when applied to CBM15. Although the dimensions of its xylan-binding site are similar to CBM22, which display limited recognition of secondary plant cell walls, CBM15 binds to numerous cell walls. Hence, the capacity of CBMs to interact with cell walls is not completely dependent on the depth of the cleft. The binding clefts of CBM2b1,2 and

CBM15 are described as twisted platforms where single aromatic amino acid side chains in the binding subsites stack only against the α faces of the pyranose rings of xylose residues n and n+2 in xylan polymers (Henshaw, et al., 2004; Szabo, et al., 2001). In contrast, the binding site of CBM22 contains a primary binding subsite with two aromatic amino acid side chains that sandwich the same sugar ring, rather like ''aromatic tongs,'' thus interacting with both the α and β faces of a single sugar (Charnock, et al., 2000; Simpson, et al., 2000). Within the complex structure of the plant cell wall, it is likely that at least one face of the xylose residues in the xylan backbone interacts with other polysaccharides, providing an explanation for why the CBMs that contain the pair of aromatic tongs do not bind to xylan when it is embedded in the plant cell wall. In contrast, it is likely that cell wall xylans will remain accessible to CBM2b1,2 and CBM15 as only a single face of the xylose residues stack against the two tryptophans in the binding site of these proteins. This biological rationale supports the observations derived from the HRP-antibody detection system, which show that CBM15 and CBM2b1,2 target grass cell walls, whereas CBM22 did not. Therefore, it was decided that CBM15 and CBM2b1,2 were promising candidates for targeting enzymes to grass stains, and thus they were employed in the subsequent research programme. As discussed above, which both CBM2b1,2 and CBM15 are potential grass stain targeting proteins, the family 15 module displays some capacity to bind cellulose, but CBM2b1,2 is more xylan specific. The amorphous regions of cellulose are similar to xylan, the polysaccharides could, potentially, bind to CBM15 and CBM2b1,2. Indeed, in CBM15 the C5 of the bound sugars are generally sufficiently exposed such that the addition of a CH_2OH group (C6 group) would not cause significant steric clashes. Presumably, this is not the case for CBM2b1,2 where glucose molecules can not be accommodated at one or more subsites. It is not really obvious why CBM15 should bind to grass, it is possible it has higher affinity for

cellulose as it may bind to cotton in a more significant way than CBM2b1,2; it is also possible when the grass stains were arrayed on cotton, there was not many grass cells (xylan) available in the stains.

The xylan-binding modules CBM15, 22 and 2b1,2 are from different sequence-based families; they present differences in their overall fold, binding site topology in the conformation of the ligand binding residues. Hence, the mechanism by which each CBM displays ligand recognition is variable, which leads to the different binding affinities displayed for plant cell walls including grasses. These factors provide the biological rationale for why CBMs from different families with evident structural diversity can display significant variation in specificities for xylans embedded in grass cell walls.

Conclusion

In conclusion, CBM22, 15 and 2b1,2 were expressed at a high level and purified to electrophoretic homogeneity. They are functional and bind to oat spelt xylan. CBM15 and CBM2b1,2 showed strong binding to grass arrayed on nitrocellulose membranes in the presence of Marvel using the peroxidase detection assay. However, there was a significant signal produced by the non-specific absorption of HRP-antibody reagent to cotton particularly in the absence of blocking reagent Marvel. Furthermore, boiled grass needed to be arrayed due to endogenous peroxidase activity. Direct fluorescent labeling of CBM15 and CBM2b1,2 with Alexa 488 was used instead, combined with image analysis using confocal microscopes and standard fluorescence microscopy, for qualitative and quantitative assessment, respectively. Given that the Alexa 488 dye did not bind to cotton non-specifically, CBM-488 was used at 1.5 μM as this was the

optimal concentration to provide the best contrast between the signals derived from grass stains and cotton. CBM2b1,2 and CBM15 both showed an increase in targeting grass stains in the presence of SLES. CBM2b1,2 appeared to be a more promising grass targeting module for use in the laundry industry. Both CBMs are fused to xylanases to study how CBM-xylanase fusions performed in detergent assays (Chapter 4).

Chapter 4. Effect of CBMs on enzyme function in detergent assay

4.1 Introduction

The cell walls in grass stains contain a high level of xylan. Thus, treatment of these stains with xylanases, which degrade the hemicellulose, may facilitate the removal of grass from textiles during the laundry process. Xylanases are located in glycosyl hydrolase, families GH10 and GH11 (Henrissat and Davies, 1997). These enzymes hydrolyse glycosidic bonds by a double displacement acid-base assisted mechanism leading to retention of the anomeric configuration (Davies and Henrissat, 1995). Consistent with their endo- mode of action, the active site of both GH10 and GH11 xylanases consists of an open cleft that extends along the length of the protein and can accommodate four or more xylose residues (Charnock, et al., 1998). Each xylose is bound to a subsite, which is given a negative or positive number dependent on whether it binds the glycone or aglycone region of the substrate, respectively, with glycosidic bond hydrolysis occurring between the -1 and +1 subsites (Davies, et al., 1997).

The GH11 xylanase, *Np*Xyn11A, from the anaerobic fungus *Neocallimastix patriciarum,* adopts a β jelly-roll fold (Fig 4.1B), the crystal structure of enzyme revealed six subsites. The -3 and +3 subsites make extensive interactions with the substrate and thus they are kinetically significant (Vardakou, et al., 2008). The extended substrate-binding cleft of *Np*Xyn11A may also explain why the *Neocallimastix* enzyme displays unusually high activity, compared to other GH11 xylanases (Vardakou, et al., 2008). The k_{cat} against oat spelt xylan is 3.5 x 10^5 min^{-1},

which is over tenfold higher than the activity of other GH11 xylanases, reported to date (Vardakou, et al., 2008).

GH10 xylanases display a (β/α) 8-barrrel fold structure (Harris, et al., 1996) (Fig 4.1A). It can accommodate from four to seven xylose residues (Charnock, et al., 1998). The crystal structure of xylanases in complex with oligosaccharides have revealed detailed information on the interaction of these enzymes with their substrate in the proximal, -2 and -1 glycone subsites (Notenboom, et al., 2001b). Among GH10 xylanases, the *Cellvibrio japonicus* enzyme, *Cj*Xyn10A is the most active described to date, with a k_{cat} value against oat spelt xylan of 6.0 x 10^4 min^{-1} (Charnock, et al., 1997). Since *Cj*Xyn10A and *Np*Xyn11A are well characterised in our laboratories, and both show unusually high catalytic activities for oat spelt xylan, the two enzymes were fused to CBM2b1,2 and CBM15 xylanases. Their capacity to remove grass stains was evaluated.

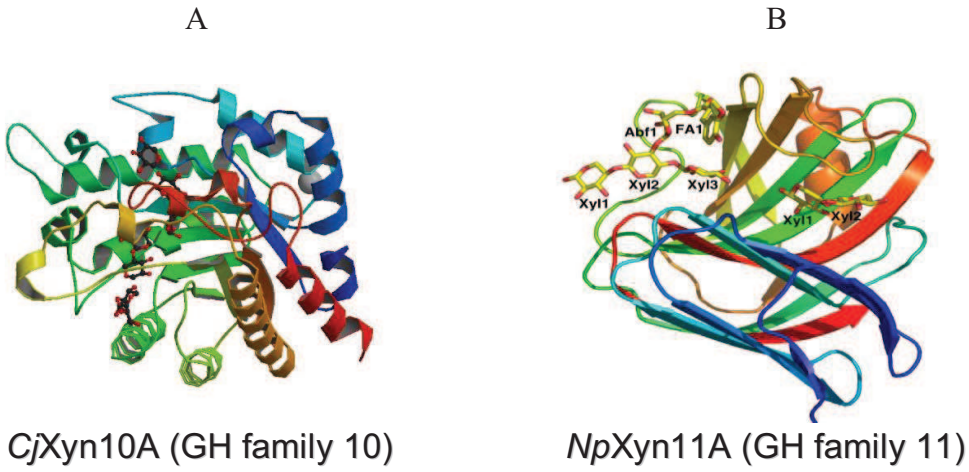


Fig 4.1 Crystal structures of *Cj*Xyn10A and *Np*Xyn11A (PDB code 1CLX for *Cj*Xyn10A; PDB code 2C1F for *Np*Xyn11A) (Harris, et al., 1996; Vardakou, et al., 2008).

4.2 Objectives

To test hypothesis that *Cj*CBM15 and *Cf*CBM2b1,2, which were shown to bind to grass stains (Chapter 3), enhance the capacity of GH10 and GH11 xylanases to remove these stains from textiles.

4.3 Results

4.3.1 Constructs of CBM-Xylanase cloned for removing grass stains

*Cj*Xyn10A is a glycoside hydrolase in family 10 and consists of 611 amino acids. It contains a catalytic module from residue 267 to 608 in a full length sequence and two cellulose-specific CBMs. *Np*Xyn11A is a glycoside hydrolase in family 11 and consists of 607 amino acids. It contains a catalytic module from residue 44 to 485 in a full length sequence and two cellulose-specific CBMs. In this project, these two catalytic modules were designated as *Cj*Xyn10A-CD (or Xyn10A) and *Np*Xyn11A-CD (or Xyn11A). They were fused to CBM15 and CBM2b1,2 by a Proline-Threonine linker (PT linker) (Fig 4.2). The CBM were at N terminus, as how it exists in bacterial enzymes in nature, i.e. normally it is at N terminus and comes in front of the catalytic module in a whole sequence of enzymes. As it is mentioned in Chapter 3, CBM15 is the family 15 CBM from *Cellvibrio japonicus* xylanase 10C, and CBM2b1,2 is the family 2b CBM from *Cellulomonas fimi* Xylanase 11A. The cloning work was carried out by Dr. A. Rogowski.

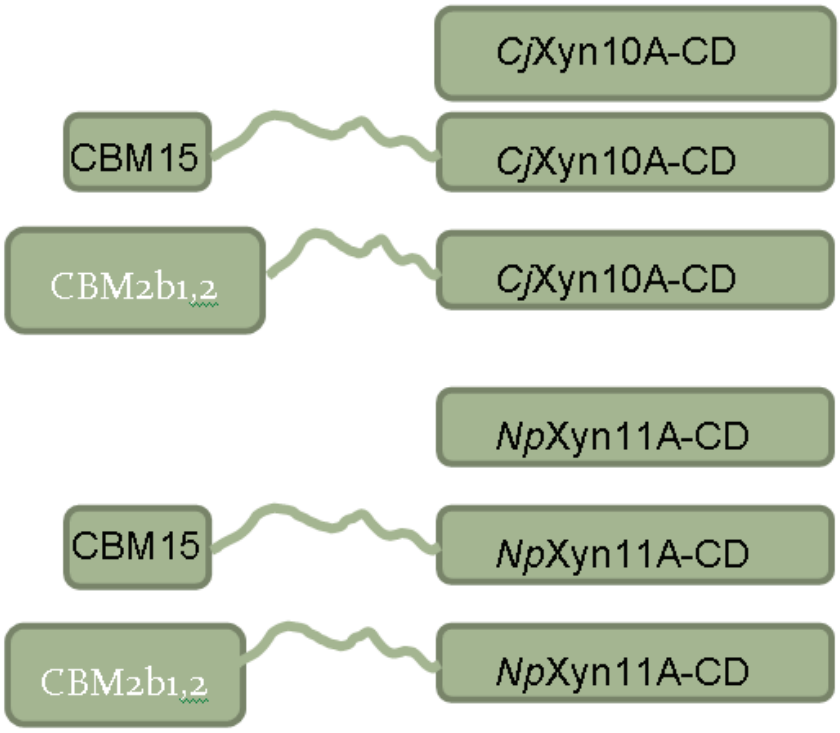


Fig 4.2 Modular architecture of CBM-Xylanase fusion constructs cloned for this project

The catalytic modules from *Cellvibrio japonicus* Xyn10A (38 KDa) and *Neocallimastix patriciarum* Xyn11A (24 KDa) are designated as *Cj*Xyn10A-CD and *Np*Xyn11A-CD. CBM15 (18 KDa) is the family 15 CBM from *Cellvibrio japonicus* xylanase 10C, and CBM2b1,2 (22 KDa) is the family 2b CBM from *Cellulomonas fimi* xylanase 11A, as mentioned in Chapter 3. The two catalytic modules were fused to CBM15 and CBM2b1,2 by a Proline-Threonine linker (PT linker), with the CBM modules at N terminus. The cloning work was carried out by Dr. A. Rogowski.

4.3.2 Over expression and purification of CBM-Xylanases constructs

*Cj*Xyn10A and *Np*Xyn11A were over-expressed in *E. coli* and, as the proteins contain His_6 tags, were purified by immobilized metal ion affinity chromatography. The proteins were purified to a high level of homogeneity (> 90 % pure), and had molecular weights of the predicted size (Table 4.1). Thus, the enzymes were not processed during expression or purification (Fig 4.3).

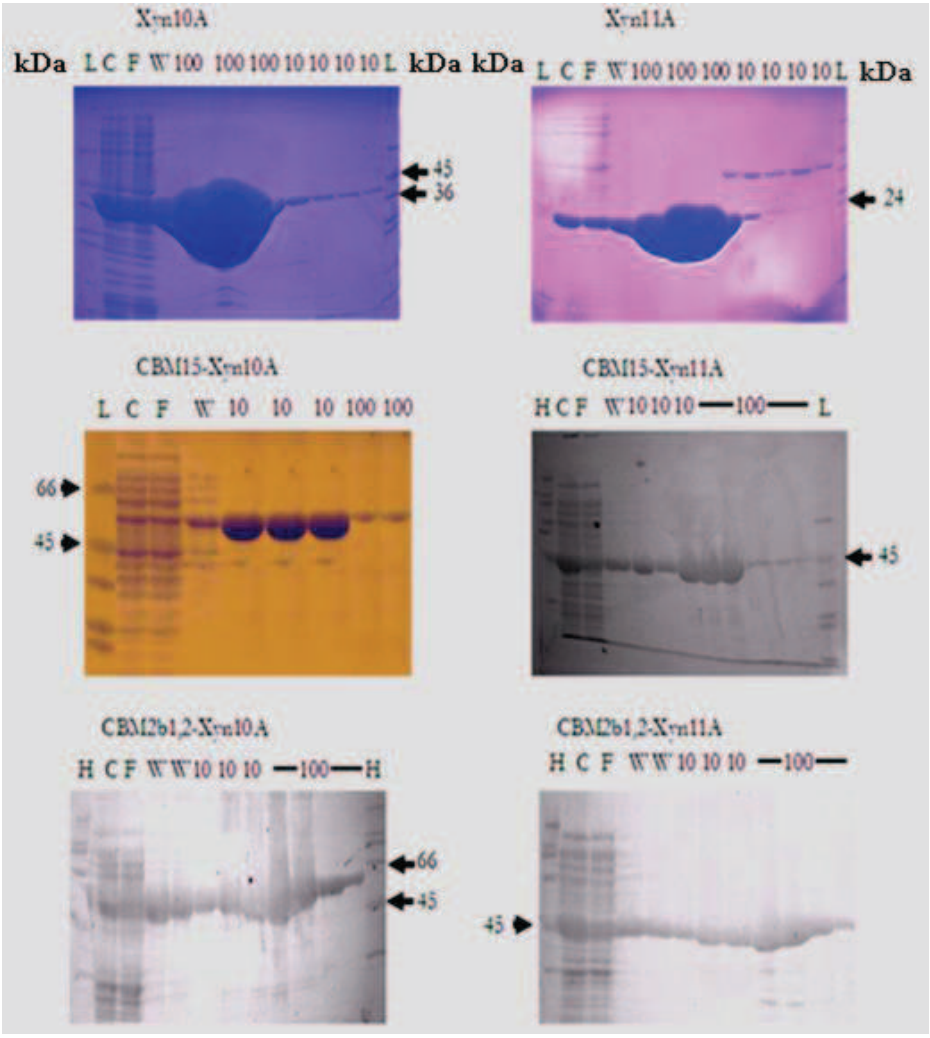


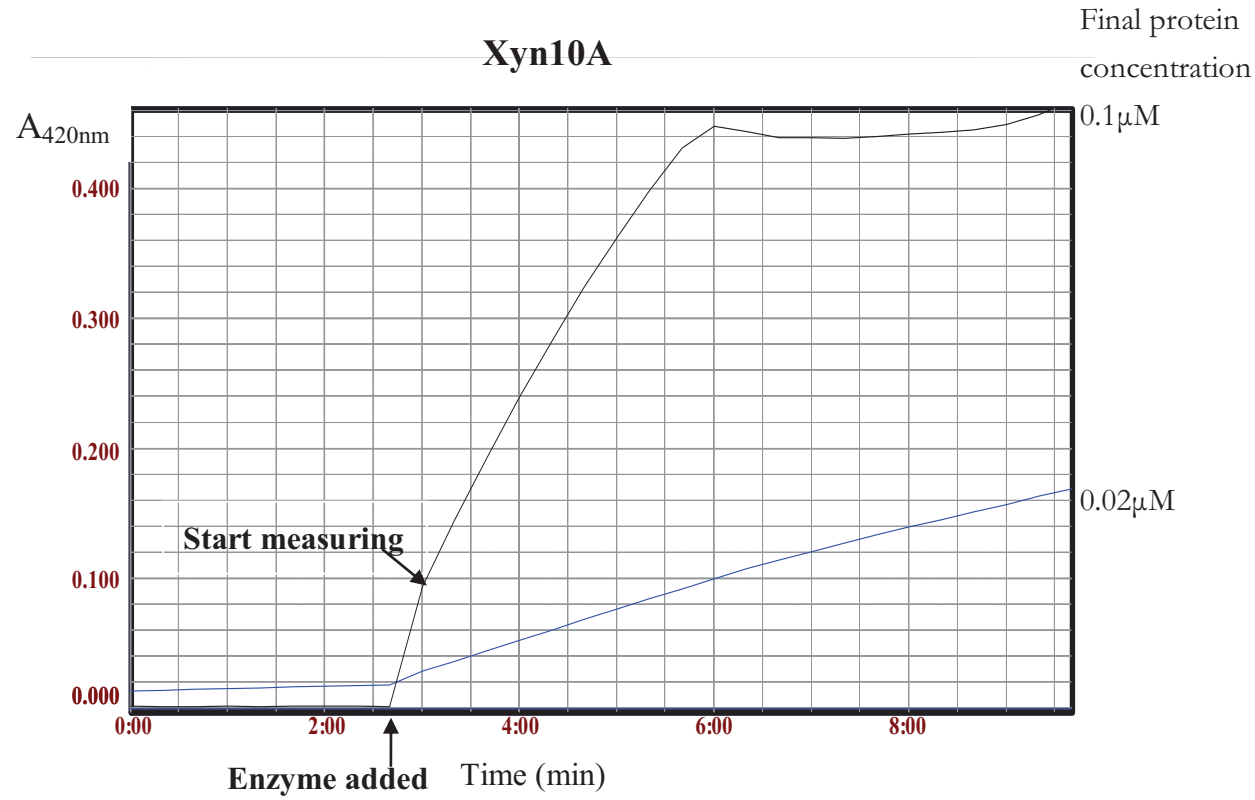
Fig 4.3 SDS-PAGE of Xyn10A, Xyn11A, CBM15-Xyn10A, CBM15-Xyn11A, CBM2b1,2-Xyn10A and CBM2b1,2-Xyn11A purified by IMAC Purification was carried out as described in Section 2.1.24. A 5 µl aliquot of cell free extract (C) and flow-through (F) was loaded, then 10 µl of elution with 20 mM Tris/HCl, pH 8.0, containing 300 mM NaCl and 10 mM imidazole (10) and 10 µl of elution with 100 mM imidazole (100) . Lane M_H contained Sigma high molecular weight standards. Analysis was performed by SDS-PAGE, using a 12.5 % (w/v) polyacrylamide gel.

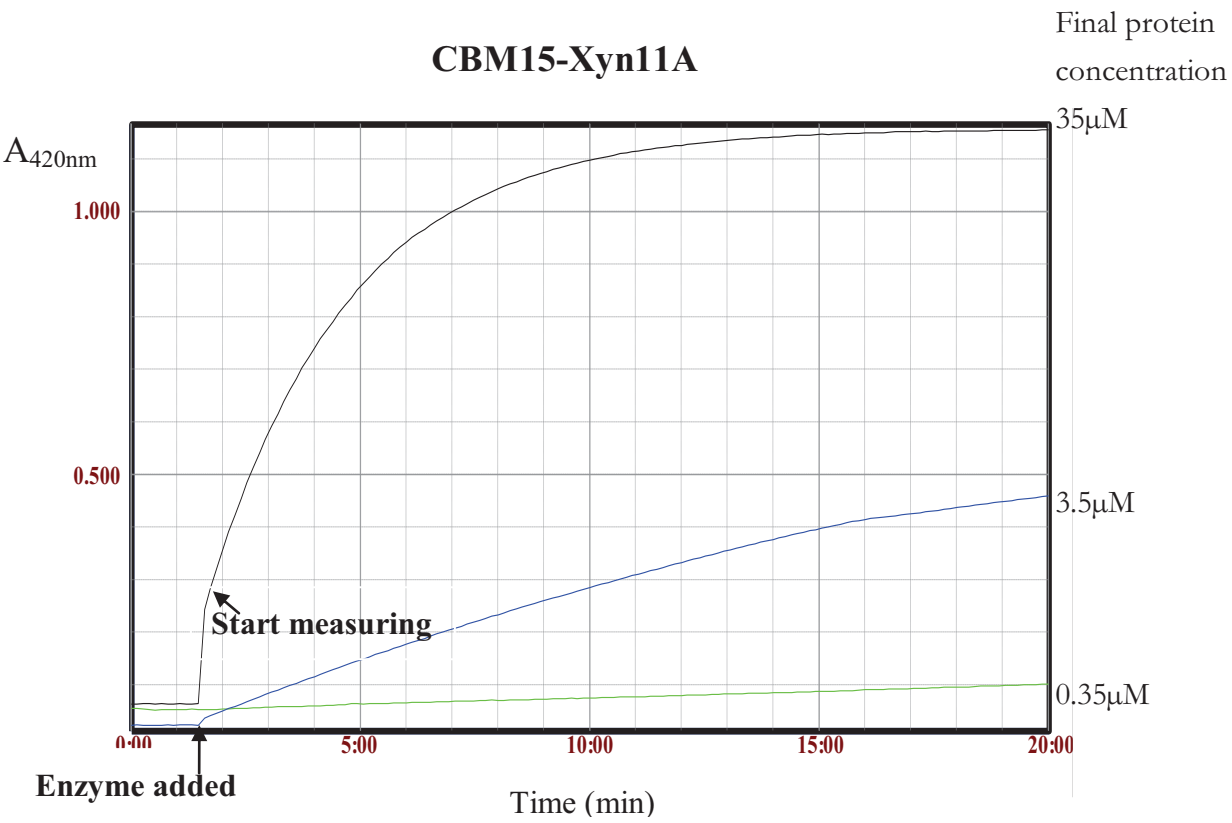
Protein	Extinction coefficient $M^{-1}cm^{-1}$	Predicted molecular Weight (kDa)	Experimentally Estimated size (kDa)
Xyn10A	60975	38.45	38
Xyn11A	61808	24.16	24
CBM15-Xyn10A	79790	57.00	57
CBM15-Xyn11A	79995	43.27	43
CBM2b1,2-Xyn10A	98695	59.69	60
CBM2b1,2-Xyn11A	99600	45.40	45

Table 4.1 Summary of extinction coefficient and molecular weight of different CBM-Xylanases The experimentally estimated sizes of the proteins were determined by SDS-PAGE of the purified enzymes (Fig 4.3).

4.3.3 Standardization of protein activities using p-nitrophenyl-ß-D-xylobiose (PNP-X_2) assay

Not all enzyme molecules are guaranteed to be active after protein expression and purification, due to minor protein aggregates or misfoldings *in vivo* (Mukhopadhyay, et. al., 1997). To determine the proportion of biologically active protein molecules, an enzyme assay using p-nitrophenyl-ß-D-xylobiose (PNP-X_2) was performed for each protein. These assays enabled the same amount of active xylanases to be included in the reactions designed to assess the capacity of the enzymes to remove grass stains from textiles. PNP-X_2 is a common substrate for both xylanase family 10 and family 11. The assay for each protein was performed in duplicates (Section 2.3.8). When enzymes were added to the reaction, the enzymes started to hydrolyse PNP-X_2. The ΔA_{420nm} was monitored as the reaction product, 4-nitrophenolate, has a maximum absorbance at 420 nm. Examples of the absorbance increase at 420 nm during the catalysis reactions of PNP-X_2 by different enzymes are shown in Fig 4.4. The assays identified the enzyme concentrations that catalyzed at a rate of ΔA_{420nm} of 0.14~0.24 min^{-1} (Table 4.2).


Xyn10A
Final protein concentration
A420nm
0.400
0.300
0.200
0.100
0.000
0.1μM
0.02μM
Start measuring
0:00
2:00
4:00
6:00
8:00
Enzyme added
Time (min)


CBM15-Xyn11A
Final protein concentration
A420nm
1.000
0.500
35μM
3.5μM
0.35μM
Start measuring
0:00
5:00
10:00
15:00
20:00
Enzyme added
Time (min)

Protein	Protein final concentration (μM)	Gradient ($\Delta A_{420nm} \cdot min^{-1}$)	*k*cat (min^{-1})
Xyn10A	0.1	0.14	140
CBM15-Xyn10A	0.5	0.14	28
CBM2b1,2-Xyn10A	0.1	0.14	140
Xyn11A	37	0.24	0.65
CBM15-Xyn11A	35	0.25	0.71
CBM2b1,2-Xyn11A	36.6	0.24	0.66

Table 4.2 The *k*cat values and the correspondent final enzyme concentrations used in the PNP-X_2 assay The reactions were carried out at 37 °C, as described in Section 2.3.8. The extinction coefficient of the product, 4-nitrophenolate is ~10,000. *k*cat are in mol of product [p-nitrophenolate]/mol of enzyme/min.

From the results summarized in Table 4.2, Xyn10A and CBM2b1,2-Xyn10A are five times more active than CBM15-Xyn10A. Thus, in the detergent assays, used to assess the removal of grass stains, the dosage of CBM15-Xyn10A was five times more than Xyn10A and CBM2b1,2-Xyn10A. The relatively low activity of CBM15-Xyn10A could reflect misfolding during post translation *in vivo* or digestion by protease *in vivo* or after being extracted and purified *in vitro*. However, as the size of the CBM15-Xyn10A is close to the predicted molecular weight, protease cleavage is unlikely. The molar ratio of Xyn10A:CBM2b1,2-Xyn10A: CBM15-Xyn10A used in the detergent assays was 1:1:5, to ensure a standard amount of functional enzyme was used.

The activity of Xyn11A, CBM15-Xyn11A and CBM2b1,2-Xyn11A were similar. The final enzyme concentrations corresponding to the specific activity of [4-nitrophenolate] $[enzyme]^{-1}$ min^{-1} at a similar rate of ~0.07 min^{-1} in the reactions were 37, 35 and 36.6 μM for Xyn11A, CBM15-Xyn11A and CBM2b1,2-Xyn11A, respectively. Therefore for the subsequent detergent assays, the same dosage of the three GH11 enzymes was used. To achieve the same enzyme activity against PNP-X_2, the concentration of the

GH11 enzymes were ~215 higher than the concentration of the GH10 xylanases. This indicates PNP-X_2 is much more easily cleaved by *Cj*Xyn10A than *Np*Xyn11A.

4.3.4 Cleaning of Grass stain in Tergo (washing machines) using CBM detergent technologies

The detergent assays were carried out in 1 L reactions in Tergo washing machines. The buffer consisted of 5 mM Tris/HCl buffer, pH 8.0, containing 0.001 % (v/v) Mono Propylene Glycerol (MPG*)*, 2.4 mM Ca^{2+}, surfactant Formulation A (also called Maradona, currently used in Persil Small & Mighty), or the company's most promising surfactant containing a high percentage of SLES (Formulation B, or high SLES blend), containing 4.2 % LAS, 4.2 % Neodal and 16.8 % SLES, were included to minimize non specific binding of protein to cotton, as described in Chapter 3.

The assays included two controls. The first one was the novel formulation currently used in the Persil 'Small & Mighty' brand, which contains three enzymes, lipase, amylase and protease (we used savinase as the proteolytic enzyme) all at 0.5 %. The second control was 'Non Bio', which lacked enzymes but contained surfactant (Surfactant Formulation A) and calcium in distilled water. The composition of each of these formulations, which were included with the xylanases, are listed in Table 4.3. Test cloths with grass stains arrayed were washed in the Tergo washing machines at 30 °C, according to Unilever's established protocol. The reflectance of the arrayed cotton at 460 nm was measured before and after wash (ΔR_{460}) using the stain monitor (Macbeth 1500/Plus). The ΔR_{460} values indicated the extent of grass stain removal. The numerical data of each treatment is displayed in Table 4.4 and the results are shown in bar charts (Fig 4.5) is based on the average of six replicates; duplicate

samples were assayed in three discrete experiments, which were carried out in the same wash cycle. The width of textiles and grass swatches arrayed in the centre are 2 and 4 cm, respectively.

Formulation	Surfactant	Lipase	Amylase	Protease	Cell wall degrader
F1 Persil Small & Mighty (control 1)	Current blend Formulation A	√	√	√	0
F2 Non Bio (control 2)	Formulation A*	√	√	0	0
F3 High SLES blend (control 3)	High SLES blend (Formulation B)	√	√	0	0
F4	Formulation A*	√	√	0	Xyn10A
F5	Formulation B				Xyn10A
F6	Formulation A*	√	√	0	CBM2b1,2-Xyn10A
F7	Formulation B	√	√	0	CBM2b1,2-Xyn10A
F8	Formulation A*	√	√	0	CBM15-Xyn10A
F9	Formulation B	√	√	0	CBM15-Xyn10A
F10	Formulation A*	√	√	0	Xyn11A
F11	Formulation B	√	√	0	Xyn11A
F12	Formulation A*	√	√	0	CBM2b1,2-Xyn11A
F13	Formulation B	√	√	0	CBM2b1,2-Xyn11A
F14	Formulation A*	√	√	0	CBM15-Xyn11A
F15	Formulation B	√	√	0	CBM15-Xyn11A

Table 4.3 Table summarizing the composition in each wash test

The concentration of Xyn10A, CBM15-Xyn10A and CBM2b1,2-Xyn10A were 1 mg/L, 5 mg/L and 1 mg/L (~0.02 μM, 0.1 μM and 0.02 μM). The concentration of Xyn11A, CBM15-Xyn11A and CBM2b1,2-Xyn11A were 3 mg/L (~0.1 μM). Formulation A contained the current blend of additives (surfactants, Ca^{2+}) and enzymes used in Persil Small & Mighty. Formulation A* contained the additives used in Persil Small & Mighty but did not include the enzymes. Ideally, the level of the GH10 and GH11 xylanases should have been added at a ratio of 1:215. However, such a loading would be unrealistic with respect to laundry processes. Thus, the experiment was designed to compare the different GH10 and GH11 variants, but no comparison across the two families were possible.

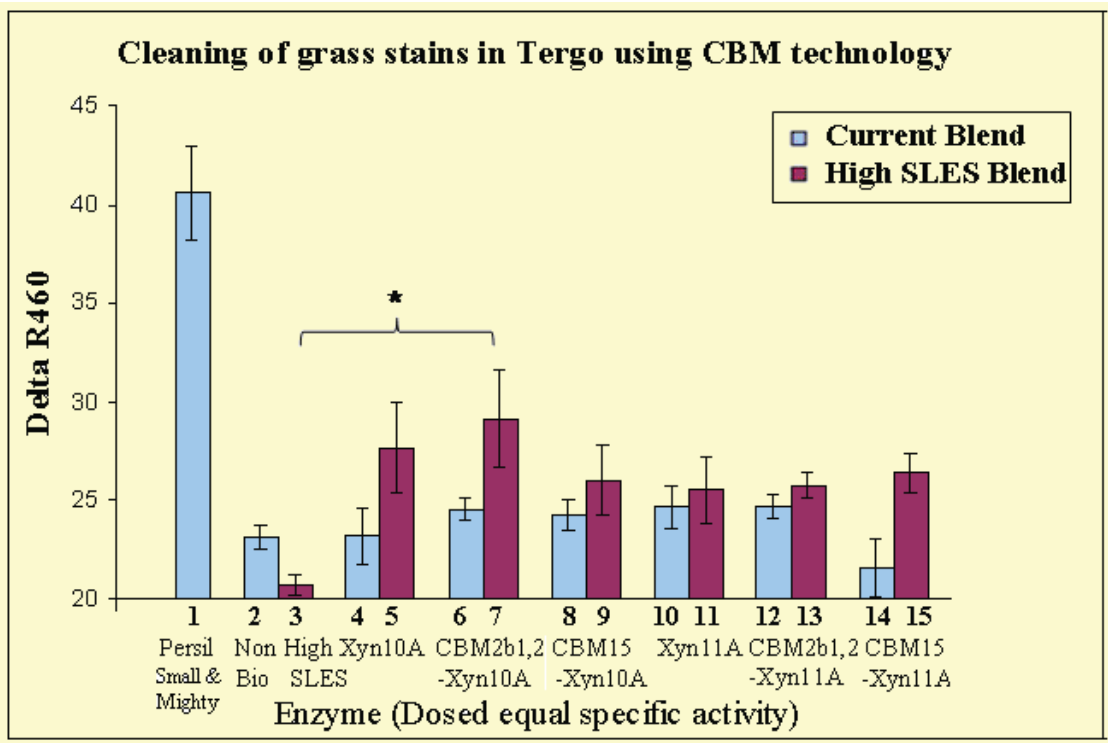


Fig 4.5 Grass stain removal from cotton through Tergo machine washing using different enzymes added at equal specificity activity

All enzyme activities were quantified using PNP-X_2 as substrate, and the enzymes dose in GH10 and 11 were standardised, respectively. Degree of stain removal was quantified as the difference in R_{460} (applied as measure of whiteness) of the cotton before and after washing. The figrure shows the data graphically; values are plotted as mean ± Standard error for n=6. All values differed significantly from "Persil Small & Mighty" according to analysis by one way ANOVA followed by Tukey's post test. *P<0.05 by one way ANOVA followed by Tukey's post test, after exclusion of the data for "Persil Small & Mighty". F3 control data was provided by Dr. Stevenson, Unilever Laundry Department (Appendix G).

Formulation	Chart name	Average Delta R_{460}	Standard Error n=6
1	**Persil Small & Mighty**	**40.62**	**2.39**
2	**Non Bio**	**23.15**	**0.58**
3	**High SLES**	**20.76**	**0.56**
4	Xyn10A	23.18	1.4
5	Xyn10A	27.68	2.27
6	CBM2b1,2-Xyn10A	24.57	0.57
7	CBM2b1,2-Xyn10A	29.14	2.47
8	CBM15-Xyn10A	24.26	0.77
9	CBM15-Xyn10A	26.04	1.78
10	Xyn11A	24.7	1.1
11	Xyn11A	25.55	1.73
12	CBM2b1,2-Xyn11A	24.7	0.6
13	CBM2b1,2-Xyn11A	25.79	0.63
14	CBM15-Xyn11A	21.56	1.48
15	CBM15-Xyn11A	26.41	1.02

Table 4.4 Grass stain removal from cotton through Tergo machine washing using different enzymes added at equal specificity activity

The Table shows the same data in Fig 4.5, in tabular format. All enzyme activities were quantified using PNP-X_2 as substrate, and the enzymes dose in GH10 and 11 were standardised, respectively. Degree of stain removal was quantified as the difference in R_{460} (applied as measure of whiteness) of the cotton before and after washing. The data of the average and the Standard error for n=6 are presented in tabular format. F3 control data was provided by Dr. Stevenson, Unilever Laundry Department (Appendix G).

The presence of High SLES blend (Formulation B) enhanced the capacity of the family 10 and family 11 xylanases, to remove grass stains, compared to the current novel blend (Formulation A*). The statistical analysis showed that the only significant difference was between the control (bar 3 in Fig 4.5) and CBM2b1,2-Xyn10A (bar 7 in Fig 4.5) in condition that contains the high SLES blend. The p value was < 0.05 (Appendix G3.1). However, although the other xylanase constructus were not significantly different, they still appeared to release an average more grass dye than the control in the presence of SLES; i.e, the xylanases presented in bar 5, 9, 11, 13 and 15 had greater average ΔR_{460}, comparing to the control bar 3. Hence, it showed they removed the grass stains more than the control with no xylanases. When looking at the Xyn10A group individually, CBM2b1,2-Xyn10A performed better than CBM15-Xyn10A on grass stain removal. In conclusion, CBM2b1,2 helps Xyn10A to target grass stains and that is consistent with Chapter 3, which shows that CBM2b1,2 module binds to grass stains better than CBM15 when SLES is present.

Unilever laundry research department has a long history of developing new technologies to increase the efficiency of laundry detergents. According to research experience of laundry specialists, many technologies tested gave a beneficial effect in small scale experiments, which when scaled up, did not improve the cleaning process. The CBM technologies used here appeared to deliver an observable increase in grass stain removal effect at the scale-up stage. It should be noted that the use of xylanases were less efficient than Persil Small & Mighty. The difference between these cocktails was that in Persil Small & Mighty the xylanase was replaced by protease, suggested that protease has a much bigger effect in stain removal than xylanases.

4.3.5 Resistance of CBM-Xylanases to savinase treatment

Stability assays were carried out in collaboration with Dr. Nic Brown (in Gilbert's group, Newcastle University) to test the resistance of CBMs, xylanase catalytic modules and the linker in CBM-Xylanase fusions to protease (savinase) attack; reflecting the presence of savinase (at 0.01 mg/ml final concentration) in Unilever laundry powder. The proteins at 0.5 mg/ml were incubated at 37 °C with savinase (at 0.01 mg/ml final concentration), in either 50 mM Tris/HCl buffer pH 8.0 or the buffer containing 1 g/L SLES; 0.2 mM $CaCl_2$ and 5 mM NaCl. Savinase was inactivated by phenylmethylsulfonyl fluoride (PMSF) at 1 mM final concentration. The degradation of proteins was assessed by SDS-PAGE and the residual xylanase activity was analyzed using the DNSA reducing sugar assay (Section 2.3.8).

The results showed that the surfactant did not influence the effect of the savinase on xylanases and CBMs. CBM15 was resistant to savinase treatment at 0.01 mg/ml final concentration, as SDS-PAGE showed that the size of the protein module did not change after incubation with the protease (Fig 4.6, Panel A and B). Furthermore, AGE (Fig 4.7) showed that the protease treated CBM15 retained its capacity to bind oat spelt xylan (OSX). CBM2b1,2 was less resistant and SDS-PAGE showed some protein degradation (Fig 4.6, Panel C and D). Xyn10A was resistant to savinase treatment, as the size and activity of the xylanase was unaffected by protease treatment (Fig 4.8, Panel A and B). Similarly, SDS-PAGE indicated that Xyn11A was not degraded by the savinase (Fig 4.8, Panel C and D). However, enzyme assays showed a ~18 % decrease in activity of Xyn11A after protease treatment suggesting that some degradation had occurred (Fig 4.9, Panel B and D). Thus, the lack of low molecular weight fragments (Fig 4.8, Panel C and D) suggests that once an initial cleavage has

occurred in Xyn11A, the resultant loss in tertiary structures leads to rapid degradation of the enzyme molecules (Fig 4.9, Panel B and D).

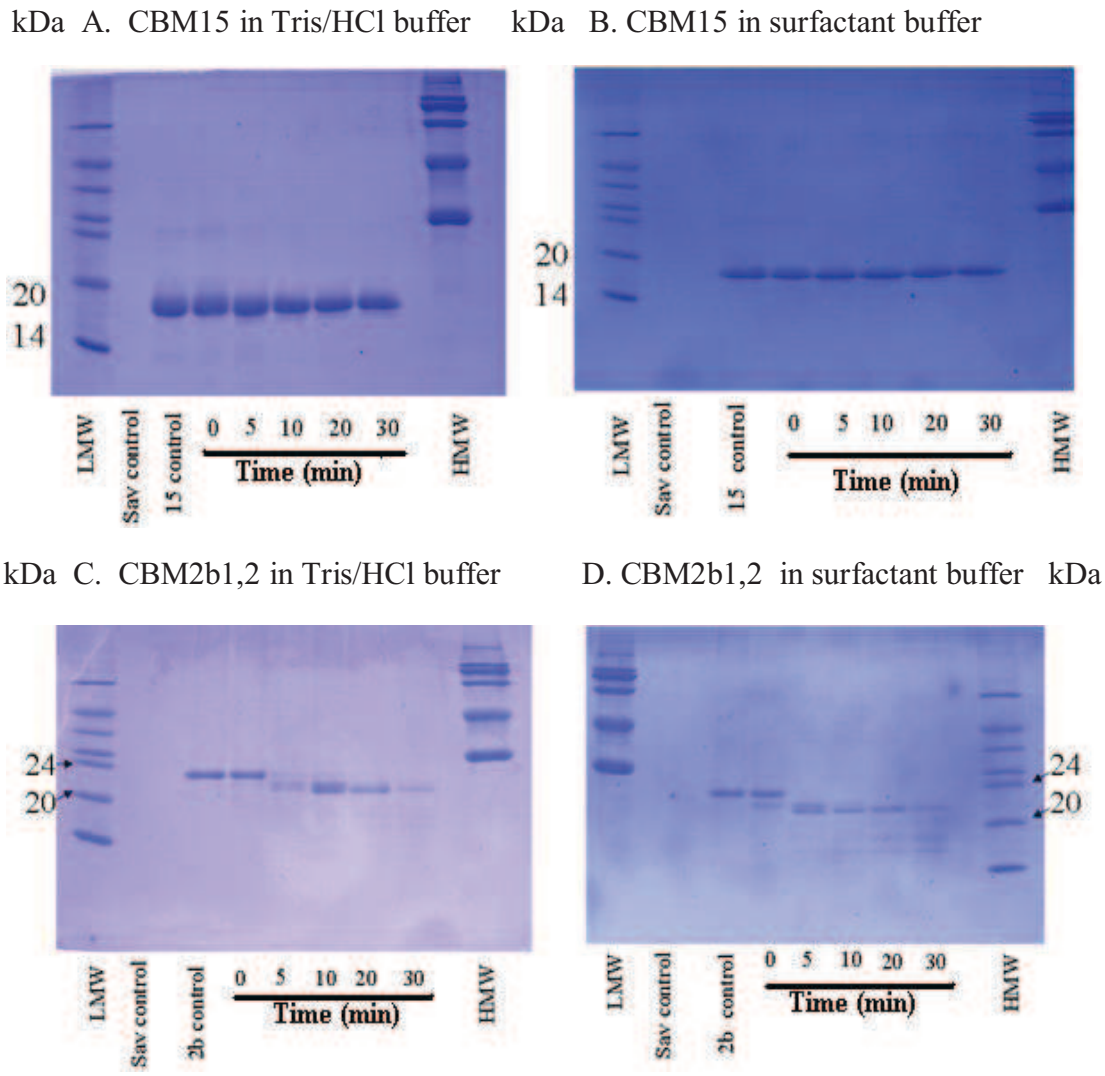


Fig 4.6 Stability assay of CBM15 and CBM2b1,2 in the presence of savinase

CBM15 and CBM2b1,2 (final concentration at 0.5 mg/ml) were incubated with savinase (final concentration at 0.01 mg/ml) at room temperature, in either 50 mM Tris/HCl buffer pH 8.0 (Panels A, C) or buffer containing the surfactant 1 g/L SLES; 0.2 mM $CaCl_2$ and 5 mM NaCl (Panels B, D). Savinase was inactivated by the inhibitor PMSF at 1 mM final concentration at 0, 10, 20 and 30 min time intervals. The degradation of proteins was assessed by SDS-PAGE using a 12.5 % gel.

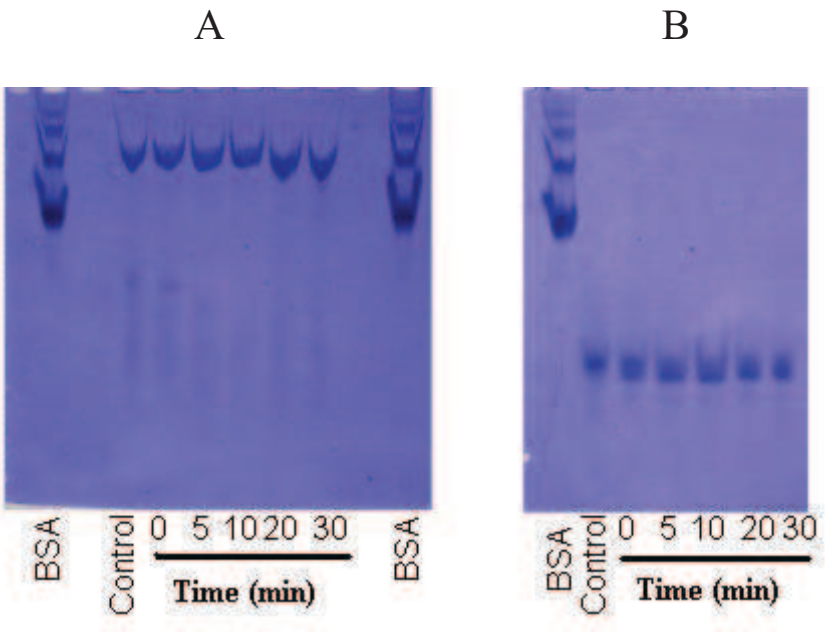


Fig 4.7 Non-denaturing affinity gel electrophoresis of CBM15

Non-denaturing affinity gel electrophoresis (AGE) was carried out, using 12.5 % (w/v) polyacrylamide gels in a Tris/glycine buffer system, pH 8.3. BSA, which was the negative control, CBM15 control incubation without savinase and CBM15 incubated with savinase at 0, 5, 10, 20 and 30 min were subjected to non-denaturing affinity gel electrophoresis in the presence (Panel A) and absence (Panel B) of 0.1 % oat spelt xylan. A 10 µg aliquot of protein was loaded into each well.

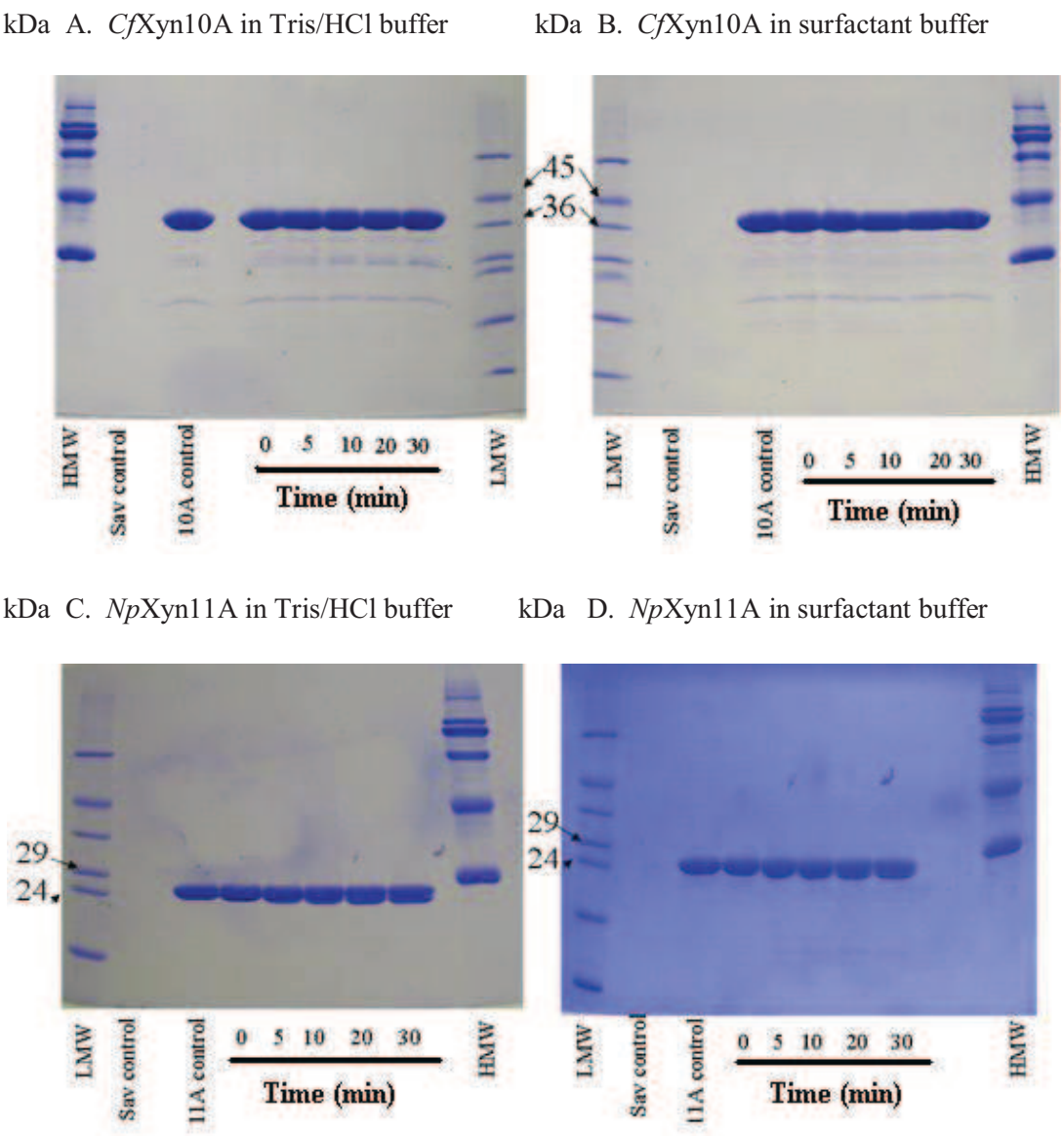
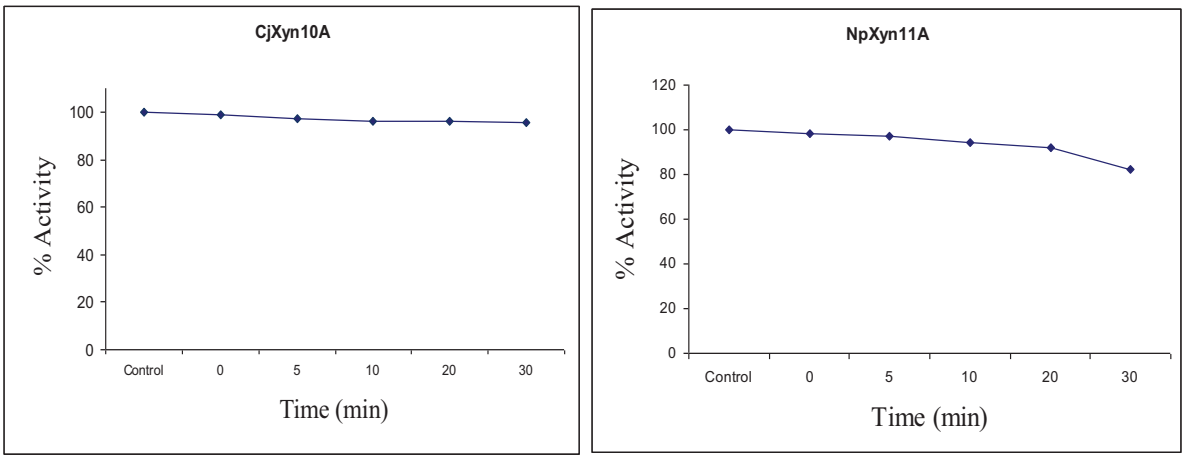


Fig 4.8 Stability assay of *Cj*Xyn10A and *Np*Xyn11A in the presence of savinase

*Cj*Xyn10A and *Np*Xyn11A (final concentration at 0.5 mg/ml) were incubated with savinase (final concentration at 0.01 mg/ml) at room temperature, in either 50 mM Tris/HCl buffer pH 8.0 (Panels A, C) or surfactant buffer containing 1 g/L SLES; 0.2 mM $CaCl_2$ and 5 mM NaCl (Panels B, D). Savinase was inactivated by its inhibitor PMSF at 1 mM final concentration at 0, 10, 20 and 30 min time intervals. The degradation of proteins was assessed by SDS-PAGE using a 12.5 % gel.

A. *Cj*Xyn10A **B. *Np*Xyn11A**



C. *Cj*Xyn10A

***Cj*Xyn10A Time (min)**	**Concentration (µg/ml)**	**% Activity *Cj*Xyn10A**
Control	159.25	100
0	157.87	99.1
5	155.25	97.5
10	153.67	96.5
20	153.09	96.1
30	152.7	95.9

D. *Np*Xyn11A

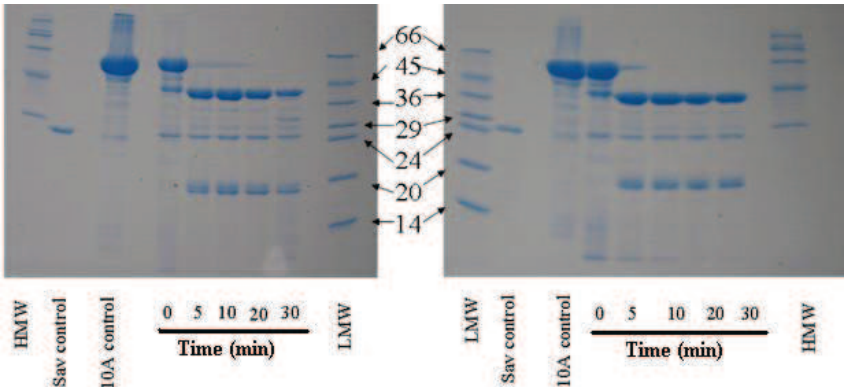
***Np*Xyn11A Time (min)**	**Concentration (µg/ml)**	**% Activity *Np*Xyn11A**
Control	212.96	100
0	209.26	98.3
5	207.5	97.4
10	200.67	94.2
20	195.8	92
30	174.93	82.1

Fig 4.9 Residual activity of *Cj*Xyn10A and *Np*Xyn11A against oat spelt xylan in the presence of savinase

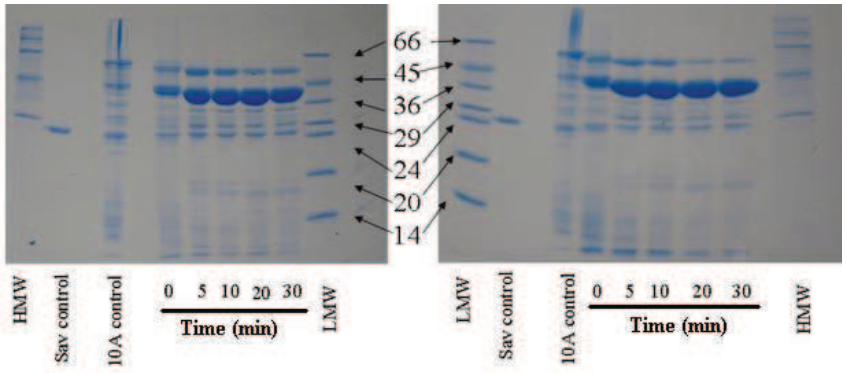
The reactions were carried in 50 mM Tris/HCl buffer pH 8.0, at 30 °C. Panel A and B display the catalytic activity of *Cj*Xyn10A and *Np*Xyn11A against oat spelt xylan in 30 minutes, respectively, in the presence savinase at 0.01 mg/ml final concentration. Panel B and D displays the concentration of reducing sugar produced at different time intervals during catalysis of oat spelt xylan by *Cj*Xyn10A and *Np*Xyn11A, respectively. The concentration of reducing sugar was assayed using DNSA, as described in Section 2.3.8, and the enzyme activity of *Cj*Xyn10A and *Np*Xyn11A in the presence of savinase in 30 minutes was also calculated and displayed in tabular format.

When CBM-Xylanase fusions were treated with savinase, the proteins were degraded after 5 min incubation with savinase. CBM15-Xyn10A was degraded into two protein species of 38 kDa and 18 kDa (Fig 4.10, Panel A and B), CBM15-Xyn11A was cleaved into 24 kDa and 18 kDa peptides (Fig 4.10, Panel E and F). CBM2b1,2-Xyn10A was degraded into protein species of 38 kDa and some fragments less than 24 kDa (Fig 4.10, Panel C and D), CBM2b1,2-Xyn11A was cleaved into 24 kDa and many fragments less than 24 kDa peptides (Fig 4.10, Panel G and H). The sizes of these proteins are similar to the predicted molecular weight of CBM15 (18 kDa), CBM2b1,2 (22 kDa) and the catalytic modules of Xyn11A (24 kDa) and Xyn10A (38 kDa). It would appear, therefore, that the savinase has cleaved the linker sequences (Fig 4.2) in CBM15-Xyn10A and CBM15-Xyn11A, generating the individual modules of these fusion proteins as discrete entities. For CBM2b1,2-Xylanase fusions, only a single polypeptide was evident, i.e. the catalytic modules were intact but CBM2b1,2 appeared to be degraded (Fig 4.10, Panel C, D, G and H). The activity of CBM15-Xyn10A, CBM2b1,2-Xyn10A, CBM15-Xyn11A and CBM2b1,2-Xyn11A decreased to 40 %, 70 %, 85 % and 90 %, respectively (Fig 4.11). CBM15-Xyn10A activity diminished following savinase treatment (Fig 4.11, Panel A and B), and the other three enzymes retained more than 70 % of their xylan degrading activities (Fig 4.11, Panel C to H).

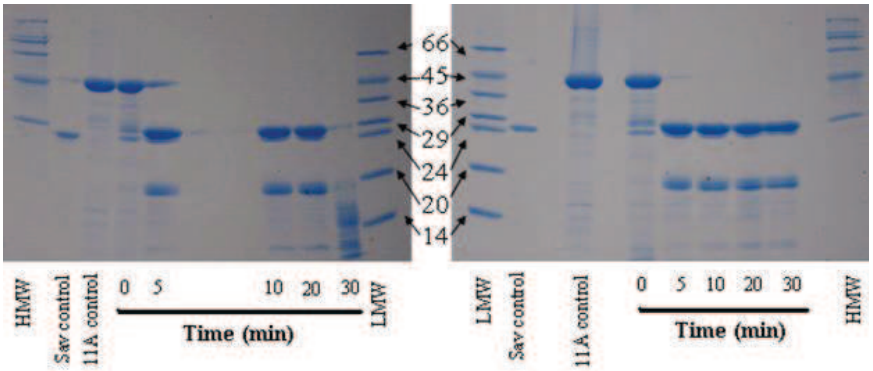
A.CBM15-Xyn10A in Tris/HCl buffer kDa B. CBM15-Xyn10A in surfactant buffer



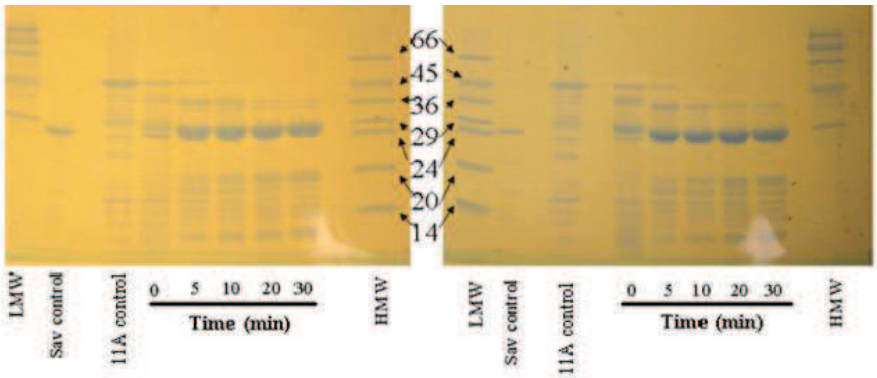
C.CBM2b1,2-Xyn10A in Tris/HCl buffer D. CBM2b1,2-Xyn10A in surfactant buffer



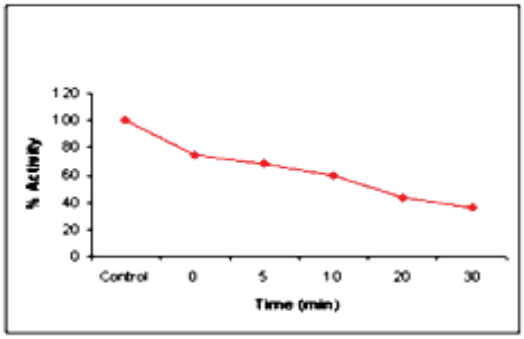
E.CBM15-Xyn11A in Tris/HCl buffer F. CBM15-Xyn11A in surfactant buffer



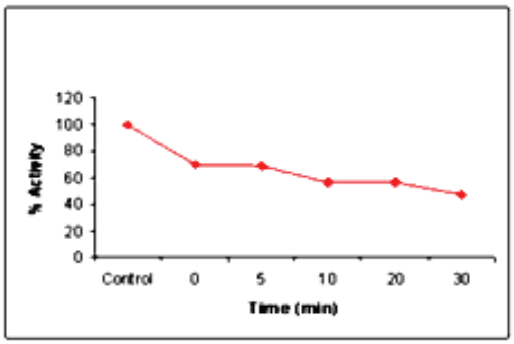
G.CBM2b1,2-Xyn11A in Tris/HCl buffer H. CBM2b1,2-Xyn11A in surfactant buffer



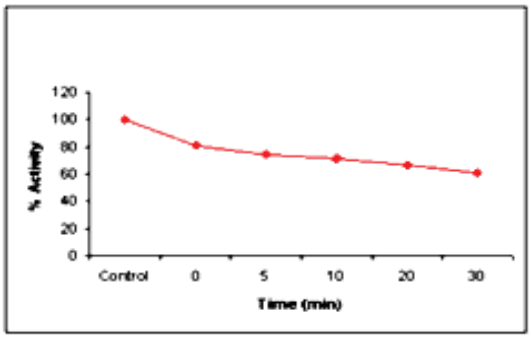
A. CBM15-Xyn10A in Tris/HCl buffer



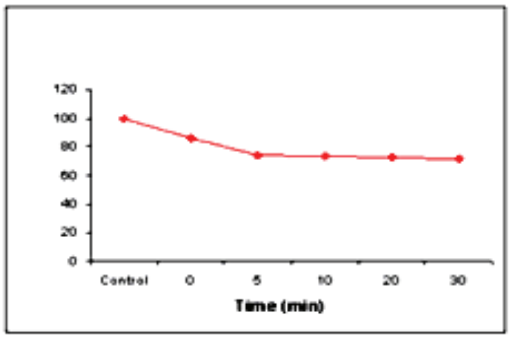
B. CBM15-Xyn10A in surfactant buffer



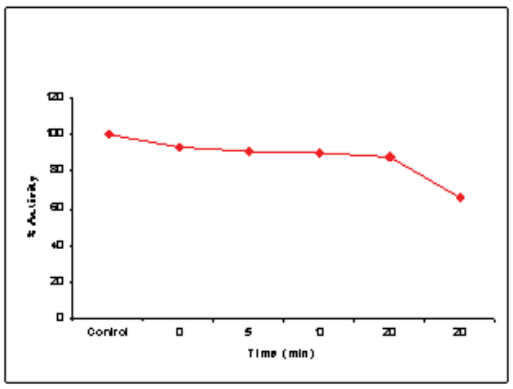
C. CBM2b1,2-Xyn10A in Tris/HCl buffer



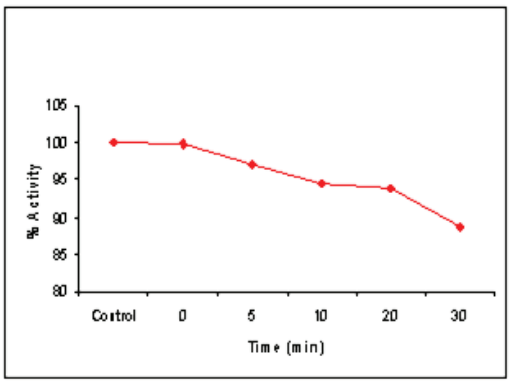
D. CBM2b1,2-Xyn10A in surfactant buffer



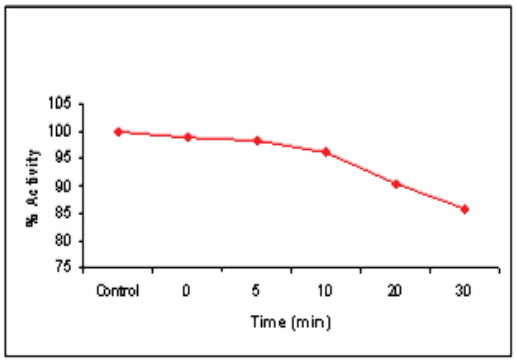
E. CBM15-Xyn11A in Tris/HCl buffer



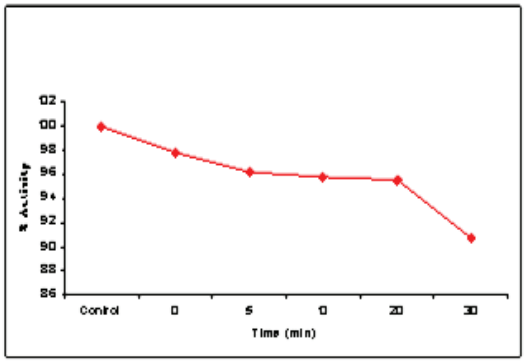
F. CBM15-Xyn11A in surfactant buffer



G. CBM2b1,2-Xyn11A in Tris/HCl buffer



H. CBM2b1,2-Xyn11A in surfactant buffer



4.4 Discussion

This chapter studied whether appending CBM15 and CBM2b1,2 to GH10 and GH11 xylanases would significantly enhance enzyme activity against the grass stains arrayed on textiles, in novel and high SLES surfactant formulations. The resistance of linker regions in CBM-Xylanase fusions and the respective modules to protease attack were also evaluated. Persil Small & Mighty showed a large increase in grass stain removal when savinase at 0.01 mg/ml, together with lipase and amylase, were added to Non-Bio surfactant solution containing lipase, amylase and savinase. The enzyme blends containing protease, lipase and amylase are routinely used as biodegradable, environmentally friendly alternatives to removing organic buildup from hard surfaces or laundry. The combined enzyme action solubilizes complex organic buildup, such as soil and grease, making it more water soluble and easier to remove by other laundry ingredient, for example, the CBM-Xylanase used here. The most important enzyme used in biological washing powder is savinase, which breaks down all proteins as it has a very relaxed specificity. By two-dimensional protein gels of cell wall preparations, it is estimated that there may be as many as several hundred proteins in the cell wall. Structural proteins, which accounts up to 5 % of the mass of plant cell walls, is a major class of proteins and includes hydroxyproline-rich glycoproteins (HRGP), arabinogalactan proteins (AGP), glycine-rich proteins (GRPs). Except GRPs, HRGPs and AGPs are glycosylated and cross linked to plant cell walls. For example, some AGPs are attached to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor. Apart from the main structural role, they can also be involved in signalling pathway and intracellular protein transport in plants during cell development. Thus, digesting these proteins by adding savinase into laundry powders will helps to open up

the interlocking polysaccharide chains, thus loosening cell walls. It will make the xylans embedded in the cell walls more accessible to xylanases, hence leading to an increase of grass stain removal.

There are two major barriers in laundry washes, one is the barrier between stain surface and water molecules; and the other is grass cell walls. The surface tension of the first barrier is reduced by surfactants (Bajpai and Tyagi, 2007) and the second barrier is targeted by xylanases. By experience of the laundry research specialists in Unilever, when protease is added, there is a possibility it enhances the grass stain removal ability of CBM-Xyn10A and CBM-Xyn11A, as protease may digest the protein composition in grass cells or and chloroplasts, thus it will loosen and expose grass cell walls and make it easier for CBM-Xylanases to access and break down grass stains.

Compared the Non-Bio control (bar 2 and 3 in Fig 4.5), although there is not a statistically significant difference between xylanase treatment and the control, there was a trend showing a small increase of grass stain removal in treatments that included the xylanases (bar 4-6, 8-13 and 15 in Fig 4.5). Grass contains polysaccharide-rich cell walls with a complex interlocking structure, in which xylan is the most abundant hemicellulose in the secondary walls. It is believed that xylan specific CBM–Xylanase fusions would attack some grass cells or chloroplasts which can be intact. Thus, GH10 and GH11 xylanases, which mediate the deconstruction of xylan in cell walls may weaken these structure and help grass stain removal.

Based on the PNP-X_2 assay results in Table 4.2, it showed GH10 enzymes were with k_{cat} at 28~140 min^{-1} and GH11 enzymes were with k_{cat} at 0.65~0.71 min^{-1}. Thus, GH10 enzymes were ~215 fold more active against the GH11 xylanases, it was not possible to

compare the enzyme performance across GH10 and GH11 families, unless the inputs of GH11 xylanases were 215 fold higher than the GH10 enzymes. In laundry formulation, protein concentrations are traditionally used at 3 mg/L. To run the experiment, two surfactant systems were studied and 6 replicates were used for each surfactant condition. Hence, 36~180 mg of each GH10 enzyme were consumed. Ideally, 7200 mg of GH11 enzyme would be needed for comparison between GH10 and GH11 families, but this amount of protein would take up to 12 weeks to prepare. Therefore, three enzymes within each family were standardized only, due to enzyme availability.

Comparing the catalytic activities of *Cj*Xyn10A and *Np*Xyn11A against PNP-X_2 to the data in literature, the *k*cat values of *Cj*Xyn10A and *Np*Xyn11A against PNP-X_2 were 5186 and 11 min^{-1}, respectively (Charnock, et al., 1998; Vardakou, et al., 2008), which are ~20 to 30 times higher than the *k*cat values obtained in this study. The reason for this is not clear. Also, it could also have been carried out to assess the catalytic activity of different xylanase constructs against xylan using DNSA assay for protein activity standardization before the detergent assay.

The *k*cat values from the literature indicated the enzyme activities of *Cj*Xyn10A and *Np*Xyn11A against xylan were 60137 and 350000 min^{-1}, thus GH10 xylanases are less active than GH11 enzymes assayed against purified xylans *in vivo* (Charnock, et al., 1998; Vardakou, et al., 2008). In contrast, recent data suggest that GH10 enzymes are more active against xylans embedded in plant cell walls than GH11 xylanases (Herve, et al., 2009). In view of the variable activity of GH10 and GH11 xylanases when substrate is presented in difference formats, and the difficulty in preparing gram quantities of the *Neocallimastix* enzyme, it was felt that only comparison of different forms of the enzymes within each family was valid in this study.

The data showed a trend in which CBM-Xyn10A fusions performed better than CBM-Xyn11A hybrid, especially when CBM2b1,2 was appended to the catalytic modules. This indicates that grass stain removal by the GH10 enzymes was potentiated by CBMs to a greater extent than the GH11 xylanases. This observation is consistent with recent work (Herve, et al., 2009), which dissected the capacity of CBM2b and CBM15 to enhance the catalytic activity of GH10 and GH11 xylanases against intact cell walls of tobacco stem sections. The data showed that appended CBM2b1,2 potentiated the activities of both xylanases by two fold, comparing to the respective catalytic modules alone, whereas appending CBM15 only enhance Xyn11A activity by around 50 %, but did not increase the catalytic activity of Xyn10A. The underlying principle is the topology of the binding clefts in CBM2b1,2 and the catalytic modules of GH10 xylanases are wider clefts than CBM15 and Xyn11A enzymes (Herve, et al., 2009), thus they are more likely to interact with xylan chains embedded in grass stains. Thus *Np*Xyn11A and CBM15 can only target highly exposed xylan chains on the stain surface. Therefore, even if the PT linker in CBM-Xylanases (Fig 4.2) are attacked by savinase, the catalytic module of Xyn10A might still be able to hydrolyze xylans that are in sparse association with other cell wall components, whereas Xyn11A might benefit from the targeting function of these modules to deconstruct such structures. It is possible that xylans in grass stains, which are available to the CBMs is at a low level, or absent. Thus, only a proportion of xylans could be accessed by Xyn11A in stains, whereas such substructure may be more suited to the topology of the Xyn10A substrate binding site.

Recent studies have shown that the enzymatic removal of pectic homogalacturonan can increase the access of CBMs to cellulose, and is required to reveal xylan and xyloglucan in discrete regions of cell walls (Herve, et al., 2009). These observations

indicate that even in cell walls exposed by sectioning, polysaccharide interactions could restrict access of enzymes to their insoluble substrates.

Besides xylanases, arabinofuranosidase, pectic lyases and mannanases also play an important role in plant cell wall degradation, as cell walls contain a range of interacting arabinoxylan, homogalacturonan and mannans (Gilbert, 2010). Arabinofuranosidases remove arabinosyl residues from arabinoxylan in wheat endosperm cell walls, and is promoted by appended xylan-binding CBMs (Herve, et al., 2009). The GH51 arabinofuranosidase, Abf51A, from *C. japonicus,* releases *O2* and *O3* linked arabinofuranose side chains from monosubstituted backbone residues in xylan and arabinan (Herve, et al., 2009). Fusing the enzyme to a xylan-binding CBM2b1,2, resulted in a substantial potentiation in activity against plant cell walls.

Cell walls substructures also contain different amounts of pectic homogalacturonan. It has been shown that removal of pectic homogalacturonan can increase the access of CBM to cellulose and reveal xylan and xyloglucan in cell walls (Herve, et al., 2009) that are susceptible to enzymes attack. Generally, pectic polymers are considered the most accessible polymers in primary cell walls, however, some are attached to cellulose and/or hemicellulose polymers in intractable cell wall regions, which are more recalcitrant (Herve, et al., 2009). These regions could become accessible to enzyme attack by appending nonpectic binding modules to appropriate pectinases (Herve, et al., 2009). Furthermore, pectin methyl esterases, which remove the methyl groups on uronic acid to generate GalA, can also be added to assist pectin degradation.

Mannan, which is present in many plant cell walls, can be degraded by mannanases. Despite mannan esterification in the secondary cell walls, and the presence of pectin

homogalacturonan in primary cell wall, which can block enzyme access pectic lyases and esterases can be incorporated into the endo- and exo-mannanse cocktail to enhance degradation of the mannose polymer. One of the core principles of Laundry formulation development is that it is a mixture of many ingredients which work, within a cocktail of enzymes and non-bio reagents, in synergy to enhance the total performance of stain removal. In future, a cocktail of arabinofuranosidase, mannanases, pectin lyases and xylanases fused to non-catalytic CBMs, such as CBM2b1,2 variants that displays no cellulose binding capacity, might be included in gras stain removal formulation.

Considering the resistance assay results, the PT linker in the CBM-Xylanase was attacked by savinase at 0.01 mg/ml, but the catalytic modules and CBM15 were almost intact. The linker sequence is more flexible and more amenable to attack by proteases than the respective modules are much more resistant to savinase attack than the linker region, thus in future work expressing the proteins in *Pichia pastoris* to investigate where O-glycosylation occurs in this yeast may protect the linker regions from protease attack. The linker region where it is being attacked could be further analyzed through Mass Spectrometry. Building a chemical synthetic linker that is resistant to savinase attack, i.e, a non amino acid linkage, can be carried out in future. For example, using an amine-to-amine crosslinker bis-succinimide ester-activated polyethylene glycol compounds, abbreviated as $BS(PEG)_x$, based on N-hydroxysuccinimide ester (NHS-ester) reactive groups for selective conjugation of primary amines on CBM modules and xylanases, at either end of the $(PEG)_x$ spacer arms. The flexible Polyethylene glycol spacer arms helps maintain conjugate solubility and forms an irreversibly and stable amide bond with protein modules at pH 7.0-9.0. This type of linker increases the stability of linkers to savinase, reduced tendency toward protein aggregation and

reduced immunogenicity with other compound. It is commercially available from Thermo Fisher Scientific®.

Unfortunately, CBM2b1,2 binding activity after savinase treatment could not to be assessed by AGE method as its PI value is equal to the pH of the buffer. ITC could be used to assay the residual binding capacity of CBM2b1,2 to soluble oat spelt xylan. In Fig 4.5, we do observe differences between the three variation of *Cj*Xyn10A and *Np*Xyn11A. CBM2b1,2 thus target the enzymes to the grass stains, resulting in an increase in stain removal. In the binding assays and the final detergent assay, CBM2b1,2 was shown to target grass stains better than CBM15, as it possesses a shallower binding cleft for xylan chains than CBM15. CBM2b1,2, however, binds weakly to cellulose and cotton textiles, which leads to a reduction of the amount of CBM2b1,2 binding grass stain, therefore reduces its efficiency in grass stain targeting. It could be optimised with the aim of abolishing its cellulose binding affinity, which is discussed in the final discussion (Chapter 6).

Chapter 5. Interaction of CBM60 and CBM35 with their ligands

5.1 Introduction

The Carbohydrate-Active Enzymes CAZy database (http://www.cazy.org/) (Coutinho and Henrissat, 1999) describes the families of structurally-related catalytic and non-catalytic carbohydrate-binding modules (CBMs) of enzymes that are active against carbohydrate polymers. Within this database, there are several hundred prokaryotic CBMs. These modules display considerable variation in primary structure and are grouped into ~60 sequence-based families. The binding specificity of CBMs differs greatly between, and sometimes also within, these families, while individual modules can show affinity for more than one type of carbohydrate target (Boraston, et al., 2006; Charnock, et al., 2002a; Najmudin, et al., 2006). In the original CAZy database, there were numerous sequence-based families of X-modules which were discovered through their association with carbohydrate active enzymes, but whose function was unknown. One of the modules, defined as X14, was identified in a glycoside hydrolase family 11 xylanase (Xyn11A from *Cellvibrio japonicus),* which catalyses the hydrolysis of 1, 4-β-D-xylosidic linkages in xylan. The X14 module was shown to bind carbohydrates and was therefore defined as a CBM. The 13 kDa module, designated *Cj*CBM60A was the founding member of family CBM60.

*Cj*CBM60A uniquely displays broad ligand specificity. It recognizes pectic galactan, which comprise a backbone of galacturonic acid and rhamnose residues decorated, principally, with β-1,4-Gal polymeric side-chains (Scheller, et al., 2007). It also binds to cellulose (β-1,4-Glc), glucans (β-1,3/1,4-Glc) and xylan with the same affinity ($k_A \approx 6 \times 10^3\ M^{-1}$) (Montanier, et.al., 2010). To evaluate whether the broad ligand specificity, exhibited by *Cj*CBM60A, is a generic feature of CBM60, attempts were made to

express other proteins from this family. Out of six homologs of *Cj*CBM60A, provided by Verenium Corporation (San Diego USA) from their in-house environmental library of GH11 xylanases, only one protein module, designated *v*CBM60, could be produced in soluble form in *E. coli*. The ligand specificity of *v*CBM60 was very similar to *Cj*CBM60A (Montanier, et.al., 2010). The protein module binds to galactan, xylans and ß-1,4-linked glucans, although the affinities for these polysaccharides are considerably lower than observed with *Cj*CBM60A.

Previous crystallography studies carried out by Dr. C. Montanier solved the structure of *v*CBM60 (C. Montanier's PhD thesis). The protein displays a β jelly roll fold consisting of eight β strands in two anti parallel β sheets; eight loops and one α helix (Fig 5.1). Moreover, it has a deep and wide cleft formed by the loops connecting the two β sheets, approximately 17.4 Å broad and 5.5 Å deep, which enables it to exhibit exceptional plasticity in ligand recognition. Ligand binding at the single binding site is Ca^{2+} dependent.

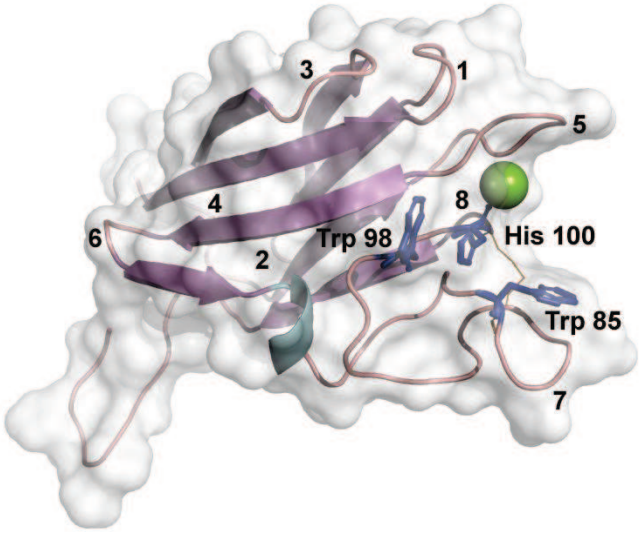


Fig 5.1 Crystal structure of *v*CBM60

Cartoon representation overlaying a surface representation of one molecule of *v*CBM60. The β-strands are shown in purple on one α-helix in grey and structural Ca^{2+} ion is colored in green. Aromatic residues are displayed as purple sticks. The structural loops are labelled 1, 2, 3, 4, 5, 6, 7 and 8. The large and deep cleft is between loops 5 and 7, while the base of this cleft comprises loop 8 (adapted from C. Montanier's PhD thesis).

The previous crystal structure of *v*CBM60 was solved with a final model which consisted of two molecules per asymmetric unit (Fig 5.2). However, as a result of space group constraints, the entrance to the binding cleft of *v*CBM60 is partially occupied by the C-terminal His_6 tag of a symmetry-related molecule. A construct encoding *v*CBM60 with no tag was made by Dr. Flint. One of the objectives of this project is to purify the *v*CBM60 with no tag and determine the crystal structure of the protein in complex with different ligands. The data were used to understand the mechanism of carbohydrate recognition in the CBM60 family of protein modules. Since *v*CBM60 is known as a CBM that can recognize, a wide variety of carbohydrates with the same affinity, it is considered a potential grass binding CBM and its targeting of grass stains was also assessed (Chapter 3).

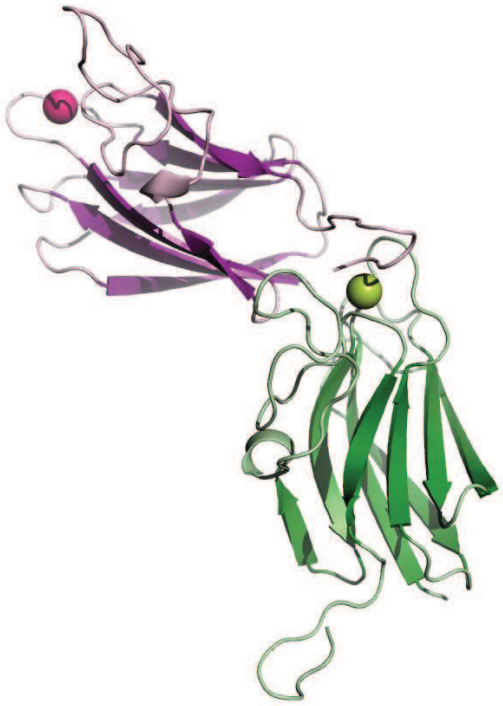


Fig 5.2 Crystal structure of two symmetry-related CBM60 molecules

Cartoon representation of two symmetry-related molecules of CBM60. Metal ion is displayed as a sphere 8 (adapted from C. Montanier's PhD thesis).

Beside CBM60, another galactose binding module, CBM35 from *Clostridium thermocellum* (designated as *Ct*CBM35-Gal), has also been intensively studied (Montanier, et al., 2009b). It targets galactoglucomannan and galactomannan the major hemicellulose in softwoods, by binding to α-galactose, and it is a component of the cellulosome, a multimodular enzyme complex that deconstructs the plant cell wall (Bayer et al, 2004). *Ct*CBM35-Gal likely functions to localize the enzymes tethered within the cellulosome to regions of galactomannan or other components of the plant cell wall that contain terminal galactose residues.

Recent studies on CBM family 35 revealed a cohort of four structurally related CBMs. These modules all recognise uronic acids (designated as CBM35-UA), Pel-CBM35 and Rhe-CBM35 recognize Δ4,5-anhydrogalacturonic acid, while Xyl-CBM and Chi-CBM35 display affinity for both D-glucuronic acid and Δ4,5-anhydrogalacturonic acid (Montanier, et al., 2009b). Another member of CBM family 35 recognizes mannan and binds to the internal regions of mannan, a beta-polymer of mannose (Tunnicliffe, et al., 2005). *Ct*CBM35-Gal and the CBM35-UAs are exo-CBMs, and display significant sequence identity (29 %) and structural similarity (Correia, et.al., 2010). It is surprisingly; therefore, that *Ct*CBM35-Gal displays subtle differences in the conformation of residues that are conserved in the ligand binding site of CBM35-UA, which likely leads to the loss of uronate recognition and the introduction of different ligand specificity. In this study, site-directed mutagenesis was used to confirm that the pocket in *Ct*CBM35-Gal, which functions as the ligand binding site in CBM35-UAs, fulfills a similar role in the galactose binding module. Mutagenesis was also used to probe the role of residues in the ligand binding site that contribute to galactose recognition.

5.2 Objectives

With respect to *v*CBM60, this project explored the mechanism of ligand recognition. With respect to *Ct*CBM35-Gal, the primary objective of this study was to use site-directed mutagenesis to identify the location of the ligand binding site, and the amino acids that contribute to galactose recognition.

5.3 Results of *v*CBM60 protein module

5.3.1 Over expression and purification

The constructs of CBM60 modules, for both wild type and mutants, were cloned by Dr. C. Montanier. Proteins were expressed in *E. coli* as described in Section 2.1.21. The recombinant strain was cultured at 37 °C and the expression was induced by the addition of 0.5 mM IPTG and incubation for 4 h at 30 °C. As *v*CBM60 (120 residues, 13 kDa) encoded by pET22b contains a *C*-terminal His_6 tag, the protein was purified by immobilised metal affinity chromatography (IMAC) as described in Section 2.1.24. The protein was electrophoretically > 90 % pure, and was used for biochemical studies (Fig 5.3). To generate *v*CBM60 that contains no His_6 tag (for crystallization with ligands, designated *v*CBM60-no tag), a stop codon was inserted in place of the Leu in the C-terminal motif LEHHHHHH (encoded by pET22a). It was expressed by the same method as CBM60 containing a His_6 tag, but purified by affinity chromatography using insoluble oat spelt xylan as the matrix (Section 2.1.24). The protein, which was electrophoretically > 90 % pure, was dialysed three times against 6L distilled water containing 5 mM Ca^{2+} at 4 °C and was subjected to another round of purification by Q12 anion exchange (Section 2.1.24), resulting in the removal of trace contaminants. An example of the expression and purification of *v*CBM60-no tag is shown (Fig 5.4).

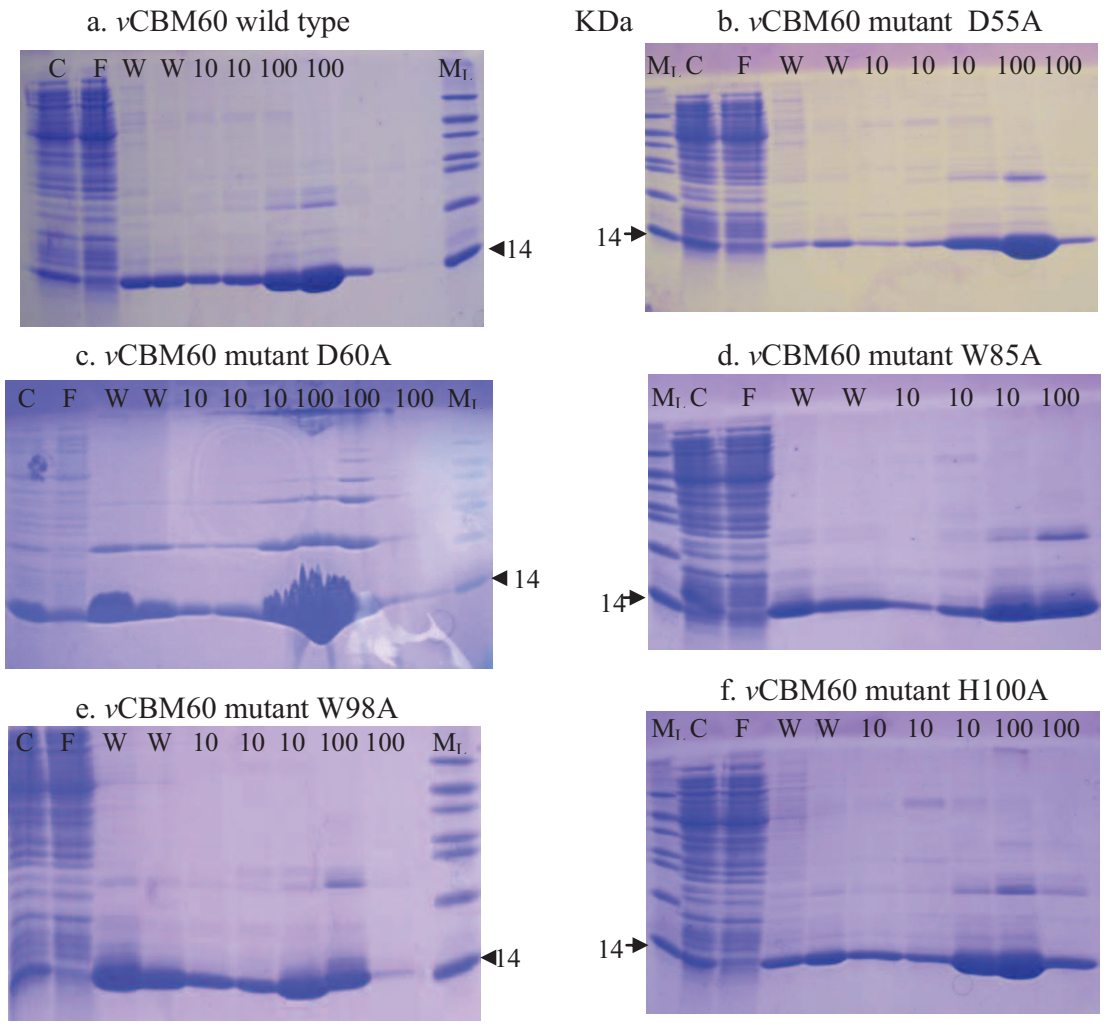


Fig 5.3 SDS-PAGE of *v*CBM60 wild type and five mutants purified by IMAC

Extraction and purification of wild type *v*CBM60 and its variants, which all contain a His_6 tag, was carried out as described in Section 2.1.24. The panels refer to the purified of the following proteins: a, *v*CBM60 wild type; b, *v*CBM60 mutant D55A; c, *v*CBM60 mutant D60A; d, *v*CBM60 mutant W85A; e, *v*CBM60 mutant W98A; f, *v*CBM60 mutant H100A. A 5 µl aliquot of cell free extract and flow-through (Lanes C and F respectively) was loaded, then 10 µl of wash with 10 mM Tris/HCl pH 8.0, containing 300 mM NaCl (W), and 10 µl of every elution with 10 mM imidazole (10) and 100 mM imidazole (100). Lanes M_L contained Sigma low molecular weight standards. Analysis was performed by SDS-PAGE, using a 12.5 % (w/v) polyacrylamide gel. The cloning work of the constructs of CBM60 modules, for both wild type and mutants, were carried out by Dr. C. Montanier.

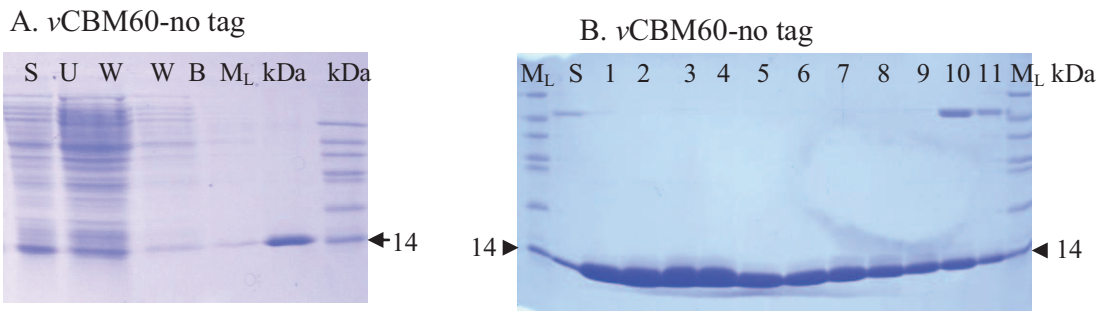


Fig 5.4 The purification of *v*CBM60 no tag

Panel A displays purification of *v*CBM60-no tag using insoluble xylan in a pull down method described in Section 2.1.24. Panel B displayed purified *v*CBM60-no tag, which was purified by the pull down method, subjected to anion exchange chromatography for further purification. All fractions were analyzed by SDS−PAGE using a 12.5 % (w/v) polyacrylamide gel. In Panel A, lanes S, U, W and B contained the cell free extract, supernatant after incubating with the insoluble ligand, the wash and the protein released from the polysaccharide by ethylene glycol, respectively. In Panel B, lane S contained purified *v*CBM60-no tag after IMAC, 1-11 are fractions eluted during anion exchange chromatography. Lanes M_L contained Sigma low molecular weight markers.

5.3.2 ITC of wild type and mutants of *v*CBM60 against polysaccharides and oligosaccharides

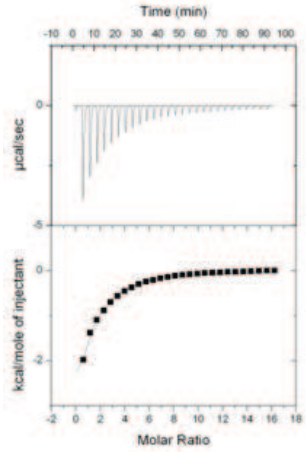
Isothermal titration calorimetry (ITC) was used to quantify the affinity of *v*CBM60 for polysaccharides and oligosaccharides. Examples of the titrations are displayed in Fig 5.5 and Fig 5.6, while the complete data sets are reported in Table 5.1. The data were fitted by non-linear regression analysis (MicroCal ORIGIN, v7.1) to a single site binding model, which resulted in an n value of ≈ 1 for oligosaccharide ligands. Titrations were carried out at 25 °C in 50 mM Na-Hepes buffer, pH 7.5 containing 5 mM $CaCl_2$, unless otherwise stated. The data yielded values for k_A and ΔH. Other thermodynamic parameters were calculated using the standard thermodynamic equation $-RT \ln k_A = \Delta G = \Delta H - T\Delta S$.

5.3.2.1 Wild type and mutants of vCBM60 against polysaccharides (β−1,4-galactans, xylans and β-glucans)

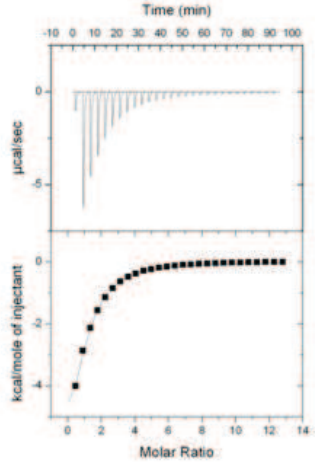
The data show that wild type *v*CBM60 binds to a range of β−1,4-galactans, xylans and β-glucans (β−1,4-glucan backbone, xyloglucan; mixed linked β−1,3-β−1,4-glucans; (Table 5.1 and Fig 5.5). In general, the binding affinities for the three complex polysaccharides were similar, irrespective of the number and type of side chains decorating of the backbone of the three polymers. Therefore, the ligand binding site *v*CBM60 appears to be able to recognise polysaccharides that are extensively decorated. *v*CBM60 recognition to all three complex ligands were driven by favorable changes in enthalpy with, generally, a small unfavorable entropic contribution.

ITC data show that binding to all three complex polysaccharides are completely abrogated by the mutations D55A, D60A and W85A, indicating that D55, D60 and W85 play a critical role in ligand recognition. Even though W98 appears to be located in the ligand bind site, this residue caused only a two-fold decrease in affinity for potato pectic galactan, oat spelt xylan and barley β-glucan. H100 is very important in the binding of *v*CBM60 to oat spelt xylan, as the k_A of the mutant H100A is 885 M^{-1} for the hemicellulosic polysaccharide, compred to 7.2 x 10^3 M^{-1} for the wild type protein. However, H100 does not appear to play an important role in the binding of *v*CBM60 to barley ß-glucan. When the imidazole ring is removed by mutation, the affinity for galactan is increased by two fold, indicating it is easier for galactan to dock into the binding cleft of the H100A mutant.

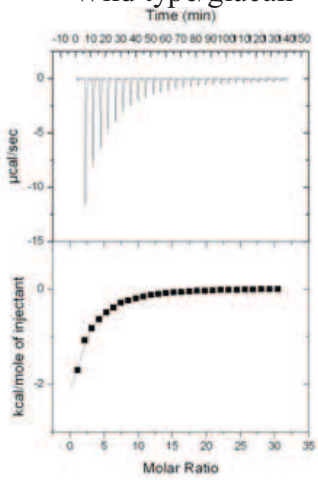
A. Wild type/xylan



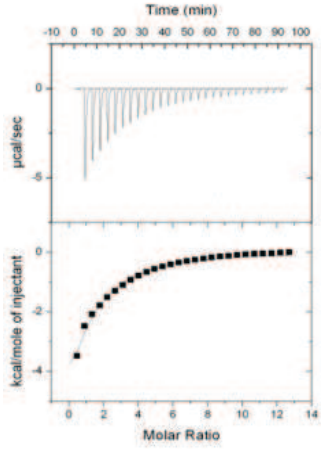
Wild type/galactan



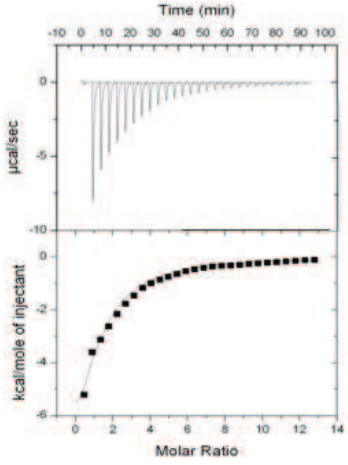
Wild type/glucan



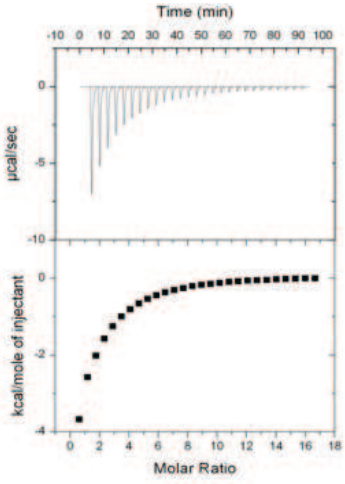
B. W98A/xylan



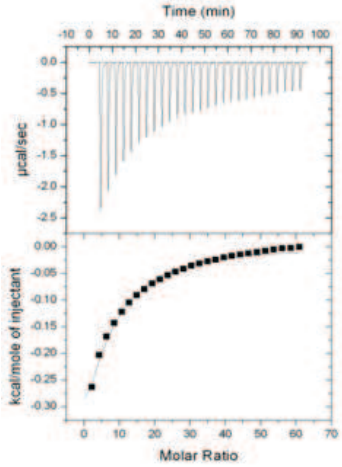
W98A/galactan



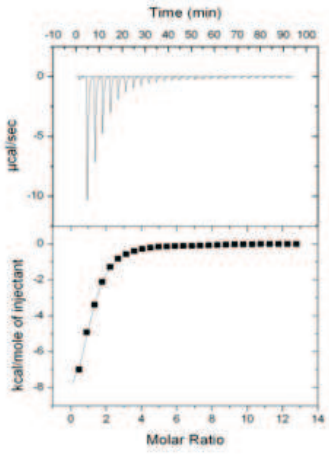
W98A/glucan



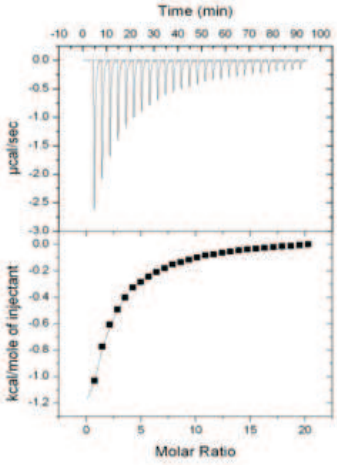
C. H100A/xylan



H100A/galactan



H100A/glucan



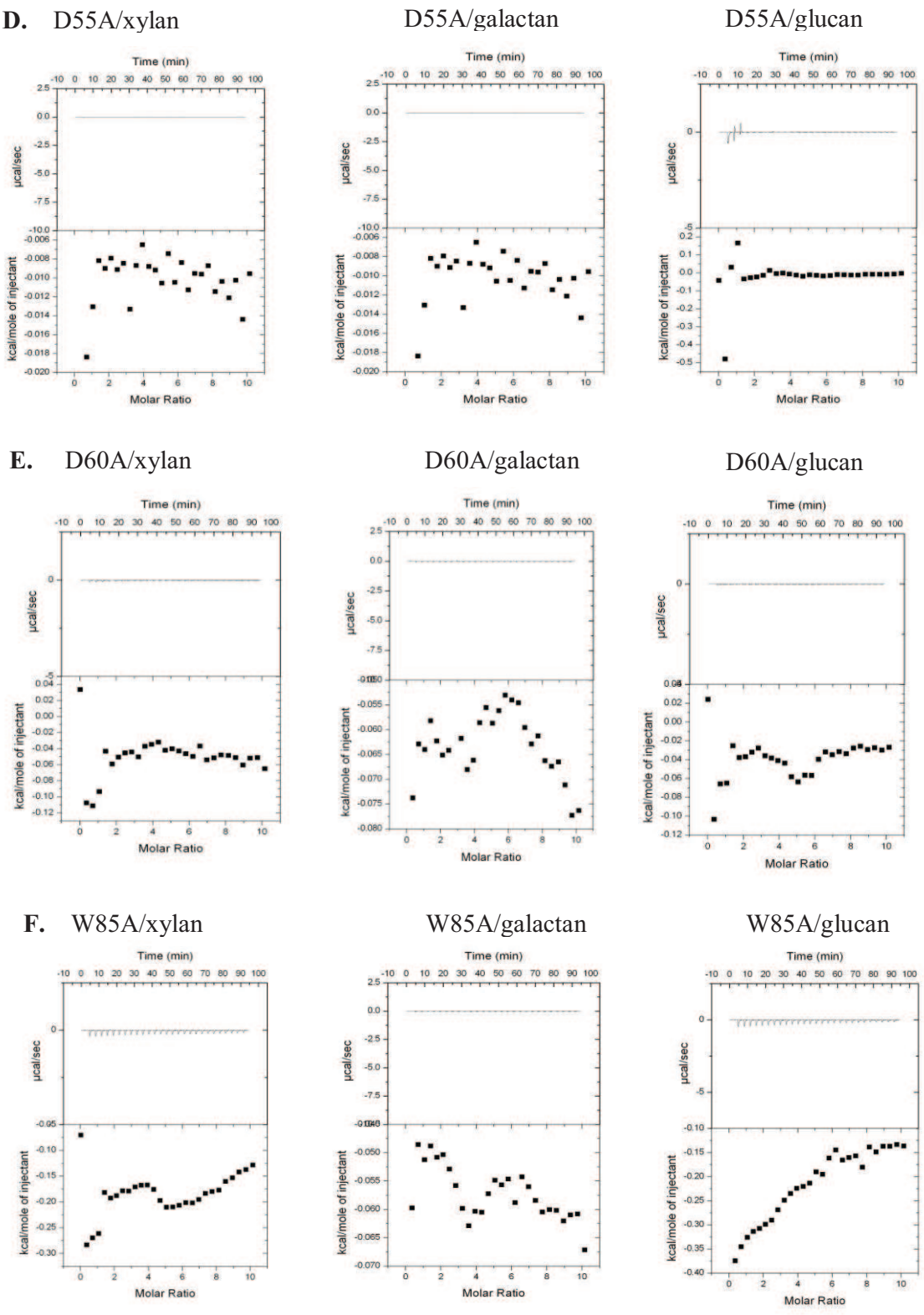


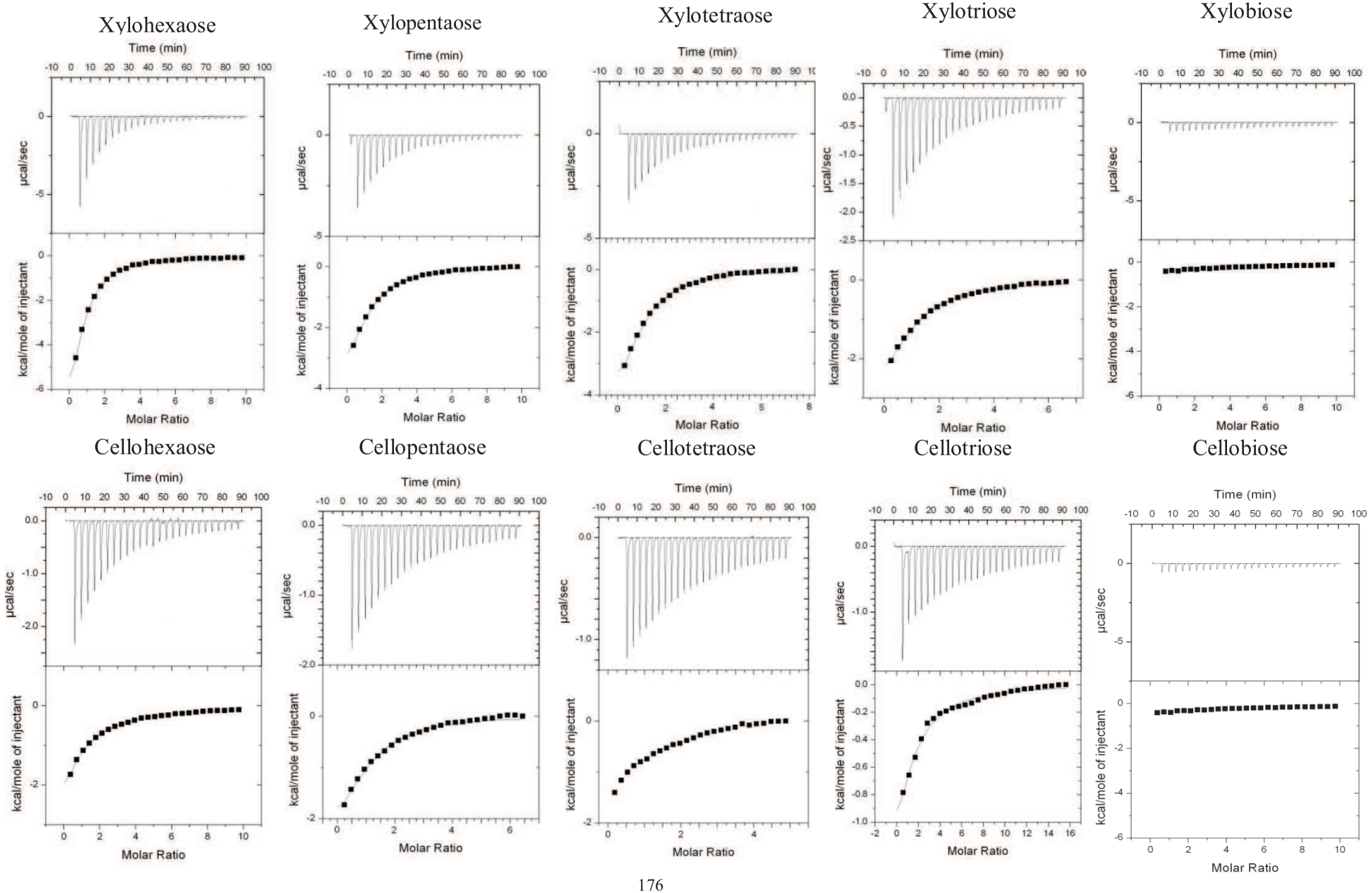
Fig 5.5 Representative ITC data of *v*CBM60 wild type and five mutants to polysaccharides. ITC was carried out in 50 mM Na/Hepes buffer, pH 7.5 containing 5 mM $CaCl_2$. Typical ITC data for wild-type (Panel A) and five mutants (Panel B to F) of *v*CBM60 titrated with 5 mg/ml oat spelt xylan, pectic galactan and barley β-glucans. Data were fitted to a single site binding module.

*v*CBM60 derivative	Ligand	Affinity (k_A, $M^{-1}x10^3$)	*ΔG* (kcal mol^{-1})	*ΔH* (kcal mol^{-1})	*TΔS* (kcal mol^{-1})	n[a]
Wild type	Oat spelt xylan	7.2 ± 0.8	-5.1	-6.3 ± 0.1	-1.1	1.00±0.18
Wild type	Potato pectic galactan	12 ± 0.6	-5.7	-8.0 ± 0.5	-2.3	0.99±0.05
Wild type	Barley ß-glucan	8.7 ± 0.9	-5.0	-7.0 ± 0.2	-2.0	1.04±0.31
W98A	Oat spelt xylan	5.3 ± 0.9	-5.1	-11.0 ± 3.3	-5.9	1.00±0.27
W98A	Potato pectic galactan	6.6 ± 1.0	-5.2	-14.3 ± 0.3	-9.0	0.98±0.20
W98A	Barley ß-glucan	5.7 ± 1.0	-5.1	-11.9 ± 0.2	-6.8	0.98±0.16
H100A	Oat spelt xylan	0.89 ± 0.11	-4.0	-1.0 ± 0.3	3.1	-
H100A	Potato pectic galactan	29 ± 2.0	-6.1	-10.6 ± 0.3	-4.5	1.00±0.02
H100A	Barley ß-glucan	3.2 ± 0.3	-4.8	-4.9 ± 1.4	-0.1	1.02±0.27
D55A, D60A or W85A	Oat spelt xylan, potato pectic galactan or Barley ß-glucan	No binding of each mutant to the three polysaccharides	-	-	-	-

Table 5.1 The binding affinity and thermodynamic parameters of wild type and mutants of *v*CBM60 to polysaccharides and oligosaccharide determined by ITC. Binding to the various polysaccharides was determined by ITC in 50 mM Na/Hepes pH 7.5 containing 5 mM $CaCl_2$. The errors are presented, for typical titrations. The ITC data were fitted to a single site binding model for all ligands. For polysaccharide ligands in which the molar concentration of binding sites is unknown, the n-value was iteratively fitted to as close as possible to one, by adjusting the molar concentration of the ligand, the rationale for this approach is described in Section 2.3.2. n.b.: no binding occurred.

5.3.2.2 Wild type vCBM60 to cello- and xylo- oligosaccharides

ITC was carried out using 100 μM of wild type *v*CBM60 titrated against 5 mM cello- and xylo- oligosaccharides, from cellohexaose (C6, six glucose units) to cellobiose (C2, two glucose units); and also from xylohexaose (X6, six xylose units) to xylobiose (X2, two xylose units). Examples of the titrations are displayed in Fig 5.6, while the complete data set are reported in Table 5.2. The affinity of *v*CBM60 for the series of oligosaccharides shows that the protein does exhibit similar affinity for the hexasaccharide and trisaccharide of cello- and xylooligosaccharides, but it does not bind to xylobiose or cellobiose. Binding to xylo-oligosaccharides is driven primarily by the release of heat, as the change in entropy is very small. To interact with cellooligosaccharides, binding is driven by enthalpy with the decrease in entropy having a negative impact on affinity. In general, *v*CBM60 displays a preference for the *xylo*- ligands.


Xylohexaose
Xylopentaose
Xylotetraose
Xylotriose
Xylobiose
Cellohexaose
Cellopentaose
Cellotetraose
Cellotriose
Cellobiose
Time (min)
µcal/sec
kcal/mole of injectant
Molar Ratio

Ligand	Affinity x 10^3 (M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n[a]
Xylohexaose	10.4±0.279	24.1	-14.8±0.682	-9.33	0.549±0.02
Xylopentaose	10.1±0.773	5.6	-5.53±0.435	-0.07	1.05±0.07
Xylotetraose	13.9±1.20	-5.6	-5.63±0.395	-0.02	1.00±0.06
Xylotriose	9.66±0.700	-3.7	-4.54±0.364	-0.89	0.96±0.06
Xylobiose	n.b	-	-	-	-
Cellohexaose	3.2±0.4	-4.8	-8.2±2.9	-3.4	1.00±0.32
Cellopentaose	14.0±2.04	-0.32	-2.99±0.33	-2.67	1.06±0.09
Cellotetraose	13.8±2.87	0.97	-2.341±0.367	-3.31	1.05±0.13
Cellotriose	4.09±0.684	-1.4	-3.10±1.19	-1.74	1.03±0.362
Cellobiose	n.b	-	-	-	-

Table 5.2 The binding affinity and thermodynamic parameters of wild type *v*CBM60 to oligosaccharides determined by ITC.

Binding to the various oligosaccharides was determined by ITC in 50 mM Na Hepes pH 7.5 containing 5 mM $CaCl_2$. The errors are presented, for typical titrations. The ITC data were fitted to a single site binding model for all ligands. The molar concentration of the ligand was 5 mM. The rationale for this approach is described in Section 2.3.2. n.b.: no binding occurred.

5.3.3 Crystallography of *v*CBM60 with ligands

5.3.3.1 Acid digestion of galactan

Although the crystal structure of *v*CBM60 has been obtained in previous studies (C. Montanier's PhD thesis), the previous crytal obtained was without the ligand, and it was as a dimer due to the presence of His_6 tag (Fig 5.2). Thus, the protein interaction between the protein and galactan was explored at a molecular level in this study. It was decided to use galactobiose (G2) or galactotriose (G3) as the ligands in *v*CBM60 crystallization experiments. As these two oligosaccharides are not commercially available, potato pectic galactan (Sigma) was partially digested with an endogalactanase and the G2 and G3 generated were purified.

5.3.3.2 Enzyme digestion of galactan

A. Over expression of galactanase A

To prepare galactobiose and galactotriose, pectic galactan was partically digested with the *C. japonicus* galactanase *Cj*Gal53A. The enzyme was expressed in *E. coli* as described previously (Braithwaite, et al., 1997), and a cell-free extract of the cultured bacteria (*Cj*Gal53A CFE) was used as the source of the galactanase.

B. Galactan digestion by Galactanase A

Galactan digestion by *Cj*Gal53A was carried out in a small scale. Initial pilot experiments identified a ratio of *Cj*Gal53A CFE : galactan that was optimal for the production of G2 and G3. Scaling up of the reaction to produce sufficient amounts of the two galacto-oligosaccharides was successful. Details of these reactions are provided in Fig 5.7. The digested samples were analyzed with TLC technique (Section 2.3.4), using a manno-oligosaccharides marker containing mannose (M1), mannbiose (M2), mannotriose (M3) each at 1 mg/ml and galactose (G1) at 20 mM, as the galacto-oligosaccharides marker are not commercially available.

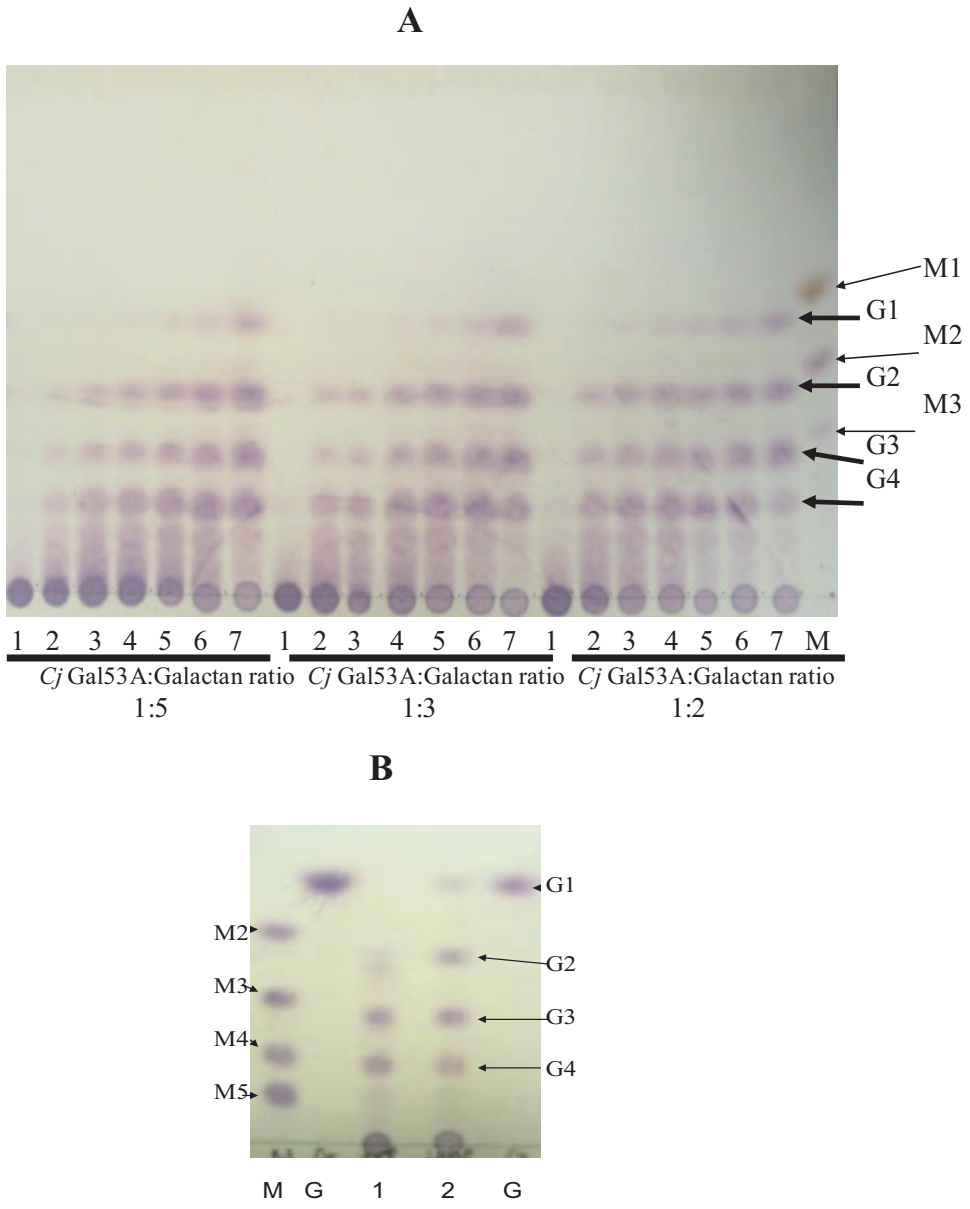


Fig 5.7 Pectic galactan digestion by galactanase at different enzyme/substrate ratios

Panel A displays small scale galactanase digestion in 50 mM sodium phosphate buffer, pH 7.0, (total reaction volume was < 5 ml) of pectic galactan (potato) at 10 mg/ml at room temperature at three enzyme:substrate ratios (v/v) 1:5, 1:3 and 1:2. An aliquot of 2 µl from each incubation was taken at 0 min (lane1), 20 min (lane 2), 40 min (lane 3), 1 h (lane 4), 2 h (lane 5), 4 h (lane 6), and 18 h (lane 7) and subjected to TLC analysis. Lane M contains 2 µl of mannose (M1), mannbiose (M2), mannotriose (M3), each at 1 mg/ml. Panel B displayed large scale digestion (total reaction volume at 120 ml) at a 1:3 enzyme:substrate ratio. Lane 1 contained a 2 µl aliquot of the small scale digestion, as a positive control. Lane 2 contained an aliquot of 2 µl of the large scale incubation at the end of digestion (19 h). M contains the same standards used in Panel A. G contains 20 mM galactose (G1). G2, G3 and G4 are assumed to consist of galactobiose, galactotriose and galactotetraose, respectively.

C. Separation of products from digested galactan using size exclusion chromatography

The mixture of galactan-derived oligosaccharides was subjected to size exclusion chromatography using two BioGel P2 columns run in series as described previously (Proctor, et. al, 2005). The eluted fractions were analyzed by TLC. The results (Fig 5.8) showed fractions 86-92 contained pure galactotriose fractions 97-106 contained pure galactobiose and fractions 108-118 contained pure galactose. The other fractions contained mixtures of galactooligosaccharides. Hence, fractions 86-92 and 97-106 containing only pure galactobiose and galactotriose, respectively, were pooled and freeze dried. The final yield of galactobiose and galactotriose obtained in pure form were 35 and 11.8 mg, respectively, starting from approximately 1 g of pectic galactan.

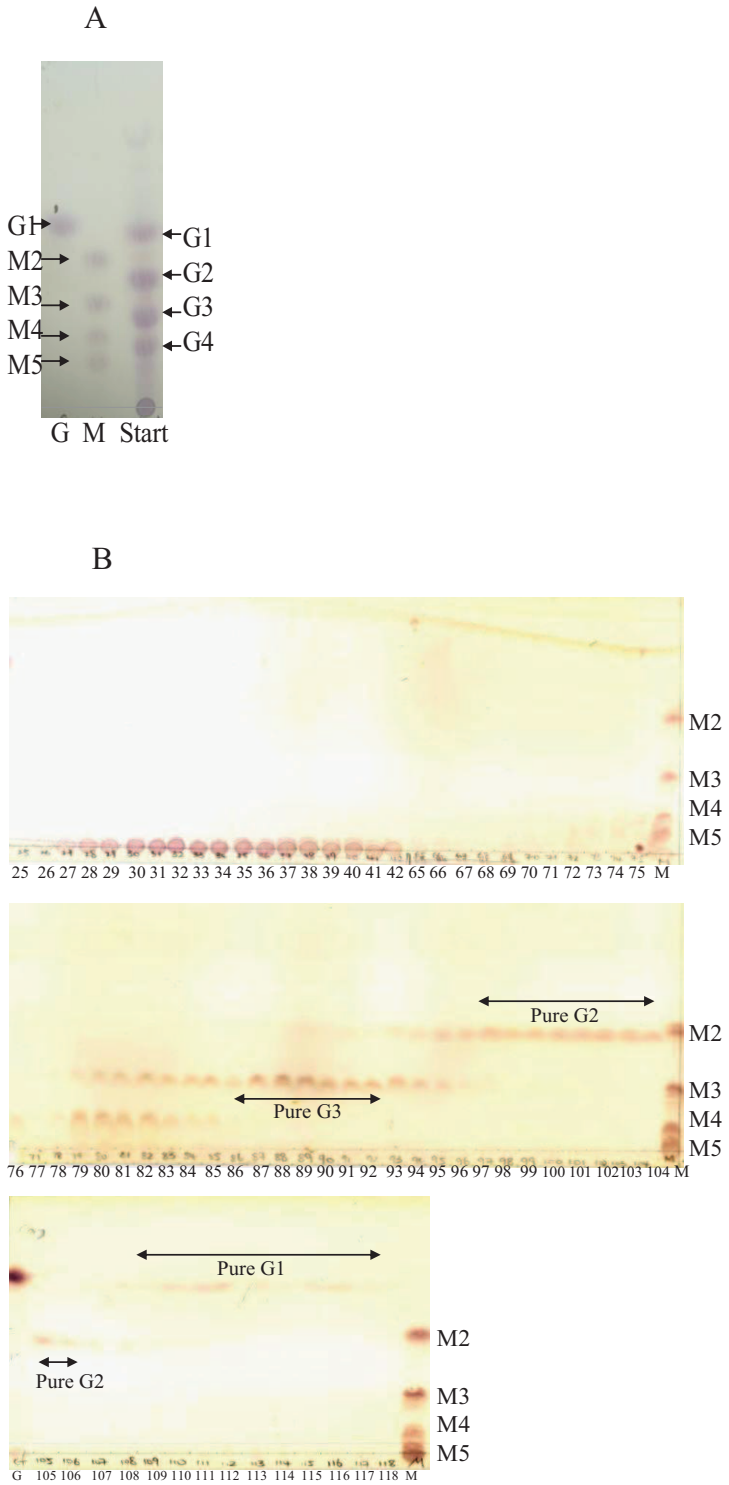


Fig 5.8 Separation of mixed digested galacto-saccharides using size exclusion chromatography Panel A displays starting material before separation, which contained a mixture of di-tri- and tetra- galacto saccharides after pectic galactan digestion. Panel B displays TLC data of fractions 25-42 and 65-118. M contains 2 µl of mannobiose, -triose, -tetraose and -pentaose (noted as M2, M3, M4 and M5, respectively) each at 1 mg/ml. G contains 20 mM galactose. Galactose, -biose and –triose were designated as G1, G2 and G3, respectively.

D. MALDI-Mass Spectrometer (MS) of galactobiose and galactotriose fractions

Pure galactobiose and galactotriose were subjected to MALDI-MS analysis and the peaks indicated galactobiose with a molecular weight of 365 and galactotriose with a mass of 527 (Fig 5.9). The estimated formular weight (Fwt) of galactobiose and galactotriose were 342 and 504, and as each was 23 (m/z unit) from the theoretical value, they were, as expected, adducts of sodium. It is noted that both galactobiose and galactotriose contained small amounts of tri- and di-saccharide contaminants, respectively.

A. Galactobiose

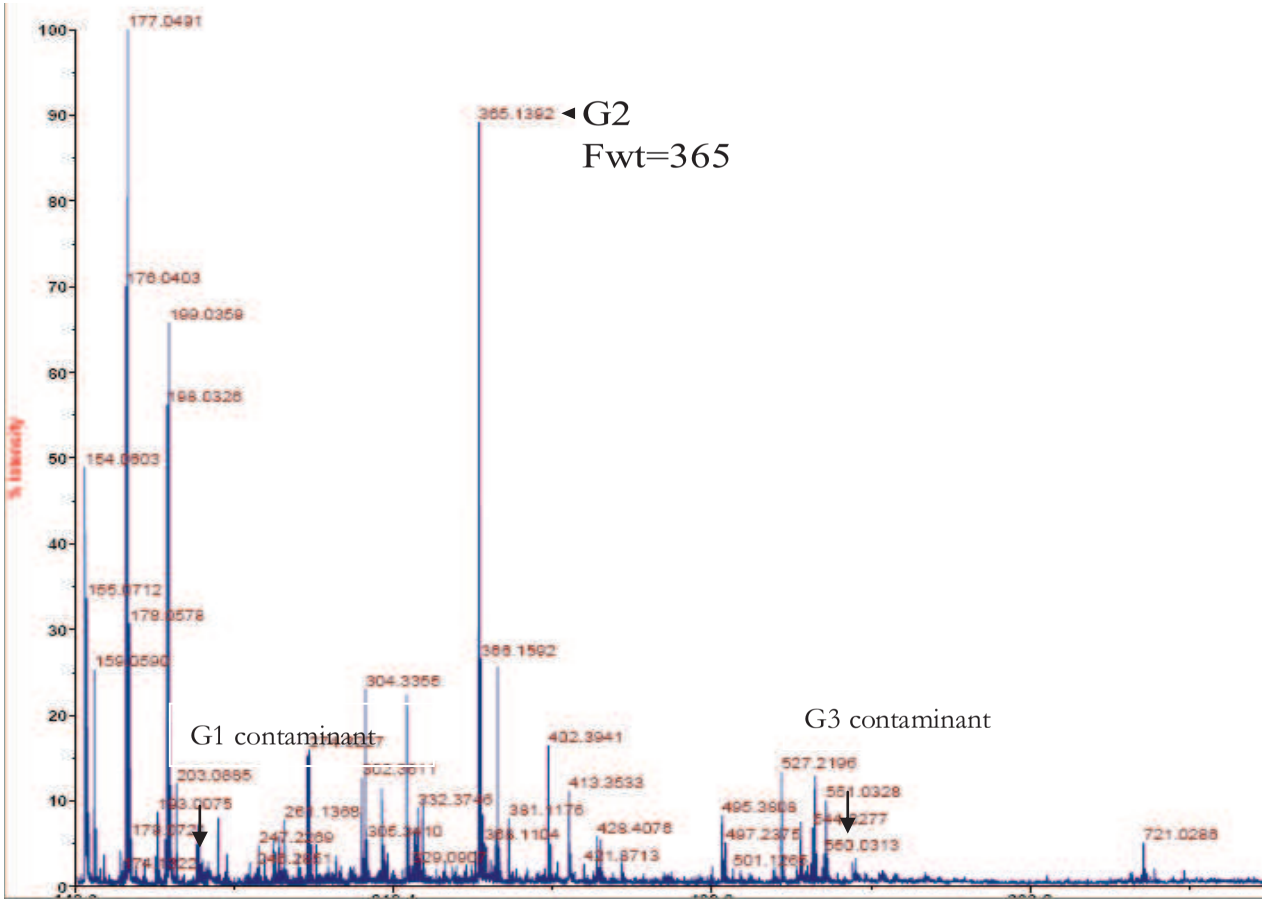


Fig 5.9 Mass spectrometry result of sample containing galactobiose (Panel A) and galactotriose (Panel B). The Biological Mass Spectrometry was conducted by Pinnacle Laboratory, Newcastle Medical School, on 15 Oct 2009. Abbreviations: G1, galactose; G2, galactobiose; G3, galactotriose.

B. Galactotriose

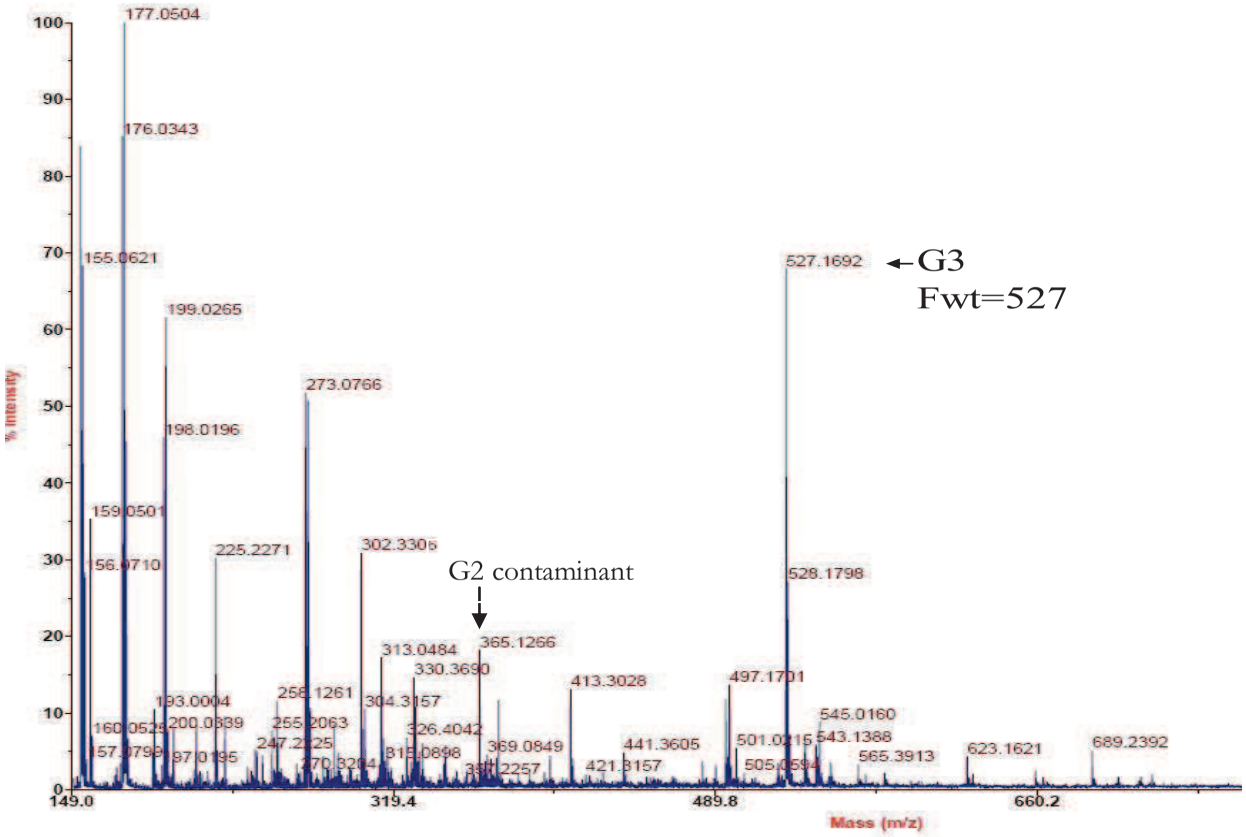


Fig 5.9 Mass spectrometry result of sample containing galactobiose (Panel A) and galactotriose (Panel B). The Biological Mass Spectrometry was conducted by Pinnacle Laboratory, Newcastle Medical School, on 15 Oct 2009. Abbreviations: G1, galactose; G2, galactobiose; G3, galactotriose.

E. ITC of *v*CBM60 with no tag with pure galactobiose and galactotriose

Purified *v*CBM60 (no tag) was titrated against pure galactobiose and galactotriose. The data are shown in Fig 5.10 and Table 5.3. *v*CBM60 (no tag) binding was driven by favourable changes in enthalpy, with an unfavorable entropic contribution. The k_A values for galactobiose and galactotriose were very similar at ~10^4 M^{-1}. It should be noted, unlike the specificity of *v*CBM60 for xylo- and cello- oligosaccharides, where only ligands with a degree of polymerization of 3 or higher bound to the protein, the CBM displayed maximum affinity for galactobiose. Therefore, *v*CBM60 (no tag) was subjected to crystallisation with galactobiose and galactotriose, respectivey.

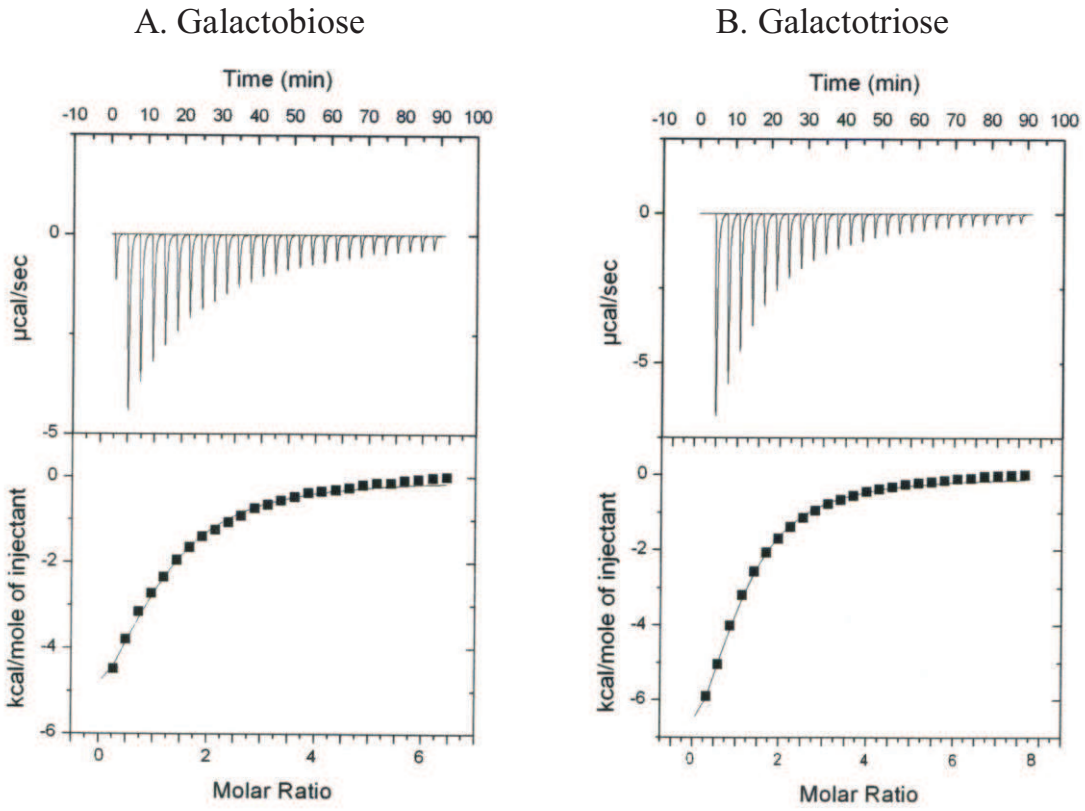


Fig 5.10 Representative ITC data of *v*CBM60 (no tag) wild type to galactobiose (Panel A) and galactotriose (Panel B). ITC was carried out in 50 mM Na/Hepes buffer, pH 7.5, containing 5 mM $CaCl_2$. *v*CBM60 at 100 µM was titrated with oligosaccharides at 5 mM.

Ligand	Affinity x 10^3 (M^{-1})	ΔG (kcalmol^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n[a]
Galactobiose	13.4±1.3	-5.6	-8.3±0.7	-2.7	1.01±0.07
Galactotriose	16.5±1.1	-5.6	-10.5±0.5	-4.9	1.00±0.04

Table 5.3 The binding affinity and thermodynamic parameters of wild type *v*CBM60 to galactobiose and galactotriose determined by ITC. Binding to galactobiose and galactotriose were determined by ITC in 50 mM Na/Hepes pH 7.5 containing 5 mM $CaCl_2$. The ITC data were fitted to a single site binding model for all ligands. The molar concentration of the ligand was 5 mM. The rationale for this approach is described in Section 2.3.2.

F. Crystal structure of *v*CBM60 with galactobiose and cellotriose

To investigate the structural basis for the unusual ligand plasticity displayed by CBM60, and to probe how *v*CBM60 interacts with galactan, cellulose and xylan, co-crystallisation trials were set up with pure galactobiose, galactotriose, cellotriose, cellotetraose, xylotriose and xylotetraose. Pure *v*CBM60, as judged by SDS-PAGE, were dialyzed against distilled water contaning 5 mM Ca^{2+} at 4 °C, and concentrated in a 10 kDa concentrator. Four screens, PACT, NCL, Classics and JSCG, were used in the crystallization robotic screens (Section 2.2.2). Negative controls of protein only with no ligands were also included.

Robot screen and optimisation

Crystals of wild type *v*CBM60 in complex with cellotriose grew ~4-6 days at 18 °C in drops containing protein:ligand mixtures at 20 mg/ml *v*CBM60 : 10 mM ligand and mixing 1:1, with 0.01 M zinc chloride, 0.1 M Tris/HCl, pH 8.0, 20 % (w/v) PEG 6000 (PACT buffer condition) (Fig 5.11A). Crystals of wild type *v*CBM60 in complex with galactobiose grew ~4-6 days at 18° C in drops containing protein:ligand mixtures at 20 mg/ml *v*CBM60: 10 mM ligand and mixing 1:1 and 1:2, with 30 % PEG4000, 0.1M Tris/HCl, pH 8.5, 0.2 M $MgCl_2$ (Classics buffer condition) and with 30 % PEG4000, 0.1 M Tris/HCl, pH 8.5, 0.2 M sodium acetate (Newcastle buffer condition). These two buffer conditions were further optimised using the hanging drop vapour diffusion method, by slightly varying the PEG and salt concentrations. The crystal complexes of

*v*CBM60 with galactobiose were shown in Fig 5.11 Panel B and Panel C. For growing *v*CBM60 in complex with galactobiose, Classics buffer condition gave the best crystals. Optimisation of *v*CBM60 in complex with cellotriose based on PACT buffer condition was carried out, but crystals only grew in mother liquor. No crystal of *v*CBM60 in complex with galactotriose was obtained.

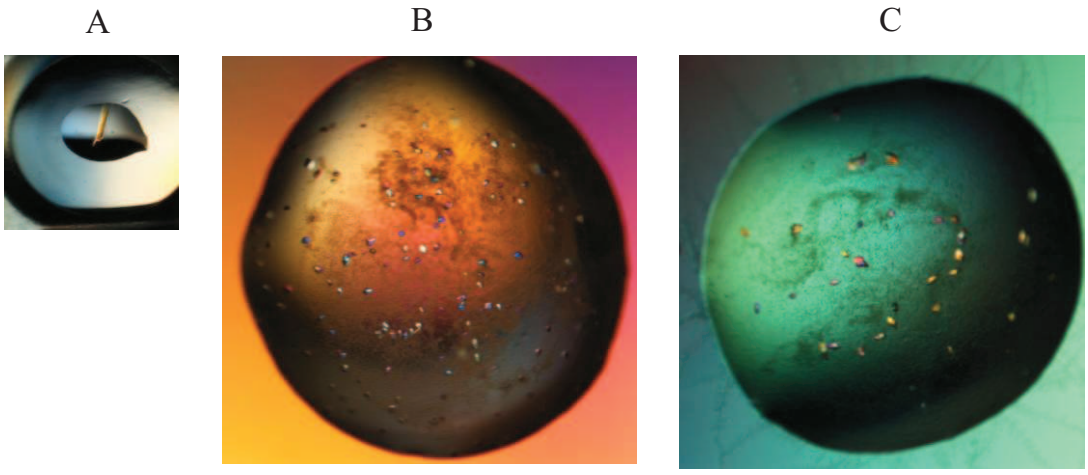


Fig 5.11 Crystals of *v*CBM60 wild type in complex with cellotriose and galactobiose. Panel A displays the crystal of *v*CBM60 in complex with cellotriose in 0.01 M zinc chloride, 0.1 M Tris/HCl pH 8.0, 20 % (w/v) PEG 6000. Panel B displays the crystals of *v*CBM60 in complex with galactobiose in 30 % PEG 4000, 0.1 M Tris/HCl, pH 8.5, 0.2 M $MgCl_2$. Panel C displays the crystals of *v*CBM60 also in complex with galactotriose in 30 % PEG4000, 0.1 M Tris/HCl buffer, pH 8.5, 0.2 M sodium acetate.

G. Collection of x-ray diffraction data, structure solution and refinement

All crystals harvested were in the form of plates and were flash frozen in liquid nitrogen prior to data collection. Collection of x-ray diffraction data, structure solution and refinement were performed by Dr. James Flint. Crystals were mounted in mother liquor supplemented with 25 % glycerol (v/v) and 10 mM appropriate ligand before being cryocooled in liquid N_2. Data for these crystals were collected on Diamond IO3 beamline to 1.2 Å for *v*CBM60-cellotriose and 1.8 Å for *v*CBM60-galactobiose. Data for the liganded complexes of *v*CBM60 were integrated and scaled using MOSFLM and SCALA. The structures of *v*CBM60- cellotriose and *v*CBM60-galactobiose were determined by molecular replacement in the CCP4 version of MOLREP using the

unliganded *v*CBM60 as the search model. The starting models for *v*CBM60-cellotriose and *v*CBM60-galactobiose were refined by rounds of manual rebuilding in COOT (Emsley and Cowtan, 2004), interspersed with restrained refinement in REFMAC (Murshudov, et al., 1997). Solvent water molecules were added using COOT, and checked manually.

5.3.4 Crystal structure of *v*CBM60 with galactobiose and cellotriose

The crystal structure of *v*CBM60 was solved previously by Dr. Cedric Montainier, using the single-wavelength anomalous dispersion (SAD) method employing selenomethionine-labelled protein and refined using data extending to 1.6 Å resolution. *v*CBM60 adopts a β-sandwich fold in which the two β-sheets, each containing four antiparallel β-strands, are connected entirely by loops in Fig 5.12A. The first β-sheet (β-sheet 1) includes β-strands β-4 (Q36-A40), β-1 (I4-G10), β-6 (V61-V68) and β-7 (Q71-Q74), whereas the second β-sheet (β-sheet 2) consists of strands β-3 (T24-T30), β-2 (S16-V21), β-5 (E47-F52) and β-8 (G103-G108). The hydrophobic core of the protein comprises 3 leucine, 3 isoleucine, 2 phenylalanine, 1 tryptophan and 8 valine residues. The calcium ion (Ca1) in *v*CBM60 that connects β−6 and β−7 with the loop linking β−7 and β-8, is highly conserved in CBMs that display a β−sandwich fold. The metal ion makes co-ordinate bonds with D64 O and Oδ1, E97 Oε1, E76 Oε1 and Oε1, Q74 Oε1, with the octahedral coordination completed by a water molecule (Fig 5.1, Section 5.1). The other calcium ion, Ca2, plays a direct role in ligand recognition and will be discussed in this context below (Fig 5.12 and Fig 5.13). The protein also contains a single disulphide bond (C90-C101) that stabilizes the extended loop between β−7 and β−8, while also contributing to the floor of the ligand binding site (Montanier, et.al., 2010).

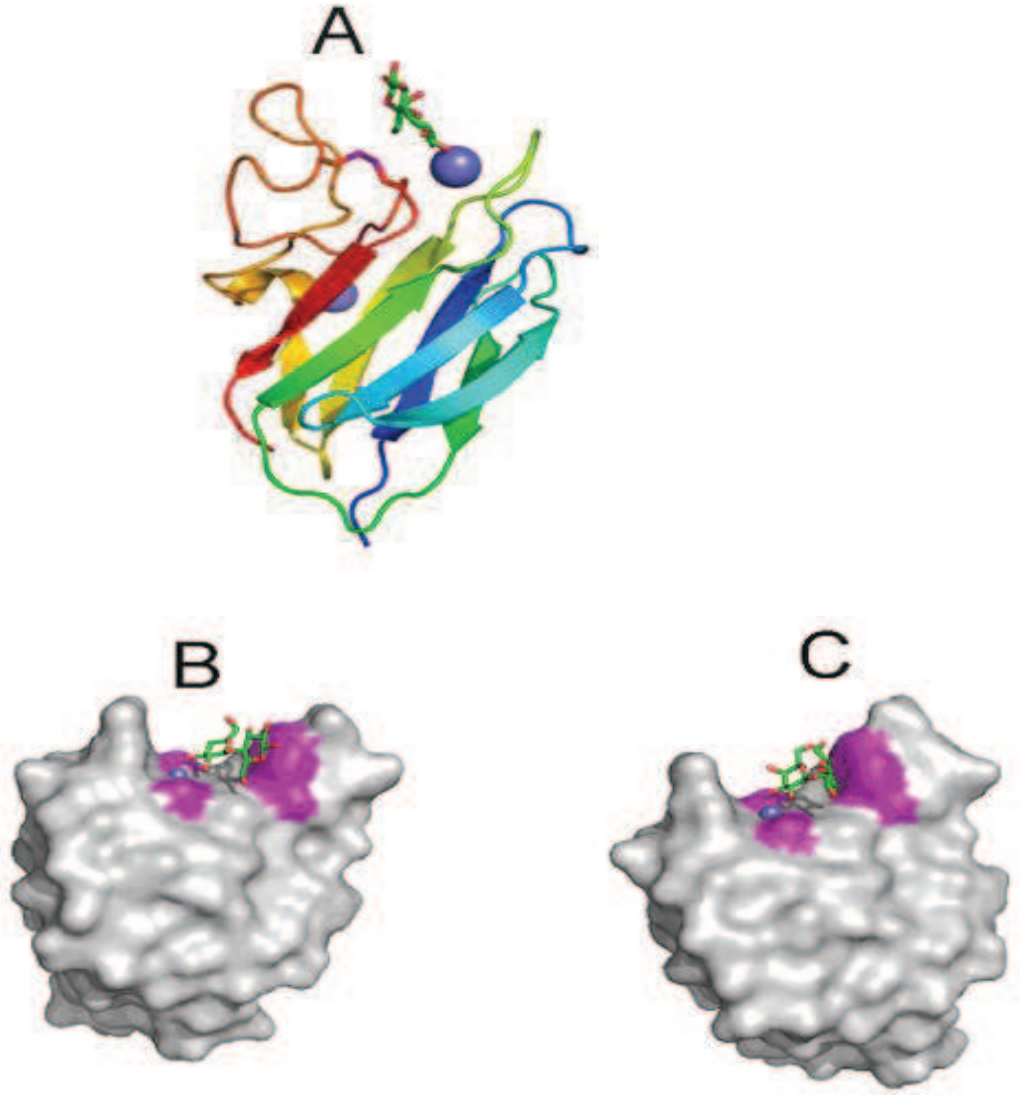


Fig 5.12 Crystal structure of *v*CBM60

Panel A displays a protein cartoon of *v*CBM60 in complex with cellobiose (green) colour ramped from N-terminus (blue) to C-terminus (red). Calcium ions are shown as spheres shaded as blue slate and cellobiose is in stick representation. The disulphide bond stabilizing the loop connecting β-7 and β-8 is shown in magenta sticks. Panel B and Panel C show the solvent accessible surface of *v*CBM60 in complex with cellobiose and galactobiose, respectively. Bound ligand is shown in green (carbon) stick representation. Amino acids whose side chains contribute to ligand recognition are coloured magenta and the ligand binding calcium as a slate blue sphere. The structure figures (Fig 5.1, Fig 5.2, Fig 5.12 and Fig 5.13), were drawn with PyMol (DeLano Scientific; http://pymol.sourceforge.net/).

The crystal structure of *v*CBM60 in complex with cellotriose and galactobiose was solved to resolutions of 1.2 Å and 1.8 Å, respectively. Electron density for both oligosaccharides were evident, and thus the mechanism for the broad ligand specificity displayed by *v*CBM60 could be explored, Fig 5.13. Both oligosaccharides occupied the broad but short cleft located in the loops that connect the two ß-sheets, confirming that it comprised the ligand binding site. With respect to galactobiose the disaccharide is twisted into a helical type structure consistent with the conformation adopted by β−1,4-Gal polymers bound to GH53 galactanases (Ryttersgaard, et al., 2004). The reducing sugar, Gal1, makes parallel hydrophobic interactions with W85, mainly with the pyrole component of the aromatic ring system, but does not make any direct polar contacts with the protein. While Gal2 also makes hydrophobic contacts with W85, primarily the benzene component of the indole ring, the *O2* and *O3* of the sugar make extensive interactions with Ca^{2+}, and with the aspartates that form coordinate bonds with the divalent metal ion. Thus, Ca2 interacts with the ligand binding site by making coordinate bonds with R59 O, H100 O, D55 Oδ1 and D60 Oδ2, while its octahedral coordination is completed through interactions with *O2* and *O3* of Gal2 and a water molecule. In addition to interacting with Ca2, *O2* and *O3* of Gal2 make hydrogen bonds with D60 Oδ2 and Asp55 Oδ2, respectively, and also interact with the backbone carbonyl of H100. It is also possible that *O2* and *O3* make indirect contact with the CBM through the water that interacts with Ca2. The structure of the *v*CBM60-cellotriose complex, Fig 5.13, reveals only two of the three Glc residues. The disaccharide adopts a two-fold screw axis in which the two Glc residues (defined as cellobiose) are orientated at 180° with respect to each other (galactobiose adopts a more helical conformation). The interactions between *v*CBM60 and cellobiose are very similar to the *v*CBM60-galactobiose complex. Trp85 makes planar hydrophobic contacts with Glc1 and Glc2, however the poor electron density displayed by Glc1 indicates that these apolar interactions are weak. The *O2* and *O3* of Glc2 make polar contacts with Ca2, the side chains of D55 and D60, the backbone carbonyl of H100 and indirect interactions through the water that coordinates with Ca2. In addition, the equatorial *O4* appears to make a polar contact with D55 Oδ2, which is not mirrored by the axial *O4* of Gal2. It was not possible to obtain a complex of *v*CBM60 with

xylooligosaccharides. However, as there are no interactions with *O6* of either galactobiose or cellobiose, it is highly likely that *xylo*-conFigd ligands will make similar, if not identical, interactions with the protein to cellobiose. It should be noted, however, that maximal binding to galactan is achieved through interactions with only two sugars, while the protein binds to trisaccharides but not disaccharides of cellulose and xylan. The reason(s) for this subtle difference in specificity is not entirely clear. It is possible that the weak hydrophobic interaction between Glc1 (and by inference Xyl1) with W85 (compared to Gal1, which makes more extensive apolar contacts with the tryptophan) is maximized by presenting a closed pyranose ring to the protein, which occurs when Glc1 is the central sugar in a trisaccharide. Indeed, previous studies have shown that oligosaccharides that extend beyond the ligand binding site of CBMs display higher affinity than oligosaccharides that have a degree of polymerization that matches the number of available sugar binding sites (Boraston, et al., 2004).

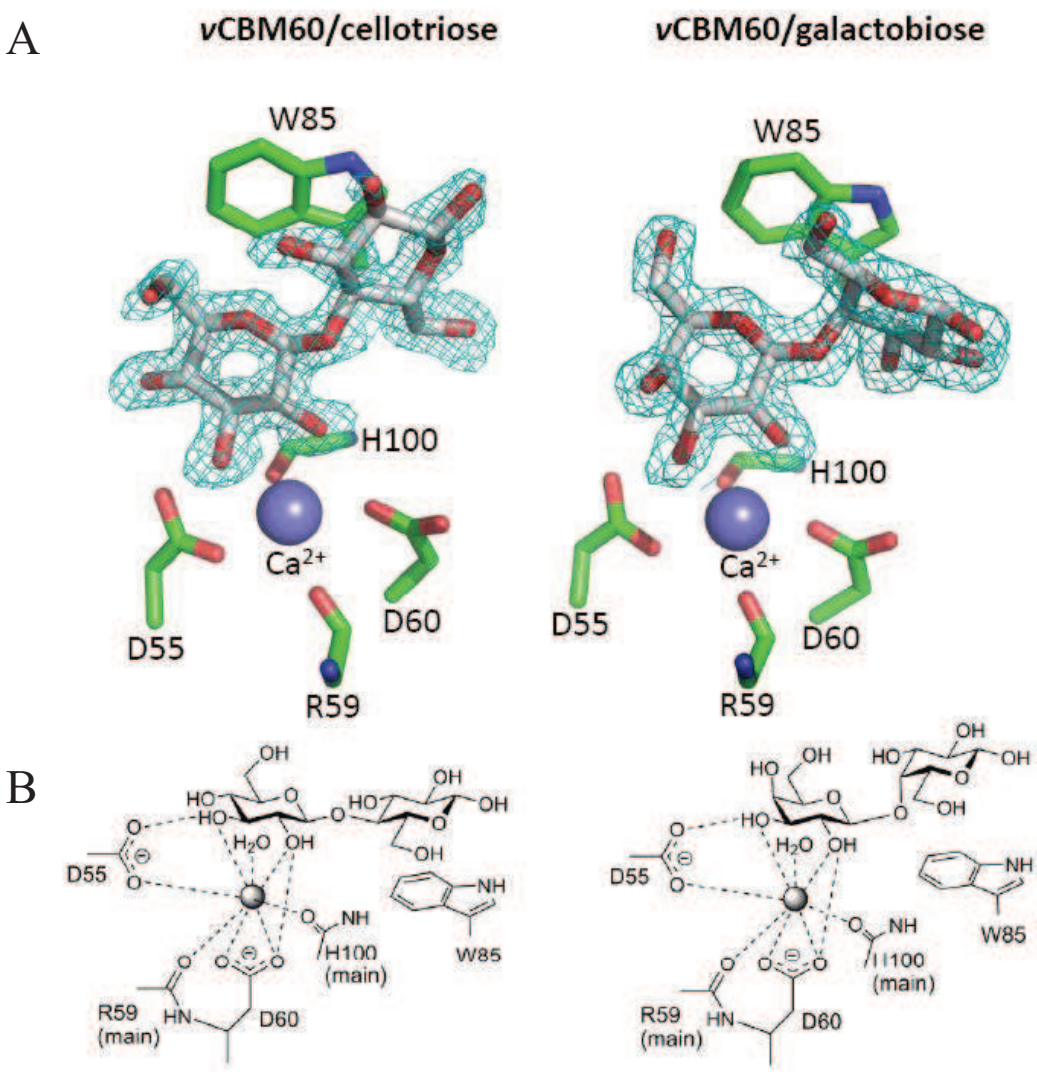


Fig 5.13 The ligand binding site of *v*CBM60

Panel A shows the 3D position of amino acids of *v*CBM60 that contribute to ligand recognition either directly, or through calcium (slate blue spheres). The amino acids (carbons in green) and carbohydrate ligands (carbon in silver) are shown. The maximum likelihood/σa-weighted 2*F*obs – *F*calc electron density map for the carbohydrate ligands is shown in *cyan mesh* and contoured at 1σ (0.61e-/Å3). Only the backbone of R59 and H100 is shown. Panel B is a schematic of Panel A. The dotted lines show polar interactions and calcium is shown as a grey sphere (Montanier, et.al., 2010).

5.4 Results of *C. thermocellum* CBM35-Gal (*CtCBM35-Gal)*

5.4.1 Selection and construction of nine mutants of *Ct*CBM35-Gal

The crystal structure prevciwas harnessed to guide a site-directed mutagenesis strategy to identify the residues that play an important role in ligand recognition. The loop area on the top of the β-jelly-roll (Fig 5.14) was the only region of the protein to display solvent aromatic residues, W40, W108 and Y37, Y137, which generally play a role in carbohydrate recognition (Correia, et.al., 2010). Several amino acids in the putative binding site which presents a shallow pocket, in addition to the aromatic residues mentioned above, could also be involved in ligand binding; which namely Q27, R86, P96 and N140. The role of these residues in ligand binding was investigated by mutagenesis in this study.

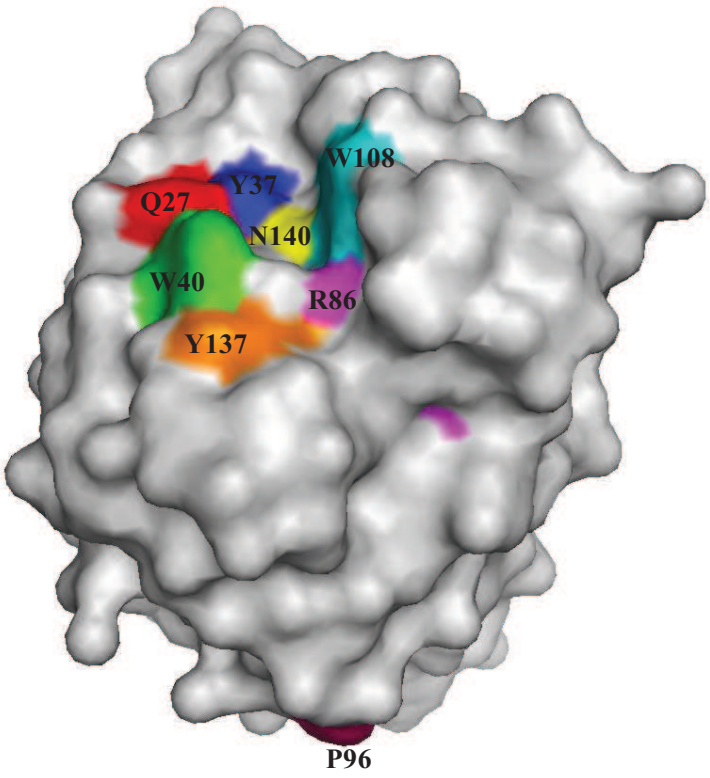


Fig 5.14 Amino-acids selected as potential key residues of *Ct*CBM35-Gal.

Charged polar residues: R86 (pink). Uncharged polar residue: Q27 (red), N140 (yellow). Hydrophobic residues: P96 (dark red), W40 (green), W108 (sky blue), Y37 (deep blue) and Y137 (orange)(also aromatic). The plasmid encoding wild type *Ct*CBM35-Gal was cloned and the crystal structure of *Ct*CBM35 was obtained by Dr. Correia (PDB code 2WZ8) (Correia, et.al., 2010).

Q27, Y37, W40, R86, P96, W108, Y137 and N140 were mutated to alanine. Table 5.4 displays the mutagenic primers. Residues P96 and Y37 were also mutated into phenylalanine (F) to evaluate the influence of the hydroxyl on ligand binding. Site-directed mutagenesis was conducted by employing a PCR-based QuikChange site-directed mutagenesis kit (Stratagene) using p*Ct*CBM35-Gal in pET28a (~5.5 kb) as the template. The PCR products were *DPN*I treated to nick the parental DNA strands, spin purified and transformed into *E. coli* Top10 strain. Three colonies for each mutant were picked and its plasmid DNA was sequenced to confirm the correct mutation had been generated. An example of a PCR and a sequencing data is shown below (Fig 5.15 and Fig 5.16). The wild type and mutants of *Ct*CBM35-Gal were purified to electrophoretic homogeneity and SDS-PAGE showed that all were expressed at a similar level to the wild type protein (data not shown).

Residues		**Primers**	
Q27A-F[a]	5’	GCGGCTGTAAGAGCAAGAGATAATGCTGC	3’
Q27A -R	3’	GCAGCATTATCTCTTGCTCTTACAGCCGC	5’
Y37A-F	5’	GCGTCAGGCGGACAAGCTGTAGGCTGGATTGGC	3’
Y37A-R	3’	GCCAATCCAGCCTACAGCTTGTCCGCCTGACGC	5’
Y37F-F	5’	GCGTCAGGCGGACAATTTGTAGGCTGGATTGGC	3’
Y37F-R	3’	GCCAATCCAGCCTACAAATTGTCCGCCTGACGC	5’
W40A-F	5’	GGACAATATGTAGGCGCGATTGGCAATGG	3’
W40A -R	3’	CCATTGCCAATCGCGCCTACATATTGTCC	5’
R86A-F	5’	GTAGTTGACGCATATTGCAGTATTAGTGTAAACGG	3’
R86A -R	3’	CCGTTTACACTAATACTGCAATATGCGTCAACTAC	5’
P96A-F	5’	GTGTAAACGGAGGAGCCGAAAAAGGGCATT	3’
P96A -R	3’	AATGCCCTTTTTCGGCTCCTCCGTTTACAC	5’
P96G-F	5’	GTAAACGGAGGAGGCGAAAAAGGGCATT	3’
P96G-R	3’	AATGCCCTTTTTCGCCTCCTCCGTTTAC	5’
W108A-F	5’	CAACACCCGTGGAGCGAATACATATCG	3’
W108A -R	3’	CGATATGTATTCGCTCCACGGGTGTTG	5’
Y137A-F	5’	GGCACATCGGGAAGTGCTGCACCGAATATTG	3’
Y137A -R	3’	CAATATTCGGTGCAGCACTTCCCGATGTGCC	5’
N140A-F	5’	GGAAGTTATGCACCGGCTATTGATAAAATAC	3’
N140A -R	3’	GCTATTTTATCAATAGCCGGTGCATAACTTCC	5’

Table 5.4 Quickchange™ mutagenetic primers for *Ct*CBM35-Gal.

[a]-F and –R denotes sense and antisense primer respectively. Mutated nucleotides are underlined.

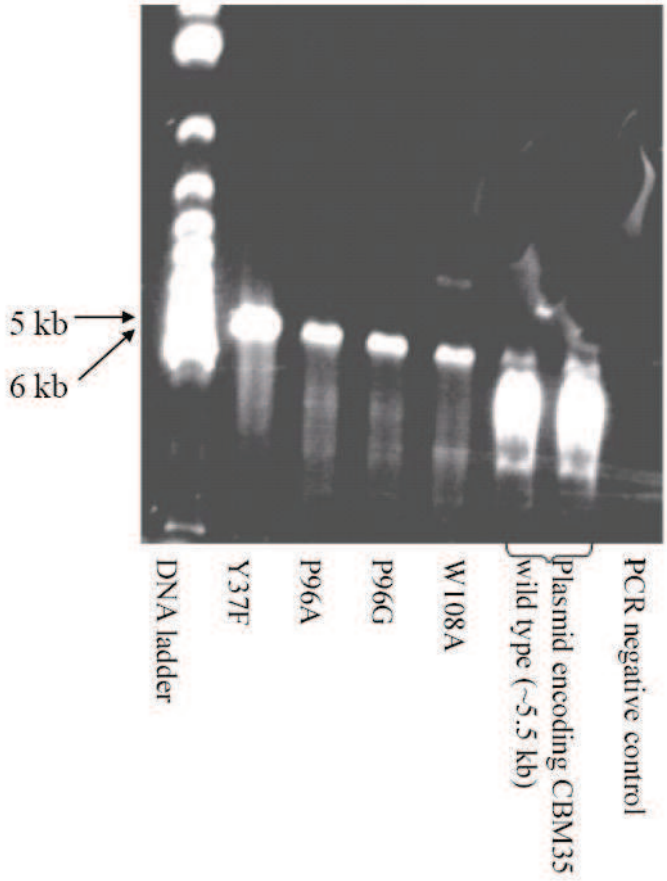


Fig 5.15 Agarose gel electrophoresis of PCR products after QuikChange site-directed mutagenesis

PCR reactions were carried out, as described in Section 2.1.18, using p*Ct*CBM35-Gal encoded in pET28a (~5.5 kb) as DNA template, and respective primers from Table 5.4. A 10 µl (~5 µg) of DNA Hyper Ladder IV and an aliquot of 5 µl PCR product (~5.5 kb) from each PCR reaction were electrophoresed using 1 % agarose gel, as described in Section 2.1.13.

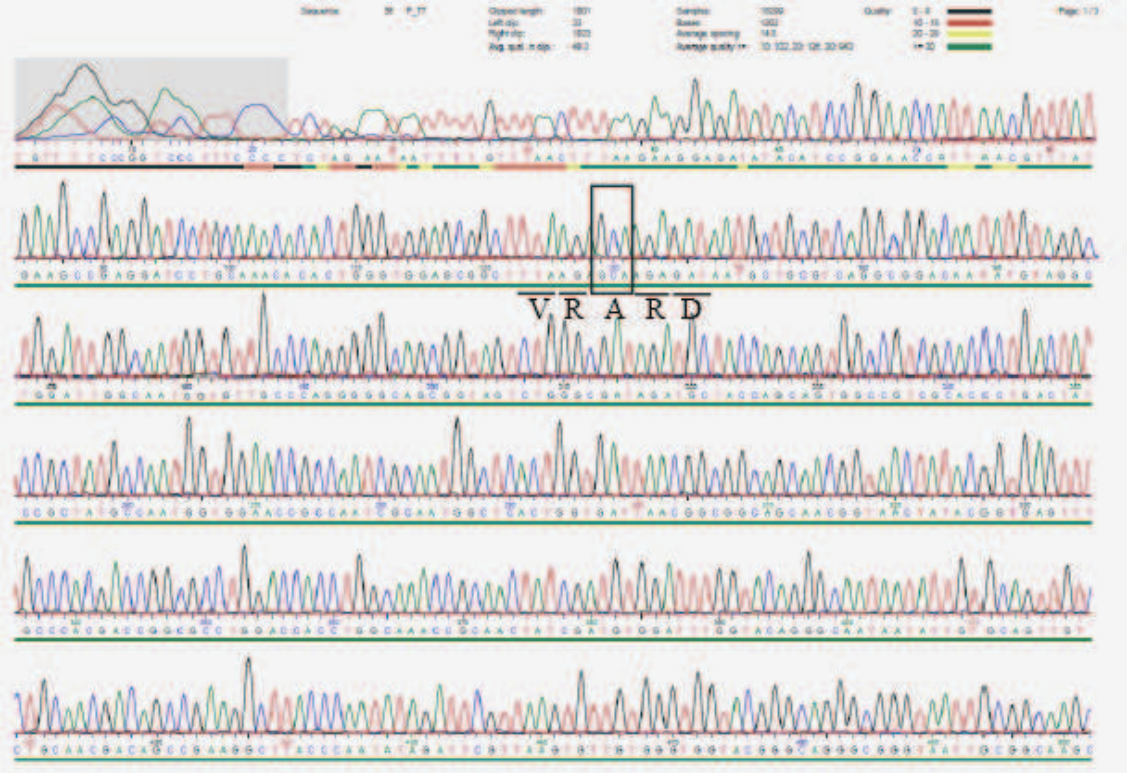
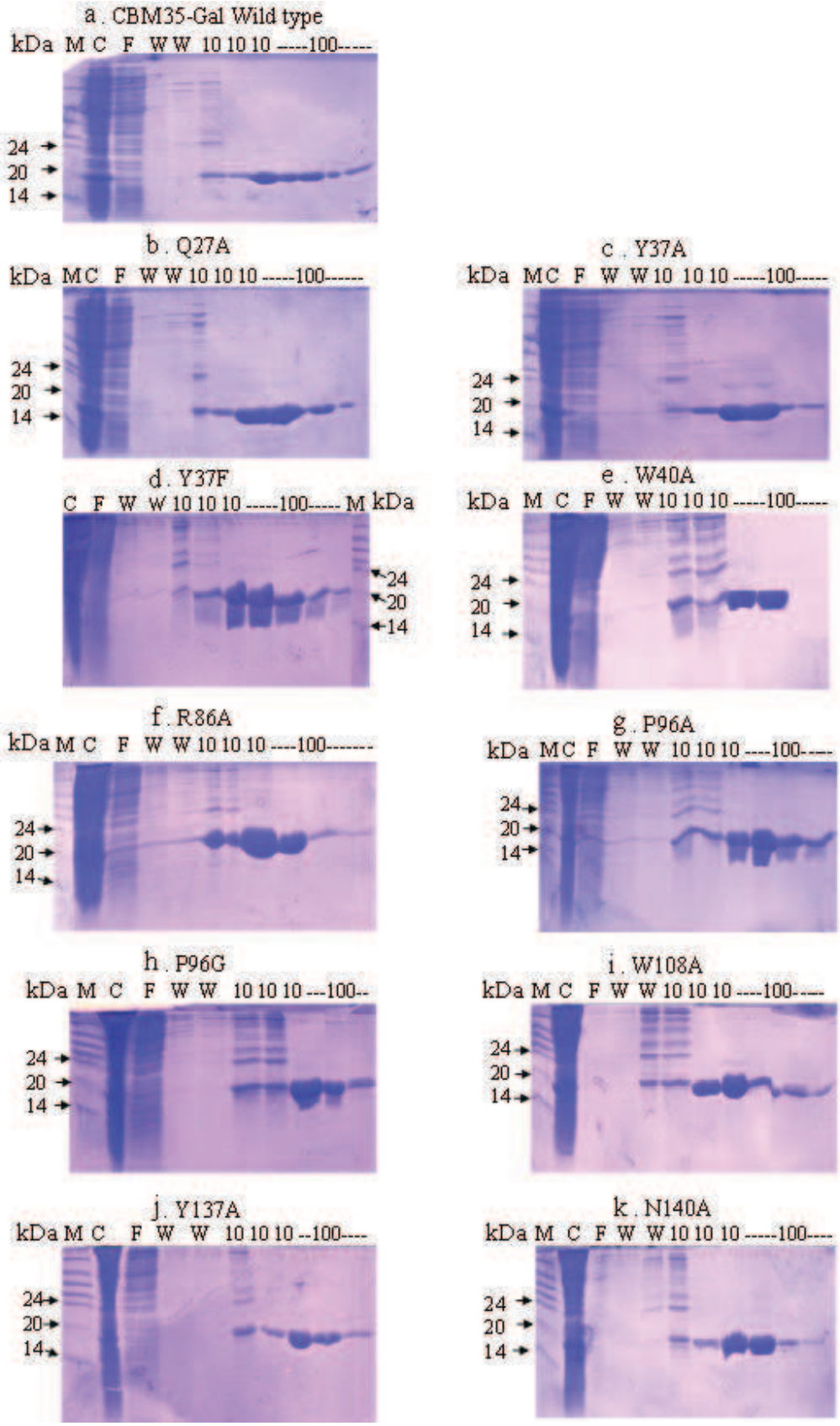


Fig 5.16 An example chromatogram of automated sequencing of the gene encoding *Ct*CBM35-Gal mutant Q27A using a T7 forward sequencing primer The nucleotide sequence CAA encoding Q27 (glutamine) in the wild type *Ct*CBM35-Gal was mutated into GCA, which encodes A (alanine). The mutation was boxed in the figure, and the amino acids encoded were indicated beneath the nucleotide sequence.

5.4.2 Over-expression and purification of *Ct*CBM35-Gal

Wild type and variants of *Ct*CBM35-Gal, were expressed in *E. coli* BL21 DE3 (Novagen) strains containing kanamycin (50 μg/mL) as described in Section 2.1.21. The recombinant strain was cultured at 37 °C to mid exponential phase (A_{600} ~0.6), at which point isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the cultures were incubated for a further 16 h at 16 °C. The recombinant proteins were purified by IMAC (Fig 5.17).


a . CBM35-Gal Wild type
kDa M C F W W 10 10 10 ----100-----
24
20
14
b . Q27A
kDa M C F W W 10 10 10 -----100------
24
20
14
c . Y37A
kDa M C F W W 10 10 10 -----100-----
24
20
14
d . Y37F
C F W W 10 10 10 -----100----- M kDa
24
20
14
e . W40A
kDa M C F W W 10 10 10 -----100-----
24
20
14
f . R86A
kDa M C F W W 10 10 10 ----100-------
24
20
14
g . P96A
kDa M C F W W 10 10 10 ----100-----
24
20
14
h . P96G
kDa M C F W W 10 10 10 ---100--
24
20
14
i . W108A
kDa M C F W W 10 10 10 ----100-----
24
20
14
j . Y137A
kDa M C F W W 10 10 10 --100----
24
20
14
k . N140A
kDa M C F W W 10 10 10 -----100----
24
20
14

5.4.3 ITC of wild type and mutants of *Ct*CBM35 to carob galactomannan (low viscosity)

The data presented in Fig 5.18 and Table 5.5, show that all the mutations prevented binding to galactomannan. Thus, the shallow pocket which houses the mutated residues comprises the ligand binding site.

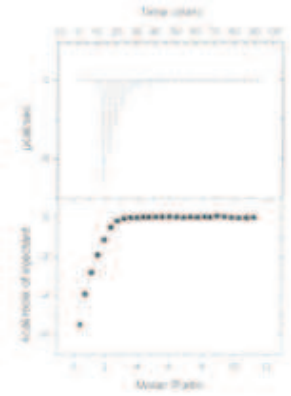
Protein	**Affinity x 10^4 (M^{-1})**	***ΔG*** **(kcal mol^{-1})**	***ΔH*** **(kcal mol^{-1})**	***TΔS*** **(kcal mol^{-1})**	**n[a]**
WT	5.10 ±1.05	-6.43	-6.90±0.47	-0.47	1.00±0.05
Q27A	binding *	-	-	-	-
Y37A, Y37F W40A, R86A, P96A, P96G W108A, Y137A, N140A	n.b	-	-	-	-

Table 5.5 The binding affinity and thermodynamic parameters of wild type and mutants of *v*CBM60 to oligosaccharide determined by ITC. ITC was carried out in 50 mM Na Hepes pH 7.5 containing 5 mM $CaCl_2$. Typical ITC data for wild-type, ten mutants of CBM35 titrated with 5 mg/ml carob galactomannan (low viscosity). The ITC data were fitted to a single site binding model for all ligands.

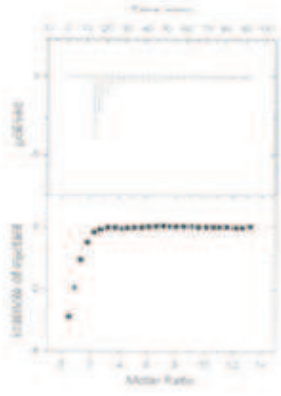
*binding detected but too weak for accurate quantification.

n.b.: no binding occurred.

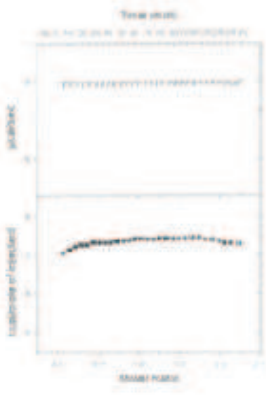
A. Wild type



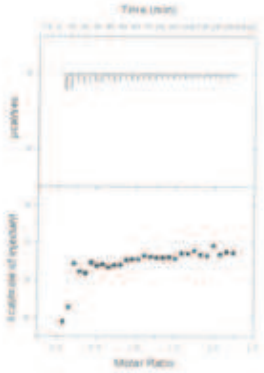
B. Q27A



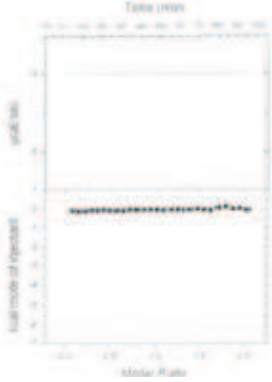
C. Y37A



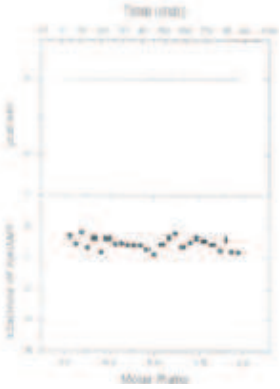
D. Y37F



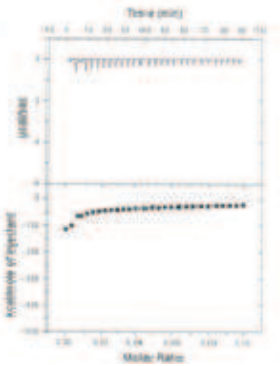
E. W40A



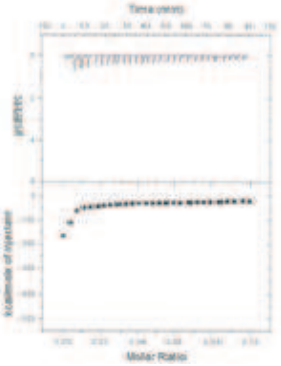
F. R86A



G. P96A

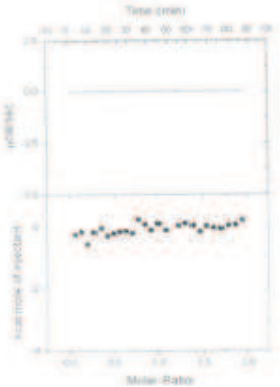


H. P96G

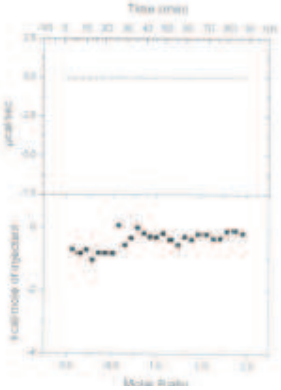


I. W108A

J. Y137A



K. N140A



5.5 Discussion

Previous ITC showed that *v*CBM60 and *Cj*CBM60A display similar ligand specificities (Montanier, et.al., 2010). The protein module *v*CBM60 binds to galactan, xylans and β−glucans (that contain β−1,4 linkages), although the actual affinities for these polysaccharides are considerably lower than observed with *Cj*CBM60A. This likely reflects the inability of *v*CBM60 to oligomerize (data not shown), precluding any benefits derived from avidity effects. Indeed, this view is supported by the observation that the artificial construct, containing two tandem copies of *v*CBM60 (designated *v*CBM60-*v*CBM60), binds, to xylan and galactan, respectively, 20 and 40-fold more tightly than the monomeric form of the protein (Montanier, et al., 2010).

In the CBM60 family, there are 15 other protein modules that display a high level of sequence similarity. For example, *Cj*CBM60A and B have an e score $<10^{-14}$ (Montanier, et al., 2010). All of these proteins were from bacteria and 14 of them were contained in GH10 and GH11 xylanases, indicating that the primary ligand for CBM60 modules is xylan. The residues coordinating the ligand binding calcium ion (D55, D60 and H100) are invariant in the 15 protein modules, and W85, which comprises the second sugar binding site, is always an aromatic side, therefore it is a functionally conserved amino acid in the family. This explains the similarity of the ligand specificity between the different members in this family. As mentioned above binding affinity is variable in CBM60. Some members contain a 10-residue C-terminal extension which might lead to protein dimerization, and thus higher affinity due to an avidity effect. Indeed a CBM60 from *Saccharophagus degradans,* a marine bacterium, contains a tandem copy of CBM60 that have the C-terminal extension that assembles multiple copies of these proteins into a dimer to mediate tight bindings to multivalent ligands.

ITC studies showed that, typical of carbohydrate-protein recognition (Boraston, et al., 2002a; Charnock, et al., 2000), binding to polysaccharides and oligosaccharides was

driven by favourable changes in enthalpy, with generally an unfavourable entropic contribution. Mutagenesis studies and ITC data (Fig 5.5, Table 5.1) are remarkably consistent with the structure of *v*CBM60 in complex with galactobiose and cellotriose. Mutating D60, D55 and W85 to alanine lead to a complete loss in binding to pectic galactan and cellulose. D55 and D60 coordinate the calcium that makes polar contacts with the O2 and O3 of the sugars in different ligands. W85 is also shown to be crucial in ligand binding. This well conserved aromatic residue is likely to make hydrophobic stacking interactions with the sugar rings of target carbohydrate ligand. W85 comprises the second sugar binding site and interacts with Glc1 and Gal1 (Fig 5.13). Mutating W85 to alanine abolished *v*CBM60 binding to all three ligands indicating the importance of the aromatic residue in driving ligand recognition. Thus, the mutagenesis data are consistent with the view that xylo-configured ligands recognize and interact with *v*CBM60 in a similar mechanism to either galactobiose or cellobiose.

W98 is perpendicular to W85 shown by crystallography, but the role of W98 in ligand recognition is much less obvious than W85, as the W98A mutation did not abolish the affinity for galactan, xylan or cellulose. It reduced the binding affinity for galactan, xylan or cellulose by 45 %, 26 % and 34 %, respectively (Table 5.1). These data are consistent with the observation that ligands channeled into the binding cleft are not in proximity to W98. From the thermodynamic parameters, the values of both enthalpy and entropy become more negative when W98 was mutated to alanine, however, the change in enthalpy was compensated by a similar change in entropy to all three complex ligands, which resulted in a similar ΔG values, compared to those of wild type. The significance for the change in the thermodynamics of ligand binding is unclear.

When H100 is mutated to alanine, the binding affinity of *v*CBM60 for galactan is increased by two fold. Since the two fold change is very modest, and the galactose unit is too far away from other residues in the binding site to make any interaction, it is very likely the mutation caused a slight change in the peptide backbone. Hence, W98 and the loop it is in came closer to the galactose unit in galactan, making a hydrophobic contact with the C5 and C6 of galactose residues. Moreover, mutating

H100 to alanine only had a small loss in binding to ß-glucan and a 10 fold decrease in binding to xylan. Thus the imidazole side chain of H100 may make weak polar or apolar interactions with xylose but not glucose residues.

The ligand binding sites in CBMs that display a β−sandwich fold comprise, typically, the concave surface presented by one of the β−sheets, or at the end of the elliptical protein, within the loops connecting these two structural elements (Boraston, et. al, 2003). Inspection of the concave surface of *v*CBM60 does not reveal a cleft–like structure that would accommodate polysaccharide ligands. By contrast, the loops connecting the two β−sheets presents a very broad but short cleft. The floor and one wall of the cleft are formed by the loop connecting β-7 with β-8, while the other wall comprises the loop linking β-5 with β-6. Confirmation that the loop-derived cleft consists of the ligand binding site is derived from mutagenesis studies. Substitution of residues on the surface of this cleft with alanine either completely abrogates, or greatly reduces, the affinity of the protein for its carbohydrate ligands (Fig 5.18). Indeed, the observation that these mutations cause a similar reduction in affinity for the gluco-, xylo- and galacto-conFigd ligands demonstrates that the protein interacts with these polymers at a common binding site.

*v*CBM60 is a calcium dependent protein module. The divalent metal ion is in the loops that link the two ß-sheets comprising the binding site. In general, calcium plays a role of mediating particularly tight ligand binding in CBMs that bind to monosaccharides (Montanier, et al.), but not in those modules that bind to more extensive ligands. It has been shown that CBMs that bind to four or more sugars in target polymers do not use Ca^{2+} in ligand recognition (Jamal, et al., 2004). As in CBM36, CBM60 interacts with only two or three internal sugars, and calcium is critical in carbohydrate binding (Jamal-Talabani, et al., 2004), and makes contact to ligands through charged dipole-dipole interactions (Fersht, et al., 1985).

5.5.1 Circular permutation of CBM36 and CBM60

Circular permutation, is a genetic event (or post-translational processing) lead to the ligation of the N and C termini of the protein and subsequent cleavage at another site generating new N and C termini. It is believed to have occurred either in CBM60 or CBM36. CBM family 60 is a structurally homologous and functionally related to the family 36 CBM from *Paenibacillus polymyxa* xylanase 43A (Montanier, et.al., 2010). The two CBMs display 17 % sequence identity. The structure of the CBM36 binding site is also a short broad cleft formed by the loops connecting the two β-sheets. The ligand binding sites in these two families are remarkably conserved (Fig 5.19); in CBM36 Xyl2 (non-reducing sugar of xylobiose) makes contacts with two tyrosines through hydrophobic interactions, while Xyn1, through O2 and O3, forms polar contacts with calcium and the two aspartates and the backbone carbonyls that coordinate the metal ion. Xyl2 sits deeper in the cleft in CBM36 than Gal1 or Glc1 in *v*CBM60, and it seems likely that the –OH on C6 of Glc2 will clash with Y40 in the CBM36 binding site, thus CBM36 is unlikely to bind to galactan or cellulose, but will recognize xylan. Thus, the structural basis for this ligand specificity is subtle. Moreover, the locations of the carbohydrate/metal binding residues in the two protein sequences are not equivalent. In CBM36, D116 and D121 and W120 which coordinate the calcium, are at the C-terminus, while W26, which makes interactions with the ligand at the second sugar binding site, is at N-terminus. In *v*CBM60, the residues in contact with the calcium ion is in the central region of the protein module, while W85, comprising the second sugar binding site is in the C-terminal region of the protein.

Therefore, the structural and sequence similarity of non-contiguous regions in CBM36 and *v*CBM60 suggests they are evolutionary linked and that a circular permutation event occurred. The overlaid 3D structure of the two modules showed a much higher degree of structural-based sequence identity than the similarity derived from linear primary sequence alignments (Fig 5.19). The genetic event leading to circular permutation is the ligation of the DNA encoding C and N terminal amino acids of the protein, followed by a cleavage at another internal site of the gene, which generated a new C and N terminus (Lindqvist and Schneider, 1997). This could happen because

*v*CBM60 and CBM36 display a stable β-sandwich fold with the N- and C- termini in close proximity. It should be noted that the strand order in CBM36 is conserved in the other ß-sandwich CBM families. It is likely, therefore, that CBM60 arose through the circular permutation of CBM36 rather than *vice versa*.

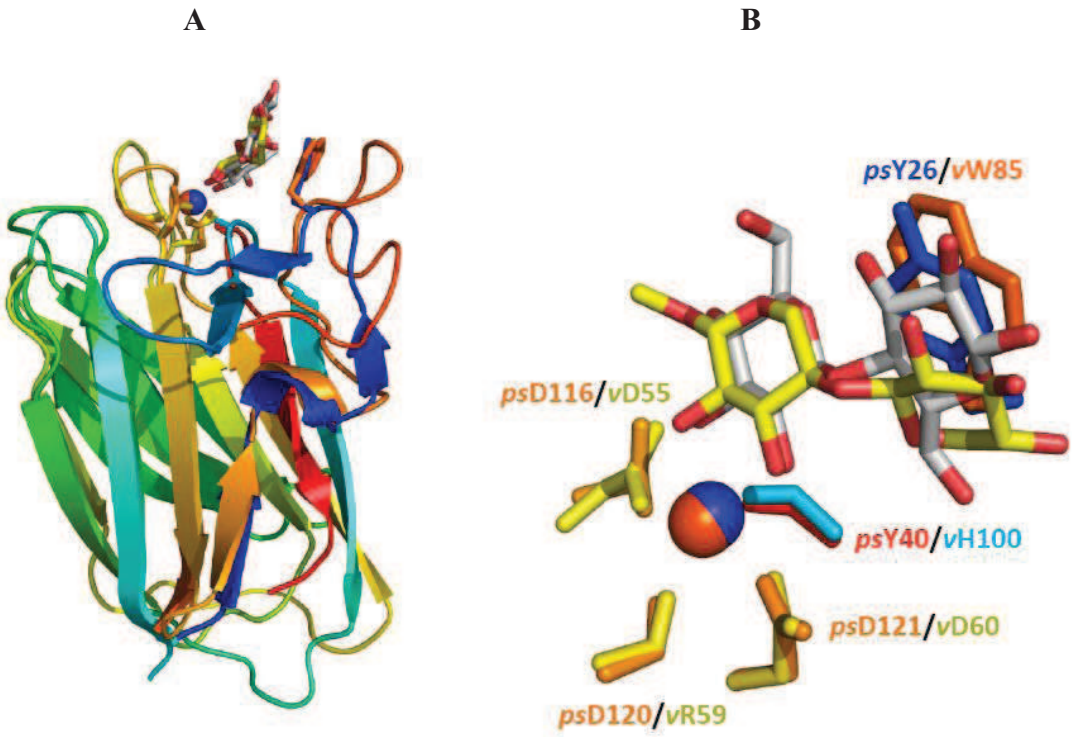


Fig 5.19 Overlay of the structures of *v*CBM60 and CBM36

The figure shows an overlap of the secondary structure of the two proteins. The calcium atoms are displayed as spheres; the ligands are xylobiose (silver) in CBM36 and cellobiose (yellow) in *v*CBM60. Panel A shows that the two proteins are colour ramped from the N-terminus (blue) to the C-terminus (red), and there is a structural conservation of the ligand binding apparatus of *v*CBM60 and CBM36. Panel B shows the locations of the carbohydrate/metal binding amino acids in the two protein sequences are not equivalent. The aspartates Asp^{116} and Asp^{121} and Trp^{120}, which coordinate the ligand-binding calcium in CBM36 (shown in orange), are at the C-terminal region of the protein; in contrast, the location of amino acids that interact with the critical ligand binding calcium in *v*CBM60 (D55, R59 and D60, shown in green) are located in the central region of the protein.

The majority of CBMs display tight specificity for their target ligands. CBM60, however, exhibits unusually broad ligand specificity, recognizing xylan, galactan and glucans. CBM60 targets the equatorial O2 and O3 of pyranose sugars as its primary ligand specificity determinant, which is a common feature of many carbohydrates. The second determinant is the geometric signature of such sugars in which the O3-O3-O2-O2 torsion angle is 60º. For example, mannnoside contains O3-O3-O2-O2 but the torsion angle is -60º and thus CBM60 can not recognize mannan. It is believed that xylan is the primary target ligand of CBM60, as the majority of CBM60 are located in xylanases (Millward-Sadler, et al., 1995). However, since the equatorial O2 and O3 of pyranose sugar is also present in galactan and cellulose, CBM60 is also able to recognize and bind to their polymers, galactan and glucans.

It is observed that CBM60-containing xylanases usually also contain xylan esterases, which target acetyl groups decorated at some of the O2 and O3 positions of xylose residues in the backbone of xylan polymers (Brett, et al., 1996). By targeting these hydroxyl groups in a single xylose, coupled to apolar contacts with the adjacent xylose, the binding length is only two xylose units. As xylan does not have decorations on every xylose position; CBM60 can bind to a vast number of decorated xylans with various degrees of decoration. This multi-modular enzyme utilises CBM60 firstly to lock onto xylan, the esterase then removes acetyl side groups on the backbone, which enables the xylanase to degrade the xylan more easily. This reinforces the importance of CBM60 binding to only a limited region of xylan, unlike many other xylan-specific binding modules that target as many as six xylose units with hydroxyls groups exposed to solvent, which enables these CBMs to recognize ligands with a high degree of decoration (Simpson, et al., 1999; Szabo, et al., 2001).

The first CBM60 identified was from *C. japonicus,* which is a bacterium that is capable of synthesizing a sophisticated xylan degrading system (Charnock, et al., 2002a). The discovery of the CBM60 family completes the modular assignment of the xylan degradation system in *C. japonicus*. The system comprises Xyn5A and Xyn10A,

which contain cellulose-binding CBMs and thus likely degrade xylan in close proximity to cellulose in plant cell walls. The bacterium also secretes Xyn10B, containing a uronic acid binding CBM35, and thus may digest xylans closely associated with pectins. Finally, *C. japonicus* produces CBM60-containing Xyn11A and Xyn11B that target xylans which are not in intimate contact with cellulose or pectin. In conclusion, the flexibility in xylan recognition displayed by CBM60 underscore the utility of this CBM family in directing enzymes to different substructure of this major hemicellulosic polysaccharide, illustrating its utility in the toolbox of biocatalysts required to deconstruct the plant cell wall.

5.5.2 *Ct*CBM35-Gal

Previous ITC studies carried out by Montainier et.al showed that *Ct*CBM35-Gal binds to D-Galactose and galactomannan, indicating that Gal is the major specificity determinant of galactomannan which contains an α-Gal decoration. However, it does not bind to xyloglucan which contains β-D-Gal epitopes, suggesting CBM35-Gal is able to distinguish between the two anomers of the hexose sugar. The affinity of CBM35-Gal for galactomannan is about 8-fold tighter than that for D-galactose. Tighter binding of CBMs to multivalent, compared to monovalent, ligands is normally associated with avidity effects, mediated through protein oligomerization (Boraston, et al., 2002a; Freelove, et al., 2001). However, the difference observed here is not adequate to indicate such an event. One explanation for the difference in the affinities is that the conformation of galatomannnan is more rigid than galactose, therefore there is a smaller entropic penalty when galactomannan binds to *Ct*CBM35-Gal, which is consistent with the observed TΔS values.

*Ct*CBM35-Gal adopts a β-sandwich fold in which the two β-sheets, containing four and five antiparallel β-strands, respectively, are connected entirely by loops (Fig 5.20). The two β-sheets are twisted, and it could be argued that the protein displays an elongated β-barrel-like structure. The first β-sheet (β-sheet 1) includes β-strands β-1 (Ile8-Glu12), β-9 (Asn140-Ala147), β-4 (Gly60-Ala69), and β-7 (Thr113-Leu120), whereas the second β-sheet (β-sheet 2) consists of strands β-2 (Gly21-Ala24), β-3 (Tyr48-Tyr55), β-

8 (Asn124-Tyr130), β-5 (Arg86-Val92), and β-6 (Lys98-Phe102). The hydrophobic core of the protein comprises four leucines, seven isoleucines, five phenylalanines, four valines, and one methionine. The calcium ion in *Ct*CBM35-Gal is positioned between the loops linking β-1 and β-2 and the loop that joins β-2 and β-3, while β-9 also contributes to the metal binding site. This cofactor is highly conserved in CBMs that display a β-sandwich fold and, in the case of Xyl-CBM4 from the thermophile *Rhodothermus marinus*, and likely other CBMs that display this fold, confers significant thermostability (Abou Hachem, et al., 2000).

From site-directed mutagenesis studies, it is seen the ligand is accommodated within the loops that connect the two beta-sheets. A model is proposed in which the orientation of the pair of aromatic residues that interact with the two faces of the Gal pyranose ring plays a pivotal role in orientating the axial O4 atom of the ligand toward N140, which is invariant in CBM35. Subtle differences in the conformation of conserved residues in the ligand binding site lead to the loss of galactomannan recognition.

Within a wider context for the CBM35 family, where the ligand binding sites are similar structures, but display divergent ligand specificity, the signature residues in their ligand binding sites are studied. The ligand binding site of the CBM35-Gal displays significant structural similarity with calcium-dependent CBM35s that target uronic acids, subtle differences in the conformation of conserved residues in the ligand binding site lead to a complete loss of galactomannan recognition. In the CBM35 ligand binding site targeting uronic acid, a key element of uronic acid recognition is through protein-bound calcium that makes a polar contact with the carboxylic acid of the ligand. Calcium recognition is conferred, in part, by a stretch of amino acids at the end of β-strand 4, consisting of an Asp/Asn-Tyr/Thr-X-Asn consensus sequence (Correia, et al.). The flanking residues Asp/Asn and Asn coordinate the metal directly through side chain oxygen atoms. The Tyr/Thr motif presents its peptidyl backbone carbonyl oxygen, regardless of functional group chemistry, and the X residue operates as a spacer to create appropriate spatial geometry for the C-terminal asparagine to be positioned for metal coordination. In contrast, *Ct*CBM35-Gal confers some

components of the calcium binding pocket of the uronic acid bonding (typically in Chitin specific CBM35, designated as Chi-CBM35) (Correia, et al., ; Montanier, et al., 2009a). There are however, two subtle structural differences which explain why *Ct*CBM35 can not use Ca^{2+} directly during ligand recognition. First, the "Asp/Asn" residue in Chi-CBM35 (Asn33) is substituted with Gly39 in *Ct*CBM35-Gal. This amino acid replacement not only removes one of the calcium ligands (Oδ1 of Asn33 in Chi-CBM35), but also decreases the torsional constraints on the next residue, Trp40, allowing the side chain to rotate 180 ºC relative to Tyr34, the equivalent residue in Chi-CBM35. This structural transition has two effects. First, the bulky side chain occupies the equivalent space used by the coordination pocket within the uronic acid binding CBM35s. Second, while the carbonyl of Tyr34 in Chi-CBM35 can coordinate with the calcium, the backbone oxygen of Trp40 in *Ct*CBM35-Gal is angled in the opposite direction, toward the core of the protein, and therefore is unable to interact with a potential ligand binding metal ion (Correia, et al., ; Montanier, et al., 2009a). The second feature of *Ct*CBM35-Gal that leads to the loss of calcium binding is the insertion of two residues (Ile41 and Gly42) between the conserved Tyr/Thr and Asn residues (Asn43 in *Ct*CBM35-Gal and Asn36 in Chi-CBM35). The insertion of an additional residue results in the translation of Asn43 out of proximity to where the metal binding site would lie. Thus, despite the high degree of sequence identity, the only components of the calcium binding site that are structurally conserved in *Ct*CBM35-Gal and Chi-CBM35 are Asn140 and Asn122, respectively. It would appear, therefore, that the inability of *Ct*CBM35-Gal to recognize uronic acids exists primarily because of the loss of calcium coordination in the ligand binding site. However, many of the remaining features of the ligand binding site of *Ct*CBM35-Gal, which are conserved in the uronate-specific CBM35s, are harnessed by CBM35-Gal for galactose recognition (Montanier, et al., 2009b).

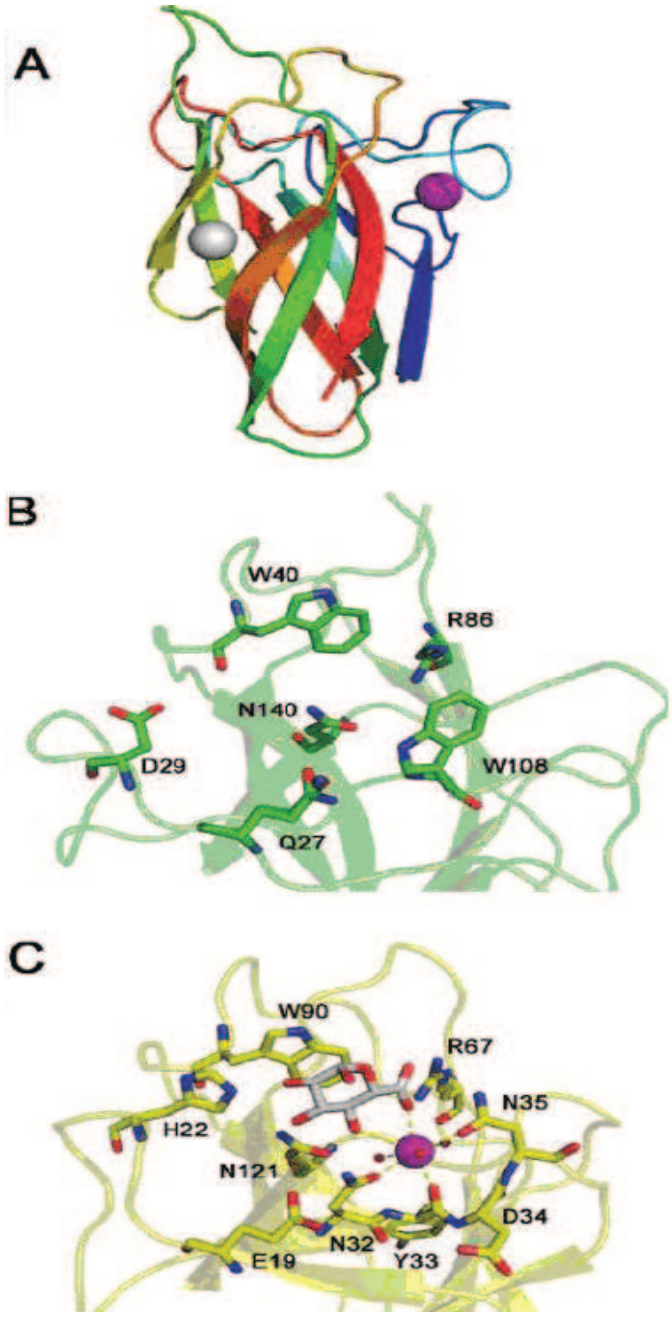


Fig 5.20 Three-dimensional crystallographic structure of *Ct*CBM35-Gal. (A) Cartoon representation of *Ct*CBM35-Gal. The color scheme is ramped from the N-terminus (blue) to the C-terminus (red). The structural calcium is shown as a magenta sphere and the bound magnesium as a silver sphere. Structural alignment of *Ct*CBM35-Gal (green) (B) and Chitin specific CBM35 (yellow) (C). Amino acids involved in binding uronic acids in Chitin specific CBM35 are labeled and aligned with binding site residues in *Ct*CBM35-Gal. The coordinated calcium in Chi-CBM35 is represented as a magenta sphere (Correia, et.al., 2010).

CBM35 is also related to CBM family 6. The ligand binding site of CBM6 has been dissected into five key molecular regions (A-E) that contribute to affinity and

specificity (Abbott, et al., 2009). Comparison of the five characterized regions (A-E) in CBM6s with the three-dimensional structures of CBM35 demonstrates that there are indeed similarities between the two families (Correia, et.al., 2010). Region B in CBM35s displays structural variability, ranging from a single tryptophan in *Ct*CBM35-Gal, to a metal coordination site in the uronate binding CBMs, to being completely absent in *Cj*CBM35-Man. Such structural diversity suggests that this region is a critical functional region that significantly contributes to ligand selectivity in this family. By contrast, region B in CBM6 displays less plasticity comprising a highly conserved aromatic residue. Indeed, the aromatic residues in regions B in CBM6 contribute to a conserved mechanism of ligand recognition in which the aromatic residues stack against the two planar faces of the bound ligand. In CBM35s, Region C, comprising an asparagine at the base of the binding site, is invariant in both CBM6 and CBM35. Depending on the orientation of the ligand, the residue makes critical hydrogen bonds with O2, O3, or O4 of the central sugar. Loops are present in the sequence of region E in each CBM35 with a known structure; however, *Ct*CBM35-Gal and *Cj*CBM35-Man contain hallmark aromatic residues, which are important for functional specialization in these regions. *Ct*CBM35-Gal contains a tyrosine residue (Tyr137) that is conserved within the binding site of CBM6 from *Clostridium cellulolyticum* (*Cc*CBM6). In both structures, these tyrosines block off the binding site cleft and contribute to the exo-specificity displayed by these proteins. This orientation of the reserved key residues is modified by flanking glycine residues, which leads to a subtle changes in protein structure and radical changes in ligand specificity.

Comparing the similarity and difference of CBM35-Gal and CBM60, it can be seen that CBM35-Gal and CBM60 share similar features. They both have structurally-related families, CBM6 and CBM36, respectively; but display subtle differences in ligand specificity, which were explained at a structural level in this chapter. Finally, they are both harnessed in the deconstruction of the plant cell wall by *Ct*CBM35-Gal targeting α-D-Galactose residues in galactomannan, or CBM60 targeting the ß-1,4-backbone in pectic galactan, xylan and cellulose.

In conclusion, CBM60 uniquely displays specificity for a wide range of polysaccharides, binding to galactans, xylans, glucans. Mutation of Ca^{2+}-coordinating residues Asp55 and Asp60, and Trp 85 leads to loss of binding to all three ligands, indicating these three residues play a key role in carbohydrate recognition. In contrast, mutating Trp98 and H100 reduced but did not completely destroy binding to all three ligands, thus these two amino acids have a smaller impact on carbohydrate recognition. Mutating H100 to alanine leads to a 2-fold increase in affinity for pectic galactan. The crystal structure of *v*CBM60, a member of the CBM60 family, displays a ß-sandwich with the ligand binding site consisting of a broad cleft formed by the loops connecting the two ß-sheets. Ligand recognition at site 1 within the cleft is, through hydrophobic interactions, while binding at site 2 is conferred by polar interactions between a protein-bound calcium and the *O2* and *O3* of the sugar. The observation that ligand recognition at site 2 requires only a sugar that contains equatorial hydroxyls at *C2* and *C3*, explains the broad ligand specificity displayed by *v*CBM60. Meanwhile, studies of another galactose binding module, *Ct*CBM35-Gal, reveal that *Ct*CBM35-Gal accommodates galactomannan within the loops that connect the two ß-sheets. All ten residues selected and mutated around the binding site abolished binding to galactomannan.

In nature, CBM families 35 and 60 play an important role in directing enzymes to target and hydrolyze the plant cell walls, to provide important substrates for bacteria to utlise. The flexibility in ligand recognition displayed by CBM60 underscores the utility of this CBM family in directing enzymes to different substructures of this major hemicellulosic polysaccharide. It illustrates that it may be utilised as biocatalysts to deconstruct plant cell walls. *Ct*CBM35-Gal is also reported that it provides evidence that CBMs can target the terminal α-D-Gal residues of complex polymers, expanding further the repertoire of carbohydrates recognized by CBMs (Montanier, et al., 2009b), thus it may be utilised for laundry industry. For example, to target human blood stains, as galactose is one of the most common antigens on the surface of human red blood cells. Overall, the deconstruction of plant cell walls by these two CBM families illustrates their utility as the toolbox of biocatalysts which may contribute to laundry field.

Chapter 6. Final Discussion

6.1 The importance of synergistic interactions between enzymes, surfactants, savinase and other detergent addictives for laundry formulation development

The first large-scale application of microbial enzymes was in detergents. Cellulases have been a part of detergents as color brightening and softening agents since the early 1990s. It is found that incorporating bioactive enzymes into detergent formulations are advantageous for enhancing the washing performance of detergents at low temperature (30 ºC) and for reducing water consumption. Thus, to remove grass stains arrayed on cotton fabrics in this project, xylan in grass stains was selected as the target ligand. The use of novel xylan binding CBM-containing xylanases and the interaction between proteins, surfactants and protease were studied intensively.

Protease is one type of important enzyme commonly contained in washing powders, since it helps to break down protein based stains, such as food stains and blood stains, thus it has a broad activity. Xylanases10A and xylanase11A show good stability against proteases and work well at low temperature, whereas CBM2b1,2 is degraded by protease to some extent, which could reduce the effectiveness of this module in the detergent setting. The data showed that surfactant SLES did not limit the protease activity against CBM2b1,2. Thus we can not use it in the applications at the moment. Tergo detergent assays showed that CBM2b1,2-Xylanase10A, derived from *C. japonicus,* gave the best washing performance of the enzymes assessed against grass stains in the presence of High SLES surfactant at 30 ºC. Taking in to account the savinase tolerance and the protein-surfactant interaction, CBM2b1,2-xylanase10A,

together with other laundry enzymes (amylase, lipase, protease), might provide a promising detergent additive in future.

Some studies showed that the anionic surfactant, including SLES, may induce protein unfolding. For example, a cellulase derived from *F. succinogenes* S85 may be damaged by detergents, based on the stability of enzymes in surfactants and commercial detergents, probably by proteolytic degradation and anionic surfactant-induced unfolding (Otzen, 2002; Otzen, et al., 1999). Interactions between proteins and surfactant monomers are shown to play a key role in determining the kinetics of the unfolding process (Stoner, et al., 2005; 2006). Further studies have reported that protease was highly active against unfolded or partially unfolded proteins (Markert, et al., 2001; Stoner, et al., 2006). This could be one of the possible reasons why CBM2b1,2 was degraded more by savinase in high SLES (16 %) buffer than in buffer containing nonionic or cationic surfactants. To solve this issue, further research using an alternative surfactant system mainly containing nonionic or cationic surfactant could be explored, and the interactions between CBM2b1,2-xylanase 10A, savinase and the surfactant buffer could also be studied. Furthermore, as CBM2b1,2 displays weak cellulose binding affinity, and thus binds to non-grass stain regions, CBM2b1,2 engineering to abolish its cellulose binding affinity could be carried out.

The pigment in grass sap may have a staining effect, thus a bleaching agent may be required to remove the green color (Markert, et al., 2001) and other types of tough soil stains, for example, mud, that adhered to the fibers (Smulders, 2002). Some studies showed that the enzymes, for example, cellulases, had no ability to thoroughly remove the grass sap stains. A bleaching agent is probably required to improve the washing performance. Bleaching usually involves either oxidative or reductive reactions that decompose stains and soils. These processes may involve the removal or modification of larger molecules containing color-bearing groups, into smaller, more soluble units which are more easily removed in the washing process. Therefore, the detergent class, enzymes properties, and mixing ratio of surfactants have to be considered when

combining enzymes and detergents for achieving a better cleaning performance in laundry formulation development.

6.2 Xylan degrading system in C. japonicus: xylanases and how CBMs help these enzymes to target and break down grass cell walls

The microbial degradation of the plant cell wall is an important biological and industrial process. *C. japonicus* has evolved a complex xylan degrading system, which enables it to utilise the enzymatic hydrolysis to break down the composite structure of plant cell walls. As the cell walls comprise the most abundant source of organic carbon in the biosphere, the photosynthetically fixed carbon could therefore be released as an energy source.

The majority of glycoside hydrolases comprise, in addition to catalytic domain, one or more ancillary non-catalytic domains, which frequently bind xylan, cellulose, mannan and β-glucans. The discovery of the CBM60 family completes the modular assignment of the enzymes that comprise the complex xylan degrading system of *C. japonicus*. A subset of three enzymes, which are xylanase *Cj*Xyn10B, esterase *Cj*CE1, and the xylan-specific arabinofuranosidase, *Cj*Abf62A, contains a cellulose-specific CBM2a and a uronic acid-specific CBM35. CBM35 targets uronic acids (Δ4,5-GalA) in cell wall regions that are attacked by pectate lyases (Montanier, et al., 2009b). Since Δ4,5-GalA, as the anhydrosugar located at the nonreducing end of the oligosaccharides generated by pectate lyase action (Moran, et al., 1968), is a signature molecule for pectin degradation, thus by targeting Δ4,5-GalA rich regions, CBM35 can direct Xyn10B to accessible regions being actively degraded to a further degradation stage. Moreover, it is interesting to note that the gene encoding the CBM35-xylanase is a component of a *C. japonicus* operon that also synthesizes arabinofuranosidase and a CE1 acetyl esterase (Kellett, et. al., 1990) (Ferreira, et. al., 1993), which both contain CBM35 targeting Δ4,5-GalA. Hence, the regions that are targeted for further degradation were not restricted to a single xylanase, but other enzymes that removed other types of side chains on xylans.

C. japonicus also synthesizes a GH10 (*Cj*Xyn10A) and a GH5 (*Cj*Xyn5A) xylanase which contain only cellulose-specific CBMs (CBM2a and CBM10 in the GH10; only CBM2a in the GH5), whereas the two GH11 xylanases both contain a CBM60 (*Cj*Xyn11A also contains a cellulose-specific CBM10 and a CE4 xylan-specific esterase) (DeBoy, et al., 2008)). An interesting theory is that CBM2a and CBM10 anchors *Cj*Xyn5A and *Cj*Xyn10A to target xylans that are in intimate contact with cellulose microfibrils. Thus, cellulose appears to be a primary receptor for the xylan-degrading apparatus of *C. japonicus*, first proposed by Kellett and colleagues (Kellett, et al., 1990). Furthermore, CBM2as are able to slide over cellulose surfaces, which would increase substate access (Jervis, et al., 1997). Similarly, the two CBM60 modules lock the two GH11 xylanases, *Cj*Xyn11A and *Cj*Xyn11B onto xylans with their extensive acetyl decorations, thus are not in intimate contact with cellulose or other polysaccharides. The removal of the acetate groups by the xylan esterase in *Cj*Xyn11A would enable the two CBM60-containing enzymes to subsequently hydrolyze the exposed xylan backbone. In conclusion, all these enzymes work synergistically, which enables xylanases to locate to xylan-rich regions for more efficient hydrolysis. Hence, for laundry formulation development, utility of multiple types of these CBM containing enzymes to target grass stains theoretically should be more efficient in plant cell wall deconstruction.

6.3 Structural basis for CBM2b1,2 cellulose binding affinity

In this project, CBM2b1,2 and to a less extent CBM15 appear to bind to grass stains arrayed on cotton, but they also may bind to cotton. Surfactant SLES increases the binding of both CBMs to grass stains, however, the non specific binding of CBMs onto cotton was not stopped by the addition of SLES. The non specific binding likely reduced the targeting efficiency of CBM2b1,2 to grass stains arrayed on cotton. It is known that CBM2b1,2 displays strong xylan binding affinity, but also a weak cellulose binding affinity. Thus, it would be ideal if one could convert CBM2b1,2 to a pure xylan binding module with no cellulose binding property.

CBMs generally bind to planer crystalline cellulose because their binding sites are flat. However, some CBMs that bind to soluble forms of cellulose and likely amorphous cellulose are able to bind helical structures as this is the conformation adopted by disordered cellulose (Bolam, et al., 2001). Hence, it is possible that the CBM is binding to other regions of cotton that are disordered. In CBM2b1 from *C. fimi* Xyn11A, the two tryptophans, W259 and W291, are approximately perpendicular to each other and are separated by 12 Å. These two amino acids are separated by a ß-turn, and are thus next to one another in the 3D structure, on the same face of adjacent ß-strands (Fig 6.1). They are ideal for binding xylan via a stacking interaction to the pyranose rings of xylose residues n and n+2 of the polysaccharide that has an approximately 120° rotation between one monomer and the next, but may also make interactions with the disordered cellulose of a slight helical structures. The sort of disordered helical cellulose although may not be not very common, but probably enough of it caused the cellulose binding to CBMs in a weak way, which lead to the CBMs non specific binding. Hence, one would hope to engineering CBM2b1,2 to create a mutant with abolished cellulose binding property but still retained strong xylan binding affinity.

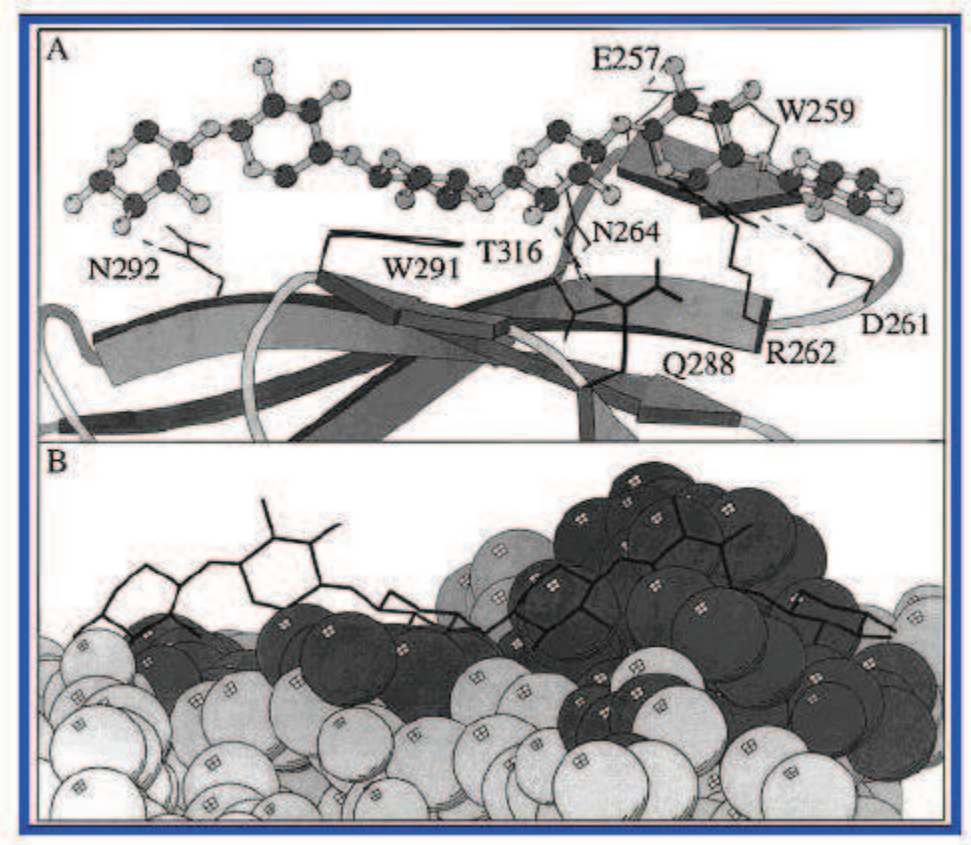


Fig 6.1 Main interactions expected between CBM2b1 and xylan. (A) Side chains implicated in the recognition of xylooligosaccharides. (B) A space-filling representation of the complex, showing the exposed surface available to much of the xylan chain (Simpson, et al., 2000).

6.4 Future work: CBM2b1,2 engineering by phage display

It has been proposed that the xylan binding cleft in CBM2b1,2 displays a weak affinity for cellulose (Bolam, et.al., 2001). To test this hypothesis, one could knockout the two aromatic residues (W259 and W291) in the cleft that bind to xylan, and assess whether cellulose binding in addition to xylan is lost. If CBM2b1,2 lost its cellulose binding afterwards, it indicates the binding cleft is actually responsible for the cellulose binding. One then could engineer the cleft by modifying the side chains. As the difference between glucose and xylose is only the presence of a CH_2OH group attached to C5, there might be a few regions of CBM2b1,2 binding site that present a planar surface that may interact with cellulose. One would hope to create a binding cleft with residue side chains sticking out of the cleft to abolish the cellulose binding, therefore some polar residues in the binding sites, for example, T316, T316 T266, T312 and

N264, could be selected and mutate to bigger side chains, such as to arginine, glutamine and lysine.

Another possibility after knocking out the two tryptophans would be the cellulose binding of CBM2b1,2 is still present. If that is the case, it indicates that the weak binding to cellulose is not happening in the binding cleft but somewhere else, i.e. the cleft is not responsible to the cellulose binding. There could be some hydrophobic patches on the protein that may bind to cellulose, thus carrying out a wider range of mutagenesis would be needed. One approach to do this is by using phage display technology as a random mutagenesis strategy to create a library of CBM2b1,2 mutants which may bind to xylans with no cellulose binding capacity. M.Ohlin and colleagues investigated the use of CBM4-2, from *Rhodothermus marinus* xylanase, as a suitable diversity-carrying scaffold and thus used phage display (Lehtio, et al., 2000) to change the specificity of the CBM. Variations were introduced at 12 positions in the carbohydrate binding site of CBM4-2. The aromatic and hydrophilic charged residues were preserved, as they are established mediators of carbohydrate-protein interactions (Boraston, et al., 1999). A combinatorial library (1.6×10^6 colonies) was created, including mutants derived from the primer design, and through mutations introduced by the polymerase reaction. Phages displaying CBM4-2 variants were selected on birchwood xylan, Avicel and ivory nut mannan, following several rounds of enrichment for the target-specific binders. This approach could be applied to CBM2b1,2 to alter its binding specificity such that it only recognised to xylan. This method will required CBM2b1,2 as a suitable diversity-carrying scaffold to be displayed on phage, which allows the subsequent selection on different ligands. Ohlin showed that CBM in comparison to antibodies, as a diversification molecule, is quite easily produced in *E. coli* and purified by IMAC, with a generally high yield in large-scale application.

Also, expressing CBM2b1,2 in another organism, such as in yeast *Streptomyces thermoviolaceus,* may express CBM2b1,2 in a glycosylated way, thus to prevent CBM2b1,2 modules from protease attack during laundry process. Artificial linker could also be chemically synthesized to build stable linkers between CBM2b1,2 and

*Cj*Xyn10A, thus to increase the specificity and stability of the protein module for xylan targeting during grass stains removal. This research could be carried out in future to potentiate the synergy displayed by the biocatalysts, savinase and surfactants in the formulation for the development of laundry products.

References

Abbott, D.W. and Boraston, A.B., (2007). 'A family 2 pectate lyase displays a rare fold and transition metal-assisted beta-elimination'. *J Biol Chem*, 282, 35328-35336.

Abbott, D.W., Ficko-Blean, E., van Bueren, A.L., Rogowski, A., Cartmell, A., Coutinho, P.M., Henrissat, B., Gilbert, H.J. and Boraston, A.B., (2009). 'Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules'. *Biochemistry*, 48, 10395-10404.

Abbott, D.W., Hrynuik, S. and Boraston, A.B., (2007). 'Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*'. *J Mol Biol*, 367, 1023-1033.

Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxa, E., Raghothama, S., Simpson, P.J., Gilbert, H.J., Williamson, M.P. and Holst, O., (2000). 'Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies'. *Biochem J*, 345, 53-60.

Akita, M., Suzuki, A., Kobayashi, T., Ito, S. and Yamane, T., (2001). 'The first structure of pectate lyase belonging to polysaccharide lyase family 3'. *Acta Crystallogr D Biol Crystallogr*, 57, 1786-1792.

Andersson, A., Persson, T., Zacchi, G., Stalbrand, H. and Jonsson, A., (2007). 'Comparison of diafiltration and size-exclusion chromatography to recover hemicellulose from process water from thermomechanical pulping of spruce'. *Appplied Biochemistry and Biotechnology*:136-140.

Antoni, D., Zverlov, V.V. and Schwarz, W.H., (2007). 'Biofuels from microbes'. *Appl Microbiol Biotechnol*, 77, 23-35.

Armand, S., Drouillard, S., Schulein, M., Henrissat, B. and Driguez, H., (1997). 'A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases'. *J Biol Chem*, 272, 2709-2713.

Aspinall, G.O., (1980). 'Chemistry of cell wall polysacharides'. In: Preiss, J. (ed). *The Biochemistry of Plants (A Comprehensive Treaty)*. New York: Academic Press, 475-500.

Atkins, E.D.T., (1992). 'Three-dimensional strucutre, interactions and properties of xylans.' *Progress in Biotechnology*, 39-50.

Bacic, A., Harris, P.J. and Stone, B.A., (1988). 'Structure and function of plant cell walls'. In: Preiss, J. (ed). *The Biochemistry of Plants*. New York: Academic Press, 297-371.

Bajpai, D. and Tyagi, V.K., (2007). 'Laundry detergents: an overview'. *J Oleo Sci*, 56, 327-340.

Banka, R.R., Mishra, S. and Ghose, T.K., (1998). 'Fibril formation from cellulose by a novel protein from *Trichoderma reesei*: a non-hydrolytic cellulolytic component?' *World J. Microb. Biotechnol.*, 4:551-558.

Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G., (1975). 'Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data'. *Nature*, 255, 609-614.

Barral, P., Suarez, C., Batanero, E., Alfonso, C., Alche Jde, D., Rodriguez-Garcia, M.I., Villalba, M., Rivas, G. and Rodriguez, R., (2005). 'An olive pollen protein with allergenic activity, Ole e 10, defines a novel family of carbohydrate-binding modules and is potentially implicated in pollen germination'. *Biochem J*, 390, 77-84.

Bayer, E.A., Morag, E. and Lamed, R., (1994). 'The cellulosome—a treasuretrove for biotechnology.' *Trends Biotechnol*, 12:379-386.

Bayer, E.A., Shimon, L.J., Shoham, Y. and Lamed, R., (1998). 'Cellulosomes-structure and ultrastructure'. *J Struct Biol*, 124, 221-234.

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

Berdichevsky, Y., Ben-Zeev, E., Lamed, R. and Benhar, I., (1999). 'Phage display of a cellulose binding domain from *Clostridium thermocellum* and its application as a tool for antibody engineering'. *J Immunol Methods*, 228, 151-162.

Biely, P., Kratky, Z. and Vrsanska, M., (1981). 'Substrate-binding site of endo-1,4-beta-xylanase of the yeast *Cryptococcus albidus*'. *Eur J Biochem*, 119, 559-564.

Blake, A.W., McCartney, L., Flint, J.E., Bolam, D.N., Boraston, A.B., Gilbert, H.J. and Knox, J.P., (2006). 'Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes'. *J Biol Chem*, 281, 29321-29329.

Boisset, C., Fraschini, C., Schulein, M., Henrissat, B. and Chanzy, H., (2000). 'Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the

cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A'. *Appl Environ Microbiol*, 66, 1444-1452.

Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J.E., Boraston, A., Hazlewood, G.P. and Gilbert, H.J., (1998). '*Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity'. *Biochem J*, 331, 775-781.

Bolam, D.N., Xie, H., White, P., Simpson, P.J., Hancock, S.M., Williamson, M.P. and Gilbert, H.J., (2001). 'Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* xylanase 11A'. *Biochemistry*, 40, 2468-2477.

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

Boraston, A.B., Creagh, A.L., Alam, M.M., Kormos, J.M., Tomme, P., Haynes, C.A., Warren, R.A. and Kilburn, D.G., (2001a). 'Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A'. *Biochemistry*, 6240-6247.

Boraston, A.B., McLean, B.W., Chen, G., Li, A., Warren, R.A. and Kilburn, D.G., (2002a). 'Co-operative binding of triplicate carbohydrate-binding modules from a thermophilic xylanase'. *Mol Microbiol*, 43, 187-194.

Boraston, A.B., McLean, B.W., Guarna, M.M., Amandaron-Akow, E. and Kilburn, D.G., (2001b). 'A family 2a carbohydrate-binding module suitable as an affinity tag for proteins produced in *Pichia pastoris*'. *Protein Expr Purif*, 21, 417-423.

Boraston, A.B., McLean, B.W., Kormos, J., Alam, M.M., Gilkes, N.R., Haynes, C.A., Tomme, P., Kilburn, D.G. and Warren, R.A., (1999). In: Gilbert, H.J., Davies, G.J., Henrissat, B. and Svensson, B. (eds). *Recent Advances in Carbohydrate Bioengineering*. Cambridge: The Royal Society of Chemistry, 202-211.

Boraston, A.B., Notenboom, V., Warren, R.A., Kilburn, D.G., Rose, D.R. and Davies, G., (2003a). 'Structure and ligand binding of carbohydrate-binding module *Cs*CBM6-3 reveals similarities with fucose-specific lectins and "galactose-binding" domains'. *J Mol Biol*, 327, 659-669.

Boraston, A.B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G. and Davies, G.J., (2002b). 'Differential oligosaccharide recognition by evolutionarily-related beta-1,4 and beta-1,3 glucan-binding modules'. *J Mol Biol*, 319, 1143-1156.

Boraston, A.B., Revett, T.J., Boraston, C.M., Nurizzo, D. and Davies, G.J., (2003b). 'Structural and thermodynamic dissection of specific mannan recognition by a carbohydrate binding module, *Tm*CBM27'. *Structure*, 11, 665-675.

Boraston, A.B., Wang, D. and Burke, R.D., (2006). 'Blood group antigen recognition by a *Streptococcus pneumoniae* virulence factor'. *J Biol Chem*, 281, 35263-35271.

Brett, C.T. and Waldren, K., (1996). *Physiology and Biochemistry of Plant Cell Walls.*: Black, M.Charlewood, B.Brett, C.T. and Waldron, K., (1990). *Physiology and biochemistry of plant cell walls*. London: Unwin Hyman.

Brett, C.T. and Waldron, K., (1996). *Physiology and biochemistry of plant cell walls*. London: Chapman, J. and Hall, J.Brown, I.E., Mallen, M.H., Charnock, S.J., Davies, G.J. and Black, G.W., (2001). 'Pectate lyase 10A from *Pseudomonas cellulosa* is a modular enzyme containing a family 2a carbohydrate-binding module'. *Biochem J*, 355, 155-165.

Brown, R.M., (1999). ' Cellulose structure and biosynthesis.' *Pure and Apllied Chemistry*, 71:204 - 212.

Carpita, N.C., (1997). 'Structure and biosynthesis of plant cell walls.' In: Dennis, D.T., Turpin, D.H., Lefebure, D.D. and Layzell, D.B. (eds). *Plant Metabolism*. Harlow: Wesley Longman, 125-147.

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

Carrard, G., Koivula, A., Soderlund, H. and Beguin, P., (2000). 'Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose'. *Proc Natl Acad Sci U S A*, 97, 10342-10347.

Cartmell, A., Topakas, E., Ducros, V.M., Suits, M.D., Davies, G.J. and Gilbert, H.J., (2008). 'The *Cellvibrio japonicus* mannanase *Cj*Man26C displays a unique exo-mode of action that is conferred by subtle changes to the distal region of the active site'. *J Biol Chem*, 283, 34403-34413.

Charnock, S.J., Bolam, D.N., Nurizzo, D., Szabo, L., McKie, V.A., Gilbert, H.J. and Davies, G.J., (2002a). 'Promiscuity in ligand-binding: The three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose'. *Proc Natl Acad Sci U S A*, 99, 14077-14082.

Charnock, S.J., Bolam, D.N., Turkenburg, J.P., Gilbert, H.J., Ferreira, L.M., Davies, G.J. and Fontes, C.M., (2000). 'The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain'. *Biochemistry*, 39, 5013-5021.

Charnock, S.J., Brown, I.E., Turkenburg, J.P., Black, G.W. and Davies, G.J., (2002b). 'Convergent evolution sheds light on the anti-beta -elimination mechanism common to family 1 and 10 polysaccharide lyases'. *Proc Natl Acad Sci U S A*, 99, 12067-12072.

Charnock, S.J., Lakey, J.H., Virden, R., Hughes, N., Sinnott, M.L., Hazlewood, G.P., Pickersgill, R. and Gilbert, H.J., (1997). 'Key residues in subsite F play a critical role in the activity of *Pseudomonas fluorescens subspecies cellulosa* xylanase A against xylooligosaccharides but not against highly polymeric substrates such as xylan'. *J Biol Chem*, 272, 2942-2951.

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

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

Correia, M.A., Abbott, D.W., Gloster, T.M., Fernandes, V.O., Prates, J.A., Montanier, C., Dumon, C., Williamson, M.P., Tunnicliffe, R.B., Liu, Z., Flint, J.E., Davies, G.J., Henrissat, B., Coutinho, P.M., Fontes, C.M. and Gilbert, H.J., (2010). 'Signature active site architectures illuminate the molecular basis for ligand specificity in family 35 carbohydrate binding module'. *Biochemistry*, 49, 6193-6205.

Cosgrove, D.J., (2000). 'Loosening of plant cell walls by expansins'. *Nature*, 407, 321-326.

Cosgrove, D.J., (2005). 'Growth of the plant cell wall'. *Nat Rev Mol Cell Biol*, 6, 850-861.

Coutinho, P.M. and Henrissat, B., (1999). 'Carbohydrate-active enzymes: an integrated database approach'. In: Gilbert, H.J., Davies, G., Henrissat, B. and Svensson, B. (eds). *Recent Advances in Carbohydrate Bioengineering*. Cambridge The Royal Society of Chemistry, 3-12.

Czjzek, M., Bolam, D.N., Mosbah, A., Allouch, J., Fontes, C.M., Ferreira, L.M., Bornet, O., Zamboni, V., Darbon, H., Smith, N.L., Black, G.W., Henrissat, B. and Gilbert, H.J., (2001). 'The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved'. *J Biol Chem*, 276, 48580-48587.

Darvill, A.G., Albersheim, P., McNeil, M., Lau, J.M., York, W.S., Stevenson, T.T., Thomas, J., Doares, S., Gollin, D.J., Chelf, P. and et al., (1985). 'Structure and function of plant cell wall polysaccharides'. *J Cell Sci Suppl*, 203-217.

Darvill, J.E., McNeil, M., Darvill, A.G. and Albersheim, P., (1980). 'Structure of Plant Cell Walls: Glucoronoarabinoxylan, a second hemicellulose in the primary cell walls of suspension-cultured sycamore cells'. *Plant Physiol*, 66:1135-1139.

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

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

de Vries, R.P. and Visser, J., (2001). '*Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides'. *Microbiol Mol Biol Rev*, 65, 497-522, table of contents.

DeBoy, R.T., Mongodin, E.F., Fouts, D.E., Tailford, L.E., Khouri, H., Emerson, J.B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H.J. and Nelson, K.E., (2008). 'Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*'. *J Bacteriol*, 190, 5455-5463.

Dhawan, S. and Kaur, J., (2007). 'Microbial mannanases: an overview of production and applications'. *Crit Rev Biotechnol*, 27, 197-216.

Din, N., Damude, H.G., Gilkes, N.R., Miller, R.C., Jr., Warren, R.A. and Kilburn, D.G., (1994). 'C1-Cx revisited: intramolecular synergism in a cellulase'. *Proc Natl Acad Sci U S A*, 91, 11383-11387.

Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K., Teeri, T.T. and Jones, T.A., (1994). 'The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*'. *Science*, 265, 524-528.

Divne, C., Stahlberg, J., Teeri, T.T. and Jones, T.A., (1998). 'High-resolution crystal structures reveal how a cellulose chain is bound in the 50 A long tunnel of cellobiohydrolase I from *Trichoderma reesei*'. *J Mol Biol*, 275, 309-325.

Dumon, C., Vardakou, M., Murray, J.W., Christakopoulos, P., Weiner, D.P., Juge, N., Lewis, R.J., Gilbert, H.J. and Flint, J.E., (2008). 'Understanding the structural basis for substrate and inhibitor recognition in eukaryotic GH11 xylanases'. *J Mol Biol*, 375, 1293-1305.

Emsley, P. and Cowtan, K., (2004). 'Coot: model-building tools for molecular graphics'. *Acta Crystallogr D Biol Crystallogr*, 60, 2126-2132.

Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M. and Winter, G., (1985). 'Hydrogen bonding and biological specificity analysed by protein engineering'. *Nature*, 314, 235-238.

Finnegan, P.M., Brumbley, S.M., O'Shea, M.G., Nevalainen, H. and Bergquist, P.L., (2005). 'Diverse dextranase genes from *Paenibacillus species*'. *Arch Microbiol*, 183, 140-147.

Freelove, A.C., Bolam, D.N., White, P., Hazlewood, G.P. and Gilbert, H.J., (2001). 'A novel carbohydrate-binding protein is a component of the plant cell wall-degrading complex of *Piromyces equi*'. *J Biol Chem*, 276, 43010-43017.

Fry, S.C., (1979). 'Phenolic components of the primary cell wall and their possible role in the hormonal regulation of growth'. *Planta*, 146:343-351.

Fry, S.C., (1982). 'Isodityrosine, a new cross-linking amino acid from plant cell-wall glycoprotein'. *Biochem J*, 204, 449-455.

Fujimoto, Z., Kaneko, S., Kuno, A., Kobayashi, H., Kusakabe, I. and Mizuno, H., (2004). 'Crystal structures of decorated xylooligosaccharides bound to a family 10 xylanase from *Streptomyces olivaceoviridis* E-86'. *J Biol Chem*, 279, 9606-9614.

Gao, P.J., Chen, G.J., Wang, T.H., Zhang, Y.S. and Liu, J., (2001). 'Non-hydrolytic disruption of crystalline structure of cellulose by cellulose binding domain and linker sequence of cellobiohydrolase I from *Penicillium janthinellum*.' *Shengwu Huaxue Yu Shengwu Wuli Xuebao*, 33:13-18.

Ghangas, G.S., Hu, Y.J. and Wilson, D.B., (1989). 'Cloning of a *Thermomonospora fusca* xylanase gene and its expression in *Escherichia coli* and *Streptomyces lividans*'. *J Bacteriol*, 171, 2963-2969.

Giardina, T., Gunning, A.P., Juge, N., Faulds, C.B., Furniss, C.S., Svensson, B., Morris, V.J. and Williamson, G., (2001). 'Both binding sites of the starch-binding domain of *Aspergillus niger* glucoamylase are essential for inducing a conformational change in amylose'. *J Mol Biol*, 313, 1149-1159.

Gibbs, M.D., Saul, D.J., Luthi, E. and Bergquist, P.L., (1992). 'The beta-mannanase from *Caldocellum saccharolyticum* is part of a multidomain enzyme'. *Appl Environ Microbiol*, 58, 3864-3867.

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

Gilbert, H.J. and Hazlewood, G.P., (1993). 'Bacterial cellulases and xylanases'. *J Gen Microbiol*, 139:187-194.

Gilbert, H.J., Hazlewood, G.P., Laurie, J.I., Orpin, C.G. and Xue, G.P., (1992). 'Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin'. *Mol Microbiol*, 6, 2065-2072.

Gilkes, N.R., Kilburn, D.G., Miller, R.C., Warren, R.A., Sugiyama, J., Chanzy, H. and Henrissat, B., (1993). 'Visualization of the adsorption of a bacterial endo-beta-1,4-glucanase and its isolated cellulose-binding domain to crystalline cellulose.' *Int J Biol Macromol*, 15:347-351.

Gloster, T.M., Ibatullin, F.M., Macauley, K., Eklof, J.M., Roberts, S., Turkenburg, J.P., Bjornvad, M.E., Jorgensen, P.L., Danielsen, S., Johansen, K.S., Borchert, T.V., Wilson, K.S., Brumer, H. and Davies, G.J., (2007). 'Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12'. *J Biol Chem*, 282, 19177-19189.

Grabber, J.H., Hatfield, R.D. and J., R., (1998). 'Diferulate cross-links impede the enzymatic degradation of non-lignified maize walls'. *Journal of the Science of Food and Agriculture*, 77, 193-200.

Greenwood, J.M., Ong, E., Gilkes, N.R., Warren, R.A., Miller, R.C., Jr. and Kilburn, D.G., (1992). 'Cellulose-binding domains: potential for purification of complex proteins'. *Protein Eng*, 5, 361-365.

Hahn-Hagerdal, B., Galbe, M., Gorwa-Grauslund, M.F., Liden, G. and Zacchi, G., (2006). 'Bio-ethanol--the fuel of tomorrow from the residues of today'. *Trends Biotechnol*, 24, 549-556.

Hall, J., Black, G.W., Ferreira, L.M., Millward-Sadler, S.J., Ali, B.R., Hazlewood, G.P. and Gilbert, H.J., (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, 749-756.

Harris, G.W., Jenkins, J.A., Connerton, I., Cummings, N., Lo Leggio, L., Scott, M., Hazlewood, G.P., Laurie, J.I., Gilbert, H.J. and Pickersgill, R.W., (1994). 'Structure of the catalytic core of the family F xylanase from *Pseudomonas fluorescens* and identification of the xylopentaose-binding sites'. *Structure*, 2, 1107-1116.

Harris, G.W., Jenkins, J.A., Connerton, I. and Pickersgill, R.W., (1996). 'Refined crystal structure of the catalytic domain of xylanase A from *Pseudomonas fluorescens* at 1.8 A resolution'. *Acta Crystallogr D Biol Crystallogr*, 52, 393-401.

Hatada, Y., Hidaka, Y., Nogi, Y., Uchimura, K., Katayama, K., Li, Z., Akita, M., Ohta, Y., Goda, S., Ito, H., Matsui, H., Ito, S. and Horikoshi, K., (2004). 'Hyper-production of an isomalto-dextranase of an *Arthrobacter sp.* by a proteases-deficient *Bacillus subtilis*: sequencing, properties, and crystallization of the recombinant enzyme'. *Appl Microbiol Biotechnol*, 65, 583-592.

Hayashi, T., Ogawa, K. and Mitsuishi, Y., (1994). 'Characterization of the adsorption of Xyloglucan to Cellulose'. *Plant Cell Physiol.*, 35, 1199-1205.

Hazlewood, G.P. and Gilbert, H.J., (1998). 'Structure and function analysis of *Pseudomonas* plant cell wall hydrolases'. *Prog Nucleic Acid Res Mol Biol*, 61:211-241.
Henrissat, B., (1991). 'A classification of glycosyl hydrolases based on amino acid sequence similarities'. *Biochem J*, 280, 309-316.

Henrissat, B. and Bairoch, A., (1993). 'New families in the classification of glycosyl hydrolases based on amino acid sequence similarities'. *Biochem J*, 293, 781-788.

Henrissat, B., Heffron, S.E., Yoder, M.D., Lietzke, S.E. and Jurnak, F., (1995). 'Functional implications of structure-based sequence alignment of proteins in the extracellular pectate lyase superfamily'. *Plant Physiol*, 107, 963-976.

Henrissat, B. and Romeu, A., (1995). 'Families, superfamilies and subfamilies of glycosyl hydrolases'. *Biochem J*, 311, 350-351.

Henrissat, B. and Bairoch, A., (1996). 'Updating the sequence-based classification of glycosyl hydrolases'. *Biochem J*, 316, 695-696.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P. and Davies, G., (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, 7090-7094.

Henrissat, B. and Davies, G., (1997). 'Structural and sequence-based classification of glycoside hydrolases'. *Current Opinion in Structural Biology*, 7, 637-644.

Henrissat, B., Teeri, T.T. and Warren, R.A., (1998). 'A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants.' *FEBS Lett*, 425, 352-354.

Henshaw, J., Horne-Bitschy, A., van Bueren, A.L., Money, V.A., Bolam, D.N., Czjzek, M., Ekborg, N.A., Weiner, R.M., Hutcheson, S.W., Davies, G.J., Boraston, A.B. and Gilbert, H.J., (2006). 'Family 6 carbohydrate binding modules in beta-agarases display exquisite selectivity for the non-reducing termini of agarose chains'. *J Biol Chem*, 281, 17099-17107.

Henshaw, J.L., Bolam, D.N., Pires, V.M., Czjzek, M., Henrissat, B., Ferreira, L.M., Fontes, C.M. and Gilbert, H.J., (2004). 'The family 6 carbohydrate binding module *Cm*CBM6-2 contains two ligand-binding sites with distinct specificities'. *J Biol Chem*, 279, 21552-21559.

Herron, S.R., Scavetta, R.D., Garrett, M., Legner, M. and Jurnak, F., (2003). 'Characterization and implications of Ca^{2+} binding to pectate lyase C'. *J Biol Chem*, 278, 12271-12277.

Herve, C., Rogowski, A., Gilbert, H.J. and Paul Knox, J., (2009). 'Enzymatic treatments reveal differential capacities for xylan recognition and degradation in primary and secondary plant cell walls'. *Plant J*, 58, 413-422.

Herve, C., Rogowski, A., Blake, A.W., Marcus, S.E., Gilbert, H.J. and Knox, J.P., (2010). 'Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects'. *Proc Natl Acad Sci U S A*, 107, 15293-15298.

Ishii, T., (1997). 'Structure and functions of feruloylated polysaccharides.' *Plant Sci*, 127:111-127.

Ishii, T. and Matsunaga, T., (2001). 'Pectic polysaccharide rhamnogalacturonan II is covalently linked to homogalacturonan'. *Phytochemistry*, 57, 969-974.

Jamal-Talabani, S., Boraston, A.B., Turkenburg, J.P., Tarbouriech, N., Ducros, V.M. and Davies, G.J., (2004). 'Ab initio structure determination and functional characterization of CBM36; a new family of calcium-dependent carbohydrate binding modules'. *Structure*, 12, 1177-1187.

Jamal, S., Nurizzo, D., Boraston, A.B. and Davies, G.J., (2004). 'X-ray crystal structure of a non-crystalline cellulose-specific carbohydrate-binding module: CBM28'. *J Mol Biol*, 339, 253-258.

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

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

Kellett, L.E., Poole, D.M., Ferreira, L.M., Durrant, A.J., Hazlewood, G.P. and Gilbert, H.J., (1990). 'Xylanase B and an arabinofuranosidase from *Pseudomonas fluorescens subsp. cellulosa* contain identical cellulose-binding domains and are encoded by adjacent genes'. *Biochem J*, 272, 369-376.

Kitago, Y., Karita, S., Watanabe, N., Kamiya, M., Aizawa, T., Sakka, K. and Tanaka, I., (2007). 'Crystal structure of Cel44A, a glycoside hydrolase family 44 endoglucanase from *Clostridium thermocellum*'. *J Biol Chem*, 282, 35703-35711.

Koshland, D.E., Jr. and Clarke, E., (1953). 'Mechanism of hydrolysis of adenosinetriphosphate catalyzed by lobster muscle'. *J Biol Chem*, 205, 917-924.

Kraulis, J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J. and Gronenborn, A.M., (1989). 'Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei.* A study

using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing'. *Biochemistry*, 28, 7241-7257.

Lam, T.B.T., Iiyama, K. and Stone, B.A., (1994). 'An approach to the estimation of ferulic acid bridges in unfractioned cell walls of wheat internodes.' *Phytochemistry*, 37:327-333.

Le Nours, J., Anderson, L., Stoll, D., Stalbrand, H. and Lo Leggio, L., (2005). 'The structure and characterization of a modular endo-beta-1,4-mannanase from *Cellulomonas fimi*'. *Biochemistry*, 44, 12700-12708.

Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M. and Teeri, T.T., (2003). 'The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules'. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 484-489.

Lerouge, P., O'Neill, M.A., Darvill, A.G. and Albersheim, P., (1993). 'Structural characterization of endo-glycanase-generated oligoglycosyl side chains of rhamnogalacturonan I'. *Carbohydr Res*, 243, 359-371.

Levy, I., Shani, Z. and Shoseyov, O., (2002). 'Modification of polysaccharides and plant cell wall by endo-1,4-beta-glucanase and cellulose-binding domains'. *Biomol Eng*, 19, 17-30.

Levy, I. and Shoseyov, O., (2002). 'Cellulose-binding domains: biotechnological applications'. *Biotechnol Adv*, 20, 191-213.

Liab, K., Azadi, P., Collins, R., Tolan, J., Kim, J.S. and Eriksson, K.L., (2000). 'Relationships between activities of xylanases and xylan structures'. *Enzyme Microb Technol*, 27, 89-94.

Lietzke, S.E., Yoder, M.D., Keen, N.T. and Jurnak, F., (1994). 'The Three-Dimensional Structure of Pectate Lyase E, a Plant Virulence Factor from *Erwinia chrysanthemi*'. *Plant Physiol*, 106, 849-862.

Lindqvist, Y. and Schneider, G., (1997). 'Circular permutations of natural protein sequences: structural evidence'. *Curr Opin Struct Biol*, 7, 422-427.

Lo Leggio, L., Kalogiannis, S., Bhat, M.K. and Pickersgill, R.W., (1999). 'High resolution structure and sequence of *T. aurantiacus* xylanase I: implications for the evolution of thermostability in family 10 xylanases and enzymes with (beta)alpha-barrel architecture'. *Proteins*, 36, 295-306.

Lymar, E.S., Li, B. and Renganathan, V., (1995). 'Purification and Characterization of a Cellulose-Binding (beta)-Glucosidase from Cellulose-Degrading Cultures of *Phanerochaete chrysosporium*'. *Appl Environ Microbiol*, 61, 2976-2980.

Margo, P., Varvaro, L., Chilosi, G., Avanzo, C. and Balestra, G.M., (1994). 'Pectinolytic enzymes produced by *Pseudomonas syringae* pv. *Glycinea*.' *FEMS Microbiol Lett*, 117:1-6.

Margolles-Clark, E., Tenkanen, M., Soderlund, H. and Penttila, M., (1996). 'Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue and a cellulose-binding domain'. *Eur J Biochem*, 237, 553-560.

Markert, Y., Koditz, J., Mansfeld, J., Arnold, U. and Ulbrich-Hofmann, R., (2001). 'Increased proteolytic resistance of ribonuclease A by protein engineering'. *Protein Eng*, 14, 791-796.

Matsunaga, T., Ishii, T., Matsumoto, S., Higuchi, M., Darvill, A., Albersheim, P. and O'Neill, M.A., (2004). 'Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes, and bryophytes. Implications for the evolution of vascular plants'. *Plant Physiol*, 134, 339-351.

McCann, M.C., Wells, B. and Roberts, K., (1990). 'Direct visualization of cross-links in the primary plant cell wall'. *J Cell Sci*, 96, 323-334.

McCarter, J.D. and Withers, S.G., (1994). 'Mechanisms of enzymatic glycoside hydrolysis'. *Curr Opin Struct Biol*, 4, 885-892.

McCartney, L., Blake, A.W., Flint, J., Bolam, D.N., Boraston, A.B., Gilbert, H.J. and Knox, J.P., (2006). 'Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules'. *Proc Natl Acad Sci U S A*, 103, 4765-4770.

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

McLean, B.W., Bray, M.R., Boraston, A.B., Gilkes, N.R., Haynes, C.A. and Kilburn, D.G., (2000). 'Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues'. *Protein Eng*, 13, 801-809.

McNeil, M., Darvill, A.G., Fry, S.C. and Albersheim, P., (1984). 'Structure and function of the primary cell walls of plants'. *Annu Rev Biochem*, 53:625-663.

Meinke, A., Damude, H.G., Tomme, P., Kwan, E., Kilburn, D.G., Miller, R.C., Jr., Warren, R.A. and Gilkes, N.R., (1995). 'Enhancement of the endo-beta-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop'. *J Biol Chem*, 270, 4383-4386.

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

Millward-Sadler, S.J., Davidson, K., Hazlewood, G.P., Black, G.W., Gilbert, H.J. and Clarke, J.H., (1995). 'Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria *Pseudomonas fluorescens subsp. cellulosa and Cellvibrio mixtus*'. *Biochem J*, 312, 39-48.

Minic, Z. and Jouanin, L., (2006). 'Plant glycoside hydrolases involved in cell wall polysaccharide degradation'. *Plant Physiol Biochem*, 44 , 435-449.

Miyanaga, A., Koseki, T., Matsuzawa, H., Wakagi, T., Shoun, H. and Fushinobu, S., (2004). 'Crystal structure of a family 54 alpha-L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose'. *J Biol Chem*, 279, 44907-44914.

Montanier, C., Flint, J.E., Bolam, D.N., Xie, H., Liu, Z., Rogowski, A., Weiner, D.P., Ratnaparkhe, S., Nurizzo, D., Roberts, S.M., Turkenburg, J.P., Davies, G.J. and Gilbert, H.J., (2010). 'Circular permutation provides an evolutionary link between two families of calcium-dependent carbohydrate binding modules'. *J Biol Chem*, 285, 31742-31754.

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

Montanier, C., van Bueren, A.L., Dumon, C., Flint, J.E., Correia, M.A., Prates, J.A., Firbank, S.J., Lewis, R.J., Grondin, G.G., Ghinet, M.G., Gloster, T.M., Herve, C., Knox, J.P., Talbot, B.G., Turkenburg, J.P., Kerovuo, J., Brzezinski, R., Fontes, C.M., Davies, G.J., Boraston, A.B. and Gilbert, H.J., (2009b). 'Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function'. *Proc Natl Acad Sci U S A*, 106, 3065-3070.

Moran, F., Nasuno, S. and Starr, M.P., (1968). 'Extracellular and intracellular polygllacturonic acid trans-eliminases of *Erwinia carotovora*'. *Arch Biochem Biophys*, 123, 298-306.

Muller, M., Hori, R., Itoh, T. and Sugiyama, J., (2002). 'X-ray microbeam and electron diffraction experiments on developing xylem cell walls'. *Biomacromolecules*, 3, 182-186.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J., (1997). 'Refinement of macromolecular structures by the maximum-likelihood method'. *Acta Crystallogr D Biol Crystallogr*, 53, 240-255.

Murzin, A.G., Lesk, A.M. and Chothia, C., (1992). 'beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors'. *J Mol Biol*, 223, 531-543.

Nagy, T., Simpson, P., Williamson, M.P., Hazlewood, G.P., Gilbert, H.J. and Orosz, L., (1998). 'All three surface tryptophans in Type IIa cellulose binding domains play a pivotal role in binding both soluble and insoluble ligands'. *FEBS Lett*, 429, 312-316.

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

Nishiyama, Y., Sugiyama, J., Chanzy, H. and Langan, P., (2003). 'Crystal structure and hydrogen bonding system in cellulose Ibeta from synchrotron X-ray and neutron fibre diffraction'. *J. Am. Chem. Soc*, 125:14300-14306.

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

Notenboom, V., Boraston, A.B., Kilburn, D.G. and Rose, D.R., (2001b). 'Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms'. *Biochemistry*, 40, 6248-6256.

Notenboom, V., Boraston, A.B., Williams, S.J., Kilburn, D.G. and Rose, D.R., (2002). 'High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding'. *Biochemistry*, 41, 4246-4254.

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

Nurizzo, D., Nagy, T., Gilbert, H.J. and Davies, G.J., (2002). 'The structural basis for catalysis and specificity of the *Pseudomonas cellulosa* alpha-glucuronidase, GlcA67A'. *Structure*, 10, 547-556.

O'Neill, M.A., Ishii, T., Albersheim, P. and Darvill, A.G., (2004). 'Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide'. *Annu Rev Plant Biol*, 55:109-139.

Ofir, K., Berdichevsky, Y., Benhar, I., Azriel-Rosenfeld, R., Lamed, R., Barak, Y., Bayer, E.A. and Morag, E., (2005). 'Versatile protein microarray based on carbohydrate-binding modules'. *Proteomics*, 5, 1806-1814.

Ogawa, K., Hayashi, T. and Okamura, K., (1990). 'Conformational analysis of xyloglucans'. *Int J Biol Macromol*, 12, 218-222.

Ohta, Y., Hatada, Y., Nogi, Y., Li, Z., Ito, S. and Horikoshi, K., (2004). 'Cloning, expression, and characterization of a glycoside hydrolase family 86 beta-agarase from a deep-sea Microbulbifer-like isolate'. *Appl Microbiol Biotechnol*, 66, 266-275.

Otzen, D.E., (2002). 'Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature'. *Biophys J*, 83, 2219-2230.

Otzen, D.E., Christiansen, L. and Schulein, M., (1999). 'A comparative study of the unfolding of the endoglucanase Cel45 from *Humicola insolens* in denaturant and surfactant'. *Protein Sci*, 8, 1878-1887.

Pauly, M., Albersheim, P., Darvill, A. and York, W.S., (1999a). 'Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants'. *Plant J*, 20, 629-639.

Pauly, M., Andersen, L.N., Kauppinen, S., Kofod, L.V., York, W.S., Albersheim, P. and Darvill, A., (1999b). 'A xyloglucan-specific endo-beta-1,4-glucanase from *Aspergillus aculeatus*: expression cloning in yeast, purification and characterization of the recombinant enzyme'. *Glycobiology*, 9, 93-100.

Pell, G., Szabo, L., Charnock, S.J., Xie, H., Gloster, T.M., Davies, G.J. and Gilbert, H.J., (2004a). 'Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases'. *J Biol Chem*, 279, 11777-11788.

Pell, G., Taylor, E.J., Gloster, T.M., Turkenburg, J.P., Fontes, C.M., Ferreira, L.M., Nagy, T., Clark, S.J., Davies, G.J. and Gilbert, H.J., (2004b). 'The mechanisms by which family 10 glycoside hydrolases bind decorated substrates'. *J Biol Chem*, 279, 9597-9605.

Pell, G., Williamson, M.P., Walters, C., Du, H., Gilbert, H.J. and Bolam, D.N., (2003). 'Importance of hydrophobic and polar residues in ligand binding in the family 15 carbohydrate-binding module from *Cellvibrio japonicus* Xyn10C'. *Biochemistry*, 42, 9316-9323.

Pickersgill, R., Jenkins, J., Harris, G., Nasser, W. and Robert-Baudouy, J., (1994). 'The structure of *Bacillus subtilis* pectate lyase in complex with calcium'. *Nat Struct Biol*, 1, 717-723.

Pires, V.M., Henshaw, J.L., Prates, J.A., Bolam, D.N., Ferreira, L.M., Fontes, C.M., Henrissat, B., Planas, A., Gilbert, H.J. and Czjzek, M., (2004). 'The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5a in

complex with oligosaccharides reveals two distinct binding sites with different ligand specificities'. *J Biol Chem*, 279, 21560-21568.

Ponyi, T., Szabo, L., Nagy, T., Orosz, L., Simpson, P.J., Williamson, M.P. and Gilbert, H.J., (2000). 'Trp22, Trp24, and Tyr8 play a pivotal role in the binding of the family 10 cellulose-binding module from *Pseudomonas* xylanase A to insoluble ligands'. *Biochemistry*, 39, 985-991.

Popper, Z.A. and Fry, S.C., (2003). 'Primary cell wall composition of Bryophytes and Charophytes'. *Ann Bot*, 91, 1-12.

Puls, J. and Schuseil, J., (1993). 'Chemistry of hemicelluloses: relationship between hemicelluose structure and enzymes required for hydrolysis. ' In: Coughlan, M.P. and Hazlewood, G.P. (eds). *Hemicellulase and Hemicellulases*. London: Portland Press, 1 - 27.

Ralet, M.C., Bonnin, E. and Thibault, J.F., (2001). 'Pectins.' In: Vandamme, E. (ed). *Biopolymers, 6*. Weinhein: Wiley-VCH Verlag, 345-380.

Richardson, J.S., (1981). 'The anatomy and taxonomy of protein structure.' *Adv. Protein Chem.*, 34:167-339.

Roberts, K. and Shirsat, A.H., (2006). 'Increased extensin levels in *Arabidopsis* affect inflorescence stem thickening and height'. *J Exp Bot*, 57, 537-545.

Rodriguez, B., Kavoosi, M., Koska, J., Creagh, A.L., Kilburn, D.G. and Haynes, C.A., (2004). 'Inexpensive and generic affinity purification of recombinant proteins using a family 2a CBM fusion tag'. *Biotechnol Prog*, 20, 1479-1489.

Rombouts, F.M. and Thibault, J.-F., (1986). 'Feruloylated pectic substances from sugar-beet pulp.' *Carbohydr Res*, 154:177-187.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K. and Jones, T.A., (1990). 'Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*'. *Science*, 249, 380-386.

Rumyantseva, N.I., (2005). 'Arabinogalactan proteins: involvement in plant growth and morphogenesis'. *Biochemistry (Mosc)*, 70, 1073-1085.

Ryabova, O., Vrsanska, M., Kaneko, S., van Zyl, W.H. and Biely, P., (2009). 'A novel family of hemicellulolytic alpha-glucuronidase'. *FEBS Lett*, 583, 1457-1462.

Ryttersgaard, C., Le Nours, J., Lo Leggio, L., Jorgensen, C.T., Christensen, L.L., Bjornvad, M. and Larsen, S., (2004). 'The structure of endo-beta-1,4-galactanase from *Bacillus licheniformis* in complex with two oligosaccharide products'. *J Mol Biol*, 341, 107-117.

Schols, H.A., Voragen, A.G. and Colquhoun, I.J., (1994). 'Isolation and characterization of rhamnogalacturonan oligomers, liberated during degradation of pectic hairy regions by rhamnogalacturonase'. *Carbohydr Res*, 256, 97-111.

Schols, H.A., Bakx, E.J., Schipper, D. and Voragen, A.G., (1995). 'A xylogalacturonan subunit present in the modified hairy regions of apple pectin.' *Carbohydr Res*, 279:265-279.

Schols, H.A. and Voragen, A.G.J., (1994). 'Occurrence of pectic hairy regions in various plant cell wall materials and their degradability by RGase.' *Carbohydr Res*, 256:93-95.

Shanmugam, V., (2005). 'Role of extracytoplasmic leucine rich repeat proteins in plant defence mechanisms'. *Microbiol Res*, 160, 83-94.

Shoseyov, O., Shani, Z. and Levy, I., (2006). 'Carbohydrate binding modules: biochemical properties and novel applications'. *Microbiol Mol Biol Rev*, 70, 283-295.

Simpson, P.J., Bolam, D.N., Cooper, A., Ciruela, A., Hazlewood, G.P., Gilbert, H.J. and Williamson, M.P., (1999). 'A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity'. *Structure*, 7, 853-864.

Simpson, P.J., Jamieson, S.J., Abou-Hachem, M., Karlsson, E.N., Gilbert, H.J., Holst, O. and Williamson, M.P., (2002). 'The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase'. *Biochemistry*, 41, 5712-5719.

Simpson, P.J., Xie, H., Bolam, D.N., Gilbert, H.J. and Williamson, M.P., (2000). 'The structural basis for the ligand specificity of family 2 carbohydrate-binding modules'. *J Biol Chem*, 275, 41137-41142.

Sinnott, M.L., (1990). 'Catalytic mechanism of enzymic glycosyl transfer'. *Chem Rev*, 90:1170 - 1202.

Smith, M.M. and Hartley, R.D., (1983). 'Occurrence and nature of ferulic acid substitution of cell wall polysaccharides in graminaceous plants.' *Carbohydr Res*, 118:65-80.

Sorensen, H.R., Jorgensen, C.T., Hansen, C.H., Jorgensen, C.I., Pedersen, S. and Meyer, A.S., (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, 850-861.

Stalbrand, H., Saloheimo, A., Vehmaanpera, J., Henrissat, B. and Penttila, M., (1995). 'Cloning and expression in *Saccharomyces cerevisiae* of a *Trichoderma reesei* beta-mannanase gene containing a cellulose binding domain'. *Appl Environ Microbiol*, 61, 1090-1097.

Stoner, M.R., Dale, D.A., Gualfetti, P.J., Becker, T. and Randolph, T.W., (2005). 'Ca^{2+}-surfactant interactions affect enzyme stability in detergent solutions'. *Biotechnol Prog*, 21, 1716-1723.

Stoner, M.R., Dale, D.A., Gualfetti, P.J., Becker, T. and Randolph, T.W., (2006). 'Surfactant-induced unfolding of cellulase: kinetic studies'. *Biotechnol Prog*, 22, 225-232.

Studier, F.W. and Moffatt, B.A., (1986). 'Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes'. *J Mol Biol*, 189 (1):113-130.

Sunna, A. and Antranikian, G., (1997). 'Xylanolytic enzymes from fungi and bacteria'. *Crit Rev Biotechnol*, 17, 39-67.

Szabo, L., Jamal, S., Xie, H., Charnock, S.J., Bolam, D.N., Gilbert, H.J. and Davies, G.J., (2001). 'Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation'. *J Biol Chem*, 276, 49061-49065.

Teeri, T.T., (1997). 'Crystalline cellulose degradation: new insight into the function of cellobiohydrolases'. *Trends in Biotechnology*, 15, 160-167.

Tenkanen, M., (1998). 'Action of *Trichoderma reesei* and *Aspergillus oryzae* esterases in the deacetylation of hemicelluloses'. *Biotechnol Appl Biochem*, 27, 19-24.

Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A.L., Sturch, K., Gilkes, N.R., Haynes, C.A., Warren, R.A. and Kilburn, D.G., (1998). 'Characterization and affinity applications of cellulose-binding domains.' *J. Chromatogr.*, 715:283-296.

Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. and Claeyssens, M., (1988). 'Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis'. *Eur J Biochem*, 170, 575-581.

Tomme, P., Warren, R.A. and Gilkes, N.R., (1995). 'Cellulose hydrolysis by bacteria and fungi'. *Adv Microb Physiol*, 37:1-81.

Tormo, J., Lamed, R., Chirino, A.J., Morag, E., Bayer, E.A., Shoham, Y. and Steitz, T.A., (1996). 'Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose'. *Embo J*, 15, 5739-5751.

Torronen, A., Harkki, A. and Rouvinen, J., (1994). 'Three-dimensional structure of endo-1,4-beta-xylanase II from *Trichoderma reesei*: two conformational states in the active site'. *Embo J*, 13, 2493-2501.

Tunnicliffe, R.B., Bolam, D.N., Pell, G., Gilbert, H.J. and Williamson, M.P., (2005). 'Structure of a mannan-specific family 35 carbohydrate-binding module: evidence for significant conformational changes upon ligand binding'. *J Mol Biol*, 347, 287-296.

Vaaje-Kolstad, G., Horn, S.J., van Aalten, D.M., Synstad, B. and Eijsink, V.G., (2005). 'The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation'. *J Biol Chem*, 280, 28492-28497.

van Bueren, A.L., Morland, C., Gilbert, H.J. and Boraston, A.B., (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, 530-537.

Vardakou, M., Dumon, C., Murray, J.W., Christakopoulos, P., Weiner, D.P., Juge, N., Lewis, R.J., Gilbert, H.J. and Flint, J.E., (2008). 'Understanding the structural basis for substrate and inhibitor recognition in eukaryotic GH11 xylanases'. *J Mol Biol*, 375, 1293-1305.

Vardakou, M., Flint, J., Christakopoulos, P., Lewis, R.J., Gilbert, H.J. and Murray, J.W., (2005). 'A family 10 *Thermoascus aurantiacus* xylanase utilizes arabinose decorations of xylan as significant substrate specificity determinants'. *J Mol Biol*, 352, 1060-1067.

Varrot, A., Frandsen, T.P., Driguez, H. and Davies, G.J., (2002). 'Structure of the *Humicola insolens* cellobiohydrolase Cel6A D416A mutant in complex with a non-hydrolysable substrate analogue, methyl cellobiosyl-4-thio-beta-cellobioside, at 1.9 A'. *Acta Crystallogr D Biol Crystallogr*, 58, 2201-2204.

Varrot, A., Hastrup, S., Schulein, M. and Davies, G.J., (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, 297-304.

Volkov, I.Y., Lunina, N.A. and Velikodvorskaya, G.A., (2004). 'Prospects for the practical application of substrate-binding modules of glycosyl hydrolases.' *Appl. Biochem. Microbiol.*, 40:427-432.

von der Osten, C., Bjornvad, M.E., Vind, J. and Rasmussen, M.D., (2000a). 'Process and composition for desizing cellulosic fabric with an enzyme hybrid.' U.S.

von der Osten, C., Bjornvad, M.E., Vind, J. and Rasmussen, M.D., (2000b). 'Process for removal or bleaching of soiling or stains from cellulosic fabric.' U.S.

Vranska, M. and Beily, P., (1992). 'The cellobiohydrolase I from *Trichoderma reesei* QM 9414: action on cellooligosaccharides.' *Carbohydr Res*, 227:19-27.

Vrsanska, M., Kolenova, K., Puchart, V. and Biely, P., (2007). 'Mode of action of glycoside hydrolase family 5 glucuronoxylan xylanohydrolase from *Erwinia chrysanthemi*'. *Febs J*, 274, 1666-1677.

Wakabayashi, K., Hoson, T. and Kamisaka, S., (1997). 'Osmotic stress suppresses cell wall stiffening and the increase in cell wall-bound ferulic and diferulic acids in wheat coleoptiles'. *Plant Physiol*, 113, 967-973.

Wang, J., Stuckey, J.A., Wishart, M.J. and Dixon, J.E., (2002). 'A unique carbohydrate binding domain targets the lafora disease phosphatase to glycogen'. *J Biol Chem*, 277, 2377-2380.

Warren, R.A., (1996). 'Microbial hydrolysis of polysaccharides'. *Annu Rev Microbiol*, 50:183-212.

Weimer, P.J., Hackney, J.M., Jung, H.J. and Hatfield, R.D., (2000). 'Fermentation of a bacterial cellulose/xylan composite by mixed ruminal microflora: implications for the role of polysaccharide matrix interactions in plant cell wall biodegradability'. *J Agric Food Chem*, 48, 1727-1733.

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

Whitney, S.E., Gothard, M.G., Mitchell, J.T. and Gidley, M.J., (1999). 'Roles of cellulose and xyloglucan in determining the mechanical properties of primary plant cell walls'. *Plant Physiol*, 121, 657-664.

Whitney, S.E.C., Brignham, J.E., Darke, A.H., Reid, J.S. and Gidley, M.J., (1998). 'Structural aspects of the interaction of mannan-based polysaccharides with bacterial cellulose'. *Carbohydr Res*, 307:299-309.

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

Williams, A.G. and Withers, S.E., (1985). 'Formation of polysaccharide depolymerase and glycoside hydrolase enzymes by *Bacteroides ruminicola subsp.ruminicola* grown in batch and continuous culture'. *Current Microbiology*, 12:79-84.

Withers, S.G., (2001). 'Mechanisms of glycosyl transferases and hydrolases.' *Carbohydrate Polymers*, 44, 325-337.

Wolfenden, R., Lu, X. and Young, G., (1998). 'Spontaneous Hydrolysis of Glycosides'. *J. Am. Chem. Soc.*, 120, 6814-6815.

Wood, T.M., (1988). 'Preparation of crystalline, amorphous, and dyed cellulose substrates'. *Methods in Enzymology*, 160:19-25.

Xiao, Z.Z., Gao, P.J., Qu, Y.B. and Wang, T.H., (2001). 'Cellulose-binding domain of endoglucanase III from *Trichoderma reesei* disrupting the structure of cellulose'. *Biotechnol. Lett.*, 23 (711-715).

Xie, H., Bolam, D.N., Nagy, T., Szabo, L., Cooper, A., Simpson, P.J., Lakey, J.H., Williamson, M.P. and Gilbert, H.J., (2001a). 'Role of hydrogen bonding in the interaction between a xylan binding module and xylan'. *Biochemistry*, 40, 5700-5707.

Xie, H., Gilbert, H.J., Charnock, S.J., Davies, G.J., Williamson, M.P., Simpson, P.J., Raghothama, S., Fontes, C.M., Dias, F.M., Ferreira, L.M. and Bolam, D.N., (2001b). '*Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding'. *Biochemistry*, 40, 9167-9176.

Xu, Z., Bae, W., Mulchandani, A., Mehra, R.K. and Chen, W., (2002). 'Heavy metal removal by novel CBD-EC20 sorbents immobilized on cellulose'. *Biomacromolecules*, 3, 462-465.

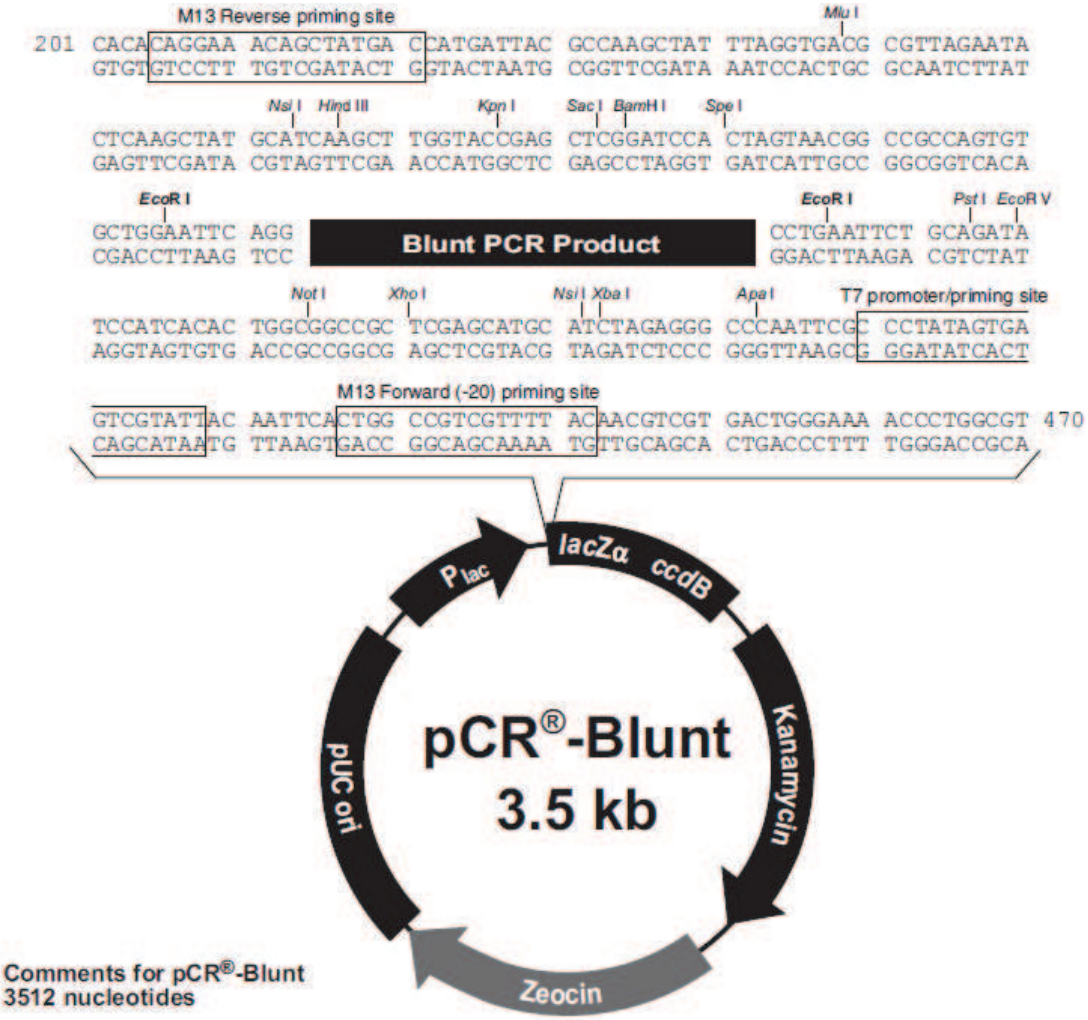
Yennawar, N.H., Li, L.C., Dudzinski, D.M., Tabuchi, A. and Cosgrove, D.J., (2006). 'Crystal structure and activities of EXPB1 (Zea m 1), a beta-expansin and group-1 pollen allergen from maize'. *Proc Natl Acad Sci U S A*, 103, 14664-14671.

Yoder, M.D. and Jurnak, F., (1995). 'The Refined Three-Dimensional Structure of Pectate Lyase C from *Erwinia chrysanthemi* at 2.2 Angstrom Resolution (Implications for an Enzymatic Mechanism)'. *Plant Physiol*, 107,:349-364.

Appendix A. Multiple cloning regions of plasmid vectors used in this study

A1. Map of pCR®-Blunt

The Fig below summarizes the features of the pCR®-Blunt vector. Restriction sites that are only found in the polylinker are shown. **The complete sequence is available for downloading from Web site (www.invitrogen.com)**



Comments for pCR®-Blunt
3512 nucleotides

Lac promoter/operator region: bases 95-216
M13 Reverse priming site: bases 205-221
*Lac*Z-alpha ORF: bases 217-570
T7 promoter priming site: bases 400-419
M13 Forward (-20) priming site: bases 427-442
Fusion joint: bases 571-579
*ccd*B lethal gene ORF: bases 580-882
Kanamycin resistance ORF: bases 1231-2025
Zeocin resistance ORF: bases 2231-2605
pUC origin: bases 2673-3386

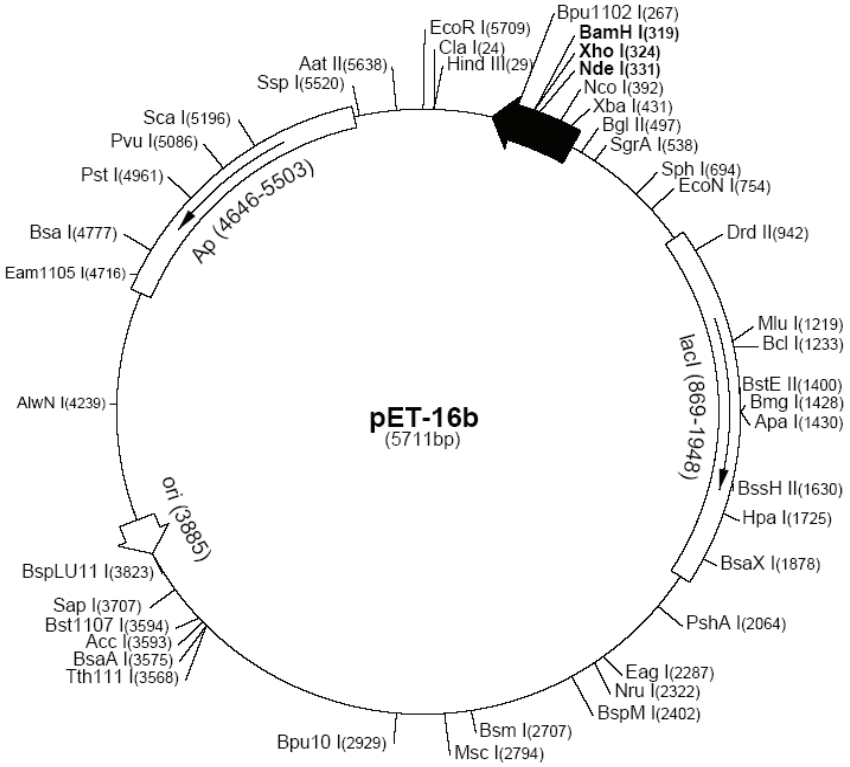
A2. pET-16b

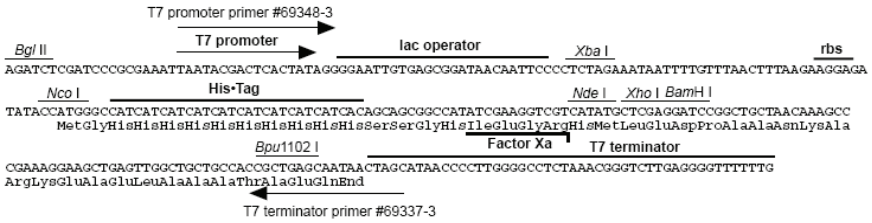
pET-16b Vector

TB046 2/00

The pET-16b vector (Cat. No. 69662-3) carries an N-terminal His•Tag® sequence followed by a Factor Xa site and three cloning sites. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below.

pET-16b sequence landmarks	
T7 promoter	466-482
T7 transcription start	465
His•Tag coding sequence	360-389
Multiple cloning sites	
(*Nde* I -*Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	869-1948
pBR322 origin	3885
bla coding sequence	4646-5503





pET-16b cloning/expression region

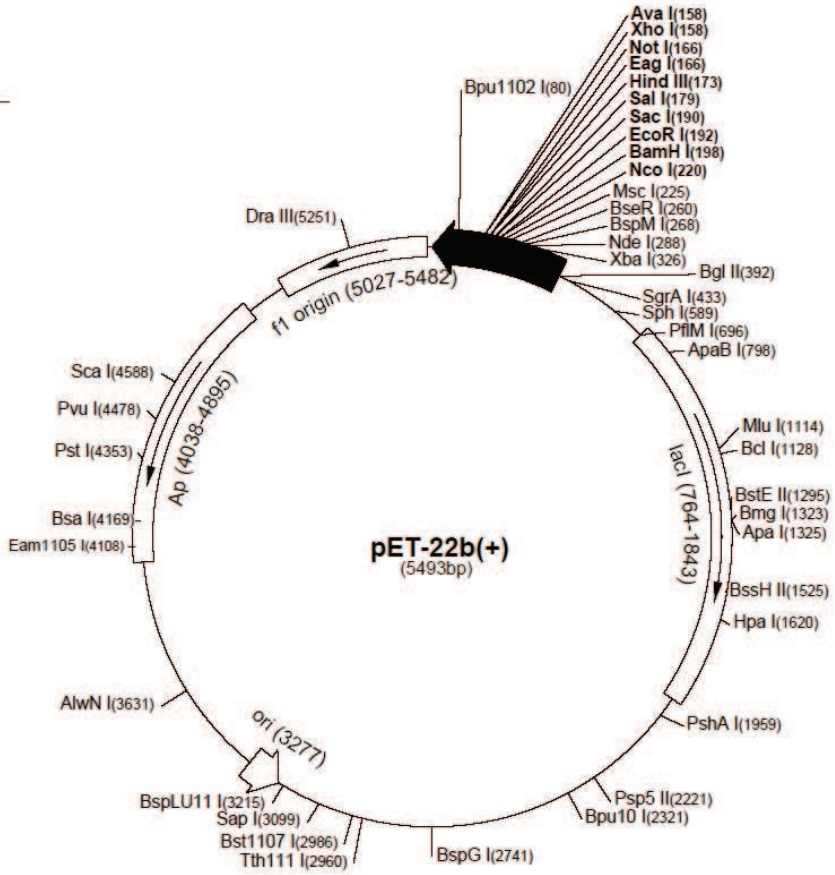
A3. pET-22b

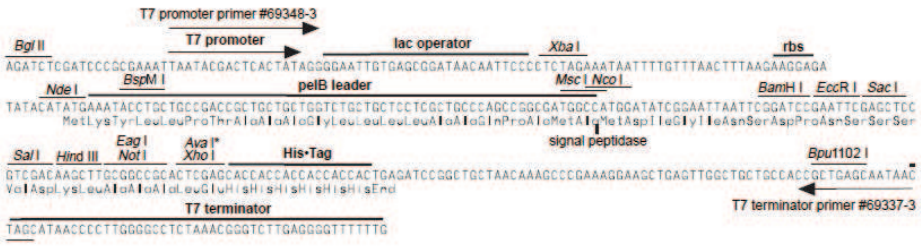
pET-22b(+) Vector

TB038 12/98

The pET-22b(+) vector (Cat. No. 69744-3) carries an N-terminal *pelB* signal sequence for potential periplasmic localization, plus optional C-terminal His•Tag® sequence. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

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

A4. pET-28a

pET-28a-c(+) Vectors

TB074 12/98

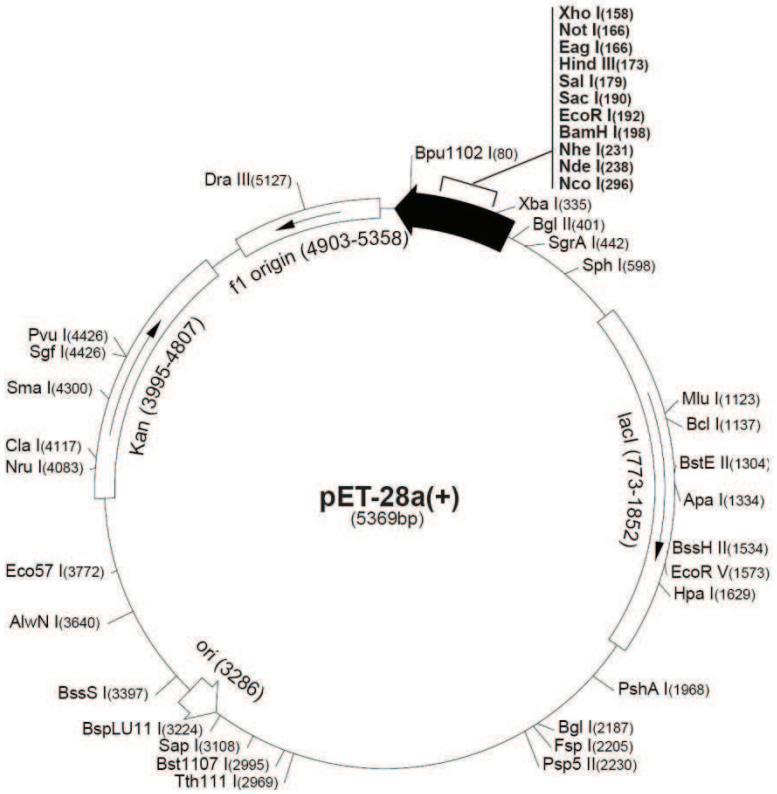
	Cat. No.
pET-28a DNA	69864-3
pET-28b DNA	69865-3
pET-28c DNA	69866-3

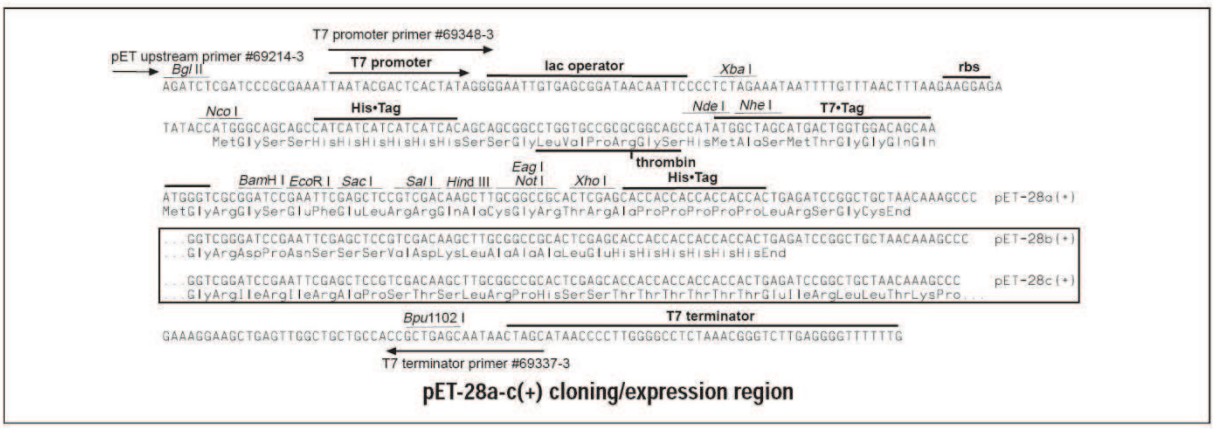
The pET-28a-c(+) vectors carry an N-terminal His• Tag®/thrombin/T7• Tag® configuration plus an optional C-terminal His• Tag sequence. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

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.





Appendix B Crystallization Screens

B1 Newcastle Screen Composition Table

Number	**Salt**	**Buffer**	**Precipitant**
1	0.2M Lithium Sulphate	0.1M Sodium Acetate pH 5.1	50% PEG 400
2		0.1M Sodium Citrate pH 5.5	20% PEG 3000
3		0.2M Diammonium Hydrogen Citrate pH 5.0	20% PEG 3350
4	0.08M Calcium Chloride	0.1M Sodium Acetate pH 4.6	30% MPD
5		0.2M Magnesium Formate pH 5.9	20% PEG 3350
6	0.2M Lithium Sulphate 0.25M Sodium Citrate	0.25M Sodium Dihydrogen Phosphate pH 4.2	20% PEG 1000
7		0.1M CHES pH 9.5	20% PEG 8000
8		0.2M Ammonium Formate pH 6.6	20% PEG 3350
9		0.2M Ammonium Chloride pH 6.3	20% PEG 3350
10		0.2M Potassium Formate pH 7.3	20% PEG 3350
11	0.2M Ammonium Dihydrogen Phosphate	0.1M Tris pH 8.5	50% MPD
12		0.2M Potassium Nitrate pH 6.9	20% PEG 3350
13	0.8M Ammonium Sulphate	0.1M Citric Acid pH 4.0	
14		0.2M Sodium Thiocyanate pH 6.9	20% PEG 3350
15		0.1M Bicine pH 9.0	20% PEG 6000
16	8% Ethylene Glycol	0.1M HEPES pH 7.5	10% PEG 8000
17		0.1M Sodium Cacodylate pH 7.0	40% MPD 5% PEG 8000
18	0.25M Sodium Citrate	0.25M Sodium Dihydrogen Phosphate pH 5.2	5% PEG 1000 40% Ethanol
19		0.1M Sodium Acetate pH 4.6	8% PEG 4000
20	0.2M Magnesium Chloride	0.1M Tris pH 7.0	10% PEG 8000
21		0.1M Citric Acid pH 5.0	20% PEG 6000
22	0.2M Magnesium Chloride	0.1M Sodium Cacodylate pH 6.6	50% PEG 200
23		1.6M Sodium Citrate pH 6.5	
24		0.2M Potassium Citrate pH 8.3	20% PEG 3350
25	0.02M Calcium Chloride	0.1M Sodium Acetate pH 4.6	30% MPD
26	0.2M Sodium Chloride 0.25M Sodium Citrate	0.25M Sodium Dihydrogen Phosphate pH 4.2	20% PEG 8000
27	1.0M Lithium Chloride	0.1M Citric Acid pH 4.0	20% PEG 6000
28		0.2M Ammonium Nitrate pH 6.3	20% PEG 3350
29		0.1M HEPES pH 7.0	10% PEG 6000
30	0.8M Ammonium Dihydrogen Phosphate 0.8M Potassium Dihydrogen Phosphate	0.1M HEPES pH 7.5	
31	0.25M Sodium Citrate	0.25M Sodium Dihydrogen Phosphate pH 5.2	40% PEG 300
32	0.2M Zinc Acetate	0.1M Sodium Acetate pH 4.5	10% PEG 3000

33		0.1M Tris pH 8.5	20% Ethanol
34	25% 1-2-Propanediol	0.1M Sodium Potassium Phosphate pH 6.8	10% Glycerol
35	2% Dioxane	0.1M Bicine pH 9.0	10% PEG 20000
36	2.0M Ammonium Sulphate	0.1M Sodium Acetate pH 4.6	
37			10% PEG 1000 10% PEG 8000
38	20% Glycerol		24% PEG 1000
39	0.2M Magnesium Chloride	0.1M HEPES pH 7.5	30% PEG 400
40	0.2M Sodium Chloride	0.1M Sodium Potassium Phosphate pH 7.2	50% PEG 200
41	0.2M Lithium Sulphate	0.1M Sodium Acetate pH 4.5	30% PEG 8000
42	0.2M Magnesium Chloride	0.1M HEPES pH 7.5	70% MPD
43		0.1M Tris pH 8.5	20% PEG 8000
44	0.2M Lithium Sulphate	0.1M Tris pH 8.4	40% PEG 400
45		0.1M Tris pH 8.0	40% MPD
46	0.17M ammonium Sulphate		25.5% PEG 4000 15% Glycerol
47	0.2M Calcium Acetate	0.1M Sodium Cacodylate pH 7.0	40% PEG 300
48	0.14M Calcium Chloride	0.07M Sodium Acetate pH 4.6	14% Isopropanol 30% Glycerol
49	0.04M Potassium Dihydrogen Phosphate		16% PEG 8000 20% Glycerol
50	1.0M sodium Citrate	0.1M Sodium Cacodylate pH 6.5	
51	0.2M Sodium Chloride	0.1M Sodium Cacodylate pH 6.5	2.0M Ammonium Sulphate
52	0.2M Sodium Chloride	0.1M HEPES pH 7.5	10% Isopropanol
53	1.26M ammonium Sulphate 0.2M Lithium Sulphate	0.1M Tris pH 8.5	
54		0.1M CAPS pH 10.1	40% MPD
55	0.2M Zinc Acetate	0.1M Imidazol pH 8.0	20% PEG 3000
56	0.2M Zinc Acetate	0.1M Sodium Cacodylate pH 6.5	10% Isopropanol
57	1.0M Diammonium Hydrogen Phosphate	0.1M Sodium Acetate pH 4.5	
58	1.6M Magnesium Sulphate	0.1M MES pH 6.5	
59	10% PEG 6000	0.1M Bicine pH 9.0	
60	0.16M Calcium Acetate	0.08M Sodium Cacodylate pH 6.5	14.4% PEG 8000 20% Glycerol
61		0.1M Imidazol pH 8.0	10% PEG 8000
62	0.05M Caesium Chloride	0.1M MES pH 6.5	30% Jeffamine
63	3.2M Ammonium Sulphate	0.1M Citric Acid pH 5.0	
64		0.1M Tris pH 8.0	20% MPD
65		0.1M HEPES pH 6.5	20% Jeffamine
66	0.2M Magnesium Chloride	0.1M Tris pH 8.5	50% Ethylene Glycol
67		0.1M Bicine pH 9.0	10% MPD
68	0.2M Ammonium Sulphate	0.1M Sodium Acetate pH 4.6	30% PEG MME 2000
69	0.2M Ammonium Sulphate	0.1M MES pH 6.5	30% PEG MME 5000
70	0.01M Zinc Suplhate	0.1M MES pH 6.5	25% PEG MME 550
71	0.01M Nickel Chloride	0.1M Tris pH 8.5	20% PEG MME 2000
72	0.1M Sodium Chloride	0.1M Bicine pH 9.0	20% PEG MME 550
73	0.005M Magnesium Chloride	0.05M HEPES pH 7.0	25% PEG MME 550
74	0.1M Potassium Chloride 0.015M Magnesium Chloride	0.05M Tris pH 7.5	10% PEG MME 550

75	0.2M Lithium Sulphate	0.1M MES pH 6.0	20% 1-4-Butandiol
76	0.2M Sodium Chloride	0.1M Imidazol pH 8.0	1M Sodium Potassium Tartrate
77		0.1M Sodium Acetate pH 4.5	20% 1-4-Butandiol
78	0.2M Lithium Sulphate	0.1M CHES pH 9.5	1M Sodium Potassium Tartrate
79		0.1M Sodium Cacodylate pH 6.5	35% Ethoxyethanol
80	35% Propanol	0.1M Tris pH 8.5	
81	3.5M Sodium Formate		
82		0.8M Succinuc Acid pH 7.0	
83		2.1M Mailc Acid pH 7.0	
84		2.4M Sodium Malonate pH 7.0	
85	0.2M Potassium Chloride	0.05M HEPES pH 7.5	35% Pentaerythritol Propoxylate
86	0.005M Ammonium Sulphate	0.05M Tris pH 6.5	30% Pentaerythritol Ethoxylate
87	0.2M Potassium Bromide	0.1M HEPES pH 7.5	25% PEG MME 2000
88	0.2M Potassium Bromide	0.1M Tris pH 8.5	8% PEG 20000 8% PEG MME 550
89	1.0M Potassium Dihydrgoen Phosphate	0.1M Sodium Citrate pH 4.6	
90	0.5M Potassium Dihydrogen Phosphate	0.1M HEPES pH 7.0	
91	0.005M Cadmium Chloride	0.1M Tris pH 8.0	20% PEG 4000
92	0.005M Nickel Chloride	0.1M MED pH 6.5	20% PEG 4000
93	0.8M Sodium Formate	0.1M Imidazol pH 8.0	10% PEG 8000 10% PEG 1000
94	0.005M Cadmium Sulphate	0.1M Sodium Cacodylate pH 6.5	15% PEG 4000
95	0.005M Cobalt Chloride	0.1M HEPES pH 7.5	20% PEG 600
96		0.1M Tris pH 8.0	2M Ammonium Sulphate 10% Jeffamine

C2. The Classics Suite™ , the JCSG+ Suite™ and The PACT Suite™ (Qiagen) Composition Tables can be found from the website:

http://www.qiagen.com/products/protein/crystallization/compositiontables/default.aspx

Appendix C

	Composition
Luria-Bertani Medium (LB)	10 g Bacto® tryptone (Difco) 10 g NaCl 5 g Yeast extract (Difco) The final volume was made upto 1 litre with double distilled water and the pH adjusted to 7.4 with NaOH, before autoclaving.

Table C.1 Growth medium

Buffer	Composition
10 x Tris.borate (TBE) (pH 8.3 checked)	108 g/l Tris base, 55 g/l Boric acid, 40 ml/l 0.5 M EDTA pH 8.0
DNA Sample Buffer	0.25 % (w/v) Orange G, 50 % (v/v) glycerol, (10x) TBE
PCR DNA Sample Buffer	0.25 % (w/v) Xylene Cyanol FF, 50 % (v/v) glycerol, (10x) TBE

Table C.2 Agarose electrophoresis solutions

Buffer	Composition
SDS Sample Buffer	25 mM Tris/HCl, pH 8.3 25 % (w/v) glycerol 10 % (w/v) SDS 25 % (v/v) β-mercaptoehtanol
Bromophenol blue Sample Buffer	0.05 % Bromophenol Blue in distilled water

Table C.3 SDS-PAGE solutions

Low molecular weight protein standards	**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	**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

Table C.4 Sigma low and high molecular weight protein standards

Appendix D

D.1 A scheme for making the plasmid p15PT-GFP encoding CBM15-GFP fusion protein with a proline threoline rich linker in between the protein modules is presented in Fig D3.1.

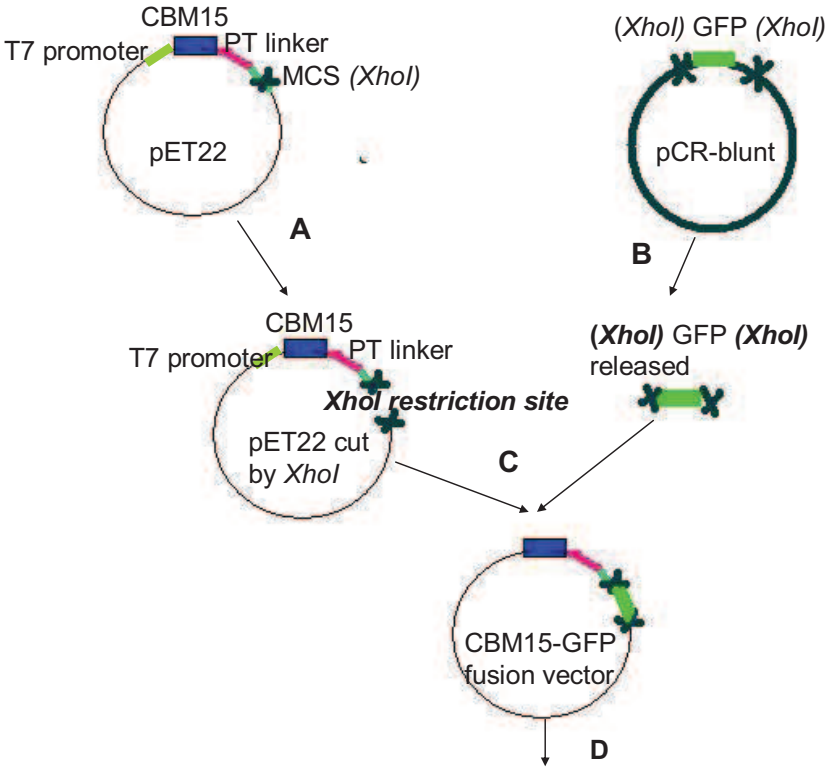


Fig D3.1 Schematic overview of the construction of p15PT-GFP which encodes the CBM15-GFP fusion protein Step A. The pCBM15-PT construct in derived from pET22 vector and encoding CBMs was cut open at *XhoI* site in the multicloning site (MCS). Step B. The pCR-blunt vector containing the gFP gene was cut by *XhoI*, releasing the gFP gene in it. Step C. The cut vector pCBM15-PT was ligated with the GFP gene insert. Step D. The ligation reaction mixture was transformed into *E.coli* Top 10 and transformed and selected on LB amp plates. Isolated colonies were selected, analyzed and sequenced. The fusion construct with the right sequence was expressed, purified and assayed.

D.2 Plasmid p15PT encoding CBM15 and a PT linker was digested by *XhoI*

To make the fusion vector, the vector (p15PT) needed to be cut open by restriction enzyme *XhoI* first. The vector p15PT was digested with *XhoI* and

subjected to agarose gel electrophoresis. It showed that uncut vector had several bands of different sizes, closed circular DNA; supercoiled DNA and different multimers. By contrast, the completely digested DNA was linearized DNA of one size (6500 bp), and thus only a single DNA band was evident.

D.3 The cut vector p15PT was dephosphorylated by calf intestinal alkaline phosphotase (CIAP) treatment.

Half of the digested p15PT vector was subjected to dephosphorylation with calf intestinal alkaline phosphatase (CIAP), which removed the phosphate end groups, thus the vector would not be re-ligated by itself. It would, anneal to and ligate with the phosphorylated GFP gene.

D.4 gFP inserts was prepared by digesting pCR-blunt vector with XhoI restriction enzyme.

The GFP gene was released from the plasmid pCR-blunt-GFP with *XhoI*. Analysis of the digested plasmid revealed a DNA band of 650 bp (the released GFP gene) and a band of 5350 bp (the remaining PCR-blunt vector). The cut gFP DNA band was extracted from the gel and spin purified.

D.5 Ligation of the cut vector p15PT and the gFP insert was carried out and transformed into Top 10 cell line.

Digested dephosphorylated and phosphorylated p15PT and the gFP gene were ligated

and transformed into *E.coli* strain Top10 competent cells (Table D3.1).

Components in the reaction mixture	colony number (approx)
1. phosphorylated cut vector p15PT, GFP insert, ligase, ligation buffer	200
2. phosphorylated cut vector p15PT, ligase, ligation buffer (no insert)	50
3. GFP insert, ligase, ligation buffer (no vector)	0
4. phosphorylated cut vector p15PT, ligation buffer (no ligase)	0
5. ligase, ligation buffer (No DNA)	0
6. cut vector p15PT (positive control reaction)	500
7. dephosphorylated cut vector p15PT, GFP insert, ligase, ligation buffer	12

Table D3.1 Ligation reactions (including negative control reactions) of cut vector 15PT and gFP insert *Reaction 1-6 used non-CIAP treated vector. Reaction 7 used CIAP treated vector. Another set of ligations was done in parallel, but using CIAP-treated vector, instead of non-CIAP treated vectors. It was observed that using non-CIAP treated vector for ligation results in four times more colonies.

D.6 Restriction enzyme analysis of colonies on plus transformation plates was carried out.

Plasmid was isolated from 60 transformants (48 from ligation with non-CIAP treated p15PT, 12 from ligation with CIAP treated p15PT) were digested with *XhoI* and subjected to agarose gel electrophoresis. Despite the lower transformation frequency of CIAP-treated p15PT, 10 out of 12 plasmids contained the gFP gene, while 3 out of 48 transformants derived from the non-CIAP treated vector contained the GFP insert. The 13 gFP-encoding plasmids were subjected to further analysis, to ensure the gFP and CBM5 genes were in the same orientation.

By using vector analysis software, a pair of restriction enzymes *NcoI* and *BglII* were chosen to digest the p15PT-gFP. If the gFP gene was in the same of the CBM15 gene, it would give two DNA fragments of 831 and 5808 bp, whereas if the gFP gene was in the opposite orientation round, the enzymes would generate two DNA fragments of 1213 and 5426 bp. Lane C5, C6, C7 (CIAP-treated vector) and 29 (non-CIAP treated vector) contained the GFP insert in the same orientation as the gFP gene (Fig D3.2). The gFP gene inserted into p15PT in the opposite orientation would not be translated into green fluorescent protein, thus the transformant would not look green fluorescent (Fig D3.3)

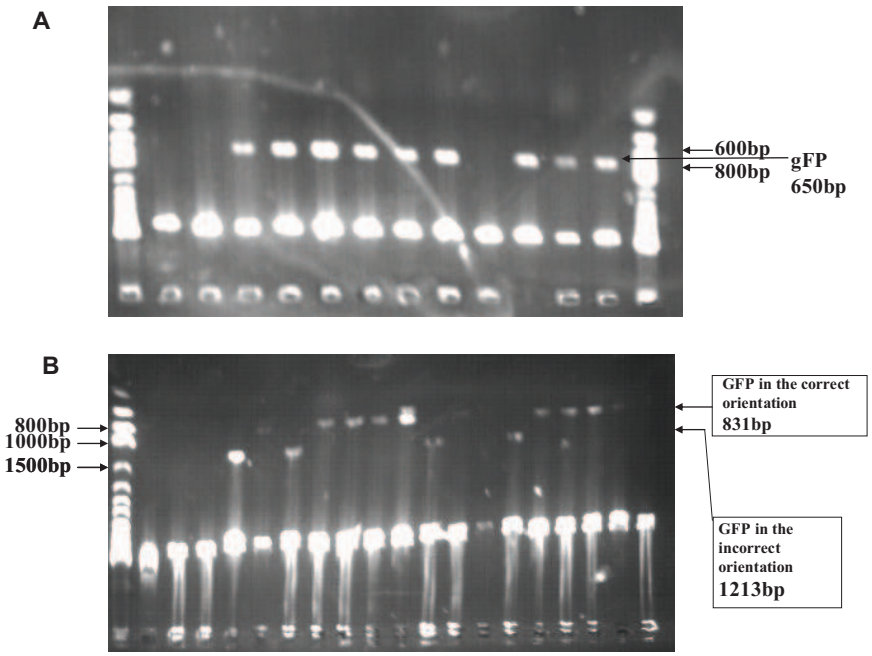
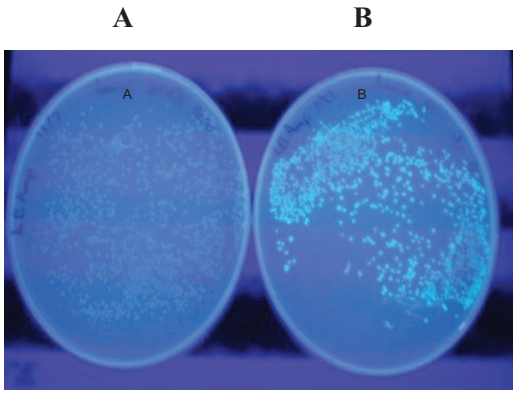


Fig D3.2 Agarose gel electrophoresis of potential fusion constructs, with gFP gene correctly and incorrectly orientated. Panel A displays the GFP gene fragment released after p15PT was digested by *XhoI*. Panel B displays the two possible DNA fragment patterns (GFP correctly and incorrectly orientated in p15PT) after *NcoI and BglII* digestion of p15PT-GFP.



C

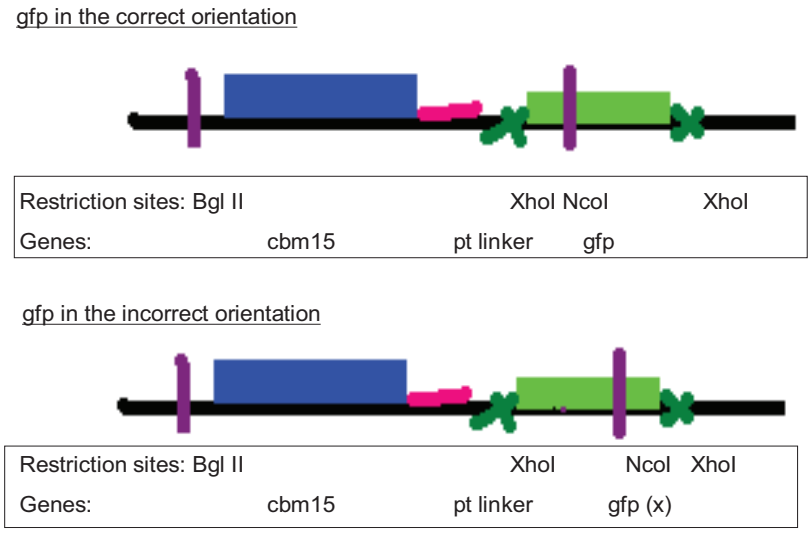


Fig D3.3 *E. coli* BL21 cells transformed with 1 µl plasmid DNA at 37°C overnight on LB^{amp} plates Panel A displays p15PT with GFP incorrectly orientated; Panel B displays p15PTwith GFP correctly oritentated; Panel C displays a schematic diagraph showing the pET22b with GFP in the correct or incorrect orientation

D.7 Over expression and purification of CBM15-GFP and CBM2b1, 2-GFP fusion proteins

E.coli strains BL21 containing p15PT-GFP variants C5, C6 and C7 and 29 encoding CBM-gFP were incubated with 1 mM IPTG for four hours to induce of the expression of the recombinant fusion protein. The CBM15-GFP fusion proteins are his-tagged, and were purified by IMAC using 10 and 100mM imidazole to elute the recombinant proteins from the Talon columns (Fig D3.4). The binding of the CBM15-GFP fusion to xylan arrayed on nitrocellulose was assessed. The protocol was established in the previous trial experiment, and the GFP-CBM15 was used at 1 mg/ml, while GFP used as a negative control was employed at the same concentration. This method involves a single incubation and a single washing step. The data showed that the CBM15 bound to insoluble oat spelt xylan strongly and weakly to Avicel and grass (Section 3.2.5.3).

Similarly, the expression vector encoding CBM2b1,2-GFP was transformed into Tuner and grown at 16 °C for overnight with the addition of 0.2 mM isopropyl-

D-thiogalactopyranoside (IPTG). It is purified by Immobilised Metal Affinity Chromatography (IMAC) Talon column and most were eluted in 10 mM imidazole Talon elution buffer (Fig D3.4).

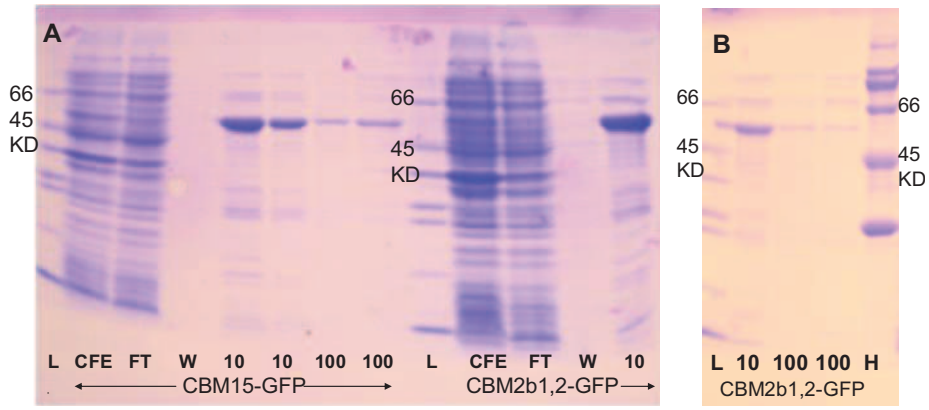


Fig D3.4 Purification of CBM15-GFP and CBM2b1,2-GFP fusion proteins
SDS-PAGE was carried out using a 12.5 % polyacrylamide gel. Panel A displays SDS-PAGE analysis of expression of CBM15-GFP and CBM2b1,2-GFP fusion proteins. Panel B displays SDS-PAGE analysis of expression of CBM2b1,2-GFP fusion proteins (Continued). *E. coli* Tuner cultures expressing the CBM15-GFP and CBM2b1,2-GFP were fractionated. CFE, 5 µl cell free extract; FT, 5 µl flow through; W, 15 µl wash; 10, elution with 10 mM imidazole; 100, elutions with 100 mM imidazole; L, low molecular weight markers. H, High molecular weight markers.

```
atgggatccgtcgctgccagcgagggcaatgttgttatagaggtggacatggcaaatggc
 M  G  S  V  A  A  S  E  G  N  V  V  I  E  V  D  M  A  N  G
tggagaggcaacgcatcaggcagtaccagccattccggtattacctacagtgccgatggc
 W  R  G  N  A  S  G  S  T  S  H  S  G  I  T  Y  S  A  D  G
gttacctttgccgcactgggtgatggcgtgggcgctgtttttgatattgcccgaccaacc
 V  T  F  A  A  L  G  D  G  V  G  A  V  F  D  I  A  R  P  T
acactggaagatgctgtgatagcaatggttgttaatgtcagcgctgaatttaaggccagt
 T  L  E  D  A  V  I  A  M  V  V  N  V  S  A  E  F  K  A  S
gaagccaacttgcagatatttgcccagttaaaagaagactggtcaaagggcgaatgggat
 E  A  N  L  Q  I  F  A  Q  L  K  E  D  W  S  K  G  E  W  D
tgtctggcggccagcagcgaactcactgcggatactgacctaaccctgacctgcaccatt
 C  L  A  A  S  S  E  L  T  A  D  T  D  L  T  L  T  C  T  I
gatgaagacgacgataaattcaaccaaacggcgcgcgatgtacaagtcggtatccaggcc
 D  E  D  D  D  K  F  N  Q  T  A  R  D  V  Q  V  G  I  Q  A
aagggaacacccgccggaactatcaccattaaaagcgtcaccattacactcgcacaggaa
 K  G  T  P  A  G  T  I  T  I  K  S  V  T  I  T  L  A  Q  E
gccaagcttggcggcggcacggcgaccccgaccccacgccgacgccgacgccggaattc
 A  K  L  G  G  G  T  A  T  P  T  P  T  P  T  P  T  P  E  F
gtcgctgccagcgagggcaatgttgttatagaggtggacatggcaaatggctggagaggc
 V  A  A  S  E  G  N  V  V  I  E  V  D  M  A  N  G  W  R  G
aacgcatcaggcagtaccagccattccggtattacctacagtgccgatggcgttaccttt
 N  A  S  G  S  T  S  H  S  G  I  T  Y  S  A  D  G  V  T  F
gccgcactgggtgatggcgtgggcgctgtttttgatattgcccgaccaaccacactggaa
 A  A  L  G  D  G  V  G  A  V  F  D  I  A  R  P  T  T  L  E
gatgctgtgatagcaatggttgttaatgtcagcgctgaatttaaggccagtgaagccaac
 D  A  V  I  A  M  V  V  N  V  S  A  E  F  K  A  S  E  A  N
ttgcagatatttgcccagttaaaagaagactggtcaaagggcgaatgggattgtctggcg
 L  Q  I  F  A  Q  L  K  E  D  W  S  K  G  E  W  D  C  L  A
gccagcagcgaactcactgcggatactgacctaaccctgacctgcaccattgatgaagac
 A  S  S  E  L  T  A  D  T  D  L  T  L  T  C  T  I  D  E  D
gacgataaattcaaccaaacggcgcgcgatgtacaagtcggtatccaggccaagggaaca
 D  D  K  F  N  Q  T  A  R  D  V  Q  V  G  I  Q  A  K  G  T
cccgccggaactatcaccattaaaagcgtcaccattacactcgcacaggaagccaagctt
 P  A  G  T  I  T  I  K  S  V  T  I  T  L  A  Q  E  A  K  L
ggcggcggcacggcgaccccgaccccacgccgacgccgacgccggagctcgtcgacctc
 G  G  G  T  A  T  P  T  P  T  P  T  P  T  P  E  L  V  D  L
gagcaccaccaccaccac
 E  H  H  H  H  H  H
```

Fig D3.5 The sequence of CBM15-CBM15 linker by a proline-threonine linker (1180 bp) with translated amino acid sequence underneath. It encodes CBM15-CBM15 of 44 kDa.

Appendix E

E.1 Fluorescence microscopy data for BSA-488 and CBM-488 binding to grass on cotton

E.1.1 BSA-488 binding data using fluorescence microscope

BSA-488-Day1	Total number of Pixels	Normalised Data
A, CBM,Grass no SLES,	2252495	1.000
D CBM,Cotton no SLES	2257986	1.002
G CBM, Grass +SLES	3605121	1.601
J CBM, Cotton +SLES	2363811	1.049
B -, Grass no SLES	137289.8	0.061
E -, Cotton no SLES	152987.2	0.068
H -, Grass +SLES	153044.8	0.068
K -Cotton +SLES	111864.2	0.050
C Free dye, Grass no SLES	3147837	1.397
F Free dye, Cotton no SLES	2830252	1.256
I Free dye,Grass +SLES	5959647	2.646
L Free dye, Cotton+SLES	1811785	0.804
BSA-488-Day2		
A, CBM,Grass no SLES,	3069056	1.000
D CBM,Cotton no SLES	3098812	1.010
G CBM, Grass +SLES	4546350	1.481
J CBM, Cotton +SLES	2334226	0.761
B -, Grass no SLES	3139177	1.023
E -, Cotton no SLES	3246581	1.058
H -, Grass +SLES	3308276	1.078
K -Cotton +SLES	1266020	0.413
C Free dye, Grass no SLES	6203099	2.021
F Free dye, Cotton no SLES	4794781	1.562
I Free dye,Grass +SLES	6511464	2.122
L Free dye, Cotton+SLES	2995530	0.976
BSA-488-Day3		
A, CBM,Grass no SLES,	3267487	1.000
D CBM,Cotton no SLES	2506222	0.767
G CBM, Grass +SLES	4736465	1.450
J CBM, Cotton +SLES	2126141	0.651
B -, Grass no SLES	3456203	1.058
E -, Cotton no SLES	3523128	1.078
H -, Grass +SLES	3822818	1.170
K -Cotton +SLES	2242112	0.686
C Free dye, Grass no SLES	5891161	1.803
F Free dye, Cotton no SLES	1534545	0.470

I Free dye,Grass +SLES	5357114	1.640
L Free dye, Cotton+SLES	1513265	0.463
BSA-488-Day4		
A, CBM,Grass no SLES,	5223297	1.000
D CBM,Cotton no SLES	3313167	0.634
G CBM, Grass +SLES	6617058	1.267
J CBM, Cotton +SLES	1409263	0.270
B -, Grass no SLES	3171669	0.607
E -, Cotton no SLES	1737845	0.333
H -, Grass +SLES	3186549	0.610
K -Cotton +SLES	2433420	0.466
C Free dye, Grass no SLES	4070264	0.779
F Free dye, Cotton no SLES	1953282	0.374
I Free dye,Grass +SLES	5272998	1.010
L Free dye, Cotton+SLES	2419101	0.463
BSA-488-Day5		
A, CBM,Grass no SLES,	3606866	1.000
D CBM,Cotton no SLES	5014890	1.390
G CBM, Grass +SLES	4172044	1.157
J CBM, Cotton +SLES	4436869	1.230
B -, Grass no SLES	2848400	0.790
E -, Cotton no SLES	2266795	0.628
H -, Grass +SLES	3063890	0.849
K -Cotton +SLES	1577751	0.437
C Free dye, Grass no SLES	2691502	0.746
F Free dye, Cotton no SLES	2299386	0.638
I Free dye,Grass +SLES	3995350	1.108
L Free dye, Cotton+SLES	2311284	0.641
BSA-488-Day6		
A, CBM,Grass no SLES,	3044057	1.000
D CBM,Cotton no SLES	2219646	0.729
G CBM, Grass +SLES	3894603	1.279
J CBM, Cotton +SLES	2406886	0.791
B -, Grass no SLES	2804712	0.921
E -, Cotton no SLES	2304247	0.757
H -, Grass +SLES	3308789	1.087
K -Cotton +SLES	1771364	0.582
C Free dye, Grass no SLES	3534480	1.161
F Free dye, Cotton no SLES	3556783	1.168
I Free dye,Grass +SLES	2620108	0.861
L Free dye, Cotton+SLES	3765067	1.237

E.1.2 CBM15-488 binding data using fluorescence microscope

CBM15-488-Day1	Total number of Pixels	Normalised Data
A, CBM,Grass no SLES,	3793678.4	1.000
D CBM,Cotton no SLES	3255819.8	0.858
G CBM, Grass +SLES	5305561.8	1.399
J CBM, Cotton +SLES	4842693.4	1.277
B -, Grass no SLES	2669642	0.704
E -, Cotton no SLES	2124657.6	0.560
H -, Grass +SLES	2723555	0.718
K -Cotton +SLES	1706019.2	0.450
C Free dye, Grass no SLES	2979688.2	0.785
F Free dye, Cotton no SLES	2980826.4	0.786
I Free dye,Grass +SLES	3574876.6	0.942
L Free dye, Cotton+SLES	2161531.2	0.570
CBM15-488-Day2		
A, CBM,Grass no SLES,	4089709.8	1.000
D CBM,Cotton no SLES	5099759.4	1.247
G CBM, Grass +SLES	4632991.8	1.133
J CBM, Cotton +SLES	4880189	1.193
B -, Grass no SLES	2334441.2	0.571
E -, Cotton no SLES	2199052.6	0.538
H -, Grass +SLES	3852289.2	0.942
K -Cotton +SLES	1602783	0.392
C Free dye, Grass no SLES	3281535.2	0.802
F Free dye, Cotton no SLES	2798278.2	0.684
I Free dye,Grass +SLES	3641743.2	0.890
L Free dye, Cotton+SLES	1944882.2	0.476
CBM15-488-Day3		
A, CBM,Grass no SLES,	8348014	1.000
D CBM,Cotton no SLES	10392488	1.245
G CBM, Grass +SLES	14952765	1.791
J CBM, Cotton +SLES	10358563.2	1.241
B -, Grass no SLES	5790386.2	0.694
E -, Cotton no SLES	4776978	0.572
H -, Grass +SLES	7843954.4	0.940
K -Cotton +SLES	3236415	0.388
C Free dye, Grass no SLES	5695929.6	0.682
F Free dye, Cotton no SLES	2816159.4	0.337
I Free dye,Grass +SLES	8610822	1.031
L Free dye, Cotton+SLES	3553959.6	0.426
CBM15-488-Day4		
A, CBM,Grass no SLES,	5967684.4	1.000

D CBM,Cotton no SLES	7092322	1.188
G CBM, Grass +SLES	6440408.8	1.079
J CBM, Cotton +SLES	7312346	1.225
B -, Grass no SLES	5164338.8	0.865
E -, Cotton no SLES	5359766	0.898
H -, Grass +SLES	6712519.2	1.125
K -Cotton +SLES	3029473.4	0.508
C Free dye, Grass no SLES	4508944.8	0.756
F Free dye, Cotton no SLES	3655992.4	0.613
I Free dye,Grass +SLES	5642175.8	0.945
L Free dye, Cotton+SLES	2569167	0.431
CBM15-488-Day5		
A, CBM,Grass no SLES,	4655309.4	1.000
D CBM,Cotton no SLES	4168302.8	0.895
G CBM, Grass +SLES	5648067.8	1.213
J CBM, Cotton +SLES	4031997.2	0.866
B -, Grass no SLES	2816643	0.605
E -, Cotton no SLES	2519613.6	0.541
H -, Grass +SLES	3931377.2	0.844
K -Cotton +SLES	1863005.2	0.400
C Free dye, Grass no SLES	3058778.4	0.657
F Free dye, Cotton no SLES	1991687.4	0.428
I Free dye,Grass +SLES	4621708.6	0.993
L Free dye, Cotton+SLES	2108551.6	0.453
CBM15-488-Day6		
A, CBM,Grass no SLES,	4612339	1.000
D CBM,Cotton no SLES	6265158	1.358
G CBM, Grass +SLES	6076087.6	1.317
J CBM, Cotton +SLES	5332430.8	1.156
B -, Grass no SLES	2737047	0.593
E -, Cotton no SLES	2436244.2	0.528
H -, Grass +SLES	3691375.8	0.800
K -Cotton +SLES	1501610	0.326
C Free dye, Grass no SLES	3252326	0.705
F Free dye, Cotton no SLES	2773687.2	0.601
I Free dye,Grass +SLES	4410910.4	0.956
L Free dye, Cotton+SLES	3331264.8	0.722

E.1.3 CBM2b1,2-488 binding data using fluorescence microscope

CBM2b1,2-488-Day1	Total number of Pixels	Normalised Data
A, CBM,Grass no SLES,	6489397	1.000
D CBM,Cotton no SLES	7879919	1.214
G CBM, Grass +SLES	10411041	1.604
J CBM, Cotton +SLES	8902424	1.372
B -, Grass no SLES	6201579	0.956
E -, Cotton no SLES	3329538	0.513
H -, Grass +SLES	7206989	1.111
K -Cotton +SLES	4218357	0.650
C Free dye, Grass no SLES	5579293	0.860
F Free dye, Cotton no SLES	8219832	1.267
I Free dye,Grass +SLES	7910489	1.219
L Free dye, Cotton+SLES	3466138	0.534
CBM2b1,2-488-Day2		
A, CBM,Grass no SLES,	7992518	1.000
D CBM,Cotton no SLES	11525047	1.442
G CBM, Grass +SLES	19680269	2.462
J CBM, Cotton +SLES	13719019	1.716
B -, Grass no SLES	4361403	0.546
E -, Cotton no SLES	3802582	0.476
H -, Grass +SLES	10920604	1.366
K -Cotton +SLES	12407047	1.552
C Free dye, Grass no SLES	4991654	0.625
F Free dye, Cotton no SLES	3022927	0.378
I Free dye,Grass +SLES	7857997	0.983
L Free dye, Cotton+SLES	4944882	0.619
CBM2b1,2-488-Day3		
A, CBM,Grass no SLES,	7102482	1.000
D CBM,Cotton no SLES	9546258	1.344
G CBM, Grass +SLES	11091216	1.562
J CBM, Cotton +SLES	6524190	0.919
B -, Grass no SLES	4901198	0.690
E -, Cotton no SLES	4531967	0.638
H -, Grass +SLES	3511805	0.494
K -Cotton +SLES	3562872	0.502
C Free dye, Grass no SLES	4826768	0.680
F Free dye, Cotton no SLES	2858788	0.403
I Free dye,Grass +SLES	3159443	0.445
L Free dye, Cotton+SLES	1742181	0.245
CBM2b1,2-488-Day4		
A, CBM,Grass no SLES,	7041054	1.000

D CBM,Cotton no SLES	7546317	1.072
G CBM, Grass +SLES	10908253	1.549
J CBM, Cotton +SLES	9461519	1.344
B -, Grass no SLES	3294379	0.468
E -, Cotton no SLES	2678388	0.380
H -, Grass +SLES	3110295	0.442
K -Cotton +SLES	916475	0.130
C Free dye, Grass no SLES	3539608	0.503
F Free dye, Cotton no SLES	2656490	0.377
I Free dye,Grass +SLES	4170457	0.592
L Free dye, Cotton+SLES	2133812	0.303
CBM2b1,2-488-Day5		
A, CBM,Grass no SLES,	6021803	1.000
D CBM,Cotton no SLES	7868584	1.307
G CBM, Grass +SLES	8541896	1.418
J CBM, Cotton +SLES	6807472	1.130
B -, Grass no SLES	2958652	0.491
E -, Cotton no SLES	1248673	0.207
H -, Grass +SLES	3472210	0.577
K -Cotton +SLES	1539084	0.256
C Free dye, Grass no SLES	3199367	0.531
F Free dye, Cotton no SLES	2184822	0.363
I Free dye,Grass +SLES	3636913	0.604
L Free dye, Cotton+SLES	1669667	0.277
CBM2b12-488-Day6		
A, CBM,Grass no SLES,	6409551	1.000
D CBM,Cotton no SLES	5931444	0.925
G CBM, Grass +SLES	7665308	1.196
J CBM, Cotton +SLES	7782853	1.214
B -, Grass no SLES	2827415	0.441
E -, Cotton no SLES	1865128	0.291
H -, Grass +SLES	4253564	0.664
K -Cotton +SLES	1670861	0.261
C Free dye, Grass no SLES	2453943	0.383
F Free dye, Cotton no SLES	2144886	0.335
I Free dye,Grass +SLES	2735742	0.427
L Free dye, Cotton+SLES	1254739	0.196

Fig E3.1 Average fluoresence intensity for different area on grass stains for each condition, recorded from Day 1 to Day 6 in tabular form to generated Fig 3.21.

E.2 Stats analysis of BSA-488 and CBM-488 binding to grass on cotton

E.2.1 BSA-488

One-way Analysis of Variance (ANOVA)

The P value is 0.0001, considered extremely significant.

Variation among column means is significantly greater than expected by chance.

Tukey-Kramer Multiple Comparisons Test

If the value of q is greater than 4.808 then the P value is less than 0.05.

Comparison	Mean Difference	q		P value
A vs B	0.2567	1.607	ns	P>0.05
A vs C	-0.3181	1.991	ns	P>0.05
A vs D	0.07778	0.4870	ns	P>0.05
A vs E	0.3463	2.168	ns	P>0.05
A vs F	0.08862	0.5549	ns	P>0.05
A vs G	-0.3724	2.332	ns	P>0.05
A vs H	0.1896	1.187	ns	P>0.05
A vs I	-0.5630	3.525	ns	P>0.05
A vs J	0.2081	1.303	ns	P>0.05
A vs K	0.5611	3.513	ns	P>0.05
A vs L	0.2360	1.477	ns	P>0.05
B vs C	-0.5747	3.599	ns	P>0.05
B vs D	-0.1789	1.120	ns	P>0.05
B vs E	0.08963	0.5612	ns	P>0.05
B vs F	-0.1681	1.052	ns	P>0.05
B vs G	-0.6291	3.939	ns	P>0.05
B vs H	-0.06708	0.4200	ns	P>0.05
B vs I	-0.8197	5.132	*	P<0.05
B vs J	-0.04857	0.3041	ns	P>0.05
B vs K	0.3044	1.906	ns	P>0.05
B vs L	-0.02072	0.1297	ns	P>0.05
C vs D	0.3958	2.478	ns	P>0.05
C vs E	0.6644	4.160	ns	P>0.05
C vs F	0.4067	2.546	ns	P>0.05
C vs G	-0.05435	0.3403	ns	P>0.05
C vs H	0.5077	3.179	ns	P>0.05
C vs I	-0.2449	1.534	ns	P>0.05
C vs J	0.5262	3.295	ns	P>0.05
C vs K	0.8791	5.505	*	P<0.05
C vs L	0.5540	3.469	ns	P>0.05

D vs E	0.2685	1.681	ns	P>0.05
D vs F	0.01083	0.06783	ns	P>0.05
D vs G	-0.4502	2.819	ns	P>0.05
D vs H	0.1118	0.7001	ns	P>0.05
D vs I	-0.6408	4.012	ns	P>0.05
D vs J	0.1303	0.8161	ns	P>0.05
D vs K	0.4833	3.026	ns	P>0.05
D vs L	0.1582	0.9905	ns	P>0.05
E vs F	-0.2577	1.614	ns	P>0.05
E vs G	-0.7187	4.500	ns	P>0.05
E vs H	-0.1567	0.9813	ns	P>0.05
E vs I	-0.9093	5.694	**	P<0.01
E vs J	-0.1382	0.8653	ns	P>0.05
E vs K	0.2148	1.345	ns	P>0.05
E vs L	-0.1104	0.6910	ns	P>0.05
F vs G	-0.4610	2.887	ns	P>0.05
F vs H	0.1010	0.6323	ns	P>0.05
F vs I	-0.6516	4.080	ns	P>0.05
F vs J	0.1195	0.7482	ns	P>0.05
F vs K	0.4725	2.958	ns	P>0.05
F vs L	0.1474	0.9226	ns	P>0.05
G vs H	0.5620	3.519	ns	P>0.05
G vs I	-0.1906	1.193	ns	P>0.05
G vs J	0.5805	3.635	ns	P>0.05
G vs K	0.9335	5.845	**	P<0.01
G vs L	0.6084	3.809	ns	P>0.05
H vs I	-0.7526	4.712	ns	P>0.05
H vs J	0.01852	0.1159	ns	P>0.05
H vs K	0.3715	2.326	ns	P>0.05
H vs L	0.04637	0.2903	ns	P>0.05
I vs J	0.7711	4.828	*	P<0.05
I vs K	1.124	7.038	***	P<0.001
I vs L	0.7990	5.003	*	P<0.05
J vs K	0.3530	2.210	ns	P>0.05
J vs L	0.02785	0.1744	ns	P>0.05
K vs L	-0.3251	2.036	ns	P>0.05

	Mean	95% Confidence Interval	
Difference	Difference	From	To
A - B	0.2567	-0.5112	1.025
A - C	-0.3181	-1.086	0.4498
A - D	0.07778	-0.6901	0.8457
A - E	0.3463	-0.4216	1.114
A - F	0.08862	-0.6793	0.8565

A - G	-0.3724	-1.140	0.3955
A - H	0.1896	-0.5783	0.9575
A - I	-0.5630	-1.331	0.2049
A - J	0.2081	-0.5598	0.9760
A - K	0.5611	-0.2068	1.329
A - L	0.2360	-0.5319	1.004
B - C	-0.5747	-1.343	0.1931
B - D	-0.1789	-0.9468	0.5890
B - E	0.08963	-0.6782	0.8575
B - F	-0.1681	-0.9359	0.5998
B - G	-0.6291	-1.397	0.1388
B - H	-0.06708	-0.8350	0.7008
B - I	-0.8197	-1.588	-0.05180
B - J	-0.04857	-0.8164	0.7193
B - K	0.3044	-0.4635	1.072
B - L	-0.02072	-0.7886	0.7472
C - D	0.3958	-0.3720	1.164
C - E	0.6644	-0.1035	1.432
C - F	0.4067	-0.3612	1.175
C - G	-0.05435	-0.8222	0.7135
C - H	0.5077	-0.2602	1.276
C - I	-0.2449	-1.013	0.5229
C - J	0.5262	-0.2417	1.294
C - K	0.8791	0.1112	1.647
C - L	0.5540	-0.2139	1.322
D - E	0.2685	-0.4993	1.036
D - F	0.01083	-0.7570	0.7787
D - G	-0.4502	-1.218	0.3177
D - H	0.1118	-0.6561	0.8797
D - I	-0.6408	-1.409	0.1271
D - J	0.1303	-0.6375	0.8982
D - K	0.4833	-0.2846	1.251
D - L	0.1582	-0.6097	0.9261
E - F	-0.2577	-1.026	0.5102
E - G	-0.7187	-1.487	0.04915
E - H	-0.1567	-0.9246	0.6112
E - I	-0.9093	-1.677	-0.1414
E - J	-0.1382	-0.9061	0.6297
E - K	0.2148	-0.5531	0.9826
E - L	-0.1104	-0.8782	0.6575
F - G	-0.4610	-1.229	0.3069
F - H	0.1010	-0.6669	0.8689
F - I	-0.6516	-1.419	0.1163
F - J	0.1195	-0.6484	0.8874
F - K	0.4725	-0.2954	1.240

F - L	0.1474	-0.6205	0.9152
G - H	0.5620	-0.2059	1.330
G - I	-0.1906	-0.9585	0.5773
G - J	0.5805	-0.1874	1.348
G - K	0.9335	0.1656	1.701
G - L	0.6084	-0.1595	1.376
H - I	-0.7526	-1.520	0.01529
H - J	0.01852	-0.7494	0.7864
H - K	0.3715	-0.3964	1.139
H - L	0.04637	-0.7215	0.8142
I - J	0.7711	0.003229	1.539
I - K	1.124	0.3562	1.892
I - L	0.7990	0.03108	1.567
J - K	0.3530	-0.4149	1.121
J - L	0.02785	-0.7400	0.7957
K - L	-0.3251	-1.093	0.4428

Assumption test: Are the standard deviations of the groups equal?
ANOVA assumes that the data are sampled from populations with identical SDs. This assumption is tested using the method of Bartlett.
Bartlett's test cannot be performed because at least one column's standard deviation is zero.
Assumption test: Are the data sampled from Gaussian distributions?
ANOVA assumes that the data are sampled from populations that follow Gaussian distributions. This assumption is tested using the method Kolmogorov and Smirnov:

Group	KS	P Value	Passed normality test?
A		>0.10	Yes
B	0.2162	>0.10	Yes
C	0.1809	>0.10	Yes
D	0.2139	>0.10	Yes
E	0.1770	>0.10	Yes
F	0.2145	>0.10	Yes
G	0.2134	>0.10	Yes
H	0.2393	>0.10	Yes
I	0.2432	>0.10	Yes
J	0.1688	>0.10	Yes
K	0.2848	>0.10	Yes
L	0.1711	>0.10	Yes

Intermediate calculations. ANOVA table

Source of	Degrees of	Sum of	Mean

variation	freedom	squares	square
Treatments (between columns)	11	6.985	0.6350
Residuals (within columns)	60	9.182	0.1530
Total	71	16.167	

F = 4.149 =(MStreatment/MSresidual)

Summary of Data

Group	Number of Points	Mean	Standard Deviation	Standard Error of Mean	Median
A	6	1.000	0.000	0.000	1.000
B	6	0.7433	0.3728	0.1522	0.8556
C	6	1.318	0.5249	0.2143	1.279
D	6	0.9222	0.2748	0.1122	0.8847
E	6	0.6537	0.4003	0.1634	0.6928
F	6	0.9114	0.4833	0.1973	0.9030
G	6	1.372	0.1650	0.06738	1.365
H	6	0.8104	0.4172	0.1703	0.9637
I	6	1.563	0.7076	0.2889	1.370
J	6	0.7919	0.3322	0.1356	0.7757
K	6	0.4389	0.2166	0.08841	0.4517
L	6	0.7640	0.3054	0.1247	0.7226

Group	Minimum	Maximum	95% Confidence Interval From	To
A	1.000	1.000	1.000	1.000
B	0.06100	1.058	0.3520	1.135
C	0.7462	2.021	0.7671	1.869
D	0.6343	1.390	0.6338	1.211
E	0.06790	1.078	0.2336	1.074
F	0.3740	1.562	0.4041	1.419
G	1.157	1.601	1.199	1.546
H	0.06790	1.170	0.3725	1.248
I	0.8607	2.646	0.8202	2.306
J	0.2698	1.230	0.4432	1.141
K	0.04970	0.6862	0.2116	0.6662
L	0.4631	1.237	0.4434	1.085

E.2.2 CBM15-488

One-way Analysis of Variance (ANOVA)

The P value is < 0.0001, considered extremely significant.

Variation among column means is significantly greater than expected by chance.

Tukey-Kramer Multiple Comparisons Test

If the value of q is greater than 4.808 then the P value is less than 0.05.

Comparison	Mean Difference	q	P value
A vs B	0.3280	5.340	* P<0.05
A vs C	0.2687	4.374	ns P>0.05
A vs D	-0.1320	2.149	ns P>0.05
A vs E	0.3938	6.410	** P<0.01
A vs F	0.4419	7.193	*** P<0.001
A vs G	-0.3220	5.242	* P<0.05
A vs H	0.1052	1.712	ns P>0.05
A vs I	0.04020	0.6544	ns P>0.05
A vs J	-0.1596	2.598	ns P>0.05
A vs K	0.4702	7.655	*** P<0.001
A vs L	0.4872	7.931	*** P<0.001
B vs C	-0.05932	0.9656	ns P>0.05
B vs D	-0.4601	7.489	*** P<0.001
B vs E	0.06573	1.070	ns P>0.05
B vs F	0.1138	1.853	ns P>0.05
B vs G	-0.6500	10.582	*** P<0.001
B vs H	-0.2228	3.628	ns P>0.05
B vs I	-0.2878	4.685	ns P>0.05
B vs J	-0.4876	7.938	*** P<0.001
B vs K	0.1422	2.315	ns P>0.05
B vs L	0.1592	2.592	ns P>0.05
C vs D	-0.4007	6.524	** P<0.01
C vs E	0.1251	2.036	ns P>0.05
C vs F	0.1732	2.819	ns P>0.05
C vs G	-0.5907	9.616	*** P<0.001
C vs H	-0.1635	2.662	ns P>0.05
C vs I	-0.2285	3.720	ns P>0.05
C vs J	-0.4283	6.973	*** P<0.001
C vs K	0.2015	3.281	ns P>0.05
C vs L	0.2185	3.557	ns P>0.05
D vs E	0.5258	8.559	*** P<0.001
D vs F	0.5739	9.342	*** P<0.001

D vs G	-0.1900	3.093	ns	P>0.05
D vs H	0.2372	3.862	ns	P>0.05
D vs I	0.1722	2.804	ns	P>0.05
D vs J	-0.02758	0.4490	ns	P>0.05
D vs K	0.6023	9.804	***	P<0.001
D vs L	0.6193	10.081	***	P<0.001
E vs F	0.04810	0.7830	ns	P>0.05
E vs G	-0.7158	11.652	***	P<0.001
E vs H	-0.2886	4.698	ns	P>0.05
E vs I	-0.3536	5.756	**	P<0.01
E vs J	-0.5534	9.008	***	P<0.001
E vs K	0.07647	1.245	ns	P>0.05
E vs L	0.09347	1.522	ns	P>0.05
F vs G	-0.7639	12.435	***	P<0.001
F vs H	-0.3367	5.481	*	P<0.05
F vs I	-0.4017	6.539	**	P<0.01
F vs J	-0.6015	9.791	***	P<0.001
F vs K	0.02837	0.4618	ns	P>0.05
F vs L	0.04537	0.7385	ns	P>0.05
G vs H	0.4272	6.954	***	P<0.001
G vs I	0.3622	5.896	**	P<0.01
G vs J	0.1624	2.643	ns	P>0.05
G vs K	0.7922	12.897	***	P<0.001
G vs L	0.8092	13.173	***	P<0.001
H vs I	-0.06498	1.058	ns	P>0.05
H vs J	-0.2648	4.311	ns	P>0.05
H vs K	0.3650	5.942	**	P<0.01
H vs L	0.3820	6.219	**	P<0.01
I vs J	-0.1998	3.253	ns	P>0.05
I vs K	0.4300	7.000	***	P<0.001
I vs L	0.4470	7.277	***	P<0.001
J vs K	0.6298	10.253	***	P<0.001
J vs L	0.6468	10.530	***	P<0.001
K vs L	0.01700	0.2767	ns	P>0.05

	Mean	95% Confidence Interval	
Difference	Difference	From	To
A - B	0.3280	0.03267	0.6234
A - C	0.2687	-0.02665	0.5640
A - D	-0.1320	-0.4274	0.1633
A - E	0.3938	0.09840	0.6891
A - F	0.4419	0.1465	0.7372
A - G	-0.3220	-0.6173	-0.02665
A - H	0.1052	-0.1902	0.4005

A - I	0.04020	-0.2551	0.3355
A - J	-0.1596	-0.4550	0.1357
A - K	0.4702	0.1749	0.7656
A - L	0.4872	0.1919	0.7826
B - C	-0.05932	-0.3547	0.2360
B - D	-0.4601	-0.7554	-0.1647
B - E	0.06573	-0.2296	0.3611
B - F	0.1138	-0.1815	0.4092
B - G	-0.6500	-0.9454	-0.3547
B - H	-0.2228	-0.5182	0.07251
B - I	-0.2878	-0.5832	0.007530
B - J	-0.4876	-0.7830	-0.1923
B - K	0.1422	-0.1531	0.4375
B - L	0.1592	-0.1361	0.4545
C - D	-0.4007	-0.6961	-0.1054
C - E	0.1251	-0.1703	0.4204
C - F	0.1732	-0.1222	0.4685
C - G	-0.5907	-0.8860	-0.2954
C - H	-0.1635	-0.4589	0.1318
C - I	-0.2285	-0.5238	0.06685
C - J	-0.4283	-0.7237	-0.1330
C - K	0.2015	-0.09383	0.4969
C - L	0.2185	-0.07683	0.5139
D - E	0.5258	0.2304	0.8211
D - F	0.5739	0.2785	0.8692
D - G	-0.1900	-0.4853	0.1054
D - H	0.2372	-0.05813	0.5326
D - I	0.1722	-0.1231	0.4676
D - J	-0.02758	-0.3229	0.2678
D - K	0.6023	0.3069	0.8976
D - L	0.6193	0.3239	0.9146
E - F	0.04810	-0.2472	0.3434
E - G	-0.7158	-1.011	-0.4204
E - H	-0.2886	-0.5839	0.006780
E - I	-0.3536	-0.6489	-0.05820
E - J	-0.5534	-0.8487	-0.2580
E - K	0.07647	-0.2189	0.3718
E - L	0.09347	-0.2019	0.3888
F - G	-0.7639	-1.059	-0.4685
F - H	-0.3367	-0.6320	-0.04132
F - I	-0.4017	-0.6970	-0.1063
F - J	-0.6015	-0.8968	-0.3061
F - K	0.02837	-0.2670	0.3237
F - L	0.04537	-0.2500	0.3407
G - H	0.4272	0.1318	0.7225

G - I	0.3622	0.06685	0.6575
G - J	0.1624	-0.1330	0.4577
G - K	0.7922	0.4969	1.088
G - L	0.8092	0.5139	1.105
H - I	-0.06498	-0.3603	0.2304
H - J	-0.2648	-0.5601	0.03055
H - K	0.3650	0.06969	0.6604
H - L	0.3820	0.08669	0.6774
I - J	-0.1998	-0.4952	0.09553
I - K	0.4300	0.1347	0.7254
I - L	0.4470	0.1517	0.7424
J - K	0.6298	0.3345	0.9252
J - L	0.6468	0.3515	0.9422
K - L	0.01700	-0.2783	0.3123

Assumption test: Are the standard deviations of the groups equal?
ANOVA assumes that the data are sampled from populations with identical SDs. This assumption is tested using the method of Bartlett.
Bartlett's test cannot be performed because at least one column's standard deviation is zero.
Assumption test: Are the data sampled from Gaussian distributions?
ANOVA assumes that the data are sampled from populations that follow Gaussian distributions. This assumption is tested using the method Kolmogorov and Smirnov:

Group	KS	P Value	Passed normality test?
A		>0.10	Yes
B	0.2299	>0.10	Yes
C	0.1725	>0.10	Yes
D	0.2749	>0.10	Yes
E	0.4269	0.0010	No
F	0.1331	>0.10	Yes
G	0.2167	>0.10	Yes
H	0.2029	>0.10	Yes
I	0.1957	>0.10	Yes
J	0.3237	0.0487	No
K	0.2503	>0.10	Yes
L	0.2932	>0.10	Yes

At least one column failed the normality test with P<0.05.
Consider using a nonparametric test or transforming the data (i.e. converting to logarithms or reciprocals).

Intermediate calculations. ANOVA table

Source of variation	Degrees of freedom	Sum of squares	Mean square

Treatments (between columns)	11	5.041	0.4583
Residuals (within columns)	60	1.358	0.02264
Total	71	6.399	

F = 20.241 =(MStreatment/MSresidual)

Summary of Data

Group	Number of Points	Mean	Standard Deviation	Standard Error of Mean	Median
A	6	1.000	0.000	0.000	1.000
B	6	0.6720	0.1093	0.04464	0.6493
C	6	0.7313	0.05862	0.02393	0.7304
D	6	1.132	0.2056	0.08394	1.217
E	6	0.6063	0.1439	0.05873	0.5507
F	6	0.5582	0.1671	0.06822	0.5570
G	6	1.322	0.2579	0.1053	1.265
H	6	0.8948	0.1413	0.05770	0.8920
I	6	0.9598	0.04804	0.01961	0.9509
J	6	1.160	0.1496	0.06105	1.209
K	6	0.5298	0.2023	0.08260	0.4955
L	6	0.5128	0.1153	0.04706	0.4643

Group	95% Confidence Interval Minimum	Maximum	From	To
A	1.000	1.000	1.000	1.000
B	0.5708	0.8654	0.5572	0.7868
C	0.6570	0.8024	0.6698	0.7928
D	0.8582	1.358	0.9162	1.348
E	0.5282	0.8981	0.4552	0.7573
F	0.3373	0.7857	0.3827	0.7336
G	1.079	1.791	1.051	1.593
H	0.7179	1.125	0.7465	1.043
I	0.8905	1.031	0.9094	1.010
J	0.8661	1.277	1.003	1.317
K	0.3256	0.8981	0.3174	0.7422
L	0.4257	0.7222	0.3918	0.6338

E.2.3 CBM2b1,2-488

One-way Analysis of Variance (ANOVA)

The P value is < 0.0001, considered extremely significant.

Variation among column means is significantly greater than expected by chance.

Tukey-Kramer Multiple Comparisons Test

If the value of q is greater than 4.808 then the P value is less than 0.05.

Comparison	Mean Difference	q	P value
A vs B	0.4015	3.301	ns P>0.05
A vs C	0.4032	3.316	ns P>0.05
A vs D	-0.2173	1.787	ns P>0.05
A vs E	0.5824	4.790	ns P>0.05
A vs F	0.4796	3.943	ns P>0.05
A vs G	-0.6319	5.196	* P<0.05
A vs H	0.2243	1.845	ns P>0.05
A vs I	0.2883	2.371	ns P>0.05
A vs J	-0.2824	2.323	ns P>0.05
A vs K	0.4415	3.630	ns P>0.05
A vs L	0.6376	5.243	* P<0.05
B vs C	0.001781	0.01465	ns P>0.05
B vs D	-0.6188	5.088	* P<0.05
B vs E	0.1810	1.488	ns P>0.05
B vs F	0.07810	0.6422	ns P>0.05
B vs G	-1.033	8.497	*** P<0.001
B vs H	-0.1771	1.457	ns P>0.05
B vs I	-0.1132	0.9306	ns P>0.05
B vs J	-0.6839	5.624	** P<0.01
B vs K	0.04002	0.3291	ns P>0.05
B vs L	0.2362	1.942	ns P>0.05
C vs D	-0.6206	5.103	* P<0.05
C vs E	0.1792	1.474	ns P>0.05
C vs F	0.07632	0.6276	ns P>0.05
C vs G	-1.035	8.512	*** P<0.001
C vs H	-0.1789	1.471	ns P>0.05
C vs I	-0.1150	0.9453	ns P>0.05
C vs J	-0.6857	5.638	** P<0.01
C vs K	0.03824	0.3144	ns P>0.05
C vs L	0.2344	1.928	ns P>0.05
D vs E	0.7998	6.577	** P<0.01
D vs F	0.6969	5.731	** P<0.01

D vs G	-0.4146	3.409	ns	P>0.05
D vs H	0.4417	3.632	ns	P>0.05
D vs I	0.5056	4.158	ns	P>0.05
D vs J	-0.06510	0.5353	ns	P>0.05
D vs K	0.6588	5.418	*	P<0.05
D vs L	0.8550	7.031	***	P<0.001
E vs F	-0.1029	0.8461	ns	P>0.05
E vs G	-1.214	9.986	***	P<0.001
E vs H	-0.3581	2.945	ns	P>0.05
E vs I	-0.2942	2.419	ns	P>0.05
E vs J	-0.8649	7.112	***	P<0.001
E vs K	-0.1410	1.159	ns	P>0.05
E vs L	0.05519	0.4539	ns	P>0.05
F vs G	-1.111	9.140	***	P<0.001
F vs H	-0.2552	2.099	ns	P>0.05
F vs I	-0.1913	1.573	ns	P>0.05
F vs J	-0.7620	6.266	**	P<0.01
F vs K	-0.03808	0.3132	ns	P>0.05
F vs L	0.1581	1.300	ns	P>0.05
G vs H	0.8562	6.204	**	P<0.01
G vs I	0.9202	7.567	***	P<0.001
G vs J	0.3495	2.874	ns	P>0.05
G vs K	1.073	8.827	***	P<0.001
G vs L	1.270	10.440	***	P<0.001
H vs I	0.06397	0.5260	ns	P>0.05
H vs J	-0.5068	4.167	ns	P>0.05
H vs K	0.2172	1.786	ns	P>0.05
H vs L	0.4133	3.399	ns	P>0.05
I vs J	-0.5707	4.693	ns	P>0.05
I vs K	0.1532	1.260	ns	P>0.05
I vs L	0.3494	2.873	ns	P>0.05
J vs K	0.7239	5.953	**	P<0.01
J vs L	0.9201	7.566	***	P<0.001
K vs L	0.1962	1.613	ns	P>0.05

	Mean	95% Confidence Interval	
Difference	Difference	From	To
A - B	0.4015	-0.1832	0.9861
A - C	0.4032	-0.1815	0.9879
A - D	-0.2173	-0.8020	0.3673
A - E	0.5824	-0.002247	1.167
A - F	0.4796	-0.1051	1.064
A - G	-0.6319	-1.217	-0.04722
A - H	0.2243	-0.3604	0.8090
A - I	0.2883	-0.2964	0.8730

A - J	-0.2824	-0.8671	0.3022
A - K	0.4415	-0.1432	1.026
A - L	0.6376	0.05295	1.222
B - C	0.001781	-0.5829	0.5865
B - D	-0.6188	-1.203	-0.03411
B - E	0.1810	-0.4037	0.7657
B - F	0.07810	-0.5066	0.6628
B - G	-1.033	-1.618	-0.4487
B - H	-0.1771	-0.7618	0.4076
B - I	-0.1132	-0.6979	0.4715
B - J	-0.6839	-1.269	-0.09920
B - K	0.04002	-0.5447	0.6247
B - L	0.2362	-0.3485	0.8209
C - D	-0.6206	-1.205	-0.03589
C - E	0.1792	-0.4055	0.7639
C - F	0.07632	-0.5084	0.6610
C - G	-1.035	-1.620	-0.4505
C - H	-0.1789	-0.7636	0.4058
C - I	-0.1150	-0.6996	0.4697
C - J	-0.6857	-1.270	-0.1010
C - K	0.03824	-0.5465	0.6229
C - L	0.2344	-0.3503	0.8191
D - E	0.7998	0.2151	1.384
D - F	0.6969	0.1122	1.282
D - G	-0.4146	-0.9993	0.1701
D - H	0.4417	-0.1430	1.026
D - I	0.5056	-0.07906	1.090
D - J	-0.06510	-0.6498	0.5196
D - K	0.6588	0.07413	1.244
D - L	0.8550	0.2703	1.440
E - F	-0.1029	-0.6876	0.4818
E - G	-1.214	-1.799	-0.6297
E - H	-0.3581	-0.9428	0.2266
E - I	-0.2942	-0.8789	0.2905
E - J	-0.8649	-1.450	-0.2802
E - K	-0.1410	-0.7257	0.4437
E - L	0.05519	-0.5295	0.6399
F - G	-1.111	-1.696	-0.5268
F - H	-0.2552	-0.8399	0.3295
F - I	-0.1913	-0.7760	0.3934
F - J	-0.7620	-1.347	-0.1773
F - K	-0.03808	-0.6228	0.5466
F - L	0.1581	-0.4266	0.7428
G - H	0.8562	0.2715	1.441
G - I	0.9202	0.3355	1.505

G - J	0.3495	-0.2352	0.9342
G - K	1.073	0.4887	1.658
G - L	1.270	0.6849	1.854
H - I	0.06397	-0.5207	0.6487
H - J	-0.5068	-1.091	0.07794
H - K	0.2172	-0.3675	0.8019
H - L	0.4133	-0.1714	0.9980
I - J	-0.5707	-1.155	0.01397
I - K	0.1532	-0.4315	0.7379
I - L	0.3494	-0.2353	0.9341
J - K	0.7239	0.1392	1.309
J - L	0.9201	0.3354	1.505
K - L	0.1962	-0.3885	0.7809

Assumption test: Are the standard deviations of the groups equal?
ANOVA assumes that the data are sampled from populations with identical SDs. This assumption is tested using the method of Bartlett.
Bartlett's test cannot be performed because at least one column's standard deviation is zero.
Assumption test: Are the data sampled from Gaussian distributions?
ANOVA assumes that the data are sampled from populations that follow Gaussian distributions. This assumption is tested using the method Kolmogorov and Smirnov:

Group	KS	P Value	Passed normality test?
A	>0.10		Yes
B	0.2729	>0.10	Yes
C	0.1551	>0.10	Yes
D	0.1811	>0.10	Yes
E	0.1448	>0.10	Yes
F	0.4596	0.0002	No
G	0.3587	0.0152	No
H	0.2837	>0.10	Yes
I	0.2988	>0.10	Yes
J	0.2029	>0.10	Yes
K	0.2638	>0.10	Yes
L	0.3018	0.0933	Yes

At least one column failed the normality test with P<0.05.
Consider using a nonparametric test or transforming the data (i.e. converting to logarithms or reciprocals).

Intermediate calculations. ANOVA table

Source of variation	Degrees of freedom	Sum of squares	Mean square
Treatments (between columns)	11	10.219	0.9290
Residuals (within columns)	60	5.324	0.08873
Total	71	15.542	

F = 10.469 =(MStreatment/MSresidual)

Summary of Data

Group	Number of Points	Mean	Standard Deviation	Standard Error of Mean	Median
A	6	1.000	0.000	0.000	1.000
B	6	0.5985	0.1960	0.08003	0.5183
C	6	0.5968	0.1647	0.06725	0.5778
D	6	1.217	0.1904	0.07772	1.261
E	6	0.4176	0.1568	0.06400	0.4281
F	6	0.5204	0.3662	0.1495	0.3778
G	6	1.632	0.4331	0.1768	1.555
H	6	0.7757	0.3751	0.1531	0.6205
I	6	0.7117	0.3191	0.1303	0.5982
J	6	1.282	0.2685	0.1096	1.279
K	6	0.5585	0.5221	0.2131	0.3813
L	6	0.3624	0.1717	0.07009	0.2900

			95% Confidence Interval	
Group	Minimum	Maximum	From	To
A	1.000	1.000	1.000	1.000
B	0.4410	0.9556	0.3928	0.8043
C	0.3830	0.8598	0.4239	0.7697
D	0.9250	1.442	1.018	1.417
E	0.2070	0.6381	0.2530	0.5821
F	0.3350	1.267	0.1360	0.9048
G	1.196	2.462	1.177	2.087
H	0.4417	1.366	0.3819	1.169
I	0.4270	1.219	0.3768	1.047
J	0.9186	1.716	1.001	1.564
K	0.1302	1.552	0.01056	1.106
L	0.1960	0.6187	0.1822	0.5426

Appendix F

LAS / gpl	$CaCl_2 2H_2O$ /mMol	NaCl / M
0	0	0.005
0	0	0.05
0	0.2	0.005
0	0.2	0.05
0.5	**0**	**0.005**
0.5	0	0.05
0.5	**0.2**	**0.005**
0.5	0.2	0.05
1.0	**0**	**0.005**
1.0	0	0.05
1.0	**0.2**	**0.005**
1.0	0.2	0.05

Table F. Full matrix of 100 % SLES system and conditions to use in grass binding assay. The priority conditions are known in red and carried out in prior, known as optimal buffer system in the presence and absence of 0.2 mM $CaCl_2$ as a function of electrolyte concentration (0.005 M and 0.05 M NaCl). Condition highlighted in yellow has been used in grass binding assay at initial stage to screen the strongest grass binding CBMs.

Appendix G

Raw data and experiment design of detergent assays

S1	Test 1 using current blend **surfactant formulation A**	S2	Test 2 using High SLES blend **surfactant formulation B**
Formulation	Enzyme composition	Formulation	Enzyme composition
F1	Control plus enzymes Lip Pro Amy		
F2	Control minus enzymes		
F3	Xylanase 10 no CBM	F9	Xylanase 10 no CBM
F4	Xylanase 10 CBM 2b1,2	F10	Xylanase 10 CBM2b1,2
F5	Xylanase 10 CBM15	F11	Xylanase 10 CBM15
F6	Xylanase 11 no CBM	F12	Xylanase 11 no CBM
F7	Xylanase 11 CBM 2b1,2	F13	Xylanase 11 CBM 2b1,2
F8	Xylanase 11 CBM 15	F14	Xylanase 11 CBM 15

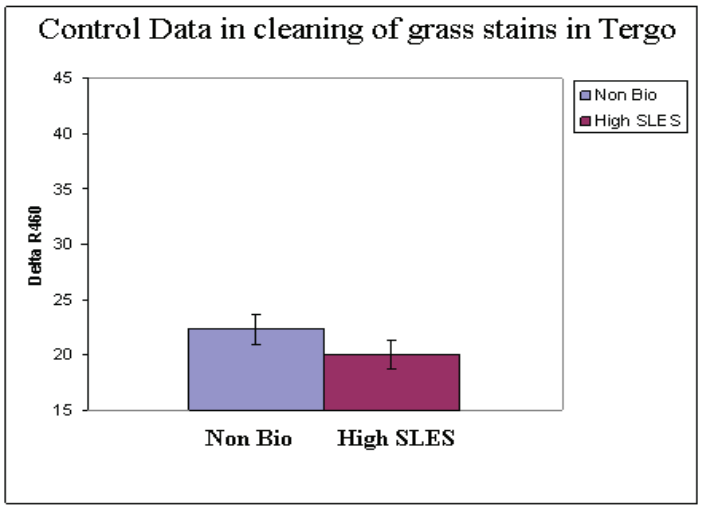
Tergo number	Formualtion	Before wash (BW) R460	After wash (AW) R460	Delta R460 =AW-BW	Xylanases with or without CBM	If With Protease (noted)	Surfactant system (S1 or S2)
White tile	None	97.53	97.61				
P6	F14	9.23	38.76	29.53	1115		S2
P6	F14	10.86	39.52	28.66	1115		S2
P5	F13	9.78	35.73	25.95	112B		S2
P5	F13	10.35	38.01	27.66	112B		S2
P4	F12	8.66	39.07	30.41	11		S2
P4	F12	9.3	34.74	25.44	11		S2
P3	F11	8.96	37.41	28.45	1015		S2
P3	F11	10.07	38.62	28.55	1015		S2
P2	F10	9.59	46.29	36.7	102B		S2
P2	F10	8.54	35.39	26.85	102B		S2
P1	F9	9.47	45.18	35.71	10		S2
P1	F9	10.21	38.15	27.94	10		S2
P6	F8	10.26	38.09	27.83	1115		S1
P6	F8	10.17	32.14	21.97	1115		S1
P5	F7	10.53	35.74	25.21	112B		S1
P5	F7	9.11	33.48	24.37	112B		S1
P4	F6	11.37	37.43	26.06	11		S1
P4	F6	9.25	34.98	25.73	11		S1
P3	F5	9.01	34.31	25.3	1015		S1
P3	F5	10.19	32.95	22.76	1015		S1
P2	F4	10.61	37.22	26.61	102B		S1
P2	F4	9.54	33.81	24.27	102B		S1
P1	F3	9.92	35.74	25.82	10		S1
P1	F3	10.09	30.52	20.43	10		S1
P6	F2	9.68	32.43	22.75	0		S1
P6	F2	10.89	35.76	24.87	0		S1
P5	F1	9.53	44.13	34.6	0	PROTEASE	S1
P5	F1	9.69	43.47	33.78	0	PROTEASE	S1
P4	F14	9.38	32.01	22.63	1115		S2
P4	F14	9.36	34.41	25.05	1115		S2
P3	F13	12.62	38.11	25.49	112B		S2
P3	F13	9.44	32.45	23.01	112B		S2
P2	F12	10.49	39.51	29.02	11		S2
P2	F12	10.47	37.82	27.35	11		S2
P1	F11	9.79	36.41	26.62	1015		S2
P1	F11	8.87	39.95	31.08	1015		S2

P6	F10	9.95	45.95	36	102B		S2
P6	F10	9.76	38.5	28.74	102B		S2
P5	F9	9.84	37.66	27.82	10		S2
P5	F9	9.53	40.6	31.07	10		S2
P4	F8	9.99	32.88	22.89	1115		S1
P4	F8	9.97	30.08	20.11	1115		S1
P3	F7	9.11	33.66	24.55	112B		S1
P3	F7	8.78	35.46	26.68	112B		S1
P2	F6	9.25	35.02	25.77	11		S1
P2	F6	12.93	40.21	27.28	11		S1
P1	F5	12.48	37.91	25.43	1015		S1
P1	F5	12.84	36.98	24.14	1015		S1
P6	F4	12.79	35.88	23.09	102B		S1
P6	F4	13.05	38.94	25.89	102B		S1
P5	F3	10.85	32.27	21.42	10		S1
P5	F3	11.09	30.55	19.46	10		S1
P4	F2	12.77	34.29	21.52	0		S1
P4	F2	11.38	35.8	24.42	0		S1
P3	F1	13.79	57.62	43.83	0	PROTEASE	S1
P3	F1	12.08	50.53	38.45	0	PROTEASE	S1
P2	F14	11.95	38.23	26.28	1115		S2
P2	F14	12.19	38.51	26.32	1115		S2
P1	F13	12.36	38.62	26.26	112B		S2
P1	F13	12.51	38.9	26.39	112B		S2
P6	F12	16.19	36.69	20.5	11		S2
P6	F12	12.06	32.61	20.55	11		S2
P5	F11	11.41	33.13	21.72	1015		S2
P5	F11	11.31	31.1	19.79	1015		S2
P4	F10	12.48	34.29	21.81	102B		S2
P4	F10	13.39	38.14	24.75	102B		S2
P3	F9	11.84	31.59	19.75	10		S2
P3	F9	12.17	35.93	23.76	10		S2
P2	F8	12.54	30.16	17.62	1115		S1
P2	F8	12.31	31.25	18.94	1115		S1
P1	F7	12.21	37.36	25.15	112B		S1
P1	F7	12.59	34.81	22.22	112B		S1
P6	F6	12.06	31.81	19.75	11		S1
P6	F6	12.17	35.76	23.59	11		S1
P5	F5	12.82	39.17	26.35	1015		S1
P5	F5	12.26	33.58	21.32	1015		S1
P4	F4	12.27	35.61	23.34	102B		S1
P4	F4	11.03	35.22	24.19	102B		S1
P3	F3	11.57	35.11	23.54	10		S1

P3	F3	12.68	41.06	28.38	10		S1
P2	F2	11.35	35.06	23.71	0		S1
P2	F2	11.64	33.25	21.61	0		S1
P1	F1	11.97	57.24	45.27	0	PROTEASE	S1
P1	F1	12.22	60.02	47.8	0	PROTEASE	S1

Table G.1 Raw data obtained from final Tergo wash (grass removal assay). The laundry compositions and surfactant system for each test in every tergo were listed in the table. The R460 before and after wash were recorded in the table for each piece of cotton coded as a number. Wavelength 460 nm is the wavelength at which the grass stain arrayed on cotton give the greatest response therefore it is used in the protocol for all grass removal assay in Laundry research department over years.

A



B

Formulation	Chart name	Average delta R460	Standard deviation
F2	**Non Bio**	**22.3**	**1.42**
F3	**High SLES**	**20.0**	**1.36**

Table G.2 Control of Non Bio and High SLES blend with surfactant system and no enzyme in Tergo wash for grass stain cleaning (provided by Dr. Paul Stevenson, Unilever Laundry Department).

G.3 Stats analysis of cleaning performance of CBM-xylanases in grass stain removal assays

G.3.1 High SLES blend

One-way Analysis of Variance (ANOVA)

The P value is 0.0427, considered significant.

Variation among column means is significantly greater than expected by chance.

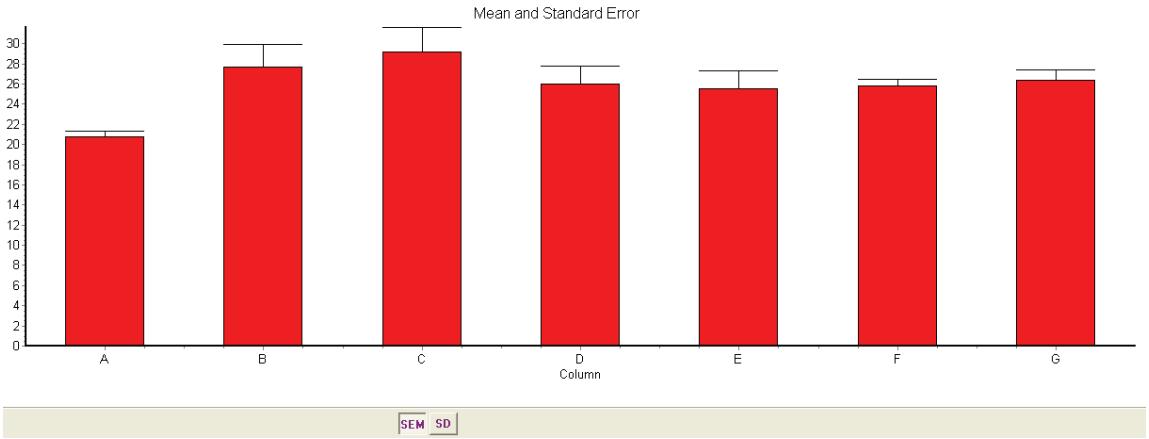
Tukey-Kramer Multiple Comparisons Test

If the value of q is greater than 4.427 then the P value is less than 0.05.

Comparison	Mean Difference	q		P value
High SLES ctrl vs Xyn10	-6.915	4.181	ns	P>0.05
High SLES ctrl vs 2B-Xyn10	-8.382	5.068	*	P<0.05
High SLES ctrl vs 15-Xyn10	-5.275	3.190	ns	P>0.05
High SLES ctrl vs Xyn11	-4.785	2.893	ns	P>0.05
High SLES ctrl vs 2B-Xyn11	-5.033	3.043	ns	P>0.05
High SLES ctrl vs 15-Xyn11	-5.652	3.417	ns	P>0.05
Xyn10 vs 2B-Xyn10	-1.467	0.8868	ns	P>0.05
Xyn10 vs 15-Xyn10	1.640	0.9916	ns	P>0.05
Xyn10 vs Xyn11	2.130	1.288	ns	P>0.05
Xyn10 vs 2B-Xyn11	1.882	1.138	ns	P>0.05
Xyn10 vs 15-Xyn11	1.263	0.7639	ns	P>0.05
2B-Xyn10 vs 15-Xyn10	3.107	1.878	ns	P>0.05
2B-Xyn10 vs Xyn11	3.597	2.175	ns	P>0.05
2B-Xyn10 vs 2B-Xyn11	3.348	2.025	ns	P>0.05
2B-Xyn10 vs 15-Xyn11	2.730	1.651	ns	P>0.05
15-Xyn10 vs Xyn11	0.4900	0.2963	ns	P>0.05
15-Xyn10 vs 2B-Xyn11	0.2417	0.1461	ns	P>0.05
15-Xyn10 vs 15-Xyn11	-0.3767	0.2278	ns	P>0.05
Xyn11 vs 2B-Xyn11	-0.2483	0.1502	ns	P>0.05
Xyn11 vs 15-Xyn11	-0.8667	0.5240	ns	P>0.05
2B-Xyn11 vs 15-Xyn11	-0.6183	0.3739	ns	P>0.05

Difference	Mean Difference	95% Confidence Interval From	To
High SLES ctrl - Xyn10	-6.915	-14.236	0.4057

High SLES ctrl - 2B-Xyn10	-8.382	-15.702	-1.061
High SLES ctrl - 15-Xyn10	-5.275	-12.596	2.046
High SLES ctrl - Xyn11	-4.785	-12.106	2.536
High SLES ctrl - 2B-Xyn11	-5.033	-12.354	2.287
High SLES ctrl - 15-Xyn11	-5.652	-12.972	1.669
Xyn10 - 2B-Xyn10	-1.467	-8.787	5.854
Xyn10 - 15-Xyn10	1.640	-5.681	8.961
Xyn10 - Xyn11	2.130	-5.191	9.451
Xyn10 - 2B-Xyn11	1.882	-5.439	9.202
Xyn10 - 15-Xyn11	1.263	-6.057	8.584
2B-Xyn10 - 15-Xyn10	3.107	-4.214	10.427
2B-Xyn10 - Xyn11	3.597	-3.724	10.917
2B-Xyn10 - 2B-Xyn11	3.348	-3.972	10.669
2B-Xyn10 - 15-Xyn11	2.730	-4.591	10.051
15-Xyn10 - Xyn11	0.4900	-6.831	7.811
15-Xyn10 - 2B-Xyn11	0.2417	-7.079	7.562
15-Xyn10 - 15-Xyn11	-0.3767	-7.697	6.944
Xyn11 - 2B-Xyn11	-0.2483	-7.569	7.072
Xyn11 - 15-Xyn11	-0.8667	-8.187	6.454
2B-Xyn11 - 15-Xyn11	-0.6183	-7.939	6.702



G.3.2 Current blend

One-way Analysis of Variance (ANOVA)

The P value is 0.2526, considered not significant.

Variation among column means is not significantly greater than expected by chance.

Tukey-Kramer Multiple Comparisons Test

If the value of q is greater than 4.427 then the P value is less than 0.05.

Comparison	Mean Difference	q	P value
current ctrl vs Xyn10	-0.02833	0.02839	ns P>0.05
current ctrl vs 2B-Xyn10	-1.418	1.421	ns P>0.05
current ctrl vs 15-Xyn10	-1.115	1.117	ns P>0.05
current ctrl vs Xyn11	-1.550	1.553	ns P>0.05
current ctrl vs 2B-Xyn11	-1.550	1.553	ns P>0.05
current ctrl vs 15-Xyn11	1.587	1.590	ns P>0.05
Xyn10 vs 2B-Xyn10	-1.390	1.393	ns P>0.05
Xyn10 vs 15-Xyn10	-1.087	1.089	ns P>0.05
Xyn10 vs Xyn11	-1.522	1.525	ns P>0.05
Xyn10 vs 2B-Xyn11	-1.522	1.525	ns P>0.05
Xyn10 vs 15-Xyn11	1.615	1.618	ns P>0.05
2B-Xyn10 vs 15-Xyn10	0.3033	0.3040	ns P>0.05
2B-Xyn10 vs Xyn11	-0.1317	0.1319	ns P>0.05
2B-Xyn10 vs 2B-Xyn11	-0.1317	0.1319	ns P>0.05
2B-Xyn10 vs 15-Xyn11	3.005	3.011	ns P>0.05
15-Xyn10 vs Xyn11	-0.4350	0.4359	ns P>0.05
15-Xyn10 vs 2B-Xyn11	-0.4350	0.4359	ns P>0.05
15-Xyn10 vs 15-Xyn11	2.702	2.707	ns P>0.05
Xyn11 vs 2B-Xyn11	-3.469E-18	3.477E-18	ns P>0.05
Xyn11 vs 15-Xyn11	3.137	3.143	ns P>0.05
2B-Xyn11 vs 15-Xyn11	3.137	3.143	ns P>0.05

Difference	Mean Difference	95% Confidence Interval From	To
current ctrl - Xyn10	-0.02833	-4.446	4.389
current ctrl - 2B-Xyn10	-1.418	-5.836	2.999
current ctrl - 15-Xyn10	-1.115	-5.532	3.302
current ctrl - Xyn11	-1.550	-5.967	2.867

current ctrl - 2B-Xyn11	-1.550	-5.967	2.867
current ctrl - 15-Xyn11	1.587	-2.831	6.004
Xyn10 - 2B-Xyn10	-1.390	-5.807	3.027
Xyn10 - 15-Xyn10	-1.087	-5.504	3.331
Xyn10 - Xyn11	-1.522	-5.939	2.896
Xyn10 - 2B-Xyn11	-1.522	-5.939	2.896
Xyn10 - 15-Xyn11	1.615	-2.802	6.032
2B-Xyn10 - 15-Xyn10	0.3033	-4.114	4.721
2B-Xyn10 - Xyn11	-0.1317	-4.549	4.286
2B-Xyn10 - 2B-Xyn11	-0.1317	-4.549	4.286
2B-Xyn10 - 15-Xyn11	3.005	-1.412	7.422
15-Xyn10 - Xyn11	-0.4350	-4.852	3.982
15-Xyn10 - 2B-Xyn11	-0.4350	-4.852	3.982
15-Xyn10 - 15-Xyn11	2.702	-1.716	7.119
Xyn11 - 2B-Xyn11	-3.469E-18	-4.417	4.417
Xyn11 - 15-Xyn11	3.137	-1.281	7.554
2B-Xyn11 - 15-Xyn11	3.137	-1.281	7.554

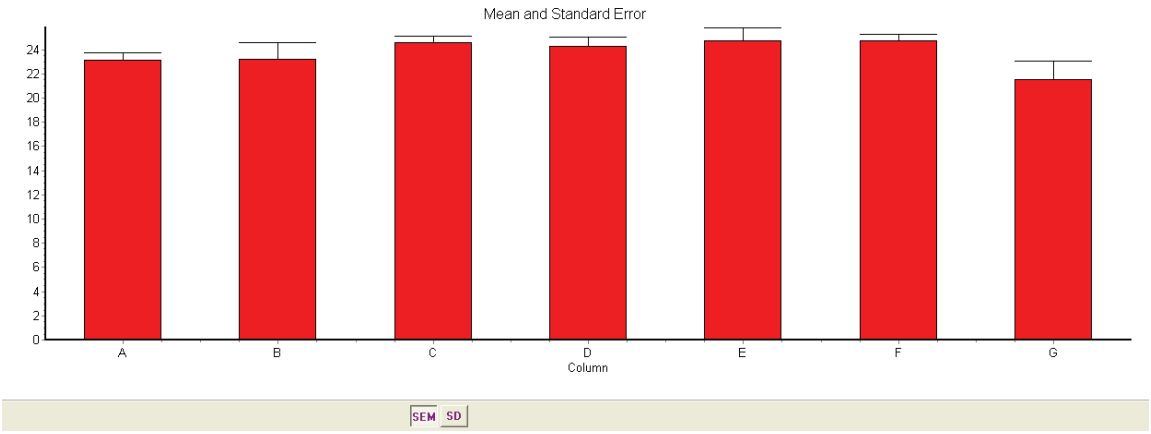


Fig 3.5 SDS-PAGE of CBM2b1,2 incubating with insoluble oat spelt xylan and grass
Purified CBM2b1,2 at 40 μM was mixed with 2.5 mg of the following insoluble ligands: a, insoluble oat spelt xylan; b, grass; c, grass treated with acetone; d, grass treated with SLES; e, BSA with insoluble oat spelt xylan (as a negative control); f, BSA with boiled grass, acetone and SLES (as negative controls). The protein was incubated with the ligand for 1 h on ice. The grass was ground in liquid nitrogen and resuspended in 10 mM Tris/HCl buffer, pH 8.0. Samples in all the fractions were analysed by SDS-PAGE, using 12.5 % (w/v) gels. Lane S, U, W and B contained the purified protein as starting materials, the supernatant after incubating with the ligand (unbound protein), the wash and the protein released from the polysaccharide by SDS, respectively. Sigma low and high molecular weight markers were on lanes M_L and M_H.

Fig 4.4 Absorbance at 420 nm monitored during the catalysis of PNP-X_2 by Xyn10A and CBM15-Xyn11A
The reactions were carried out at 37 °C, as described in Section 2.3.8. PNP-X_2 was used at 0.5 mM final concentration. Xyn10A at final concentration of 0.1 μM and 0.02 μM were used; CBM15-Xyn11A at final concentration of 35 μM, 3.5 μM and 0.35 μM were used.

Fig 4.10 Stability assay of CBM-Xylanases in the presence of savinase
CBM2b1,2 or CBM15 fused to Xyn10A and Xyn11A at 0.5 mg/ml were incubated with savinase (0.01 mg/ml) in either 50 mM Tris/HCl buffer pH 8.0 (Panels A, C, E and G) or surfactant buffer containing 1 g/L SLES; 0.2 mM $CaCl_2$ and 5 mM NaCl (Panels B, D, F and H). Savinase was inactivated by its inhibitor PMSF at 1 mM final concentration at 0, 10, 20 and 30 min time intervals. The degradation of proteins was assessed by 12.5 % SDS-PAGE.

Fig 4.11 Residual activity of CBM-Xylanases against oat spelt xylan in the presence of savinase
The reactions were carried out in either 50 mM Tris/HCl buffer pH 8.0, or surfactant buffer containing 1 g/L SLES; 0.2 mM $CaCl_2$ and 5 mM NaCl at 30 °C. Savinase was incubated at 0.01 mg/ml final concentration with the following CBM-Xylanases: Panels A and B, CBM15-Xyn10A; Panels C and D, CBM2b1,2-Xyn10A; Panels E and F, CBM15-Xyn11A; Panels G and H, CBM2b1,2-Xyn11A. The concentration of reducing sugar at different time intervals were monitored, as described in Section 2.3.8.

Fig 5.6 Representative ITC data of *v*CBM60 wild type to oligosaccharides
ITC was carried out in 50 mM Na Hepes pH 7.5 containing 5 mM $CaCl_2$. Typical ITC data for wild-type *v*CBM60 titrated with 5 mM xylo- and cello- oligosaccharides, from xylobiose to xylohexaose and from cellobiose to cellohexaose.

Fig 5.17 SDS-PAGE of CBM35 wild type (Panel A) and ten mutants (Panels B-K) purified by IMAC Extraction and purification of *Ct*CBM35 wild type and ten mutants was carried out as described in Section 2.1.24. The panels refer to the purified of the following proteins: a, CBM35-Gal wild type; b, CBM35 mutant Q27A; c, CBM35 mutant Y37A; d, CBM35 mutant Y37F; e, CBM35 mutant W40A; f, CBM35 mutant R86A; g, CBM35 mutant P96A; h, CBM35 mutant P96G; i, CBM35 mutant W108A; j, CBM35 mutant Y137A; k, CBM35 mutant N140A. A 5 µl aliquot of cell free extract and flow-through (Lanes C and F respectively) was loaded on a gel, then 10 µl of wash with 20 mM Tris/HCl pH 8.0, containing 300 mM NaCl (W) and 10 µl of every elution; with 10 mM imidazole (10) and 100 mM imidazole (100). Lanes M_L contained Sigma low molecular weight standards. Analysis was performed by SDS-PAGE, using a 12.5 % (w/v) polyacrylamide gel.

Fig 5.18 Representative ITC data of *Ct*CBM35 wild type (Panel A) and ten mutants (Panels B-K) to carob galactomannan (low viscosity). ITC was carried out in 50 mM Na Hepes pH 7.5 containing 5 mM $CaCl_2$. Typical ITC data for wild-type (A), ten mutants (B-K) of CBM35 titrated with 5 mg/ml carob galactomannan (low viscosity) with continuous stirring. The top half of each panel shows the raw ITC heats; the bottom half of each panel shows the integrated peak areas fitted using a one-binding site model with MicroCal Origin.