

**Functional and structural characterisation of a
multimodular xylanase from the hindgut
metagenome of the snouted harvester termite**

Robert Friedrich Krüger

Submitted in partial fulfilment of the requirements for the degree

Master of Science Biochemistry

in the

Department of Biochemistry

at the

University of Pretoria

Supervisor:

Prof. Wolf-Dieter Schubert

October 2017

KEYWORDS

Xylanase, temperature/pH profiles, functional/structural characterisation, carbohydrate-binding modules

ABSTRACT

The degradation of hemicellulose is a vital step in the efficient utilisation of plant biomass, an abundant source of carbohydrates in nature. Xylan, the most prevalent hemicellulose, is digested by xylanases and much work has gone into improving the efficiency and stability of these enzymes. Interest has recently shifted to non-catalytic, accessory domains found in many xylanases, including carbohydrate binding modules (CBMs). CBMs may enhance enzyme properties, such as thermostability and efficiency. This effect has, however, not yet been fully explained. In this study, a novel multimodular xylanase isolated from the hindgut metagenome of the snouted harvester termite (*Trinervitermes trinervoides*) was characterised and the properties of its two identical CBMs analysed. Xylanase variants were produced in *E. coli* by deleting none, one or both CBMs. The variants were functionally characterised using a modified DNS assay in a thermocycler and structurally modelled using computational techniques. Deleting one CBM shifted the pH profile, slightly increasing activity at lower pH whereas removing both CBMs decreased activity at 60°C from 90% (SD = 1.3) to 56% (SD = 2.9). Removal of the CBMs did not statistically significantly affect the Michaelis constants (K_M), but did reduce the turnover number (k_{cat}). Structure modelling failed to accurately predict the structure of the complete enzyme and in particular any interactions between domains. The origin of the stabilising effect of the second CBM and hence the evolutionary benefit of two CBMs remains unresolved.

Date

DECLARATION

I declare that *Functional and structural characterisation of a multimodular xylanase from the hindgut metagenome of the snouted harvester termite* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full Name

Signature

Date

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my supervisor, Professor Wolf-Dieter Schubert, who has continuously supported me during the course of my master's degree and taught me countless valuable lessons. I would also like to thank my mentor during my internship at the Council for Scientific and Industrial Research, Doctor Konanani J. Rashamuse for his guidance and allowing me to perform my research under his supervision.

I would like to acknowledge the Council for Scientific and Industrial Research and National Research Foundation, the Department of Science and Technology and the University of Pretoria for their financial support.

Finally, I would like to thank my friends and family for constantly encouraging me and being there for me, especially my colleagues at the Council for Scientific and Industrial Research and the University of Pretoria, whom I could always depend on when I needed help.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS.....	iii
LIST OF FIGURES.....	iv
LIST OF TABLES.....	v
CHAPTER 1 - Introduction	1
1.1 Lignocellulose Biomass.....	1
1.2 Glycoside hydrolases and carbohydrate binding module	4
1.3 Industrial applications of Xylanases.....	5
1.4 Xylanase Catalytic Mechanism	7
1.5 Aim and objectives	9
CHAPTER 2 – Materials and Methods.....	10
2.1 Materials	10
2.1.1 Standard protocols for preparation of materials	13
2.2 Methods	15
2.2.1 Sequencing and gene annotation	15
2.2.2 Development of expression constructs	15
2.2.3 Transformation of chemically competent cells	16
2.2.4 Recombinant production and purification	17
2.2.5 Functional characterisation.....	18
2.2.6 Structural characterisation	20
CHAPTER 3 – Results.....	22

3.1	Sequencing and gene annotation	22
3.2	Development of expression constructs	23
3.3	Recombinant protein production and purification.....	25
3.4	Functional characterisation.....	27
3.5	Structural characterisation	29
	CHAPTER 4 – Discussion.....	34
4.1	Sequencing and gene annotation	34
4.2	Recombinant production and purification	35
4.3	Functional characterisation.....	36
4.4	Structural characterisation	37
	CHAPTER 5 – Conclusion	40
	References.....	41

LIST OF ABBREVIATIONS

Abbreviation	Meaning
A ₅₄₀	Absorbance at 540 nm
aa	Amino acids
AIDA	<i>Ab initio</i> domain assembly
AMHB	Acetate, MES, HEPES and bicine buffer
CBM	Carbohydrate binding module
CD	Catalytic domain
CE4	Family 4 carboxyesterase
ddH ₂ O	Double-distilled water
DNS	3,5-Dinitrosalicylic acid
DTT	Dithiothreitol
GH	Glycoside hydrolase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
I-TASSER	Iterative threading assembly refinement
IPTG	Isopropyl- β -D-thiogalactoside
k _{cat}	Turnover number
K _M	Michaelis constant
LA	Lysogeny agar
LB	Lysogeny broth
LEW	Lysis, Elution and Wash buffer
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
OD ₆₀₀	Optical density at 600 nm
PCR	Polymerase chain reaction
PDB	Protein Data Bank
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate poly-acrylamide gel electrophoresis
TBE	Tris/Borate/EDTA buffer
TEMED	Tetramethylethylenediamine

LIST OF FIGURES

- Figure 1: Summary of the major components of plant cell walls in the multi-coat model.
- Figure 2: Cartoon and surface representation of the only published crystal structure of CBM36.
- Figure 3: The catalytic mechanism of glycosidic bond hydrolysis as proposed by Koshland (1953).
- Figure 4: Agarose gel of PCR products after electrophoresis.
- Figure 5: Analysis of XhoI/NdeI digested plasmids by agarose gel electrophoresis after expression cell transformation.
- Figure 6: SDS-PAGE analysis of small-scale GH11 protein production in the different expression strains.
- Figure 7: Chromatograms and SDS-PAGE analysis of the IMAC purification of GH11-CD, GH11-1CBM and GH11-2CBM.
- Figure 8: Graphs summarising the results of DNS assays to determine temperature profiles, pH profiles and thermostabilities of the variants.
- Figure 9: Michaelis-Menten graphs of initial rate versus beechwood xylan concentration.
- Figure 10: Micrographs of protein crystals in a 2 μ l hanging drop at 150 x magnification.
- Figure 11: Cartoon representations of derived GH11 domain folds.
- Figure 12: A hexa-xylose substrate modelled into the active site of the CD using Auto-Dock Vina.
- Figure 13: Cartoon representations of predicted structures of GH11-2CBM.

LIST OF TABLES

Table 1: Schematic overview of the xylanase variants derived from full-length GH11.

Table 2: List of *E. coli* cells used for plasmid propagation and protein production

Table 3: PCR program for gene amplification of the truncation variants.

Table 4: Primers for gene truncation and sequencing by PCR.

Table 5: Description of conserved domains database search results to design the protein constructs.

Table 6: Enzyme kinetics parameters for GH11 variants

CHAPTER 1 - INTRODUCTION

The increased application of enzymes to industrial processes as biological catalysts is a natural consequence of humanity's need to address ever more complex global issues. Advances in modern biotechnology have allowed pure, well-characterised enzymes to find a use in virtually every industry.

Natural enzymes often contain multiple domains each with their own function. While one or more catalytic domains perform the main function of the enzyme, accessory domains potentially broaden the range of substrates or increase stability under varying conditions. The purpose of accessory domains in enzymes is not always clear and similar domains may serve unrelated functions in different enzymes. Glycoside hydrolases (GHs) are enzymes that catalyse the hydrolysis of the glycosidic bond between sugars in carbohydrates. They offer an interesting model to study multimodularity as many contain accessory domains. GHs are of biotechnological interest to convert biomass to bioethanol to provide a renewable fuel source. To develop the full potential of these industrially relevant enzymes not only the GHs but also their accessory domains need to be studied.

1.1 LIGNOCELLULOSE BIOMASS

Plant dry matter (biomass) is overwhelmingly comprised of lignin, closely associated with cellulose and hemicellulose carbohydrates. The combination is referred to as lignocellulose, and derives mainly from secondary plant cell walls. The exact composition of lignocellulose differs between plant species, but the carbohydrate content ranges from 50 to 80% of the total biomass (Lee et al., 2007, Limayem and Ricke, 2012).

Cellulose is the most valuable component of lignocellulose in particular as a feedstock for the paper and for the biofuel industry. It is a linear polysaccharide of β -1,4-linked D-

glucose monomers. The degradability of cellulose correlates with the surface area accessible to enzymes (Zhang and Lynd, 2004, Rollin et al., 2011). Plants have evolved many mechanisms to deter structural cellulose digestion.

Hemicellulose is an amorphous heterogeneous polymer. It consists of highly branched chains of pentoses (xylose, arabinose), hexoses (mannose, glucose, galactose, fucose, rhamnose), and sugar acids (D-glucuronic, D-galacturonic and methylgalacturonic acids). Hemicellulose composition and branch frequency varies with plant species and growth stages (Sorek et al., 2014). Its fibres are shorter than those of cellulose, with ~70-200 sugar monomers (Zhao et al., 2012a). Hemicellulose backbones mainly consist of xylans and xyloglucans and their digestion is critical in the degradation of lignocellulose (Scheller and Ulvskov, 2010).

Although hemicellulose is more easily digested than cellulose due to its amorphous nature, it nevertheless slows the degradation of lignocellulose biomass by limiting physical access

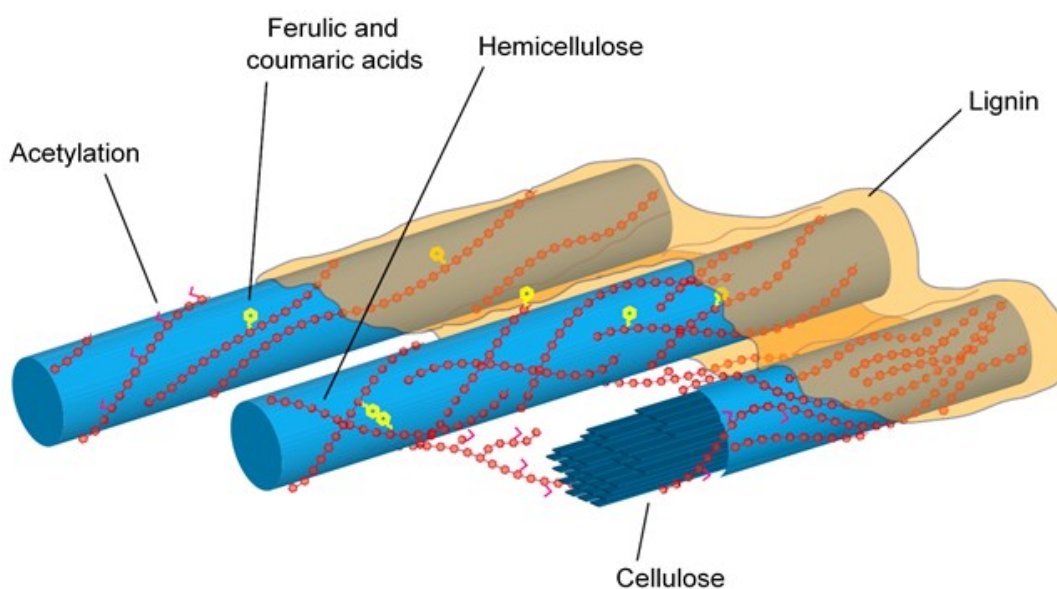


Figure 1: Summary of the major components of plant cell walls in the multi-coat model. The phenolics, ferulic and coumaric acid, are especially common in grass cell wall xylans and also interact strongly with lignin (Sorek et al., 2014).

to cellulose (Zhao et al., 2012b) through ester bonds to lignin, a rigid and complex aromatic polymer (Fig. 1) (Chabannes et al., 2001) or to phenolic acids as in lignin monomers (Crepin et al., 2004). Hemicellulose disrupts the crystalline structure of cellulose by intermittently being trapped in the latter during synthesis and subsequent crystallisation. It binds microfibrils through extensive hydrogen bonding networks tethering plant cell wall components and maintaining its strength (Scheller and Ulvskov, 2010).

Many attempts have been undertaken to optimise biofuel production from lignocellulose (see Zhao et al. (2012b) for references). Its degradation is slowed by the inherent cell wall structure retained in raw biomass where cellulose is imbedded in a dense, hydrophobic aggregate of lignin and hemicellulose with access for enzymes being limited (Mosier et al., 2005). Lignocellulose thus needs to be pretreated to increase the accessibility to cellulose (Limayem and Ricke, 2012) through physical, chemical and biological modification – the most expensive step in lignocellulose conversion (Mosier et al., 2005). Pretreatment aims to decrease cellulose crystallinity, acetylation and polymerisation, displace lignin and hemicellulose, increase biomass porosity, and decrease particle size (Gong et al., 1999, Sun and Cheng, 2002, Mosier et al., 2005). Enzymatic removal of xylan is effective, but is not yet well developed (Scharf and Boucias, 2010). Termites and their microbiota are expert lignocellulose digesters. Investigating their degradation mechanisms may reveal new ways of approaching this problem.

The higher termite *Trinervitermise trinervoides* (Termitidae: Nasutitermitinae), commonly referred to as the snouted harvester termite, typically occurs in the arid and semi-arid grasslands of southern Africa (Coaton, 1948, Adam et al., 2008). It is a foresting specialist that feeds on most available grass species, including sugarcane, but prefers green grass litter, thinner and softer leaves and tends to avoid the stem (Adam et al., 2005, Adam et al.,

2008, Sanyika et al., 2012). Its ability to cut down standing grass makes it a serious rangeland pest (Coaton, 1948, Hartwig, 1955). The hindgut microbial diversity of *T. trinervoides* is similar to that of other grass-feeding higher termites as grasses tend to contain more hemicellulose and xylan than other sources of biomass (Zhao et al., 2012a). Diet, however, does not seem to significantly affect microbial diversity in *T. trinervoides* (Sanyika et al., 2012). The xylanase that forms the focus of the current study was among the enzymes discovered in a metagenomic library of the *T. trinervoides* hindgut (Rashamuse et al., 2017).

1.2 GLYCOSIDE HYDROLASES AND CARBOHYDRATE BINDING MODULE

Plant cell walls contain a wide variety of complex carbohydrates in the form of cellulose, pectins and hemicelluloses. They differ in composition, backbone or side chain stereochemistry, and inclusion of hydroxides between sugar monomers. Accordingly, carbohydrate-active enzymes or GHs that digest the glycosidic bonds are highly diverse. GH classification is mostly based on amino acid sequence conservation (Henrissat, 1991).

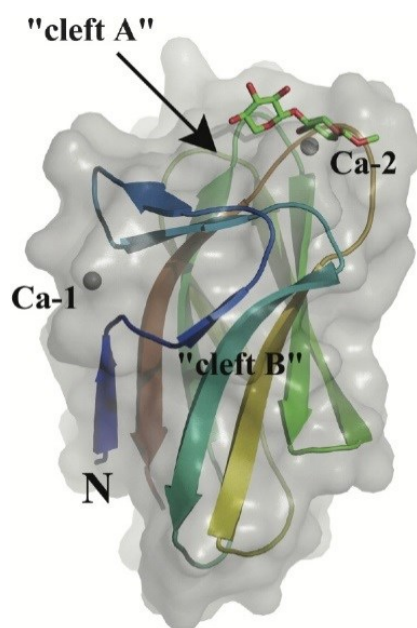


Figure 2: Cartoon and surface representation of the only published crystal structure of CBM36. It shares the β -jelly-roll fold with anti-parallel β -sheets of other type B CBMs. Clefts A and B are likely binding sites for xylan (Jamal-Talabani et al., 2004).

The carbohydrate-active enzymes database (CAZy, accessible from <http://www.cazy.org/>) provides comprehensive data on the 133 currently recognised GH families (Lombard et al., 2014). GH family members mostly degrade similar substrates, though some families contain GHs that recognise quite distinct substrates (Henrissat, 1991) preventing a simple functional assignment of GH families as e.g. required for the Enzyme Commission numerical system (Webb et al., 1992). Protein architectures are largely conserved within GH families (Schwarz, 2001) allowing enzyme mechanisms and structures to be predicted from the GH classification (Davies and Henrissat, 1995, Henrissat and Davies, 1997).

The modular nature of GHs, that combines a catalytic domain (CD) with flexibly linked non-catalytic accessory domains, further complicates a sequence-based classification (Gilkes et al., 1991). Non-catalytic domains are mostly carbohydrate binding modules (CBMs) that allow the enzymes to target insoluble substrates or disrupt the tight substrate packing (Tomme et al., 1988, Din et al., 1991, Bolam et al., 1998, Schwarz, 2001, Montanier et al., 2009, Hervé et al., 2010). The CAZy database (<http://www.cazy.org/Carbohydrate-Binding-Modules.html>) lists 81 CBM families grouped by sequence similarity that reflect the diversity of carbohydrate ligands (Boraston et al., 1999). Three structurally distinct CBM types include “planar” (type A), “endo” (B) and “exo” (C) domains. Hydrophobic aromatic residues line the binding pockets of CBMs to match the recognized carbohydrates in size and shape. Binding may involve metal ions such as Ca^{2+} in xylan-binding CBM36 (Fig. 2) increasing thermostability and affinity (Jamal-Talabani et al., 2004).

1.3 INDUSTRIAL APPLICATIONS OF XYLANASES

Industrial applications of xylanases mirror their natural functions. Examples include:

- Conversion of biomass to chemical products including its saccharification by xylanases and cellulases to ferment sugars to lactic acid (Heikkila et al., 2003, Burdette et al., 2004).
- Release of xylose from lignocellulose by xylanases for xylitol production (Sinner et al., 1988).
- Hydrolysis of xylooligomers to reduce their inhibition of cellulose saccharification after lignocellulose pre-treatment (Kumar and Wyman, 2009).
- Using xylanase in food industries as e.g. in bread dough to enhance its volume, colour and crumb structure (Van Gorcom et al., 2003).
- Increasing the digestibility of cereals and canola seeds (after oil extraction) for poultry and cattle feed (Clarkson et al., 1999, Cheng et al., 2000).
- Fermentation of xylose to lactic acid to improve the stability, digestibility and nutrition of silage (Evans et al., 1995).
- Using xylanases for pulp bleaching in the paper industry on pigments deriving from hemicellulose and lignin (Jeffries et al., 1994).
- Removing plant-based stains by laundry detergents using cellulose-linked xylanases (Smets et al., 2002).

These industrial applications require more active, specific, stable and tolerant xylanases than found in nature – achievable only by genetic modification (Juturu and Wu, 2012).

CBMs also have distinct applications including as cellulose-based affinity tags in recombinant fusion proteins to strengthen cellulose fibres in paper and cotton (Levy et al., 2003, Kavooosi et al., 2004, Zhang et al., 2011). CBMs in GHs appear to increase their efficiency and specificity in lignocellulose digestion and pulp processing (Ravalason et al., 2009, Gourlay et al., 2012, Reyes-Ortiz et al., 2013).

1.4 XYLANASE CATALYTIC MECHANISM

Enzymatic hydrolysis of O-glycosidic linkages in xylan typically involves a nucleophilic substitution at the acetal carbon of a sugar monomer resulting in the retention or inversion of stereochemistry at the anomeric carbon (Koshland, 1953, Sinnott, 1990). Family 11 GHs typically invert the configuration (Gebler et al., 1992). A dyad of active-site aspartate or glutamate residues, separated by $\sim 5 \text{ \AA}$ (McCarter and Withers, 1994), function as proton donor and nucleophile (Fig. 3). The inverting mechanism begins with the deprotonation of the donor (acid) and its nucleophilic attack on a water hydrogen with the released bond substituting the glycosidic bond at the anomeric carbon.

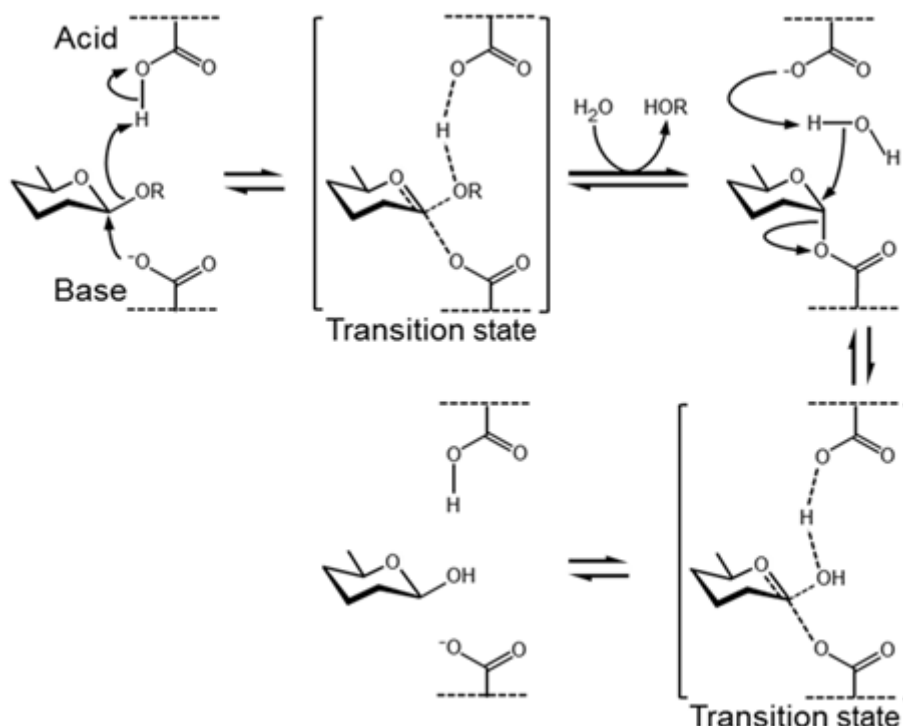


Figure 3: The catalytic mechanism of glycosidic bond hydrolysis as proposed by Koshland (1953) involves nucleophilic substitution and the retention of stereochemistry in a two-step, double-displacement reaction. Each step includes an oxocarbenium ion-like transition state. “Acid” and “Base” labels refer to amino acid side-chains in the enzyme active site (active site dyad) usually $\sim 5 \text{ \AA}$ apart (McCarter and Withers, 1994).

Enzyme thermostability and rigidity contribute to its thermophilicity (Vieille and Zeikus, 2001) where thermostability is mainly achieved through an increasing number of salt-

bridges, better amino acid packing and decreased solvent accessibility (Robinson-Rechavi et al., 2006). Some CBMs are thermostabilising and also increase stability at low pH (Nakamura et al., 2002, Paës et al., 2012). Alternatively, removal of N- or C-terminal CBMs increases thermostability in GH10 and GH11 xylanases (Blanco et al., 1999). Adding or removing CBMs would alter the termini which could explain some of these effects. Hydrophobic binding sites on CBMs may lead these to interact with other domains stabilizing the protein (Meng et al., 2015).

A single CBM36 attached to a GH11 xylanase was found to decrease thermostability and specific activity, while two family-22 CBMs attached to a GH10 xylanase had the opposite effect (Meng et al., 2015). Duplicated CBMs could increase substrate affinity due to multivalency as described for a GH11 enzyme with tandem family-2 CBMs (Bolam et al., 2001, Boraston et al., 2002) though examples of second CBMs not increasing substrate affinity have also been described (Tomme et al., 1996, Ponyi et al., 2000, Meng et al., 2015). This lack of consistency reveals a gap in the understanding of the underlying mechanisms of these vital domains.

The hypothesis of this study was that a second CBM36 would improve thermophilicity, stability and kinetics, specifically decrease the Michaelis constant K_M .

1.5 AIM AND OBJECTIVES

The aim of this study was to determine the function and structure of the domains of a GH11 xylanase from the hindgut of *T. trinervitermes*.

The objectives were:

- The production of truncated variants of the xylanase lacking both, one, or none of the CBM36 domains
- The functional characterisation of the truncated variants to determine the effect of CBM36 domains on enzyme function
- The structural characterisation by X-ray crystallography and computational modelling to elucidate the basis of the effects of the CBM36 domains

CHAPTER 2 – MATERIALS AND METHODS

2.1 MATERIALS

The gene of a 787 amino acid, multimodular GH11 xylanase (protein accession number AMO13186) was previously functionally identified in a metagenomic library of the hindgut symbionts of *T. trinervoides* (Rashamuse et al., 2017). The gene was provided in a Novagen pET20b expression vector (Merck Millipore, USA) for subcloning.

Table 1: Schematic overview of the xylanase variants derived from full-length GH11. Variants expressed from pET20b vectors had a C-terminal His₆-tag, those from pET28a both N- and C-terminal His₆-tags as reflected in their mass and gene lengths. Domains were delineated according to the NCBI conserved domains database (Marchler-Bauer et al., 2014). The colour scheme is: GH11-CD = green, CBM36-1 = red, CBM36-2 = blue, linker = yellow, CE4 = orange, and will be used consistently in all figures:





Construct (expression vector)	Length (amino acids [aa])	Molecular weight and gene length
	----- ----- ----- ----- ----- ----- -----	
	0 100 200 300 400 500 600 700 787	
	Domains and interval	
GH11 (pET20b)	 1-787 aa	84.6 kDa 2361 bp
GH11-2CBM (pET28a)	 29-559 aa	60.1 kDa 1590 bp
GH11-1CBM (pET20b)	 29-427 aa	43.7 kDa 1194 bp
GH11-CD (pET20b)	 29-246 aa	25.7 kDa 651 bp

Table 2: List of *E. coli* cells used for plasmid propagation and protein production

<i>E. coli</i> cell line	Features
DH5 α	Chemically competent, endA1 mutation (plasmid yield and quality), recA1 mutation (insert stability)
BL21(DE3)	Chemically competent, T7 RNA polymerase-IPTG induction
Rosetta-gami 2(DE3) (Merck Millipore, USA)	Chemically competent, trxB gor mutation (disulphide bond formation), pRARE2 plasmid (supplies 7 rare codons), T7 RNA polymerase-IPTG induction
BL21(DE3)-Gro7	Chemically competent, pGro7 carrying groES-groEL chaperone inducible gene (Takara Bio, Japan), T7 RNA polymerase-IPTG induction

Table 3: PCR program for gene amplification of the truncation variants. An increasing annealing temperature was based on the theoretical primer melting temperatures.

Step	Temperature	Duration	Cycles
Initial denaturation	95°C	3 min	1
Denaturation	98°C	20 s	35
Annealing	From 55 to 65°C (+0.3°C per cycle)	15 s	
Extension	72°C	1 min/kb	
Final extension	72°C	1 min/kb	1

Table 4: Primers for gene truncation and sequencing by PCR. The GH11-CBM1-R primer was used to amplify the GH11-CD and GH11-1CBM variants. All primers were synthesised by Inqaba Biotec, South Africa.

Primer	Sequence and number of bases
GH11-CD-F (forward primer)	5'- <u>CATATGGCGACCACGCTGTATGAAAATAAAAC</u> -3' 32 nucleotides
GH11-CBM1-R (reverse primer)	5'- <u>CTCGAGTGC</u> GTTGCCTTCATTGCCACCCTG-3' 30 nucleotides
GH11-CBM2-R (reverse primer)	5'-AACGCA <u>CTCGAGG</u> CCACCGGCACCGGCACC-3' 30 nucleotides
T7 promoter primer (forward primer)	5'-TAATACGACTCACTATAGGG-3' 20 nucleotides
T7 terminator primer (reverse primer)	5'-GCTAGTTATTGCTCAGCGG-3' 19 nucleotides

2.1.1 Standard protocols for preparation of materials

A) Competent cells

Chemically competent cells were prepared by inoculating 10 mL LB with a single bacterial colony and incubating the culture overnight at 37°C in a shaking incubator. This starter culture was used to inoculate 1 L of LB again incubated at 37°C in a shaking incubator. When the optical density at 600 nm (OD₆₀₀) of the culture reached 0.35-0.4, as measured by a NanoDrop ND1000 (Thermo Fisher Scientific, USA), the culture was chilled on ice for 20 min. Cells were harvested by centrifugation at 3000 x g for 15 min at 4°C. The cell pellet was suspended in 400 mL ice cold 100 mM MgCl₂. The suspended cells were centrifuged at 2000 x g for 15 min at 4°C and the cell pellet was resuspended in 200 mL ice cold 100 mM CaCl₂. This suspension was chilled on ice for 20 min. Cells were then harvested by centrifugation at 2000 x g for 15 min at 4°C. The cell pellet was resuspended

in 50 mL ice cold 85 mM CaCl₂ with 15% glycerol. Cells were again pelleted by centrifugation at 1000 x g for 15 min at 4°C. Finally, the cell pellet was resuspended in 2 mL ice cold 85 mM CaCl₂ with 15% glycerol and aliquoted. Aliquots of competent cells were flash frozen in liquid nitrogen and stored at -80°C.

B) Agarose gel electrophoresis

DNA products were analysed by agarose gel electrophoresis. Agarose gels were prepared by adding 1% (w/v) agarose (Sigma-Aldrich, USA) to 1 x TBE (89 mM Tris pH 7.6, 89 mM boric acid and 2 mM EDTA) buffer and heating the mixture to boiling in a microwave. Pronasafe nucleic acid stain (Conda, Spain) was added to the still liquid agarose mixture which was then allowed to solidify.

DNA samples were mixed with a NEB purple loading dye (New England Biolabs, USA) and loaded into wells in the gel. A potential difference of 140 V (75 W) was applied to the tank containing the gel in 1 x TBE for 6 min/cm of gel. O'GeneRuler 1 kb and 100 bp Plus were used as DNA ladders (Thermo Fisher Scientific, USA).

C) Sodium dodecyl sulphate poly-acrylamide gel electrophoresis (SDS-PAGE)

Protein products were analysed by SDS-PAGE with an upper stacking and a lower separating gel. For four separating gels 6 mL 40% (v/v) acrylamide/bis-acrylamide (final concentration: 12%), 5 mL 1.5 M Tris-HCl pH 8.8 buffer (0.375 M), 200 µL 10% (w/v) SDS (0.1%), 20 µL tetramethylethylenediamine (TEMED) and ddH₂O to a total volume of 20 mL were combined. Corresponding stacking gels were prepared by combining 0.5 mL 40% acrylamide/bis-acrylamide (final concentration 1%), 1.25 mL 0.5 M Tris-HCl pH 6.8 buffer (0.125 M), 50 µL 10% (w/v) SDS (0.1%), 20 µL TEMED and ddH₂O to a total

volume of 5 mL. Polymerisation of the stacking and separating gels was initiated by adding 50 μ L and 200 μ L 10% (w/v) ammonium persulfate, respectively. The separating gel solution was poured to fill the space between four glass plate pairs, 1 mm apart. As soon as the separating gel had polymerized, the stacking gel solution was poured on top and combs were inserted to generate sample wells. Protein samples were mixed with a 5 x sample buffer containing 10% (w/v) SDS, 10 mM DTT, 20% (v/v) glycerol, 0.2 M Tris-HCl, pH 6.8 and 0.05% (w/v) bromophenolblue, heated to 95°C for 5 min and cooled again. Protein samples were loaded into the wells and electrophoresed for 50 min at constant current of 50 mA (300 V). Pierce Unstained Protein MW Marker (Thermo Fisher Scientific, USA) was prepared in the same way as the protein samples and added as a molecular weight marker.

2.2 METHODS

2.2.1 Sequencing and gene annotation

The gene sequence of the GH11 xylanase was determined by Sanger sequencing of the expression constructs (Table 1) at the Inqaba sequencing facility (Inqaba Biotec, South Africa) using the T7 primer pair (Table 4). Raw sequencing data was analysed using Applied Biosystems Sequence Scanner Software (Thermo Fisher Scientific, USA) and the sequences were viewed in MEGA6 (Tamura et al., 2013). Domains were annotated by searching the NCBI conserved domains database (Marchler-Bauer et al., 2014). The signal peptide was predicted by SignalP 4.1 (Petersen et al., 2011).

2.2.2 Development of expression constructs

Expression constructs containing truncated variants of the GH11 gene were created as outlined in Table 1. The different domain combinations were amplified using polymerase chain reaction (PCR) amplification with a KAPA HiFi DNA polymerase (Kapa Biosystems, USA) and subcloned into Novagen pET20b and pET28a expression vectors. The pET20b vector codes for a C-terminal His₆-tag for affinity chromatography purification of the variants, while the pET28a vector codes for N- and C-terminal His₆-tags. pET20b and pET28a further contain ampicillin and kanamycin resistance genes for positive selection of transformants. The full-length GH11 construct was used as the template with forward and reverse primers (Inqaba Biotec, South Africa) introducing 5' NdeI and 3' XhoI restriction sites, respectively. The forward primers were designed to eliminate the signal peptide in all variants. Reverse primers targeted distinct interdomain linkers resulting in the gene constructs listed in Table 1. Start codons were introduced as part of the NdeI restriction site (CATATG) for pET20b. The start codon of the pET28a construct preceded the codons for the N-terminal His₆-tag. The same PCR program (Table 3) was used to amplify all variants, with a stepwise increase in annealing temperature used to encompass all melting temperatures of the primer pairs (Table 4). The PCR reaction was composed of 1 x KAPA HiFi Buffer, 0.3 mM of each dNTP from KAPA dNTP Mix, 0.3 μM of forward and reverse primers, respectively, between 10 and 100 ng template DNA, 1 Weiss Unit of KAPA HiFi DNA polymerase and ddH₂O. PCR products were purified from an agarose gel using the GeneJET gel extraction kit (Thermo Fisher Scientific, USA). Purified PCR products and plasmid vectors were double digested by NdeI and XhoI restriction enzymes (New England Biolabs, USA) in a 1 x CutSmart buffer (50 mM KAc, 20 mM tris-acetate, pH 7.9, 10 mM MgAc, 100 μg/mL BSA) for 3 h at 37°C. The digested inserts and plasmids were analysed by agarose gel electrophoresis and

again purified from the gel. Finally, the purified inserts and vectors were ligated overnight by a T4 DNA ligase (Thermo Fisher Scientific, USA) at 16°C.

2.2.3 Transformation of chemically competent cells

The ligated constructs were used to transform chemically competent DH5 α *E. coli* cells (Thermo Fisher Scientific, USA) for plasmid propagation. A 30 μ L volume of competent cells was transformed after incubation for 30 min on ice together with 5 to 20 ng plasmid, followed by a 90 s heat shock at 42°C. Cells were returned to ice for 2 minutes and then mixed with 500 μ L LB followed by shaking incubation at 200 rpm and 37°C for 60 min to allow for culture growth. Finally, 150 μ L of the culture was spread onto 1.5% LA plates, followed by incubation at 37°C overnight.

Single colonies were picked and used to inoculate 2 mL LB with either 100 μ g/mL ampicillin or 50 μ g/mL kanamycin for cells containing pET20b or pET28a constructs, respectively. These pre-cultures were incubated overnight in a shaking incubator at 37°C and pelleted by centrifugation for 2 min at 6 800 x g. Plasmids were isolated using a GeneJET plasmid miniprep kit according to manufacturer's instructions (Thermo Fisher Scientific, USA). Extracted plasmids were sequenced (Inqaba Biotec, South Africa) to verify the cloning and used to transform BL21(DE3) (New England Biolabs, USA), Rosetta-gami 2(DE3) (Merck Millipore, USA) and BL21(DE3)-Gro7 (Takara Bio, Japan) by heat shock for expression (Table 2). The pGro7 chaperone plasmid in the latter codes for the groES-groEL molecular chaperone. Transformed BL21(DE3) cells were then plated on 1.5% LA with either 100 μ g/mL ampicillin or 50 μ g/mL kanamycin and incubated overnight at 37°C. Rosetta-gami 2(DE3) and BL21(DE3)-Gro7 were plated on 1.5% LA with either 100 μ g/mL ampicillin or 50 μ g/mL kanamycin, as well as 20 μ g/ml chloramphenicol and incubated overnight at 37°C

2.2.4 Recombinant production and purification

Single colonies of transformed production cells were picked and suspended in 5 mL LB with the appropriate antibiotic: 100 µg/mL ampicillin for pET20b and 50 µg/mL kanamycin for pET28a, as well as as well as 20 µg/ml chloramphenicol for Rosetta-gami 2(DE3) and BL21(DE3)-Gro7. Pre-cultures were incubated overnight at 37°C and used to inoculate larger volumes of LB containing the appropriate antibiotics, as described above, incubated at 37°C with 180 rpm shaking. BL21(DE3)-Gro7 cultures were incubated in LB with 0.2 mg/mL L-arabinose. Once the OD₆₀₀ of the medium was between 0.4 and 0.6 measured by NanoDrop ND1000, it was cooled to 25°C and induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG). Cultures were incubated at 25°C overnight with shaking at 180 rpm. Cells were harvested by centrifugation at 5000 x g for 5 min and re-suspended in 1 x LEW buffer pH 8.0. Cells were lysed by sonication in a Q500 sonicator with 15 s on/off pulses for 3 minutes using a ½ inch probe tip (Qsonica, USA). The insoluble cellular fraction was pelleted by centrifugation at 15 000 x g for 15 min. The supernatant (soluble fraction) was filtered using a 0.45 µm Nalgene syringe filter (Thermo Fisher Scientific, USA).

His₆-tagged target proteins in the soluble fractions were enriched by affinity chromatography using Protino Nickel-TED resin (Macherey-Nagel, Germany) in a 9 x 150 mm² K9 column (GE Healthcare, USA) on an ÄKTA Avant 150 FPLC system (GE Healthcare, USA) with a flow-rate of 1 mL/min and maximum pressure of 1 bar. After applying the protein sample volume, the column was washed with 8 volumes of 1 x LEW buffer. The purified target protein was eluted with 3 volumes of a continuous imidazole gradient to a maximum concentration of 250 mM in 1 x LEW buffer. Samples taken before and after induction of expression, of the soluble and insoluble fractions, as well as samples

from the flow-through, wash step and the final elution after purification were analysed by SDS-PAGE.

The GH11-CD, GH11-1CBM and GH11-2CBM variants were concentrated to 20 mg/mL, 7.6 mg/mL and 3.5 mg/mL, respectively, in Pierce protein concentrators with appropriate molecular weight cut-offs (Thermo Fisher Scientific, USA). Protein concentrations were determined by measuring direct absorbance using a NanoDrop ND1000.

2.2.5 Functional characterisation

The GH11 variants were functionally characterised with respect to their temperature, pH and thermostability profiles as well as enzyme kinetics. The xylanase activity of the variants was verified by a dinitrosalicylic acid (DNS) assay in a thermocycler (Miller, 1959). DNS changes from yellow to red when reduced, here by the reducing ends of xylan upon enzymatic cleavage. Enzyme activity was assessed by measuring the absorbance of each reaction at 540 nm (A_{540}) in a C1000 Touch thermocycler (Bio-Rad Laboratories, USA). Substrate was prepared by dissolving 0.5% (w/v) beechwood xylan (Sigma-Aldrich, USA) in 1 x AMHB containing 100 mM acetate, MES, HEPES and Bicine each, as well as 2 mM CaCl_2 at pH 6 for CBM36 binding and stability. Reactions were initiated by adding 500 nmol xylanase variant in 5 μL 1 x AMHB pH 6 to 45 μL of 0.5% (w/v) beechwood xylan substrate and incubated for 10 min at 50°C, whereupon 50 μL 1% DNS reagent was added. The mixture was heated to 95°C for 15 min and allowed to cool to room temperature. Samples were diluted with 100 μL ddH₂O and A_{540} was measured using a KC4 96-well plate reader (Bio-Tek Instruments, USA).

Temperature and pH profiles for the enzymes were similarly determined: For the pH profile, the enzyme variant and substrate were diluted in 1 x AMHB pH 4, 5, 6, 7, 8 or 9, ten minutes prior to adding substrate. For the temperature profile, enzyme and substrate

were incubated separately at 10, 20, 30, 40, 50, 60, 70 or 80°C for 10 min before adding substrate to the enzyme and incubating for 10 min. Variant xylanase activities were compared as percentages of the maximum A_{540} of each variant in triplicate. The means of individual variant triplicate measurements were compared by Welch's unpaired t-test with unequal variance, as biochemical differences between variants could prevent an equal variance. A t-test was chosen to compare CBM-containing variants to GH11-CD, rather than to each other.

Kinetic parameters of enzyme variants were determined by altering initial substrate concentrations from 2 to 35 mg/mL beechwood xylan in the reducing sugar assay. Reactions were performed at optimal temperature and pH using equivalent amounts of enzyme variants in 1 x AMHB buffer. As only relative reaction rates were required, A_{540} values after 10 min were compared, representing the initial rates of the reaction. Kinetic parameters were calculated using computational nonlinear regression analysis with the GraphPad Prism programme (GraphPad Software, USA) using a Michaelis-Menten best-fit curve for an exponential decay.

The thermostability of the xylanase variants was compared by incubating equivalent amounts and concentrations of enzymes for 3 h at temperatures from 10 to 70°C before performing the assays at the optimal temperature and pH.

2.2.6 Structural characterisation

The atomic structures of the variants were to be determined by X-ray crystallography. The buffer was exchanged by diluting GH11-CD, GH11-1CBM and GH11-2CBM in 50 mM acetate pH 5 and concentrating in Pierce protein concentrators to 10 mg/mL, 7 mg/mL and 3 mg/mL, respectively. Initial crystallisation conditions were screened by the hanging drop method with 500 μ L reservoir solution using the HR2-110 Crystal Screen and HR2-112

Crystal Screen 2 kits (Hampton Research, USA). Droplets on the cover slides consisted of 1 μ L protein solution and 1 μ L reservoir solution and the screening experiments were incubated at 12°C. Conditions that led to macromolecular crystal formation in the droplet were optimized by systematically altering salt concentration, pH and temperature.

The best GH11-CD crystals were obtained using a reservoir solution with 0.4 M $(\text{NH}_4)_2\text{SO}_4$ and 50 mM sodium acetate pH 4.8 at room temperature. GH11-1CBM and GH11-2CBM protein microcrystals were obtained by incubating at 12°C using a reservoir solution with 1 M Li_2SO_4 and 10 mM tris-HCl pH 7 or 50 mM tris-HCl pH 6 respectively.

Due to a lack of time, experimental structure determination was substituted by computational structure prediction. The crystal structure of a CBM36 with 47 and 50% sequence identity to CBM36-1 and CBM36-2 from this study was available from the RCSB Protein Data Bank (PDB) (Jamal-Talabani et al., 2004), PDB code 1W0N. Models of the two CBM domains could thus be derived by homology modelling or protein threading yielding corresponding β -jellyroll and β -sandwich folds. Whereas xylanase-CBM structures are available from PDB, structures of GHs with two family-36 CBMs have as yet been deposited. Predicting the location and orientation of the CBM domains relative to the xylanase domain therefore requires *ab initio* modelling. The I-TASSER server combines these methods (Zhang, 2008, Roy et al., 2010, Yang et al., 2015) beginning with a threading step to combine fragments from PDB structures and reassembling these into models by “replica-exchange” Monte Carlo simulations. Regions such as loops not forming part of a domain fold are added by *ab initio* modelling. Each cycle generates a “decoy”, and a final model is generated by combining the lowest-energy decoys with optimised hydrogen-bonding network.

Here domains of the enzyme were first modelled individually to check that the predicted fold match those of known structures. Modelling of the GH11-2CBM variant was then

attempted with a hexa-xylose substrate docked into the active site of the GH11-2CBM model with AutoDock Vina (Trott and Olson, 2010). Domain interactions were modelled by *ab initio* domain assembly (AIDA) (Xu et al., 2015) using fold potential and energy minimisation of domain interactions to guide simulations of the positions of domains. Molecular structures were visualised using the PyMOL Molecular Graphics System (Version 1.6 Schrödinger, LLC).

CHAPTER 3 – RESULTS

3.1 SEQUENCING AND GENE ANNOTATION

To study the function of the CBM36 domains, the GH11 gene had to be annotated to define the start and end points of individual domains (Table 5). Interestingly, the search for a catalytic domain identified a family 4 carboxyesterase (CE4) towards the 3'-end of the gene. Two CBM36 domains of similar length and with 89% sequence identity were found in the middle of the sequence each preceded by a serine/glycine-rich linker region. Interdomain linker regions constitute 146 of 766 residues of the protein or 19%. The first 28 aa of the GH11 were predicted as a secretion signal peptide. As secretion was not required and could complicate heterologous expression in *E. coli*, this region was not included in the gene fragments of individual constructs.

Table 5: Description of conserved domains database search results used to design the protein constructs using full-length 787 amino acid GH11 as a query sequence. Positions of the domains in the GH11 gene are reported in terms of amino acid ranges.

Name	Accession	Description	Position	E-value
CE4_CtAXE_like	cd10954	Catalytic NodB homology domain of <i>Clostridium thermocellum</i> acetylxy lan esterase and its bacterial homologs	585-766	2.05e-83
Glyco_hydro_11	pfam00457	Glycosyl hydrolases family 11	41-221	1.32e-80
CBM36_xylanase-like	cd04078	Carbohydrate Binding Module family 36 (CBM36)	264-380	2.41e-58
CBM36_xylanase-like	cd04078	Carbohydrate Binding Module family 36 (CBM36)	438-553	1.26e-56

3.2 DEVELOPMENT OF EXPRESSION CONSTRUCTS

To identify possible functions of the CBM36 domains, fragments of the GH11 xylanase gene were amplified by PCR and subcloned using restriction digestion to contain either zero, one or two CBMs. PCR primers were designed to target the regions between the genes. Due to the similarity of the linker regions preceding CBM36 domains, one primer was used to amplify both the GH11-CD and GH11-1CBM encoding gene fragments. PCR products were invariably excised from an agarose gel (Fig. 4)

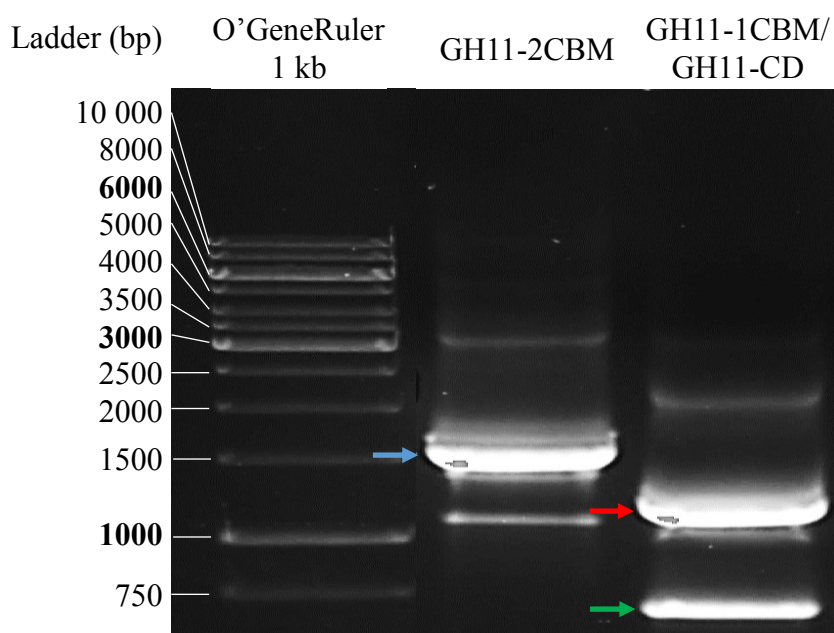


Figure 4: Agarose gel of PCR products after electrophoresis. Gene fragments marked by coloured arrows were excised from this gel, purified and digested for ligation into expression vectors. Colour code: green - GH11-CD , red - GH11-CBM, blue - GH11-2CBM.

Additional bands are visible in PCR product lanes of the agarose gel presumably due to secondary annealing sites of the amplification primers (Fig. 4). After ligation and transformation of DH5 α cells for plasmid propagation, the plasmid constructs were isolated and their correctness verified by sequencing. After transformation of expression

cells, the plasmids were isolated, digested by XhoI and NdeI restriction enzymes and analysed by agarose gel electrophoresis to confirm successful transformation (Fig. 5).

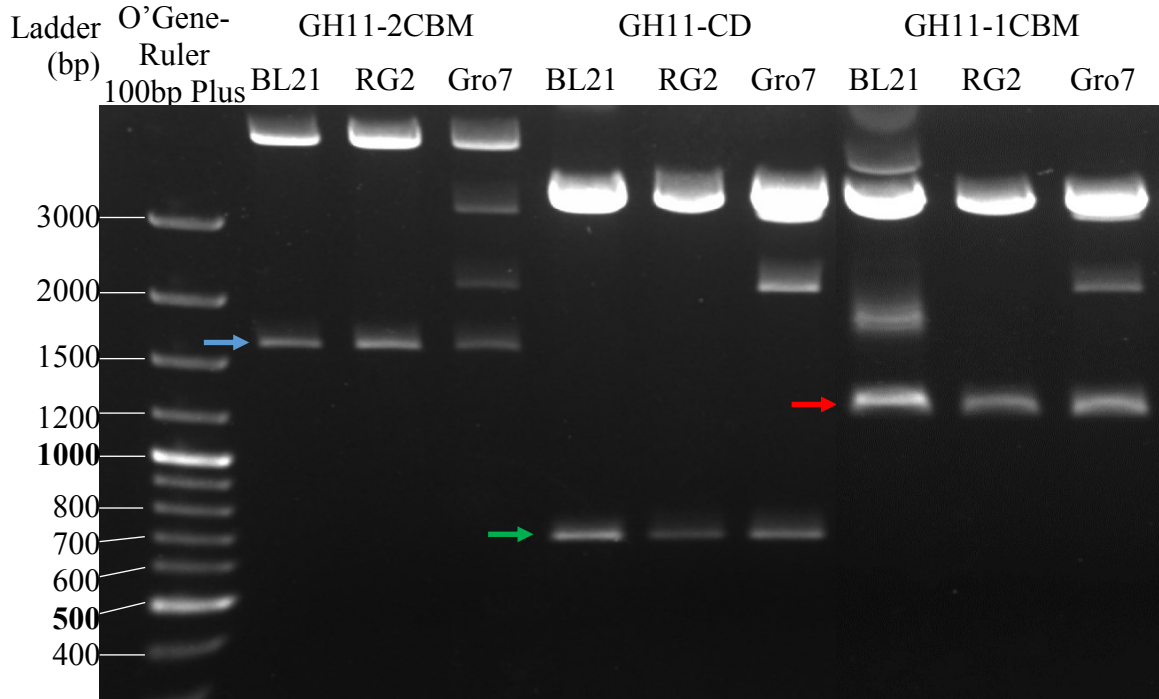


Figure 5: Analysis of XhoI/NdeI digested plasmids by agarose gel electrophoresis after expression cell transformation. Gene fragments are marked by coloured arrows: green - GH11-CD, red - GH11-CBM, blue - GH11-2CBM. Cells: BL21 = BL21(DE3), RG2 = Rosetta gami-2(DE3), Gro7 = BL21(DE3)-Gro7.

3.3 RECOMBINANT PROTEIN PRODUCTION AND PURIFICATION

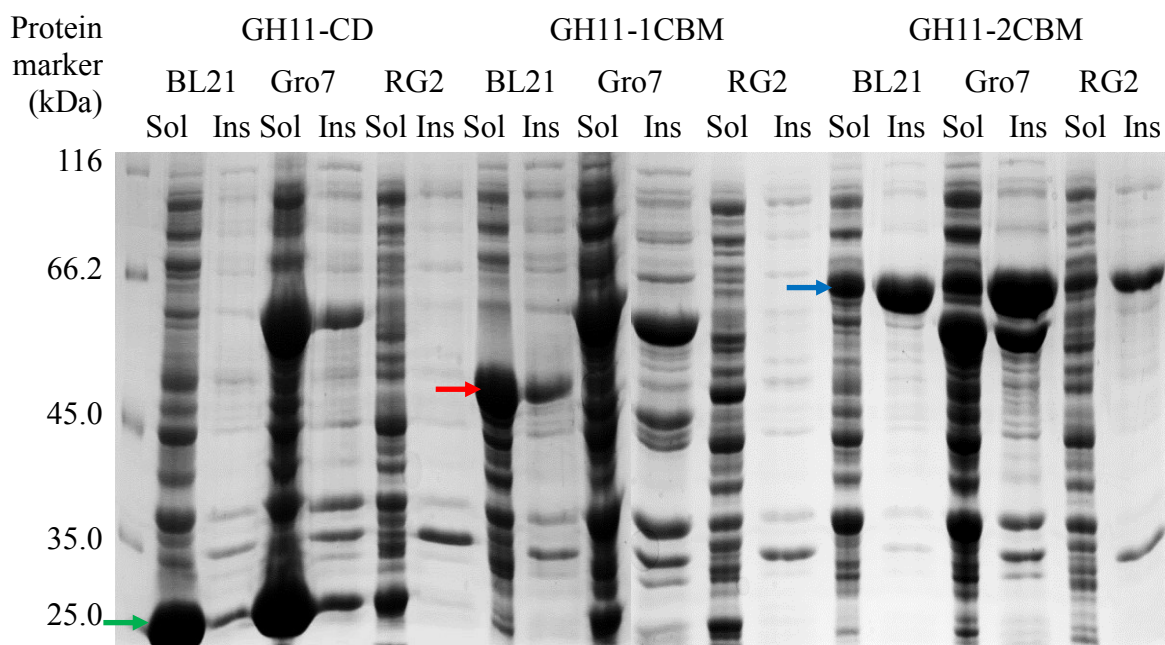


Figure 6: SDS-PAGE analysis of small-scale GH11 protein production in the different expression strains. “Sol” and “Ins” refer to the soluble and insoluble fractions, respectively. These were separated by centrifugation after cell lysis. Arrow colour code: green – GH11-CD, red – GH11-CBM, blue – GH11-2CBM.

Production of GH11 domain constructs was analysed in small-scale to determine which cells produce the largest amount of soluble protein (Fig. 6). The amount of soluble and insoluble GH11 target proteins was compared based on relative band thickness in SDS-PAGE. Interestingly, all expression cells produced all GH11 protein constructs, except BL21(DE3)-Gro7, which did not produce soluble GH11-1CBM and BL21(DE3)-Gro7 and BL21(DE3) produced comparable amounts of soluble protein, while Rosetta gami-2(DE3) produced less soluble protein. BL21(DE3) was chosen for large-scale production, because it produced the slightly more target protein, based on the target band thickness.

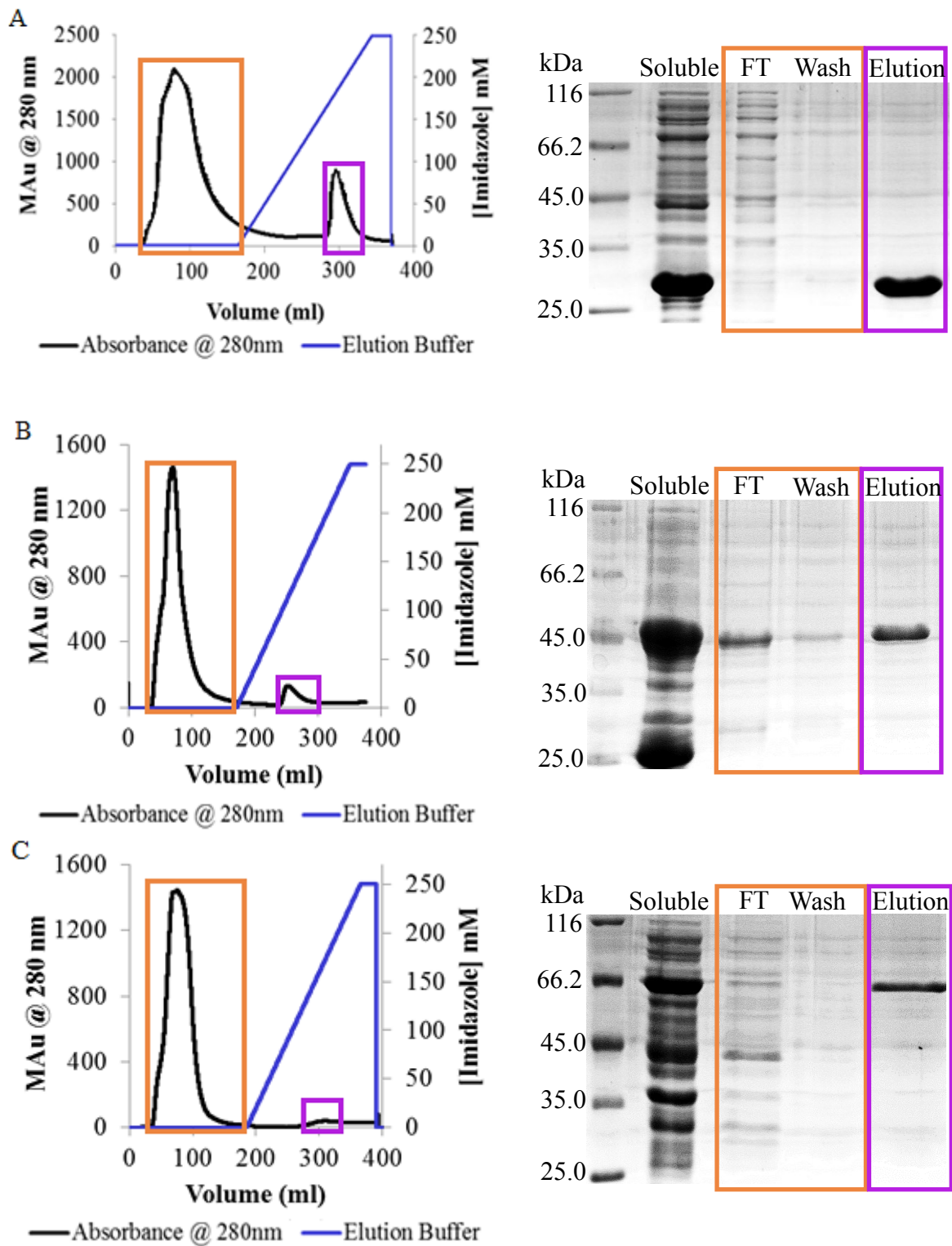
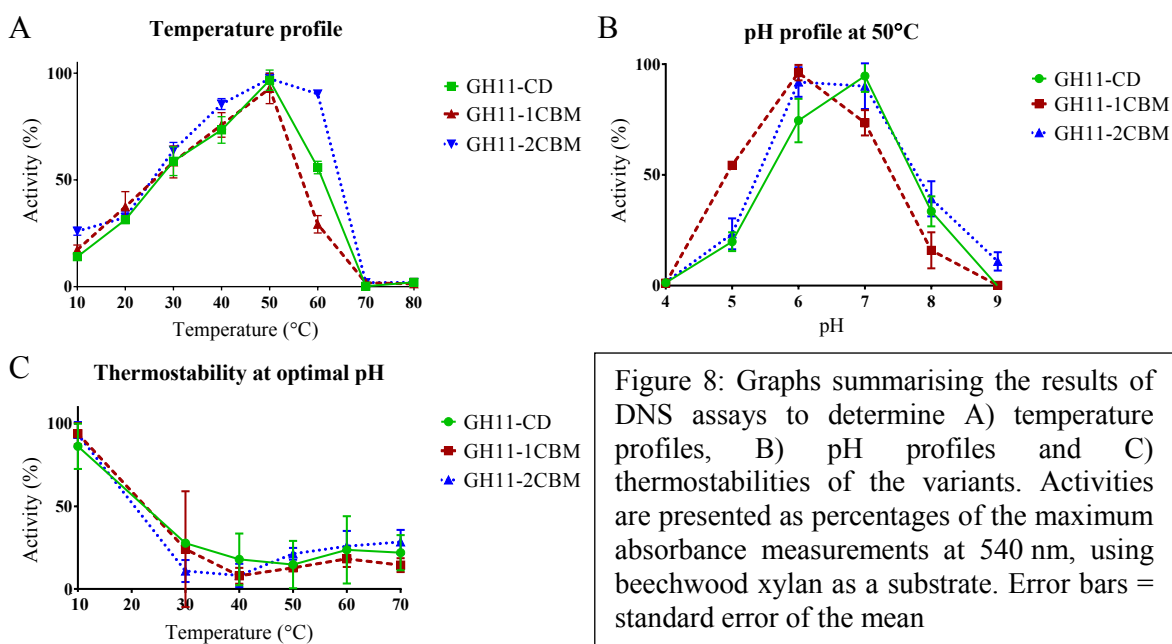


Figure 7: Chromatograms and SDS-PAGE analysis of the IMAC purification of A) GH11-CD, B) GH11-1CBM and C) GH11-2CBM. The chromatogram depicts absorbance at 280 nm indicating where the protein is being eluted.

Large-scale production in BL21(DE3) cells and purification was analysed by SDS-PAGE (Fig. 7). Fractions from different stages of the automated affinity chromatography

purification were also analysed on the same gel. Fractions included the “flow through” (FT) containing proteins not binding to the column, the “wash” containing non-specifically bound proteins, and the “elution” fraction, in which the target protein was released from the column using imidazole. Only minor additional bands were visible in the various elution fractions (Fig. 7) implying they could be used directly for functional characterisation. Adding 2 mM Ca^{2+} to the concentrated protein solutions immediately precipitated GH11 constructs with CBM36 domains, in contrast to the CD alone. Adding 20 mM EDTA allowed the protein to revert to a soluble state. Correspondingly, Ca^{2+} was only added to diluted proteins.

3.4 FUNCTIONAL CHARACTERISATION



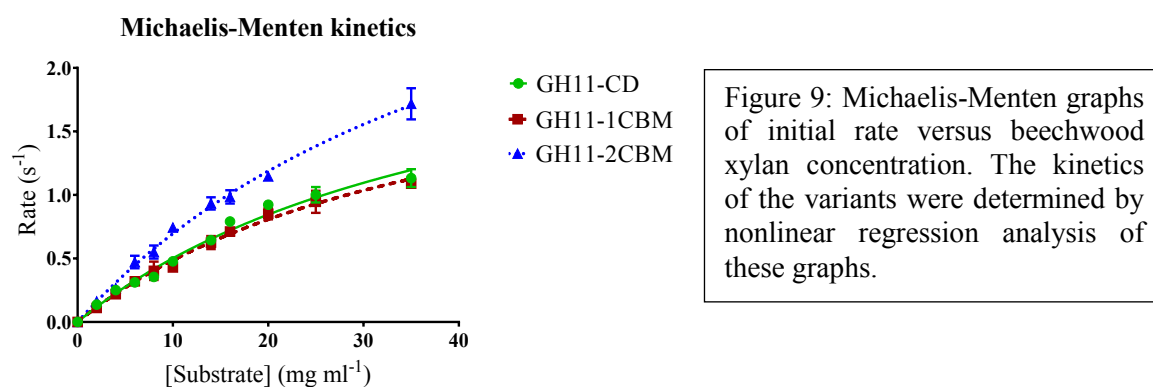
Temperature and pH optima, as well as protein thermostability were inferred by plotting enzyme activity or relative absorption values at 450 nm for DNS assays against the experimental temperature or pH (Fig. 8). Each experiment was performed in triplicate. All

GH11 constructs shared a temperature optimum of 50°C (Fig. 8, A). At 60°C, GH11-CD, GH11-2CBM and GH11-1CBM respectively retained 56% (SD = 2.9), 29% (4.1) and 90% (1.3) activity. These differences were statistically significant ($p < 0.01$) and reveal that CBM36 affects the thermostability of the GH11 xylanase.

The optimal pH of GH11-CD was found to be 7, of GH11-1CBM 6, and of GH11-2CBM between pH 6 and 7 (Fig. 8, B). At pH 5, GH11-1CBM retained 54% (SD = 1.0) activity compared to 20% (SD = 4.3) for GH11-CD, which is a statistically significant difference ($p < 0.05$). At pH 7, the activity of GH11-1CBM at 74% (SD = 5.7) was significantly ($p < 0.05$) lower than GH11-CD at 95% (SD = 7.2).

No significant difference was observed in the thermostability (Fig. 8, C) with all constructs retaining residual activity after 3 h incubation at 70°C – though this was a dramatic decrease from a similar incubation at 10°C.

Based on the activity of the different constructs, the CBM36-1 domain appears to stabilise the enzyme at low pH but destabilises it at higher temperatures. Tandem CBM36 stabilise the enzyme at high temperature but have little effect on the pH profile of the enzyme.



Michaelis-Menten kinetics were determined by nonlinear regression analysis of a rate versus substrate concentration graph (Fig. 9). The K_M or the enzyme affinities for substrate of the different constructs was indistinguishable, while the relative turnover number (k_{cat}) of GH11-2CBM was significantly higher than for GH11-CD and GH11-1CBM (Table 6).

Thus, family 36 CBMs do not appear to have a measurable effect on substrate affinity.

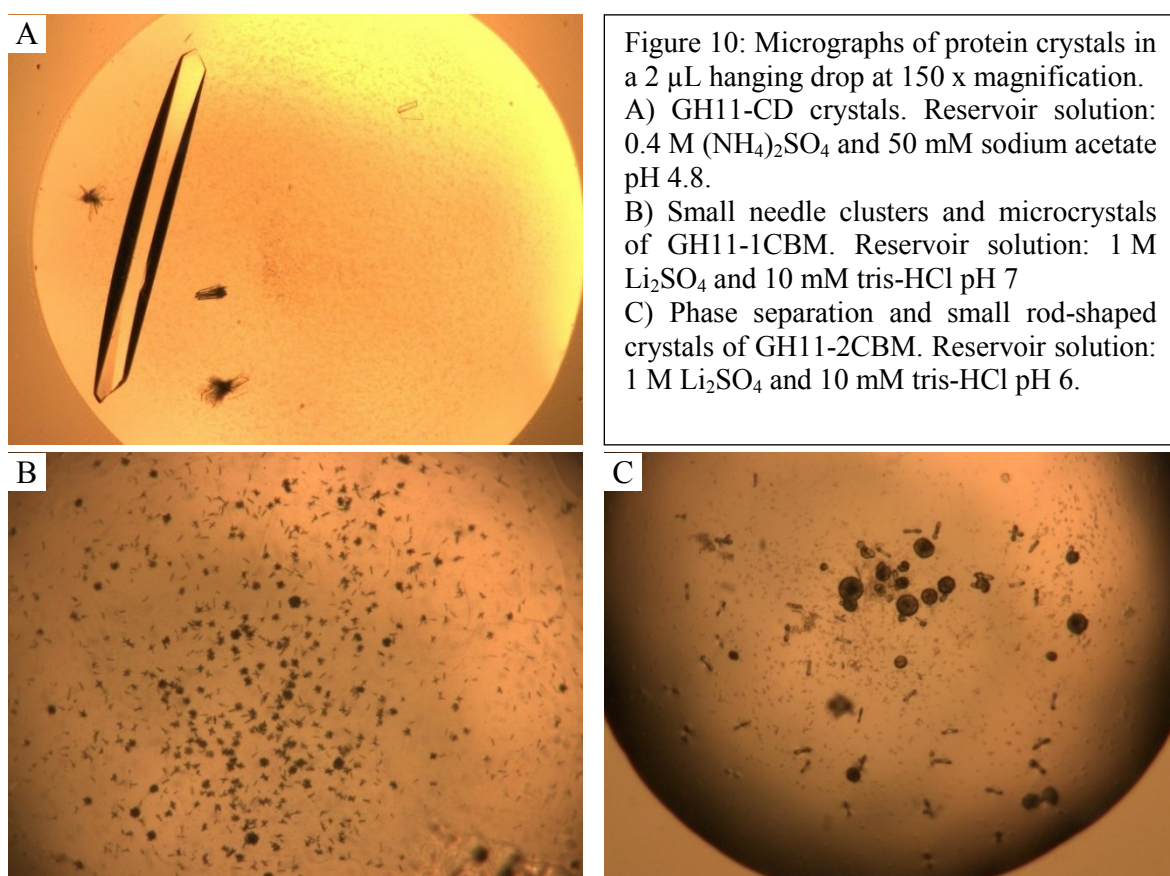
Larger differences are observed in the temperature and pH profiles.

Table 6: Enzyme kinetics parameters for GH11 variants: k_{cat} values are only relative, to ensure similar enzyme concentrations. Standard error and 95% confidence intervals (CI) are derived from 95% confidence bands around the nonlinear regression.

	GH11-CD	GH11-1CBM	GH11-2CBM
k_{cat} (min^{-1}) (Std. error; 95% CI)	2.7 (0.20; 2.3 to 3.1)	2.4 (0.17; 2.1 to 2.8)	4.1 (0.27; 3.6 to 4.7)
K_M (mg mL^{-1}) (Std. error; 95% CI)	43.4 (4.68; 35.5 to 54.1)	39.0 (4.14; 31.9 to 48.5)	48.7 (4.36; 40.9 to 58.8)

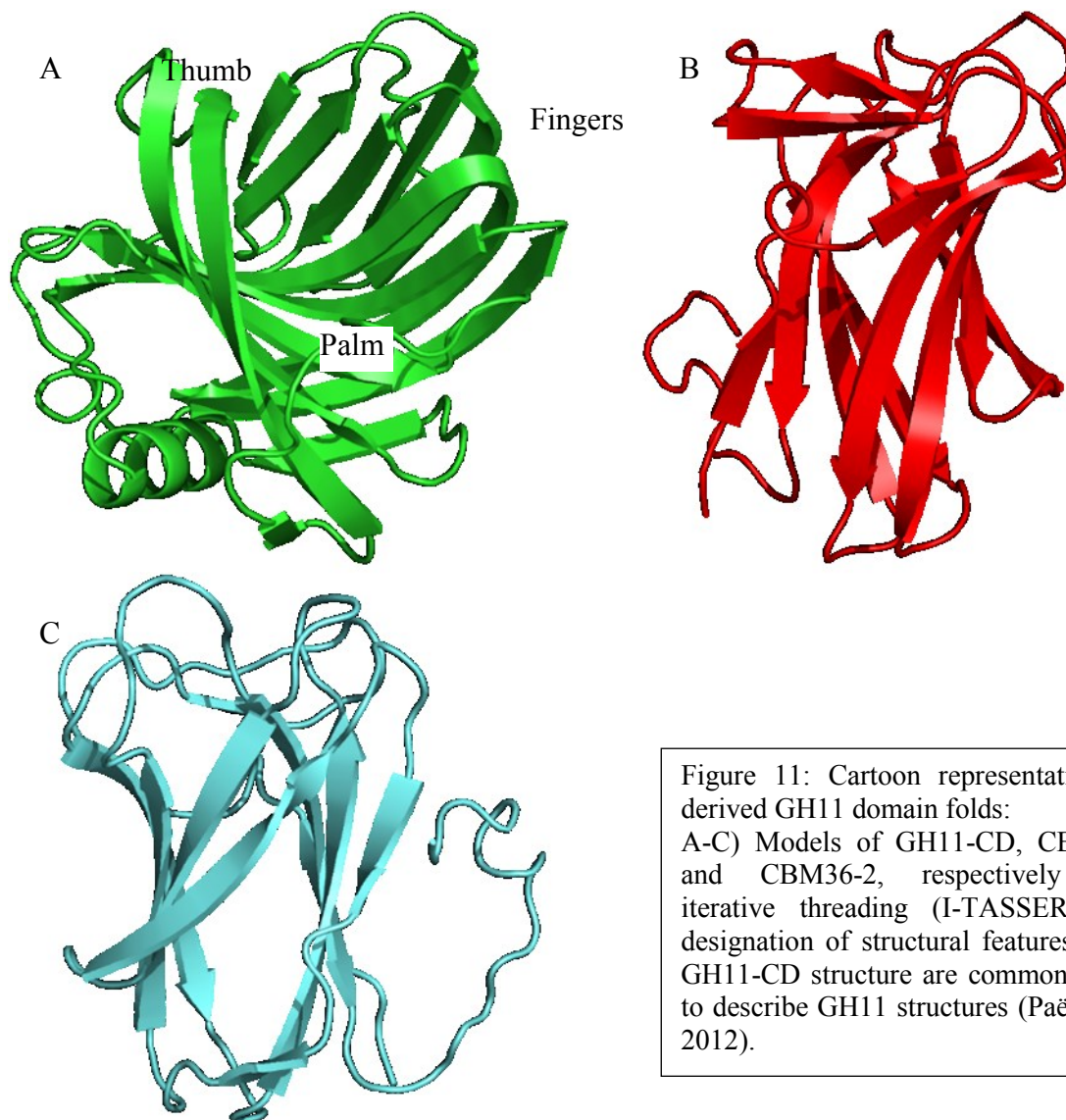
3.5 STRUCTURAL CHARACTERISATION

To assess possible structural reasons for the observed differences in key enzyme parameters, the structures of the variants were to be determined. This was first attempted by X-ray crystallography. Conditions for protein crystallisation were screened and then



optimised. Screening revealed that GH11-CD crystallised in 0.5 M and 2.0 M $(\text{NH}_4)_2\text{SO}_4$. A relatively large GH11-CD crystal was grown at room temperature with a reservoir solution consisting of 0.4 M $(\text{NH}_4)_2\text{SO}_4$ and 50 mM sodium acetate pH 4.8 with a protein concentration of 10 mg/mL (Fig. 10, A). GH11-1CBM and GH11-2CBM only formed microcrystals, small needle clusters or resulted in phase separation when incubated with a reservoir solution consisting of 1 M Li_2SO_4 and 10 mM tris-HCl pH 7 or 10 mM tris-HCl, pH 6 (Fig. 10, B and C).

Only GH11-CD produced large enough crystals for diffraction. As crystal structures of a number of related domains have been determined experimentally, computational techniques were used to model the structures of the domains alone and together (Fig. 11 and 12). Individual domains clearly retain the overall fold of related domains. The catalytic domain has a β -jelly-roll architecture that is often compared to a human, left hand (Fig. 11A) in which a β -sandwich forms an extended protrusion or “fingers”, a two-stranded β -sheet creates a “thumb” and an extended β -sheet creates the active site or “palm”. This fold is highly conserved for all CDs of GH11 xylanases with root-mean square deviations (RMSD) of less than 1.5 Å between deposited structures (Paës et al., 2012). Similarly, the computationally determined structure of the CBM36 domains (Fig. 11, B and C) closely matches the fold of the experimentally determined CBM36 structure, confirming this to be a realistic model (Fig. 2). Correspondingly, the RMSDs between the experimental structure and the CBM36-1 and CBM36-2 models is only 1.11 Å and 1.06 Å, respectively.



Docking a hexa-xylose substrate molecule into the catalytic domain confirms the presence of a glutamate dyad consisting of Glu122 and Glu212 in the active site responsible for catalytic cleavage (Fig. 12). The functional groups of these residues are ~ 8.5 Å apart and are closest acidic residues to the substrate. The distance between the carboxylate hydrogens and the hemicellulose substrate oxygens was predicted to be between 5.4 and 8.2 Å.

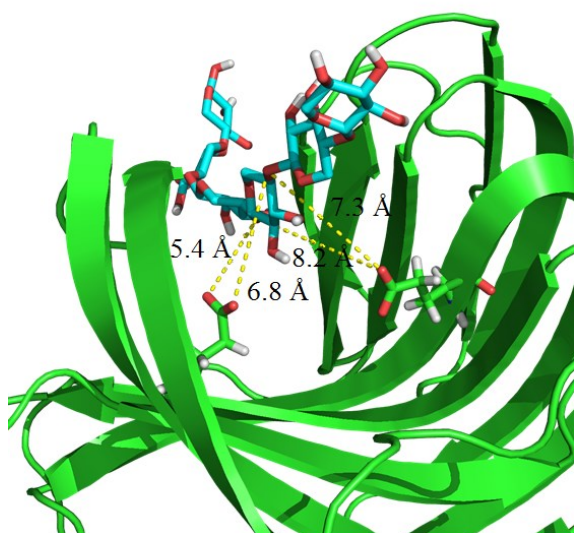


Figure 12: A hexa-xylose substrate modelled into the active site of the CD using Auto-Dock Vina. It confirms a catalytic dyad of Glu122 and Glu212 separated by 8.5 Å. The residues are between 5.4 Å and 8.2 Å from the substrate.

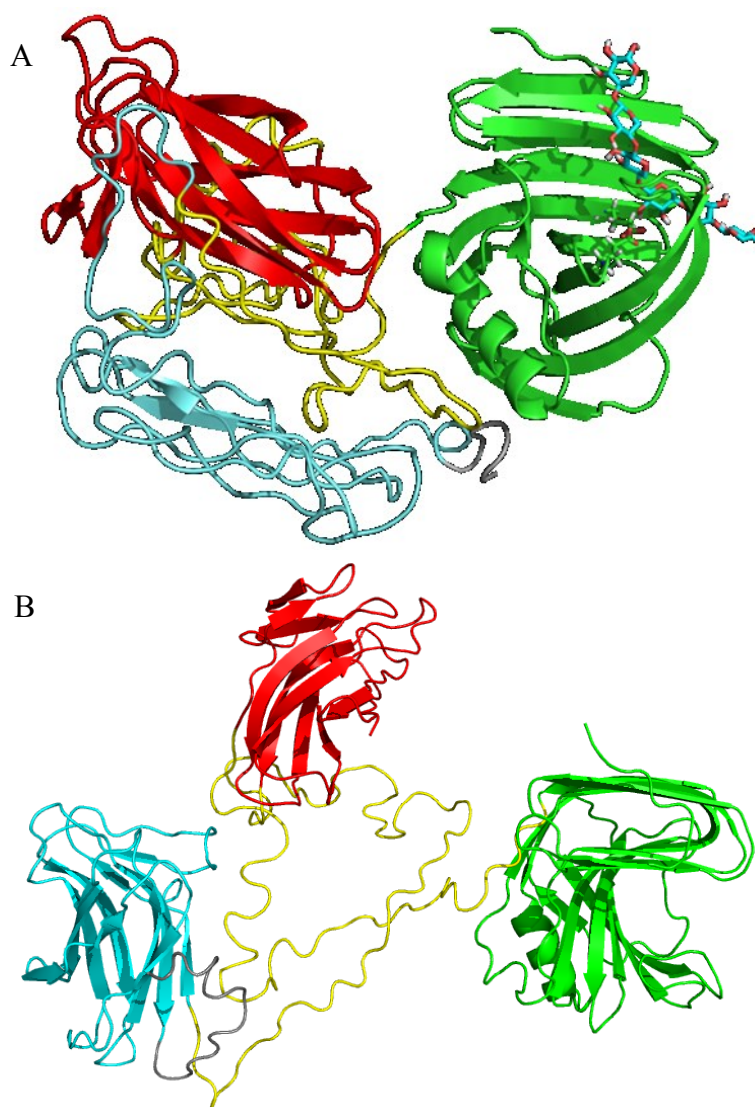


Figure 13: Cartoon representations of predicted structures of GH11-2CBM. A) A model of GH11-2CBM36 produced by iterative threading the amino acid sequences of domains by I-TASSER. Also present is a hexa-xylose substrate molecule docked into the CD active site. B) Predicted structure of GH11-2CBM from *ab initio* domain assembly by AIDA with all domains retaining their expected folds

Ab initio modelling of GH11-2CBM correctly reproduced the folds of the CD and the first CBM36 domain. The second CBM36 domain was, however, modelled with a fold different

to the expected β -sandwich, despite 89% sequence identity between the two CBM36 domains (Fig. 13, A). This deviation presumably reflects the limitations of the underlying algorithms rather than a true deviation of the domain fold from the norm. An attempt to reveal the relative orientation of the GH11-2CBM domains using their known conserved folds and the amino acid sequence by AIDA did not reveal any identifiable association between domains (Fig. 13, B).

CHAPTER 4 – DISCUSSION

Many industrial applications of xylanases require enzymes to be stable at elevated temperatures and non-physiological pH. Producing xylitol from lignocellulose using an endo-xylanases, for example, requires sustained enzyme activity for 15 hours at 40°C (Sinner et al., 1988). The enzymatic depolymerisation of lignocellulose to produce bioethanol using cellulases and xylanases similarly requires high temperatures and a dynamic pH range (Paës et al., 2012). Developing novel xylanases for industrial processes thus presupposes stable enzymes which may be achieved by molecular optimization. In this study the stabilizing effect of CBM36 domains was studied.

Various functions have been proposed for CBM domains of GHs. Most involve substrate binding. Functions such as thermostabilisation have been proposed but have not been confirmed experimentally. The main purpose of this study was to establish whether single and duplicated CBM36 domains will affect GH11 xylanase activity by stabilizing the enzyme. This is only the second functional study of tandem CBM36s.

4.1 SEQUENCING AND GENE ANNOTATION

The source organism of the GH11 in this study was an uncultured symbiont found in the third proctodeal segment of the *T. trinervoides* hindgut. Additional GHs from this hindgut metagenomic library, including GH5 and GH11 xylanases, were characterised in a parallel study (Rashamuse et al., 2017). Another GH11 xylanase had a similar domain architecture as the GH11 under investigation and was characterised in the same laboratory. The identification of two homologous GH11 xylanases in *T. trinervoides* hindgut symbionts indicates a potential benefit of tandem CBM36 under these environmental conditions and the mainly grass-based diet of the harvester termite. The production and characterisation of

the GH11 in the current study also served to enhance the view of the diversity of GH11 enzymes in the *T. trinervitermes* hindgut.

A novel feature of the current GH11 is the C-terminal CE4 domain. A similar domain structure has been reported for a GH11 xylanase from rumen *Butyrivibrio*-related bacteria (Čepeljnik et al., 2004). The CE4 domain involved in xylan deacetylation was not produced in that study leaving the esterase activity unconfirmed. Similarly, this study did not include an analysis of this domain in its aim and hypothesis. Nevertheless, such domains are not unexpected as many plant carbohydrates, especially from grasses, are acetylated to inhibit enzyme recognition and degradation (Zhao et al., 2012a).

4.2 RECOMBINANT PRODUCTION AND PURIFICATION

This study revealed that increasing numbers of domains decrease overall protein solubility as band sizes of soluble fractions decreased in SDS-PAGE analyses (Fig. 7). The amount of recovered pure protein similarly decreased in CBM36-containing constructs. The decrease in solubility may be due to trace amounts of Ca^{2+} or other divalent metal ions in the culture medium as even low concentration of Ca^{2+} were observed to precipitate protein constructs with CBM36 domains. Ca^{2+} levels had not effect on the catalytic domain, implicating CBM36 domains in Ca^{2+} binding. Interestingly, Ca^{2+} binding was previously linked to increased domain solubility and thermostability (Jamal-Talabani et al., 2004). Nevertheless, calcium ions clearly affect solubility and possibly stability of CBM36 domains – an important consideration for future studies of GHs with CBM36 domains. Another aspect to consider is that the termite P3 hindgut is anoxic, such that producing variants in aerated media could affect their solubility (Brune, 2014).

4.3 FUNCTIONAL CHARACTERISATION

Functional analysis of different GH11 constructs indicate only small effects on stability with a difference in activity at 60°C being most notable. CBM domains were previously described to destabilize xylanases and cellulases. Removing CBMs could hence yield enzymes better suited to industrial applications (Pakarinen et al., 2014, Meng et al., 2015). CBMs were proposed to strongly bind residual substrate after hydrolysis, reducing the overall efficiency of the enzyme. In this study, no measurable differences in the K_M of the different constructs were observed, indicating that the CBMs do not bind to processed soluble xylan (beechwood xylan) more tightly than the catalytic domain. A parallel study on a similar GH11 indicated that tandem CBM36 domains may slightly increase activity with insoluble oat-spelt xylan as a substrate compared to constructs with fewer CBMs (Rashamuse et al., 2017) demonstrating a potential role of tandem CBM36s in the binding insoluble xylan. Overall, the activity of all variants was, however, less than 20% compared to soluble xylan in that study.

Though statistically significant, the differences in pH profiles of the three constructs did not indicate a dramatically improved pH stability. All three constructs retained activity across a relatively broad pH range, with only a slight shift in pH optima. A single CBM36 stabilises the enzyme at low pH compared to GH11-CD and GH11-2CBM. It is likely that if the termite hindgut had a lower pH, such as in lower termites, the GH11 with tandem CBM36 domains might not have had the same selective pressure to evolve and instead a single CBM36 may have been favoured.

Similarly, the kinetics of the enzyme variants did not point to an increase in binding affinity with either one or two CBM36 domains. The substrate concentrations for Michaelis-Menten kinetic measurements did not reach saturation, because the substrate was not sufficiently soluble above 35 mg/ml to ensure reproducibility of results. Saturation

concentrations had to be inferred by the nonlinear regression analysis. The increased k_{cat} of GH11-2CBM could also be explained by an increase in enzyme concentration. With the means available to measure protein concentration and low concentrations used (nanomolar range) it is very difficult to confidently ensure equimolar concentrations. Thus, k_{cat} was not considered as a factor in characterising the variants in this study.

The thermostability measurements after 3 h incubations at different temperatures indicated that the GH11 constructs are less stable than would perhaps have been expected. The activity decreases appreciably for incubation temperatures above 10°C meaning that lower temperatures and shorter incubation times would have better demonstrated biochemical differences between the variants. These results may have also been an experimental artefact caused by the long incubation times, as indicated by the large standard deviation within the data. This would explain the apparent inconsistency with the initial rate measurements of activity at 60°C. The activity at 60°C was therefore used instead as an indicator of thermostability. The results appear to imply that a second CBM36 greatly increases the stability and thermophilicity compared to a single CBM36 and the CD alone. This could be achieved through increased intramolecular interactions particularly between domains and would be consistent with the interaction that was measured between tandem CBM22 domains (Meng et al., 2015). As with the pH, the stabilising effect of the second CBM36 may have resulted from selective pressure at higher temperatures in the termite hindgut.

4.4 STRUCTURAL CHARACTERISATION

The GH11-2CBM demonstrated a higher degree of thermophilicity than GH11-CD alone. It is likely that this thermophilicity is a result of increased stability rather than an

adaptation to an extreme environment. Compactness is a distinguishing feature of thermophilic enzymes (Robinson-Rechavi et al., 2006). A tighter binding between the three domains compared to only two domains would explain the increase in thermophilicity. This tighter binding could be caused by electrostatic interactions between the domains.

Attempts at crystallising the variants for X-ray crystallographic structure determination could not be completed. Large enough crystals for diffraction were only obtained for GH11-CD. Resolving the crystal structure of GH11-CD without the other variants was unlikely to address any of the biological questions. Larger GH11-1CBM and GH11-2CBM crystals could potentially have been produced by adding a chemically synthesised oligo-xylose substrate or by screening additional conditions and additives.

The structure prediction confirmed the expected folds for each of the domains. Domain-domain interactions could, however, not be reliably modelled. The structure model for GH11-2CBM did not include the predicated fold for CBM36-2, yielding a model largely without secondary or tertiary structure instead. The structure of CBM36-2 was thus only predicted as a stand-alone domain, but not as part of a larger structure. The true domain fold may have been eliminated as energetically unfavourable during the clustering of templates. This highlights an issue with structural prediction of complex, multimodular enzymes using threading and homology modelling (Xu et al., 2015). Specifically, a crystal structure revealing CBM36 interactions with other domains is not yet available. A major problem in predicting domain interactions is the complexity of the problem and the lack of experimental constraints. Another limitation of the structural prediction was that only a single CBM36 domain has been determined experimentally to date. Apart from their calcium binding, little is known about their biochemistry (Jamal-Talabani et al., 2004).

Chances of AIDA predicting domains interactions was seen to be higher, as it would incorporate the correct domain fold already. The AIDA model, however, did not identify significant domain interactions. As solvent interactions were not taken into account, the accuracy of the model is uncertain. CBM substrate binding sites are hydrophobic allowing for potential interactions through exclusion of water molecules. Interdomain interactions through these sites could explain why the CBM36 domains only marginally affect enzyme kinetics. Long interdomain linkers provide no significant constraints permitting essentially any – or no – interactions between domains. Partial unfolding of compact CBM domains has been observed, implying that tandem CBMs may form entirely novel interactions (Summers et al., 2016).

CHAPTER 5 – CONCLUSION

The main finding in the functional study of the GH11 enzyme under investigation is that CBM36-containing variants biochemically differ only marginally from the CD alone. No strong evidence could thus be found that the CBM36 domains provide an enzymatic advantage to the GH11 domain – the central hypothesis of this project. The payoff for the increased energy expended by the symbiont in expressing the longer variants therefore currently remains elusive. Potentially, the CBM36 domains may guide the enzyme towards acetylated substrates of the CE4 domain – something that was not investigated in this project. This unique domain architecture is clearly worth characterising further. A crystallographic study of the available multidomain constructs could yet reveal novel CBM interactions, explaining the biochemical activity observed in this study. The pH and temperature adaptation of single and tandem CBM36 domains align with previous studies of these and similar domains.

REFERENCES

- ADAM, R., MITCHELL, J. & VAN DER WESTHUIZEN, M. 2005. Food preferences in laboratory colonies of the harvester termite, *Trinervitermes trinervoides* (Sjöstedt)(Termitidae: Nasutitermitinae). *African entomology*, 13, 193-200.
- ADAM, R., MITCHELL, J. & VAN DER WESTHUIZEN, M. 2008. Aspects of foraging in the harvester termite, *Trinervitermes trinervoides* (Sjöstedt)(Termitidae: Nasutitermitinae). *African Entomology*, 16, 153-161.
- BLANCO, A., DIAZ, P., ZUECO, J., PARASCANDOLA, P. & PASTOR, F. J. 1999. A multidomain xylanase from a *Bacillus sp.* with a region homologous to thermostabilizing domains of thermophilic enzymes. *Microbiology*, 145, 2163-2170.
- BOLAM, D. N., CIRUELA, A., MCQUEEN-MASON, S., SIMPSON, P., WILLIAMSON, M. P., RIXON, J. E., BORASTON, A., HAZLEWOOD, G. P. & GILBERT, H. J. 1998. *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochemical journal*, 331, 775-781.
- BOLAM, D. N., XIE, H., WHITE, P., SIMPSON, P. J., HANCOCK, S. M., WILLIAMSON, M. P. & GILBERT, H. J. 2001. Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* xylanase 11A. *Biochemistry*, 40, 2468-2477.
- BORASTON, A., MCLEAN, B., KORMOS, J., ALAM, M., GILKES, N., HAYNES, C., TOMME, P., KILBURN, D. & WARREN, R. 1999. Carbohydrate-binding modules: diversity of structure and function. *Special Publication-Royal Society of Chemistry*, 246, 202-211.
- BORASTON, A. B., MCLEAN, B. W., CHEN, G., LI, A., WARREN, R. A. J. & KILBURN, D. G. 2002. Co-operative binding of triplicate carbohydrate-binding modules from a thermophilic xylanase. *Molecular microbiology*, 43, 187-194.
- BRUNE, A. 2014. Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology*, 12, 168-180.
- BURDETTE, J., BERG, B., CARR, B., DUCK, N., KOZIEL, M., CAROZZI, N. & PATEL, P. 2004. Methods to enhance the activity of lignocellulose-degrading enzymes. *U.S. Patent Application No. 10/795,102*.
- ČEPELJNIK, T., KRIŽAJ, I. & MARINŠEK-LOGAR, R. 2004. Isolation and characterization of the *Pseudobutyrvibrio xylanivorans* Mz5 T xylanase XynT—the first family 11 endoxylanase from rumen *Butyrvibrio*-related bacteria. *Enzyme and microbial technology*, 34, 219-227.
- CHABANNES, M., RUEL, K., YOSHINAGA, A., CHABBERT, B., JAUNEAU, A., JOSELEAU, J. P. & BOUDET, A. M. 2001. *In situ* analysis of lignins in transgenic tobacco reveals a differential impact of individual transformations on the spatial patterns of lignin deposition at the cellular and subcellular levels. *The Plant Journal*, 28, 271-282.
- CHENG, K.-J., SELINGER, L. B., LIU, J.-H., HU, Y., FORSBERG, C. W. & MOLONEY, M. M. 2000. Xylanase obtained from an anaerobic fungus. *U.S. Patent 6,137,032*.

- CLARKSON, K. A., MORGAN, A. J. & WANG, Z. C. 1999. Xylanase from *Acidothermus cellulolyticus*. *U.S. Patent 5,902,581*.
- COATON, W. 1948. Trinervitermes Species-the Snouted Harvester Termites. *Bulletin of the Department of Agriculture of South Africa*, 261, 1-19.
- CREPIN, V., FAULDS, C. & CONNERTON, I. 2004. Functional classification of the microbial feruloyl esterases. *Applied Microbiology and Biotechnology*, 63, 647-652.
- DAVIES, G. & HENRISSAT, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure*, 3, 853-859.
- DIN, N., GILKES, N. R., TEKANT, B., MILLER, R. C., WARREN, R. A. J. & KILBURN, D. G. 1991. Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. *Nature Biotechnology*, 9, 1096-1099.
- EVANS, C. T., MANN, S. P., CHARLEY, R. C. & PARFITT, D. 1995. Formulation for treating silage containing β -1, 4-xylanase and β -1, 3-xylosidase but essentially free of β -1, 4-glucanase and β -1, 4-cellobiohydrolase, and one or more lactic acid-producing bacteria. *U.S. Patent 5,432,074*.
- GEBLER, J., GILKES, N. R., CLAEYSSSENS, M., WILSON, D. B., BÉGUIN, P., WAKARCHUK, W. W., KILBURN, D. G., MILLER, R. C., WARREN, R. & WITHERS, S. G. 1992. Stereoselective hydrolysis catalyzed by related beta-1, 4-glucanases and beta-1, 4-xylanases. *Journal of Biological Chemistry*, 267, 12559-12561.
- GILKES, N., HENRISSAT, B., KILBURN, D., MILLER, R. & WARREN, R. 1991. Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiological reviews*, 55, 303-315.
- GONG, C., CAO, N., DU, J. & TSAO, G. 1999. Ethanol production from renewable resources. *Recent progress in bioconversion of lignocellulosics*. Springer.
- GOURLAY, K., ARANTES, V. & SADDLER, J. N. 2012. Use of substructure-specific carbohydrate binding modules to track changes in cellulose accessibility and surface morphology during the amorphogenesis step of enzymatic hydrolysis. *Biotechnology for biofuels*, 5, 51.
- HARTWIG, E. 1955. Control of snouted harvester termites. *Farming in South Africa*, 30, 361-366.
- HEIKKILA, H., SARKKI, M.-L., RAVANKO, V. & LINDROOS, M. 2003. Process for the production of xylose from a paper-grade hardwood pulp. *U.S. Patent 6,512,110*.
- HENRISSAT, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*, 280, 309-316.
- HENRISSAT, B. & DAVIES, G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Current opinion in structural biology*, 7, 637-644.
- HERVÉ, C., ROGOWSKI, A., BLAKE, A. W., MARCUS, S. E., GILBERT, H. J. & KNOX, J. P. 2010. Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proceedings of the National Academy of Sciences*, 107, 15293-15298.
- JAMAL-TALABANI, S., BORASTON, A. B., TURKENBURG, J. P., TARBOURIECH, N., DUCROS, V. M.-A. & 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.
- JEFFRIES, T. W., GRABSKI, A. C. & PATEL, R. N. 1994. Xylanase from *Streptomyces roseiscleroticus* NRRL-11019 for removing color from kraft wood pulps. *U.S. Patent 5,369,024*.
- JUTURU, V. & WU, J. C. 2012. Microbial xylanases: engineering, production and industrial applications. *Biotechnology advances*, 30, 1219-1227.
- KAVOOSI, M., MEIJER, J., KWAN, E., CREAGH, A. L., KILBURN, D. G. & HAYNES, C. A. 2004. Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T. maritima*. *Journal of Chromatography B*, 807, 87-94.
- KOSHLAND, D. E. 1953. Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews*, 28, 416-436.
- KUMAR, R. & WYMAN, C. E. 2009. Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pretreatment technologies. *Bioresource Technology*, 100, 4203-4213.
- LEE, D., OWENS, V., BOE, A. & JERANYAMA, P. 2007. Composition of herbaceous biomass feedstocks. *South Dakota State University Publication, SGINCI-07*. Brookings, SD.
- LEVY, G. N., SCHINDEL, R. & KRUTH, J.-P. 2003. Rapid manufacturing and rapid tooling with layer manufacturing (LM) technologies, state of the art and future perspectives. *CIRP Annals-Manufacturing Technology*, 52, 589-609.
- LIMAYEM, A. & RICKE, S. C. 2012. Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects. *Progress in Energy and Combustion Science*, 38, 449-467.
- LOMBARD, V., RAMULU, H. G., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research*, 42, D490-D495.
- MARCHLER-BAUER, A., DERBYSHIRE, M. K., GONZALES, N. R., LU, S., CHITSAZ, F., GEER, L. Y., GEER, R. C., HE, J., GWADZ, M. & HURWITZ, D. I. 2014. CDD: NCBI's conserved domain database. *Nucleic acids research*, 43(D1), D222-D226.
- MCCARTER, J. D. & WITHERS, G. S. 1994. Mechanisms of enzymatic glycoside hydrolysis. *Current opinion in structural biology*, 4, 885-892.
- MENG, D.-D., YING, Y., CHEN, X.-H., LU, M., NING, K., WANG, L.-S. & LI, F.-L. 2015. Distinct roles for carbohydrate-binding modules of glycoside hydrolase 10 (GH10) and GH11 xylanases from *Caldicellulosiruptor sp.* strain F32 in thermostability and catalytic efficiency. *Applied and environmental microbiology*, 81, 2006-2014.
- MILLER, G. L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical chemistry*, 31, 426-428.
- 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.

2009. Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proceedings of the National Academy of Sciences*, 106, 3065-3070.
- MOSIER, N., WYMAN, C., DALE, B., ELANDER, R., LEE, Y., HOLTZAPPLE, M. & LADISCH, M. 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource technology*, 96, 673-686.
- NAKAMURA, M., NAGAMINE, T., TAKENAKA, A., AMINOV, R. I., OGATA, K., TAJIMA, K., MATSUI, H., BENNO, Y. & ITABASHI, H. 2002. Molecular cloning, nucleotide sequence and characteristics of a xylanase gene (xynA) from *Ruminococcus albus* 7. *Animal science journal*, 73, 347-352.
- PAËS, G., BERRIN, J.-G. & BEAUGRAND, J. 2012. GH11 xylanases: structure/function/properties relationships and applications. *Biotechnology advances*, 30, 564-592.
- PAKARINEN, A., HAVEN, M. Ø., DJAJADI, D. T., VÁRNASI, A., PURANEN, T. & VIKARI, L. 2014. Cellulases without carbohydrate-binding modules in high consistency ethanol production process. *Biotechnology for biofuels*, 7, 27.
- PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, 8, 785-786.
- PONYI, T., SZABÓ, L., NAGY, T., OROSZ, L., SIMPSON, P. J., WILLIAMSON, M. P. & 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.
- RASHAMUSE, K., SANYIKA TENDAI, W., MATHIBA, K., NGCOBO, T., MTIMKA, S. & BRADY, D. 2017. Metagenomic mining of glycoside hydrolases from the hindgut bacterial symbionts of a termite (*Trinervitermes trinervoides*) and the characterization of a multimodular β -1, 4-xylanase (GH11). *Biotechnology and applied biochemistry*, 64(2), 174-186.
- RAVALASON, H., HERPOËL-GIMBERT, I., RECORD, E., BERTAUD, F., GRISEL, S., DE WEERT, S., VAN DEN HONDEL, C. A., ASTHER, M., PETIT-CONIL, M. & SIGOILLOT, J.-C. 2009. Fusion of a family 1 carbohydrate binding module of *Aspergillus niger* to the *Pycnoporus cinnabarinus* laccase for efficient softwood kraft pulp biobleaching. *Journal of biotechnology*, 142, 220-226.
- REYES-ORTIZ, V., HEINS, R. A., CHENG, G., KIM, E. Y., VERNON, B. C., ELANDT, R. B., ADAMS, P. D., SALE, K. L., HADI, M. Z. & SIMMONS, B. A. 2013. Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates. *Biotechnology for biofuels*, 6, 93.
- ROBINSON-RECHAVI, M., ALIBÉS, A. & GODZIK, A. 2006. Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of *Thermotoga maritima*. *Journal of molecular biology*, 356, 547-557.
- ROLLIN, J. A., ZHU, Z., SATHITSUKSANO, N. & ZHANG, Y. H. P. 2011. Increasing cellulose accessibility is more important than removing lignin: A comparison of cellulose solvent-based lignocellulose fractionation and soaking in aqueous ammonia. *Biotechnology and bioengineering*, 108, 22-30.

- ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*, 5, 725-738.
- SANYIKA, T. W., RASHAMUSE, K. J., HENNESSY, F. & BRADY, D. 2012. Luminal hindgut bacterial diversities of the grass and sugarcane feeding termite *Trinervitermes trinervoides*. *African Journal of Microbiology Research*, 6, 2639-2648.
- SCHARF, M. E. & BOUCIAS, D. G. 2010. Potential of termite-based biomass pre-treatment strategies for use in bioethanol production. *Insect Science*, 17, 166-174.
- SCHELLER, H. V. & ULVSKOV, P. 2010. Hemicelluloses. *Annual review of plant biology*, 61, 263-289.
- SCHWARZ, W. 2001. The cellulosome and cellulose degradation by anaerobic bacteria. *Applied microbiology and biotechnology*, 56, 634-649.
- SINNER, M., DIETRICH, H.-H., PULS, J., SCHWEERS, W. & BRACHTHAUSER, K.-H. 1988. Process for production of xylitol from lignocellulosic raw materials. *U.S. Patent 4,742,814*.
- SINNOTT, M. L. 1990. Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews*, 90, 1171-1202.
- SMETS, J., BETTIOL, J.-L. P., BOYER, S. L. & BUSCH, A. 2002. Laundry detergent and/or fabric care compositions comprising a modified enzyme. *U.S. Patent 6,468,955*.
- SOREK, N., YEATS, T. H., SZEMENYEI, H., YOUNGS, H. & SOMERVILLE, C. R. 2014. The Implications of lignocellulosic biomass chemical composition for the production of advanced biofuels. *BioScience*, 64, 192-201.
- SUMMERS, E. L., MOON, C. D., ATUA, R. & ARCUS, V. L. 2016. The structure of a glycoside hydrolase 29 family member from a rumen bacterium reveals unique, dual carbohydrate-binding domains. *Acta Crystallographica Section F: Structural Biology Communications*, 72, 750-761.
- SUN, Y. & CHENG, J. 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource technology*, 83, 1-11.
- TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, 30, 2725-2729.
- TOMME, P., GILKES, N., GUARNA, M., HAYNES, C., HASENWINKLE, D., JERVIS, E., JOHNSON, P., MCINTOSH, L., WARREN, R. & KILBURN, D. 1996. Cellulose-binding domains. *Annals of the New York Academy of Sciences*, 799, 418-424.
- TOMME, P., TILBEURGH, H., PETTERSSON, G., DAMME, J., VANDEKERCKHOVE, J., KNOWLES, J., TEERI, T. & CLAEYSSSENS, M. 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. *European Journal of Biochemistry*, 170, 575-581.
- TROTT, O. & OLSON, A. J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31, 455-461.
- VAN GORCOM, R. F., HESSING, J. G., MAAT, J., ROZA, M. & VERBAKEL, J. M. A. 2003. Xylanase production. *U.S. Patent 6,586,209*.

- VIEILLE, C. & ZEIKUS, G. J. 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and molecular biology reviews*, 65, 1-43.
- WEBB, E. C. 1992. Enzyme nomenclature. *Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology*. Academic Press.
- XU, D., JAROSZEWSKI, L., LI, Z. & GODZIK, A. 2015. AIDA: *ab initio* domain assembly for automated multi-domain protein structure prediction and domain–domain interaction prediction. *Bioinformatics*, 31, 2098-2105.
- YANG, J., YAN, R., ROY, A., XU, D., POISSON, J. & ZHANG, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nature methods*, 12, 7-8.
- ZHANG, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, 9, 40.
- ZHANG, Y., CHEN, S., HE, M., WU, J., CHEN, J. & WANG, Q. 2011. Effects of *Thermobifida fusca* cutinase-carbohydrate-binding module fusion proteins on cotton bioscouring. *Biotechnology and Bioprocess Engineering*, 16, 645.
- ZHANG, Y. H. P. & LYND, L. R. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnology and bioengineering*, 88, 797-824.
- ZHAO, X., ZHANG, L. & LIU, D. 2012a. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. *Biofuels, Bioproducts and Biorefining*, 6, 465-482.
- ZHAO, X., ZHANG, L. & LIU, D. 2012b. Biomass recalcitrance. Part II: Fundamentals of different pre-treatments to increase the enzymatic digestibility of lignocellulose. *Biofuels, Bioproducts and Biorefining*, 6, 561-579.