Structural and biophysical characterization of a multidomain xylanase Xyl

by

Valentine Anye

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy (PhD) Biochemistry

Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

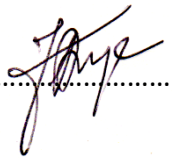
10 February 2024

Supervisor: Prof Wolf-Dieter Schubert

Dedicated to my beloved wife, Constance Bih Anye

DECLARATION

I **Valentine Anye** declare that this thesis, which I hereby submit for the degree of Doctor of Philosophy (PhD) in Biochemistry at the University of Pretoria, is my own original work and has not been previously submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: …… 

DATE: 10 February 2024

ACKNOWLEDGEMENTS

I extend my heartfelt gratitude to all those who have supported and contributed to my academic journey and research. I am deeply thankful for the invaluable assistance provided by various individuals, institutions, and family members who have made this journey possible.

Supervisor, Prof Wolf-Dieter Schubert: Your steadfast, meticulous, insightful, and financial support have been instrumental in my academic development and success. **Prof Trevor Sewell:** Thank you for your guidance in X-ray data collection and preliminary data analysis. **Dr. Precious Motshwene**: Your early-stage research support has been invaluable. **European Synchrotron Radiation Facility**: I appreciate the use of your synchrotron radiation facilities and beamtime on beamline ID29. **Aaron Klug Centre for Imaging and Analysis, University of Cape Town**: Thank you and all staff for allowing me to freely use your Rigaku Rotating Anode X-ray diffractometer. **University of Pretoria DNA Sequencing Facility**: Thank you for your support in sequencing DNA fragments. **University of Pretoria International Student Department**: Your effective coordination with international student concerns is greatly appreciated. **South African National Research Foundation**: Your financial support has been crucial to my progress. **Members of the Structural Biology Research Group**: Your critical analysis and support during our weekly meetings have been invaluable. Special thanks to Adewumi Adeyeye and Clifford Ntui. **Administrative staff at the Department of Biochemistry**: Thank you for your contributions to the smooth running of the department. **Religious Groups:** I am grateful for your prayers, encouragement, and celebration of milestones, including Petra Full Gospel Church, Arcadia; Hope Baptist Church, Sunnyside; and the "Prayer Cell Group."

Family Acknowledgments:

Beloved wife, Constance Anye: Your unwavering support and sacrifices have been my cornerstone. **Mr. and Mrs. Ngang Nche-Fambo**: Thank you for hosting and tolerating me and my son during our time in South Africa. **Dr. Azinwi Nche-Fambo epse Ngwnue**: Your understanding and companionship during this journey is deeply appreciated. **Mother, Doris Nchang Chi**: I recognize your dedication in taking care of my children during my studies. **Brother Henry Nji** and **Sisters Henrieta Akwa and Sidoni Zee**: Your unwavering support and encouragement have been a source of strength. **Children**: My beloved children, **Cornelius Anye, Akuma-Brighten Anye, Elyana Anye, and Daniella Anye**, I apologize for the time spent away from you. Your endurance and support have meant the world to me. Akuma-Brighten, your sacrifice of spending 5 months away from your mother at a young age to be with me in South Africa is truly remarkable.

Finally, I offer my heartfelt thanks and gratitude to God Almighty for granting me the ability to learn and achieve this significant milestone.

ABSTRACT

The depletion of fossil fuels, associated pollution, and resulting health hazards are of concern worldwide. Woody biomass provides an alternative source of cleaner and renewable energy by being converted into biofuels through processes like fermentation and saccharification. The efficient use of woody biomass relies on effective xylan depolymerization. Xylans, heteropolysaccharides with 1,4-linked β-D-xylopyranose units, form the main constituent of hemicellulose, the second most abundant wood component. However, their complex structure impedes enzymatic breakdown, elevating biomass recalcitrance and hindering biofuel production efficiency. Successful xylan depolymerization is essential for optimizing the utilization of woody biomass as a sustainable and efficient energy source. Xylan depolymerization involves hemicellulolytic xylanases from glycoside hydrolase (GH) families 5, 8, 10, 11, 30, 43, 62, and 98 in the CAZy database, with xylanases playing a key but not exclusive role in this enzymatic process. We analysed a multidomain xylanase (Xyl) from the hindgut metagenome of the snouted harvester termite *Trinervitermes trinervoides* that releases xylobiose and xylotriose from beech and birch xylan and wheat arabinoxylan. The four domains of Xyl include an N-terminal GH11 xylanase domain, two family 36-like carbohydrate-binding domains Xyl-CBM36-1 and 2, and a C-terminal CE4 acetylxylan esterase domain. This study aimed to explore the structure, function, and biophysical properties of the multidomain xylanase Xyl, examining both individual domains and their combinations. Various molecular biology methods to include gene cloning, protein production and purification and biochemical and crystallography methods were employed. The crystal structures of Xyl-GH11, Xyl-CBM36-1 and Xyl-CE4 domains were solved alongside the crystal structure of the two-domain construct, GH11-CBM36-1. The two-domain crystal structure revealed extensive similarity to known GH11 domain structures, however, there was no electron density corresponding to the Xyl-CBM36-1 domain, suggesting a random orientation of the Xyl-CBM36-1 domain relative to the Xyl-GH11 domain. Isothermal titration calorimetry (ITC) experiments similarly did not reveal specific interactions between the individual Xyl domains, implying a “beads-on-a-string” model for Xyl and its domains.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

3D	Three dimensional
Å	Ångström, 1 Å = 0.1 nm
AXE	Acetyl xylan esterase
bp	Base pair
CASP14	14th Critical Assessment of Structure Prediction
CAZy	Carbohydrate-Active enzymes
CBHI and CBHII	Cellobiohydrolase types I and II
CBM	Carbohydrate-binding module
CCD	Charge-coupled device
CD	Circular dichroism
CDA	Chitin deacetylases
CE	Carbohydrate esterase
*Cf*CBM4-1	CBM4-1 module of *Cellulomonas fimi* endoglucanase CenC
CiCDA	*Colletotrichum lindemuthianum* chitin deacetylases
Coot	Crystallographic Object-Oriented Toolkit
CtCE4	Family 4 carbohydrate esterase from *Clostridium thermocellum*
CV	Column volume
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylene diaminetetra acetic acid
EISA	Energy Independence and Security Act
ESRF	European Synchrotron Radiation Facility
EU	European Union
FPLC	Fast protein liquid chromatography
GH	Glycoside hydrolase
GH11-2CBM	Recombinant construct of domains GH11, CBM36-1 and CBM36-2 of Xyl
GH11-CBM36-1	Recombinant construct of domains GH11 and CBM36-1 of Xyl
GlcNAc	N-acetylglucosamine
GS	Glutathione sepharose
His_6	Hexahistidine tag

IEC	Ion exchange chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRENA	International Renewable Energy Agency
ITC	Isothermal titration calorimetry
IUBMB	International Union of Biochemistry and Molecular Biology
kDa	Kilodalton
LB	Lysogeny broth
LEW	Lysis-Equilibration-Wash buffer
MR	Molecular replacement
MSAs	Multiple sequence alignments
MurNAc	N-acetylmuramic acid
NCBI	National center for biotechnology information
NTA	Nitrilotriacetic acid
N-terminus	Amino terminus
OD_{600}	Optical density at 600 nm
PAE	Predicted alignment error
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
Phenix	Python-based hierarchical environment for integrated X-ray crystallography
pLDDT	Local distance difference test
PNAG	Poly-β-1,6-N-acetyl-D-glucosamine
p-NP	para-Nitrophenol
rmsd	Root-mean-square deviation
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SlCE4	Family 4 carbohydrate esterase from *Streptomyces lividans*
TAE	Tris acetate EDTA
*Tm*CBM4	Family 4 carbohydrate-binding module *Thermotoga maritima*
UNFCCC	United Nations Framework Convention on Climate Change
UV	Ultraviolet

VcCDA	*Vibrio cholerae* chitin deacetylases
w/v	Weight per volume (concentration)
XDS	X-ray Detector Software
Xyl	Xylanase
Xyl-CBM36-1	Carbohydrate-binding domain 1 of Xyl
Xyl-CBM36-2	Carbohydrate-binding domain 2 of Xyl
Xyl-CE4	Family 4 carbohydrate esterase domain of Xyl
Xyl-GH11	Family 11 glycoside hydrolase domain of Xyl

1. Introduction

1.1. Challenges in sustainable energy

Mineralised organic compounds of biological origin remain the primary source of energy for human economies. The continued global demand for these compounds in the form of fossil fuels, and the increased adverse effects of fossil fuel consumption on human health and the environment, have concerned individuals, researchers, grass-root organisations, governments and intergovernmental organisations for decades with increasing urgency (IEA, 2020; Johnsson et al., 2018). Governments have therefore increased their support for research into renewable and sustainable sources of energy. The 2009 European Union Renewable Energy Directive, for example, recommended that by 2020, 20 % or more of all energy consumed within the EU should be from renewable sources, with 10 % of fuels for road transport provided by biofuels (Fivga et al, 2019; ICCT, 2018). In 2018, the targets were elevated to 32 % and 14 % respectively by 2030. First generation biofuels should constitute less than 50 % of the latter, with second and third generation biofuels providing the balance (ICCT, 2018). First-generation biofuels referred to traditional biofuels produced from food crops or plant-based feedstocks whereas, second-generation biofuels are those derived from non-food crops, agricultural residues, and lignocellulosic biomass (Kowalski et al., 2022). The United States of America Energy Independence and Security Act (EISA) of 2007 similarly proposed eliminating 15 %, ~ 3-million barrels of oil used daily, or 8 % of annual projected energy use and 10 % of annual carbon dioxide emissions (50 million tons) by 2030 (Ungar et al., 2015). EISA also forecast an annual biofuel production of 860 million barrels of oil by the end of 2023.

African nations are not excluded in race of developing renewable and sustainable energy sources (Oyewo et al., 2023; Amoah et al., 2022). Their aim, however, is mainly to avoid the vagaries of crude oil exports, curb air pollution, improve energy security, create employment and reduce dependence on fossil fuels (Okafor et al., 2021; de Vries et al., 2007). According to the 2021 annual report of the United Nations Framework Convention on Climate Change (UNFCCC), Africa is particularly vulnerable to climate change due to rapid population growth, extensive subsistence agriculture, and growing water crises. Paradoxically, Africa provides

11 % of world oil, second only to the Middle East, yet consumes only 4 % of oil globally (Aliyu et al., 2018).

South Africa, Africa's second largest economy, started the transition to renewable energy in 2003. Its 2003 Renewable Energy White Paper defined targets for renewable energy usage and identified hydropower, wind, solar and biomass as the four types of renewable energy (Aliyu et al., 2018). The country already consumes biomass energy with paper packaging and sugar mills burning bagasse and biomass in generating ~210 GWh of electricity per year. Regions particularly involved include Kwazulu-Natal and Mpumalanga with ~4300 km^2 of sugar cane plantations and 13 000 km^2 of planted forests (Aliyu et al., 2018).

Reliable data on sustainable bioenergy sources for developing countries are rare (Sharma et al., 2020). Nevertheless, the International Renewable Energy Agency (IRENA) has projected that 58 % of worldwide transportation fuel will be "renewable" by 2050 (IEA, 2020; Fivga et al., 2019).

1.2. Classification of biofuels

1.2.1. First generation biofuels

A reduced reliance on fossil fuels depends on sustainable and renewable sources of energy. The primary first-generation biofuel is bioethanol, which can serve as the sole fuel or be blended with gasoline for existing gasoline engines (Safarian and Unnthorsson, 2018). Bioethanol has a higher oxygen content than gasoline, yielding a higher combustion efficiency and octane number, and allowing engines to operate at higher compression ratios (Branco et al., 2019). Maize, sugarcane, grains, and sugar beets have all been converted to bioethanol. Brazil and the USA currently contribute 59 % and 27 % of global biofuel production (Branco et al., 2019; Tran et al., 2019), respectively derived from sugarcane and corn starch (Devarapalli et al., 2015; Tran et al., 2019; Rosales-Calsderon and Arantes, 2019). Bioethanol production

in the USA is expected to increase to 515 billion litres by the end of 2023 (Menon and Rao, 2012; Tran et al., 2019). Figure 1.1 depicts fuel production from food crops.

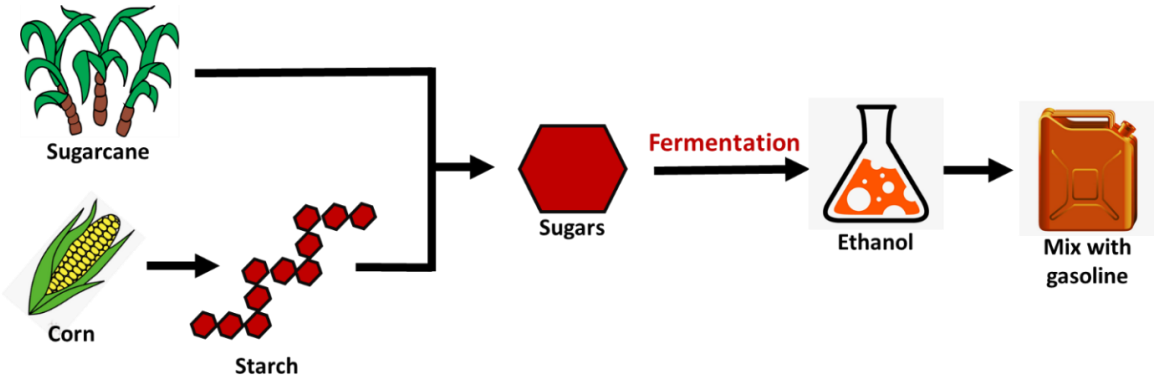


Figure 1.1: First-generation biofuel production

First-generation biofuels, while renewable, are based on food crops as feedstock (Dahman et al., 2019). This raises multiple ethical, socio-economic and environmental concerns primarily due to biofuel production competing with food production for water and arable land, inflating food and livestock feed prices, and also adversely affecting ground water and soil quality *inter alia* through over-fertilization (Buijs, Siewers and Nielsen, 2013).

1.2.2. Second generation biofuels

Second generation biofuels are generated from non-edible biomaterials such as agricultural and forest waste rich in lignocellulose, complemented by, industrial by-products including whey and crude glycerol (Kowalski et al., 2022; Thompson and Meyer, 2013; Balat et al., 2008). Figure 1.2 depicts the valorisation process of non-edible biomaterials.

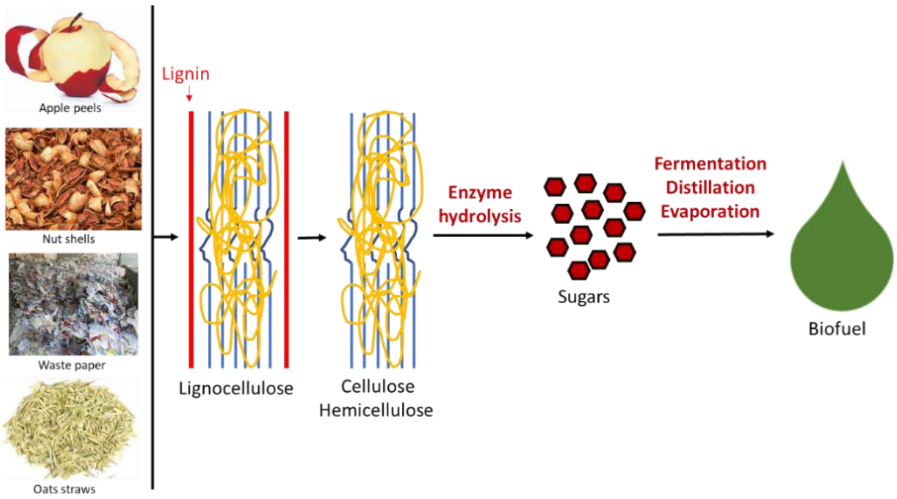


Figure 1.2: Conversion of lignocellulose biomass to liquid biofuels

Second generation feedstocks are rich in lignocellulose, the most abundant biomaterial on earth by mass (Zoghlami & Paës, 2019; Sánchez & Cardona, 2008). While lignocellulose is readily available, renewable and sustainable, its depolymerisation to simple fermentable sugars presents a significant challenge to the biofuels industry. Crude glycerol, a by-product of industrial biodiesel production, is converted to phosphoenolpyruvate or pyruvate before yielding bioethanol through pathways other than fermentation (Viegas et al., 2022; Chen & Liu, 2016; Yazdani and Gonzalez, 2007). Overall, lignocellulosic biofuels are environmentally less damaging than first generation biofuels (Rodionova et al., 2022; Hahn-Hägerdal et al., 2006).

1.2.3. Third generation biofuels

Third generation biofuels are derived from microalgae growing on sediments or in columns of fresh or marine water (Figure 1.3). Microalgae grow rapidly, are photosynthetically highly active, rapidly incorporate carbon dioxide, have a high lipid content, and grow at high densities on small footprints. These allow for continuous harvesting of biomaterials ideal for large-scale biofuel production.

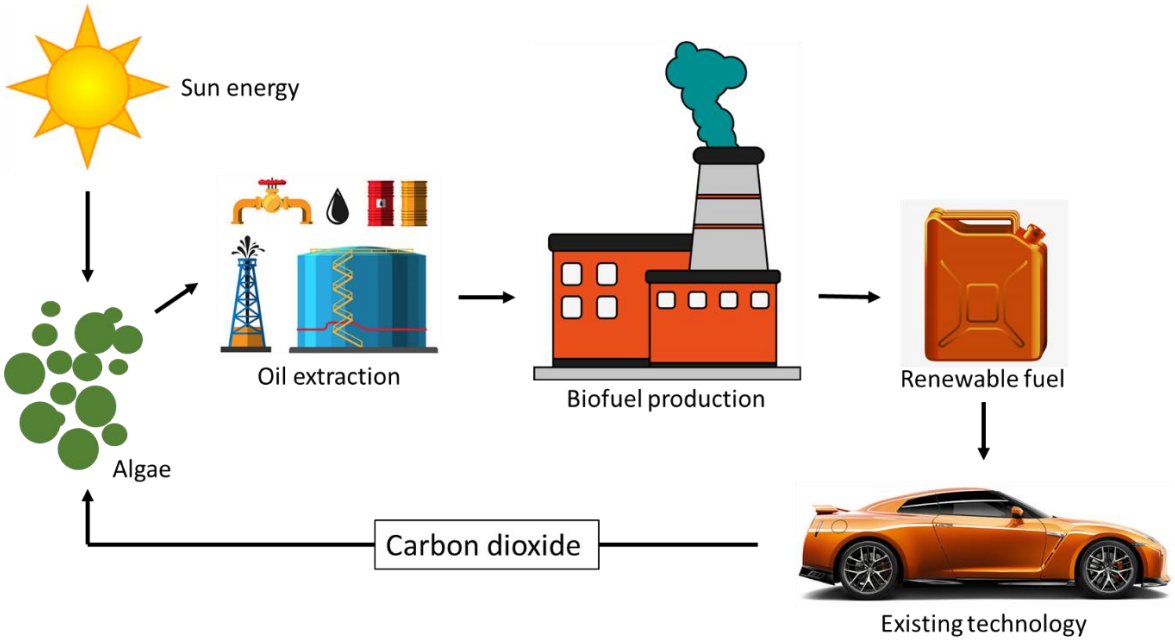


Figure 1.3: Biofuel production from microalgae

Environmental benefits of microalgae-based biofuels include: 1) Microalgae extract nutrients from industrial waste, cleaning industrial waste streams and reducing hazards. 2) Microalgae sequester carbon dioxide reducing atmospheric green-house gases. 3) Some microalgae naturally produce bioethanol while others produce mono- and polyunsaturated

hydrocarbons that can be converted to gasoline-like fuels (Sivaramakrishnan et al., 2022; Koller et al., 2012). 4) Microalgal oils can replace vegetable oils in the production of biodiesel (Zhang et al., 2022; Yazdani and Gonzalez, 2007).

1.3. Lignocellulose feedstock for biofuel production

Lignocellulose comprises non-edible plant materials. It is composed of cellulose, hemi-cellulose, and lignin (Himmel et al., 2007), generated by photosynthetic fixation of carbon dioxide and stored in plant cell walls. The global annual production of lignocellulose biomass is ~180 billion tonnes (Dahmen et al., 2019) providing a reliable, renewable, and sustainable feedstock for biofuel production. Cellulose constitutes 35 to 50 % of the lignocellulose dry mass, is the main form of renewable carbon on earth, and exists in both crystalline and amorphous forms. The cellulose backbone is a linear chain of β-1,4 linked glycosyl molecules laterally stabilised by hydrogen bonds into microfibrils (Bayer et al., 1998; Nishiyama et al., 2002). The microfibrils assemble into crystalline structures that are only solubilised at temperatures over 320 °C and pressures of over 2.5 MPa (Taylor et al., 2019).

At 20 to 40 % of lignocellulose dry mass, hemicellulose is next to cellulose as a source of renewable carbon (Ostadjoo et al., 2019). Its molecular backbone consists of branched heteropolysaccharide chains of pentose (D-xylose and L-arabinose), hexose (D-mannose, D-galactose, and D-glucose) and uronic acid (D-glucuronic acid and D-galacturonic acid) units (Ebringerová et al., 2005; Fraser-Reid et al., 2008; Saha, 2003; Scheller and Ulvskov, 2010).

Lignin, constituting 10 to 25 % of lignocellulose dry weight, is a highly branched and complex phenolic macromolecule derived from three hydroxylcinnamyl alcohol monomers: p-coumaryl, coniferyl and sinapyl alcohol (Figure 1.4) (Boerjan et al., 2003; Lapier re et al., 1986).

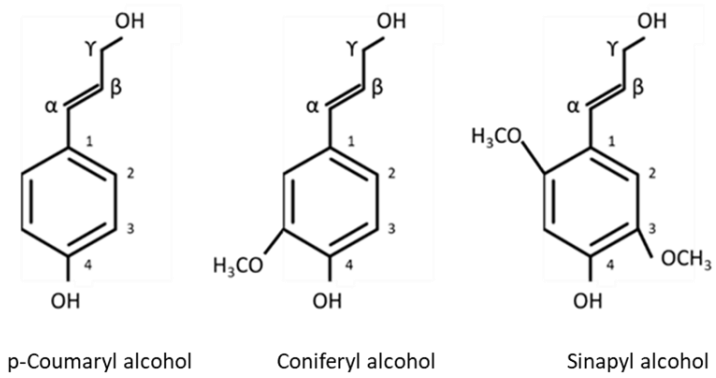


Figure 1.4: Lignin monomers

Three phenylpropanoid units are transformed into three monolignols, p-hydroxyphenyl, guaiacyl, and syringyl, which together create the three-dimensional amorphous lignin polymer. Lignin in primary cell walls provides structural support, impermeability, and protection against microbial assault (Saha, 2003). Within the cell wall, lignin is invariably covalently bonded to hemicellulose α-carbons via a benzene C-4 (Buranov and Mazza, 2008; Ralph et al., 1995). The heterogeneous linkages within lignin and between lignin and surrounding polymers complicates its isolation for structural analysis. This is further complicated by the tight packing of cellulose fibres and crosslinks to hemicellulose and lignin, together with hydrogen bonding between the chains. As a result lignocellulose is exceptionally rigid and recalcitrant to depolymerisation (Behera et al., 2014; Yamada et al., 2013).

1.3.1. Categorization of Hemicelluloses

Hemicelluloses are categorized into several groups, including xyloglucan, xylans, mannans, glucomannans, and β-(1→3,1→4)-glucans (Scheller & Ulvskov, 2010). Polysaccharides such as galactans, arabinans, and arabinogalactans are excluded from the hemicellulose group due to their association with pectin molecules and differing structural configurations (Lara-Espinoza et al., 2018; Scheller & Ulvskov, 2010). Calloses are also excluded from hemicellulose due to their distinct β-(1→3) linked glucose backbones (Wang et al., 2022; Chen & Kim, 2009).

The term "hemicelluloses" has been used historically to describe polysaccharides with β-(1→4)-linked backbones of glucose, mannose, or xylose, which are neither cellulose nor pectin. Despite its archaic nature and suggestions for alternative terminology, such as "cross-linking glycans," the term hemicelluloses remains widely used in the scientific litereature for its convenience in describing this group of wall polysaccharides (Scheller & Ulvskov, 2010).

1.3.1.1. Xyloglucan (XyG)

Xyloglucan (XyG) is ubiquitous in land plants, except for Charophytes. It constitutes a significant portion of primary walls in spermatophytes, with variations in structure (Schultink et al., 2014). While the basic structure of XyG involves repetitive units, its side chains exhibit considerable diversity, denoted using a special one-letter code (Pauly & Keegstra et al., 2016). XyG structures vary across plant families, with differences in branching patterns and charge.

The presence of charged side chains is observed in mosses and liverworts, while vascular plants typically exhibit neutral XyG (Scheller & Ulvskov, 2010).

1.3.1.2. Xylans

Xylans represent a diverse group of polysaccharides characterized by a β-(1→4)-linked xylose backbone. Common modifications include substitution with glucuronosyl and 4-O-methyl glucuronosyl residues (Saroj et al., 2018; Smith et al., 2017). In dicots, xylans are often glucuronoxylans, while commelinid monocots primarily contain arabinoxylans and glucuronoarabinoxylans in their primary walls. Xylan structures vary widely across plant species, with branching patterns correlating with taxonomy (Curry et al., 2023). Grass xylans are notable for their ferulate esters, which contribute to cross-linking and wall recalcitrance (Zhang et al., 2023).

The complexity of xylans arises from their diverse structures, which can vary depending on the source and biological function (Curry et al., 2023). In this section, we delve into the various types of xylans, including arabinoxylans (AX), glucuronoarabinoxylans (GAX), glucuronoxylans (GX), and their respective structures, shedding light on their significance in biological processes and industrial applications.

A) Arabinoxylans

Arabinoxylans (AX) are one of the most abundant hemicelluloses in the plant kingdom, primarily found in cereals such as wheat, rye, barley, and rice. Structurally, AX consists of a linear backbone of β-(1→4)-linked xylose units, with arabinose side chains attached predominantly at the O-3 position of the xylose residues (Navarro et al., 2019). The degree and pattern of arabinose substitution can vary among different plant species and tissues (Pang et al., 2023). Additionally, ferulic acid residues may esterify the arabinose side chains, contributing to cross-linking within the cell wall matrix (de O Buanafina et al., 2020).

B) Glucuronoarabinoxylans

Glucuronoarabinoxylans (GAX) represent a subgroup of xylans characterized by the presence of both arabinose and glucuronic acid residues in their side chains (Smith et al., 2017). These polysaccharides are commonly found in grasses, including cereals and certain hardwoods. The backbone structure of GAX resembles that of AX, consisting of β-(1→4)-linked xylose units (Kmezik et al., 2020). Arabinose residues are attached to the xylose backbone at the O-3

position, similar to AX. Additionally, glucuronic acid residues can be present in the side chains, usually attached to the arabinose residues (Zannini et al., 2022). The presence of glucuronic acid imparts unique properties to GAX, influencing its solubility, interaction with other cell wall components, and biological functions (Rudjito et al., 2023).

C) Glucuronoxylans

Glucuronoxylans (GX) are characterized by the presence of glucuronic acid residues attached directly to the xylose backbone, distinguishing them from GAX. GX is prevalent in various hardwoods and non-graminaceous plants. Structurally, GX comprises a linear chain of β-(1→4)-linked xylose units, with glucuronic acid substituents attached at the O-2 position of the xylose residues (Smith et al., 2017). The distribution and frequency of glucuronic acid substitutions can vary, influencing the physicochemical properties and functionalities of GX in plant cell walls.

D) Arabino-glucuronoxylans

Some xylans exhibit mixed substitutions of both arabinose and glucuronic acid residues, resulting in arabino-glucuronoxylans (AGX) or glucuronoarabinoxylans (GAX) depending on the predominant type of side chain substitution (Smith et al., 2017). These hybrid structures are commonly found in various plant species and contribute to the complexity of hemicellulose composition in cell walls (Huang et al 2021; Scheller & Ulvskov, 2010).

Understanding the diverse structures and compositions of xylans, including AX, GAX, GX, and AGX, is crucial for elucidating their roles in plant physiology, as well as for harnessing their potential in various industrial applications, such as food processing, biofuel production, and biomaterials development.

1.3.1.3. Mannans and Glucomannans

β-(1→4)-linked polysaccharides containing mannose are widely distributed and play essential roles, particularly in early land plants, mosses, and lycophytes (Mikkelsen et al., 2014; Schröder et al., 2009). In spermatophytes, mannans and glucomannans are less abundant but still present in various cell walls, with roles in seed storage and structural integrity (Fradera-Soler et al., 2022).

1.3.1.4. β-(1→3,1→4)-Glucans

Found primarily in grasses, β-(1→3,1→4)-glucans are characterized by interspersed single β-(1→3) linkages (Chang et al., 2021). These glucans are growth-stage dependent and play roles in cell expansion. Recent studies suggest their presence in diverse taxa, indicating a potentially ancient trait (Purushotham et al., 2022).

1.4. Pre-treatment of lignocellulosic biomass

Industrial utilization of lignocellulose starts with a pre-treatment to reduce cell-wall complexity and increase enzyme accessibility for improved hydrolysis (Ragauskas et al., 2006). Pre-treatment, which includes solid processing, liquid handling and inhibitory by- and co-products treatment is an essential but costly process (Menon and Rao, 2012). Effective pre-treatment must accommodate all types of lignocellulosic biomass, recover sufficient fermentable sugars and lignin for downstream processes, generate minimal co-products or inhibitors, and minimize energy input and costs (Prasad et al., 2016; Kumar and Sharma, 2017).

1.4.1. Pre-treatment technologies

The technology used for lignocellulose pre-treatment depends on the lignocellulose biomass type, and considerations on the environmental impact (Liyamen and Ricke, 2012; Menon and Rao 2012). Available pre-treatment methods are summarised in Table 1.1.

Table 1.1: Lignocellulose pre-treatment methods

Biomass				
Forest wood	Herbaceous biomass	Agricultural residues	Municipal solid waste	Microalgae
Pre-treatment methods				
Chemical • Acid • Alkali • Oxidative	**Physical** • Milling • Extrusion • Freezing • Microwave • Ultrasound	**Physio-chemical** • Steam explosion • Ammonia fibre explosion • Liquid hot water • Hydrothermal liquefaction		**Biological** • Fungi • Bacteria
Output				
Cellulose		Hemicellulose		

1.4.1.1. Physical or mechanical pre-treatment

Physical or mechanical forms of pre-treatment improve the rate and lower the costs of hydrolysis and include milling, extrusion, freezing, ultrasound, and microwave exposure. Pre-treatment reduces cellulose crystallinity enhancing enzyme and solvent accessibility.

1.4.1.2. Chemical pre-treatment

Most pre-treatment regiments include the addition of chemicals to delignify cellulose during pulp production. These regiments include strong and weak acids and bases, pH controlled hot water and ionic liquids.

a) Acid pre-treatment

Sulphuric, nitric, phosphoric, and hydrochloric acids are generally combined with temperatures above 180 °C and reaction times of 1 to 5 min or with temperatures below 120 °C and reaction times of 30 to 90 min (Kumar & Sharma 2017). For concentrated acids, temperature below 100 °C may suffice whereas dilute acids require temperatures between 100 and 250 °C. Acids are highly corrosive and create inhibitory by-products (Kumar and Sharma, 2017) such as hydroxymethylofurfural and acetic acid (Jędrzejczyk et al., 2019). Acid treatment preceded by alkali pre-treatment effectively removes hemicellulose and lignin, yielding pure cellulose.

b) Alkaline pre-treatment

Alkaline solutions disrupt ester and glycosidic side chains and interfere with cellulose recrystallization and swelling. Common bases include NaOH, KOH, NH_4OH and $Ca(OH)_2$ of which the $Ca(OH)_2$ is the least expensive. The combination of NH_4OH with dilute sulphuric acid removes up to 97 % of lignin and 91 % of cellulose from rice straw (Kim et al., 2011), respectively, while $Ca(OH)_2$ removed the lignin component from corn cob residues doubling the yield of biogas relative to untreated cobs (Baruah et al., 2018).

1.4.1.3. Physico-chemical pre-treatment

Physical and chemical treatment (Figure 1.5) of lignocellulose improves yields. The biomass is physically chipped, ground and processed to reduce cellulose crystallinity and increase its surface area for better enzymatic hydrolysis. Heat, moisture, pressure, and chemicals are used to increase cellulose availability. Physically milling corn stover and exposing it to alkaline

improved enzymatic hydrolysis by 110 % (Kumar and Sharma, 2017). Physico-chemical pre-treatment can also combine ammonia fibre explosion, ammonia recycle percolation, steam explosion, carbon dioxide explosive, microwave-chemical, and hot water (Baruah et al., 2018; Kim, 2018). Combining physical and chemical pre-treatment is universally more efficient, less time consuming and produces fewer inhibitory products than applying each method separately.

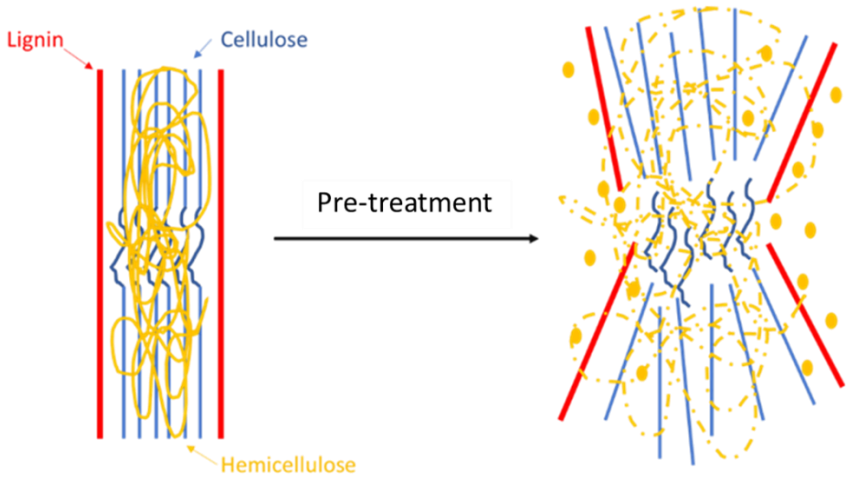


Figure 1.5: Lignocellulose pre-treatment disrupts its tightly packed structure.

1.4.1.4. Biological pre-treatment

Biological pre-treatment involves the use of microorganisms like fungi and bacteria that directly modify lignocellulosic biomass and degrade it to simpler substrates for hydrolysis by lignin and/or hemicellulose-degrading enzymes. Biological pre-treatment has the advantages of low energy requirements, (almost) no chemical treatment, mild reaction conditions, high specificity, and no fermentation inhibitors. Biological pre-treatment is environmentally friendlier than all other techniques. Nevertheless, it is a slow process that requires close monitoring of growth conditions. Additionally, lignin-feeding microorganisms also feed on hemicellulose and cellulose (Cheah et al., 2020).

Cellulolytic and hemicellulolytic microorganisms hydrolyse cellulose and hemicellulose to their monomeric sugars (Sharma et al., 2017; Koupaie et al., 2019). Lignin is similarly hydrolysed by lignin-degrading enzymes from white, brown and soft rot fungi. White rot fungi are commonly used in industry due to their higher sugar yields (Baruah et al., 2018; Waghmare et al., 2018). Microorganisms degrade and ferment lignocellulose to biofuels such as bioethanol, hydrogen and methane, as well as biomaterials including lactate, acetate and

organic acid (Sharma et al., 2017; Xu et al., 2017). Apart from fungi and bacteria, worms (Devi et al., 2019), insects (Varelas and Langton, 2017) and gastropods (Trincone, 2018) have been used for biomass pre-treatment. These organisms have well developed mechanisms for cellulosic biomass digestion that include mechanical and enzymatic processing, gut flora as well as adapted physiologies.

Physical, chemical, or biological factors affect biological pre-treatment, where physical factors include reaction temperature, moisture, incubation time, aeration, subtract size, and accessible surface area. Chemical factors involve the pH, composition of culture media, while biological factors include sources of carbon and nitrogen. Microbial species, microbial consortia, microbial interactions, and the nature of microbial competition also affect biological factors that influence pre-treatment (Vasco-Correa et al., 2019).

1.5. Enzymes in lignocellulose biodegradation

Cellulases and hemicellulases are the principal enzymes used in lignocellulose hydrolysis. Cellulases and hemicellulases mostly hydrolyse cellulose and hemicellulose, though some cellullytic enzymes mixtures like Celluclast from *Trichoderma reesei* (Silva et al., 2020) also break down hemicelluloses. The enzymes are added to reactors as crude or purified extracts or produced *in situ* by hemicellulolytic microorganisms.

1.5.1. Cellulases

Three cellulase sub-groups help to convert cellulose to glucose monomers: endoglucanases (EC 3.2.1.74), exoglucanases (EC 3.2.1.91) and β-glucosidases (EC 3.2.1.21) (Ezeilo et al., 2017; Szydlowski et al., 2015; Chang, 2007). Endoglucanases hydrolyse internal β-1,4 glycosidic bonds in amorphous cellulose and release oligosaccharides of varying lengths. Exoglucanases, by contrast, remove terminal cellobiosyl residues and to a lesser degree, glucose and other cello-oligosaccharides. Two groups of exoglucanases, CBHI and CBHII, respectively degrade the reducing and non-reducing ends of the oligosaccharides. Exoglucanases can also slowly degrade microfibril structures, decreasing cellulose polymerization. β-glucosidases digest cellobiose and cellodextrins, releasing glucose monomers (Kim et al., 2022).

Most cellulases combine a catalytic domain with a carbohydrate binding module (CBM) connected by a linker (Ye et al., 2022; Kari et al., 2021). The CBM presumably allows the

enzyme to directly bind to cellulose, placing the cellulase domain in proximity to its insoluble substrate (Kari et al., 2021). Enzymes function well in synergy such that the combined activity of enzymes in a cellulolytic enzyme system exceeds the sum of individual activities. This synergy is linked to 1) cooperation of CBM and catalytic domain in substrate binding; 2) cooperation between endoglucanases and exoglucanases in breaking down cellulose into oligomers; 3) CBHI and CBHII exo-exo synergism in cleaving cellulose at both ends, 4) synergy between exoglucananases and β-glucosidases in converting cellobiose into glucose (Ye et al., 2022; Kari et al., 2021).

1.5.2. Hemicellulases

Maximum utilization of biomass must account for the hemicellulosic component of lignocellulose (Østby & Várnai, 2023). However, the structural complexity of the latter calls for the application of diverse hemicellulosic enzymes for efficient depolymerisation. Three groups of hemicellulytic enzymes include:

(1) Depolymerising hemicellulases release short chain oligosaccharides from hemicellulose backbones. Examples include endo-1,4-β-xylanase and endo-1,4-β-mannanase that respectively hydrolyse heteroxylans and heteromannans.
(2) Accessory hemicellulases reduce spatial hindrance of main chains to enhance the activity of depolymerisation hemicellulases. Accessory hemicellulases include α-glucuronidase, α-galactosidases, β-glucosidases and esterases.
(3) Oligosaccharide-cleaving hemicellulases that breakdown oligo- into monosaccharides. Examples include β-xylosidases, β-mannosidases, and α-L-arabinofuranosidases that respectively produce xylose, mannose, and arabinose from their substrates.

Hemicellulose and cellulose are closely knit in lignocellulose such that the depolymerisation of one enhances accessibility of the other (Liu & Qu 2021). Thus, hemicellulases function in synergy both within and with other lignocellulytic enzymes (Benatti et al., 2023).

1.5.2.1. Heteroxylan hydrolysis by hemicellulases

The enzymes endo-1,4-β-xylanase (EC3.2.1.8), α-L-arabinofuranosidase (EC 3.2.1.55), and β-xylosidase (EC 3.2.1.37) (Figure 1.6) are key in hydrolysing heteroxylan. Endo-1,4-β-xylanases selectively hydrolyse β-1,4-linkages between xylosyl units from low to moderately substituted regions in xylan to initially produce short xylo-oligosaccharides followed by xylobiose and

xylotriose after prolonged incubation. Debranching α-L-arabinofuranosidases release α-(1→2) and/or α-(1→3) linked arabinofuranose residues from soluble products of endo-1,4-β-xylanases (Østby & Várnai, 2023; Linares-Pastén et al., 2021). Finally, β-xylosidases attack the non-reducing ends of xylooligosaccharides to liberate terminal xylosyl residues, helping to decrease cellulases inhibition by xylooligosaccharides (Rohman et al., 2019).

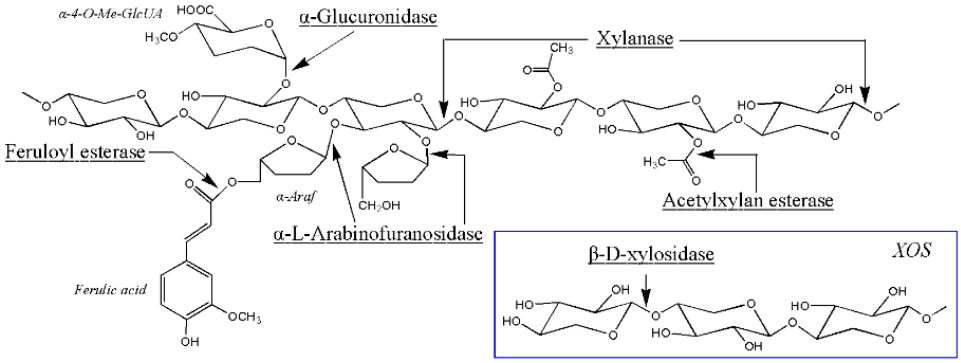


Figure 1.6: Hydrolysis of heteroxylan by hemicellulases

Enzyme names are underlined. Adapted from (Dodd and Cann, 2009)

A plethora of accessory hemicellulases aid heteroxylan hydrolysis. These accessory enzymes include α-glucuronidases (EC 3.2.1.139) that remove α-1,2-linked glucuronic acid residues in glucuronoxylan and glucurono-arabinoxylan, acetyl-xylan esterases (EC 3.1.1.72) that cleave acetyl group substitutions at C-2 and C-3 positions of xylosyl backbones, and ferulic acid/ p-coumaric acid esterases (EC 3.1.1.73) that remove phenolic acids linked to arabinosyl residues (Yang et al., 2023; Yagi et al., 2017; Lin et al., 2023 Gao et al., 2011).

In nature, many bacteria and fungi are xylanolytic (Li et al., 2022; Chakdar et al., 2016). Various genes encoding hemicellulases have been cloned into *S. cerevisiae* to create recombinant strains suitable for industrial applications (Cunha et al., 2020; Salazar-Cerezo et al., 2023). The xylanosome, a complex xylanolytic system, is found in some anaerobic bacteria. The extracellular xylanosome from *Butyvibrio fibrisolvens* H17c consists of eleven subunits with xylanase activity and three subunits displaying endoglucanase activity (Lin and Thomson, 1991). Likewise, the xylanosome in *Streptomyces olivaceoviridis* comprises of at least four xylanase units, and one xylan binding module (Jiang et al., 2006).

Three enzyme systems are essential recognized in the comprehensive deconstruction of biomass. In the free extracellular enzymes system, enzymes operate independently in the

extracellular milieu, catalyzing the breakdown of cellulose and hemicellulose (Zhao et al., 2020; Shi et al., 2019). The cellulo or xylanosomes system encompasses enzyme complexes anchored to microbial cell surfaces, enhancing substrate proximity and enzymatic efficiency (Tsai et al., 2022; Dou et al., 2018; Srikrishnan et al., 2013). Lastly, the multifunctional or fusion enzymes system integrate various catalytic activities within a single polypeptide chain, streamlining the degradation process (Wang et al., 2023a; Wang et al., 2023b).

1.6. Endo-β-1,4-xylanases

Endo-1,4-β-xylanases (EC 3.2.1.8) constitute a well characterised group of enzymes frequently referred to as xylanases. They randomly depolymerize β-1,4 linked heteroxylan to short chain xylooligosaccharides of varying lengths. In some cases, endo-1,4-β-xylanases release xylose as their main product (Mendonça et al., 2023; Pavarina et al., 2021) . Endoxylanases release either xylose and xylobiose, or xylooligosaccharides rather than arabinose. They do not hydrolyse xylotriose or xylobiose. Debranching or arabinose-liberating enzymes either hydrolyse branch points to produce xylooligosaccharides and arabinose or cleave xylan main chains and branch points to produce xylobiose, xylose, and arabinose (Nordberg Karlsson et al., 2018; Wang et al., 2022).

Endo-1,4-β-xylanases are prevalent among the bacterial and fungal genera *Bacillus*, *Aspergillus*, *Penicillium*, and *Streptomyces*. Xylanase production is mostly induced by xylan in natural hosts or by xylan derivatives in culture medium (Bhardwaj et al., 2019). Compared to their bacterial counterparts, fungal xylanolytic organisms thrive better under acidic conditions (Verma & Satyanarayana, 2020). Some fungal species produce multiple xylanases with fifteen and thirteen extracellular xylanases identified in *Aspergillus niger* and *Trichoderma viride* cultures, respectively (Bhardwaj et al., 2019).

Post-translational glycosylation is most common in eukaryotic endoxylanases compared to their prokaryotic counterparts. Nevertheless, both *Bacillus sp*. and *Clostridium stercorarium* produce glycosylated xylanases. Post-translational xylanase glycosylation improves enzyme stability and protection in extreme environments (Kiribayeya et al.,2022).

1.6.1. Classification of endo-β-1,4-xylanases

The most detailed classification of endo-β-1,4-xylanases is provided by the Carbohydrate-Active Enzymes (CAZy) database (www.CAZy.org). The database organizes enzymes into families by amino acid sequence similarity (Drula et al., 2021; Henrissat and Bairoch, 1993). While the IUBMB enzyme classification system focuses on the chemical reaction, CAZy considers enzyme structure and predicts fundamental mechanistic and structural information for families based on biochemical properties of at least one member. Importantly, CAZy links related enzymes with distinct substrate specificities related by divergent evolution. Presently 173 GH families have been defined. CAZy also defines clans composed of CAZy families with related 3D architectures. Clans thus also indicate divergent evolution (Drula et al., 2021; Henrissat and Bairoch, 1996). The current 14 clans and their associated folds are summarized in Table 1.2.

Endo-1,4-β-xylanases are part of GH families 8, 10, 11, 30, 43, 62, and 98, all with single catalytic domains. GH 5 and 10 from clan GH-A share a common architecture. Families GH 8 and 11 belong to clans GH-M and -C, respectively (Table 1.2). Xylanases of GH 10 and 11 are particularly relevant industrially and have been most studied (Zhang et al., 2021). Bifunctional enzymes with xylanase activity form part of GH-16, 43, 52 and 62 (Li et al., 2022; Drula et al., 2021).

Table 1.2: Clans of glycoside hydrolases

Clan	Structural fold	GH family
GH-A	$(\beta/\alpha)_8$	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113, 128, 140, 147, 148, 157, 158, 164, 167
GH-B	β-jelly roll	7, 16
GH-C	β-jelly roll	11, 12
GH-D	$(\beta/\alpha)_8$	27, 31, 36
GH-E	6-fold β-propeller	33, 34, 83, 93
GH-F	5-fold β-propeller	43, 62, 117
GH-G	$(\alpha/\alpha)_6$	37, 63, 100, 125
GH-H	$(\beta/\alpha)_8$	13, 70 77
GH-I	α+β	24, 80
GH-J	5-fold β-propeller	32, 68
GH-K	$(\beta/\alpha)_8$	18, 20, 85
GH-L	$(\alpha/\alpha)_6$	15, 65
GH-M	$(\alpha/\alpha)_6$	8, 48
GH-N	β-helix	28, 49
GH-O	$(\alpha/\alpha)_6$	52, 116

GH-P	$(\alpha/\alpha)_6$	127, 146
GH-Q	$(\alpha/\alpha)_6$	94, 149, 161
GH-R	$(\beta/\alpha)_8$	29, 107

1.6.2. Active site and catalytic mechanisms of Endo-β-1,4-xylanases

The active site of endo-1,4-β-xylanases contains a pair of acidic residues within the substrate binding site acting as a catalytic acid and base, or electrophile and nucleophile (Zvelebil and Sternberg, 1988). The substrate binding site consists of sub-sites for individual glycoside units denoted by integer values where positive and negative numbers lie on the reducing (aglycon) and the non-reducing (glycon) sides, respectively (Figure 1.7) (Davies et al., 1997). The site of hydrolysis lies between units +1 and -1.

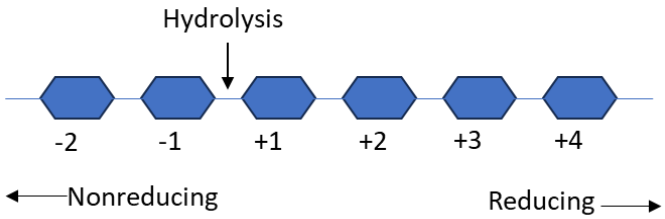


Figure 1.7: Substrate-binding sub-sites in glycosidases

1.6.3. Catalytic mechanism

Glycoside hydrolases either retain or invert the configuration of the anomeric carbon during catalysis (Akintola et al., 2021).

1.6.3.1. Retaining mechanism

The catalytic mechanism of configuration retaining hydrolases employs two carboxylic acid residues (aspartate or glutamate) located about 5.5 Å apart that respectively function as catalytic acid/base and nucleophile. The residues facilitate a two-step catalytic or double displacement mechanism (Figure 1.8A).

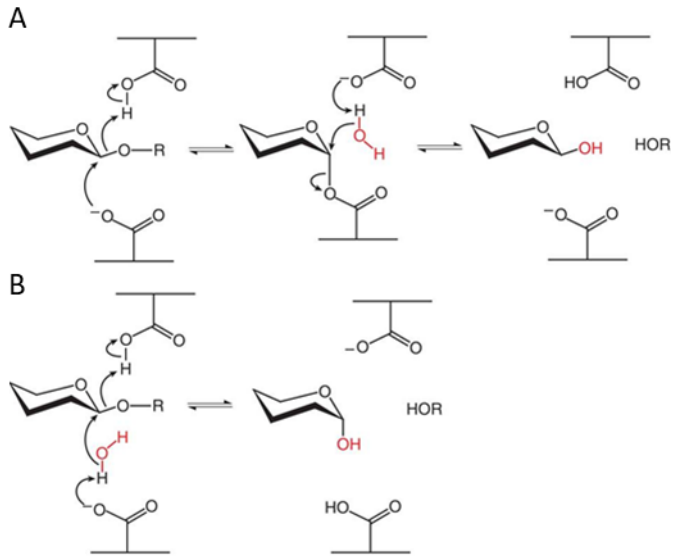


Figure 1.8: Retaining (A) and inverting (B) reaction mechanisms of glycosidases

First, the two carboxylic acid residues simultaneously form a covalent enzyme-glycosyl intermediate with an inverted anomeric carbon and an aglycon leaving group. One residue acts as a Lewis acid and the other as a nucleophile. In a second step, the now deprotonated carboxylate, acts as a general base to capture a proton from an incoming water molecule. The water molecule simultaneously hydroxylates the anomeric carbon of the covalent glycosyl-enzyme intermediate, re-inverting the anomeric configuration. At this point, the glycon product whose anomeric carbon configuration is unchanged from that of the substrate is released and the enzyme returns to its ground state (Pengthaisong et al., 2021, McCarter & Withers, 1994).

1.6.3.2. Inverting mechanism

Like retaining glycoside hydrolases, inverting enzymes also depend on two carboxylic acid residues for catalysis (Figure 1.8B). The reaction is, however, accomplished in a single step. As before, one carboxylic residue acts as an acid/base inducing the leaving group departure, whereas the other provides base-assistance to a nucleophilic attack by a water molecule. The nucleophilic attack occurs on the opposite side of the sugar ring cleaving the glycosidic bond and inverting the anomeric configuration.

Retaining and inverting enzymes typically differ with respect to the distance between the two catalytic carboxylic acids residues, with a distance of ~5.5 Å in retaining enzymes and ~6.5 to 9.5 Å in inverting enzymes (Pengthaisong et al., 2021; McCarter & Withers, 1994).

Xylanases in GH families 5, 10, 11 and 16 are retaining enzymes using two catalytic glutamate residues whereas GH 8 and 43 xylanases are inverting enzymes using a glutamate and an aspartate residue (Kim et al., 2023; Bhardwaj et al., 2019).

1.6.4. Industrial application of xylanases

Xylanases are industrially very useful enzymes. Paper and pulp producers for instance use xylanases to bleach pulp - a cheaper and environmentally less harmful process compared to chlorine (Vries & Visser, 2001). In animal feed production, xylanases help to break down arabinoxylans, improving nutrient accessibility (Polizeli et al., 2005). Xylanases in food companies are useful in the preparation of bread, beer, wine, juices, and some sweeteners (Machado De Castro et al., 2018). Xylanases increase bread volume by solubilising the arabinoxylan component of the dough (De Vries and Visser, 2001). They improve the aroma of wine, filter beer and juices by hydrolysing arabinoxylan to xylo-oligosaccharides. Yeast xylanases work in synergy with xylose reductases to reduce xylose to xylitol, a well-known non-carcinogenic sweetener for diabetic and obese individuals (Ahuja et al., 2020). In the textile industry, cellulase-free xylanases process plant fibres to linen and hessian (Polizeli et al. 2005). The products of xylan hydrolysis are used in the production of fuel ethanol. Biofuel production is a strong research focus as it can contribute to curbing global green-house gas emissions. Xylanolytic enzymes are important in the breakdown of lignocellulosic biomass to simple sugars necessary for biofuel production.

1.6.5. Xylanases from GH11 family

GH11 xylanases include endo-1,4-β-xylanases that hydrolyze xylans. These enzymes range in size between 20 and 26 kDa, have relatively high pI and a wide pH optimum of 2 to 9. Fungal xylanases mostly have acidic pH optima (Dey & Roy,2018; Ito et al., 1992). GH11 enzymes hydrolyse β-1,4 linkages between D-xylosyl with a configuration retaining mechanism.

1.6.5.1. General structure of GH11 xylanases

To date, 37 GH11 xylanase crystal structures have been deposited in the Protein Data Bank (PDB). The catalytic domain of GH11 xylanases consists of an α-helix and β-sheets A and B, respectively consisting of six and nine antiparallel β-strands (Figure 1.9). In some GH11 enzymes a shorter N-terminal region replaces the first strand of sheet A by a loop. β-Sheets A and B fold around each other to form a β-jelly roll often compared to human right hand (Figure

1.9) where the palm encompasses the α-helix and a twisted part of β-sheet B, the fingers consist of β-sheet A and parts of β-sheet B, and the loop connecting β-strands B8 and B9 form the thumb. An extended loop between β-strands B6 and B7 is called the "cord" (Purmonen et al., 2007). The substrate-binding cleft lies on the inner surface of the palm and is partially enclosed by the fingers, cord and thumb regions (Ludwiczek et al., 2007; Purmonen et al., 2007).

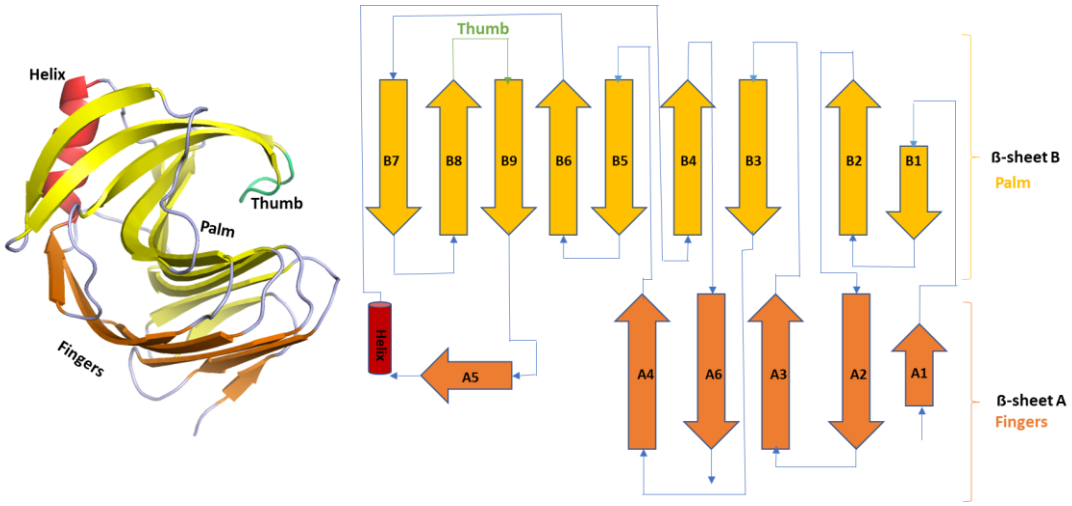


Figure 1.9: Representations of GH11 xylanase structure

Left: A cartoon style representation of the GH11 xylanase crystal structure.
Right: The GH11 xylanase topology. β-Strands are numbered A1 to 6 and B1 to 9 for β-sheets A and B, respectively. Thumb, palm and fingers regions of the tertiary structure are labelled as such (Paës, 2005; Purmonen et al., 2007).

1.7. Xylanases from metagenomes

Xylanases of varying properties have been isolated from cultured microorganisms. However, most were poorly suited to the extreme temperatures and alkaline conditions of industrial processes. Culture-independent methods have hence been used to isolate novel xylanases. Metagenomics, one such culture-independent approach, allows genes encoding useful enzymes to be retrieved directly from environmental samples. Metagenomics has been used to explore carbohydrate active enzymes in soils and extreme environments such as the rumen of herbivores and the gut of arthropods (Engel & Moran, 2013; Al-Masaudi et al., 2017).

1.7.1. Carbohydrate active enzymes from termite gut metagenomes

Termites can feed almost exclusively on lignocellulose, efficiently breaking it down and utilising up to 95 % of its cellulose and 80 % of its hemicellulose content. The ability of

termites to degrade lignocellulose links them to the carbon cycle and creates a natural source of biocatalysts for lignocellulose digestion (Wenzel et al., 2002).

Higher termites have established a symbiotic relationship with fungi for food provision, while a predominantly prokaryotic hindgut microbiome, functioning in a mutualistic manner, digests the lignocellulose. (Brune, 2014; da Costa et al., 2019; Cragg et al., 2015). Glycoside hydrolases found in the hindgut include exoglucanases, β-glucosidases, and hemicellulose-digesting enzymes, such as xylanases, arabinosidases, mannosidases, and arabinofuranosidases. The enzymes plus the extended residence time of food in the termite hindgut, ensure complete digestion. While representative hindgut enzymes have been purified and their genes heterologously expressed, most have only been identified by meta-transcriptomic analyses (He et al., 2019; Terry et al., 2019; Kariuki et al., 2023).

1.8. Xyl xylanase from the hindgut metagenome of *T. trinervoides*

Like other termites, the snouted harvester termite (*T. trinervoides*) relies on GH enzymes secreted by gastrointestinal symbionts to hydrolyse polysaccharides for digestible sugars. This work describes functional and structural aspects of a xylanase (Xyl) from the hindgut of the snouted harvester termite. Xyl is a multidomain enzyme consisting of four domains separated by structural linkers (Rashamuse et al., 2017). An N-terminal signal peptide is followed by a GH11 catalytic domain, two family 36 carbohydrate binding modules (CBM), and a C-terminal, type 4, carbohydrate esterase domain (CE4) (Figure 1.10). Recent insights into family 36 carbohydrate binding modules and family 4 esterase are reviewed below.

Figure 1.10: Domain structure of multidomain Xyl enzyme

An N-terminal secretion signal peptide (SP, white) is followed by a GH11 catalytic domain (blue), two carbohydrate binding modules CBM36-1 and CBM36-2 (orange), and an esterase CE4 catalytic domain (green). Interdomain linker regions are depicted as a blue line.

1.8.1. Carbohydrate binding modules

Carbohydrate binding modules (CBM) facilitate the physical attachment of carbohydrate active enzymes to substrates to enhance their catalytic efficiency. The CAZy database recognizes 93, sequence derived CBM families and seven structural folds. Many folds involve

a β-sandwich consisting of two overlapping β-sheets of three to six antiparallel β-strands. β-Sandwich CBMs bind ligands using either a single curved β-sheet (Figure 1.11A) or a variable β-sheet loop region (Figure 1.11B). Some β-sandwich CBMs like CBM6 and CBM20 exhibit dual binding sites (Pires et al., 2004; Lawson et al., 1994). Other common CBM fold include the β-trefoil fold, cysteine knot, OB fold, the hevein and hevein-like folds, as well as other unique folds.

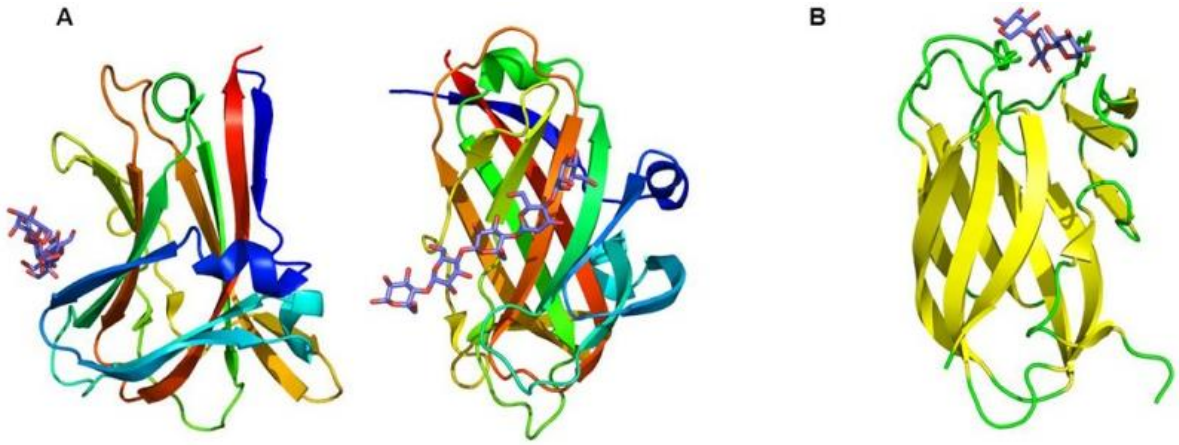


Figure 1.11: Classical carbohydrate binding modules with β -sandwich folds

A) Family CBM27 from *Thermotoga maritima* mannanase with a ligand bound to a single β-sheet, side and front view (PDB ID 1OF4) (Boraston et al., 2003a).
B) CBM6 from *Clostridium stercorarium* xylanase (PDB ID 1NAE) (Boraston et al., 2003b) with ligand binding sites in the loop region of the β-sandwich. Structures prepared and presented using the PyMoL Molecular Graphics Program.

Based on the architecture of their binding sites, CBMs are assigned to types A, B and C (Figure 1.12). Type-A CBMs bind hydrophobic ligands such as chitin and the hydrophobic plane the crystallites of cellulose through aromatic surface residues. Type B CBMs bind single glycan chains in open grooves or clefts. Type C CBMs are often “lectin-like” and bind mono-, di- or tri-saccharides in surface pockets and indentations (Boraston et al., 2000).

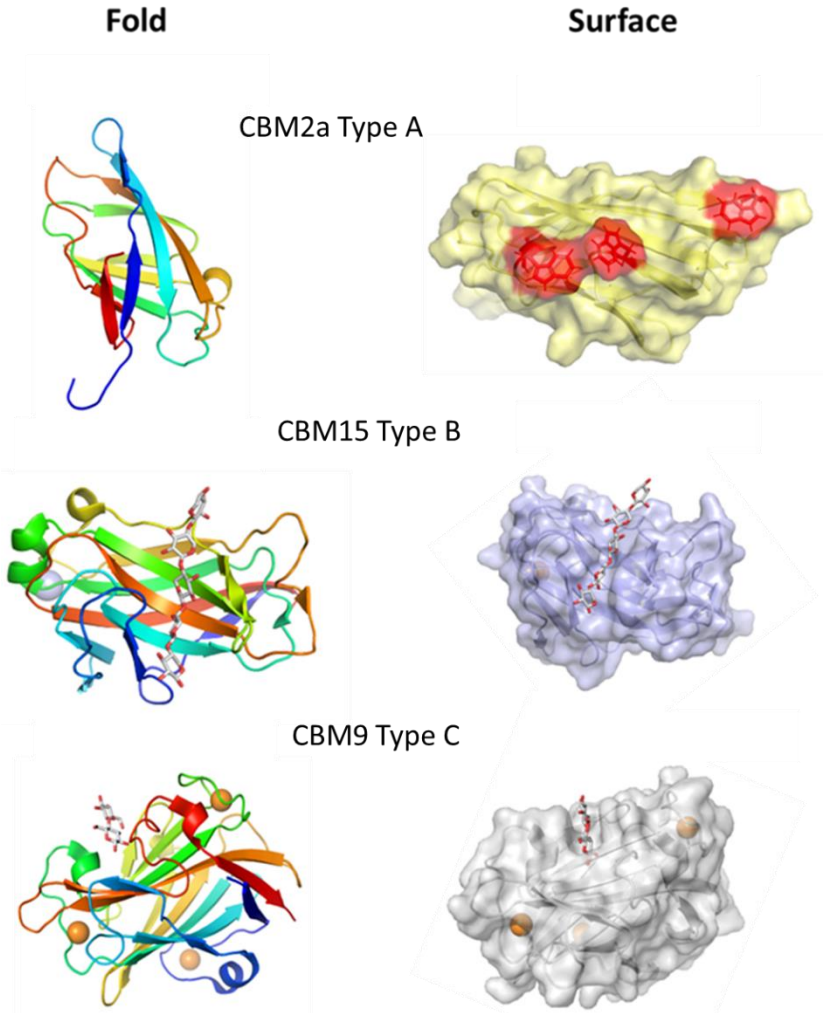


Figure 1.12: Carbohydrate binding modules of types A, B, and C

CBM2a is a part of *Cellulomonas fimi* xylanase Xyn10A (PDB ID 1EXG; Xu et al., 1995), CBM15 of *Cellvibrio japonicas* xylanase Xyn10C (PDB ID 1GNY), and CBM9 of *Thermotoga maritima* GH10 xylanase (PDB ID 1I82). The folds are colour ramped from a blue N- to a red C-terminus. Three aromatic residues forming a nonpolar, ligand-binding surface in the CBM2a module are displayed in red sticks within the molecular surface and fold depictions of the protein. Structures prepared and presented using the PyMoL Molecular Graphics Program.

1.8.1.1. The role of CBMs in catalysis

How CBMs facilitate catalysis is not entirely evident. They may disrupt hydrogen-bonding networks in crystalline cellulose, bringing the surface chains into close contact with the carbohydrate degrading domains (Liu et al., 2021). Alternatively, they may change the surface structure of cellulose to change the rate of catalysis of insoluble substrates (Moser et al., 2008, Wang et al., 2008; DeChellis et al., 2023). CBMs may also enhance glycoside hydrolase activity of insoluble substrates by placing the catalytic domain in sustained, close physical contact with the substrate.

1.8.1.2. Topology of CBM ligand-binding cleft influences specificity

The topology of the ligand binding cleft is critical in defining the CBM-ligand specificity. CBMs that bind internal polysaccharides generally have an open binding cleft, though the details of binding depend on the binding cleft shape. Some CBM4 domains thus bind linear β-1,4-polysaccharides, whereas others bind helical β-1,3-glucan (Boraston et al., 2002a). *Cf*CBM4-1 is an example of a CBM4 module with a binding cleft topology that can bind linear glucan chains. Both ends of the binding cleft are open for easy binding of the linear polysaccharide. In *Tm*CBM4-2 on the other hand, two inserted loops in the binding cleft create a bent groove that binds the helical β-1,3-glucan. In CBM6, the binding cleft shape adapts to the incoming substrate. To accommodate polysaccharide chain ends the cleft assumes a pocket-like structure (Pires et al., 2004), while an open conformation binds internal regions of xylan (Czjzek et al., 2001). CBMs thus reveal variable recognition sites tailored to the respective ligands, potentially minimizing energy penalties for the enzyme.

1.8.1.3. Calcium-assisted ligand recognition in CBMs

Direct and indirect roles of calcium ions in ligand recognition have been demonstrated in CBM families with a conserved jelly roll fold. In Aga16B-CBM6-2Aga, for example, Ca^{2+} mediates the interaction between the neoagarose substrate and a tyrosine residue (Henshaw et al., 2006). In CBM36, by contrast, Ca^{2+} is bound by O2 and O3 atoms of xylose (Jamal-Talabani et al., 2004). CBM35 domains bind uronic acids through electrostatic interactions with its C6 carboxylate via Ca^{2+} (Montanier et al., 2009b). The affinity of CBMs for plant-derived substrates is often low, with K_D of ~100 µM (Boraston et al., 2004). Multiple CBMs with similar substrate affinities are observed in some enzymes, presumably to increase total affinity (Bolam et al., 2001; Freelove et al., 2001; Boraston et al., 2002a).

1.8.1.4. Multiplicity of CBMs

CBMs are often duplicated rather than evolving a single, higher affinity domain. Also, CBM ligands are often tightly linked to other parts of the cell wall, sterically preventing additional substrate interactions to increase overall affinity. This thesis describes the structure and function of two CBM36 modules linked to a GH11 catalytic domain.

1.8.2. Acetylxylan esterases

Carbohydrate esterases (CEs) are enzymes that catalyse the O- or N-deacetylation of acetylated polymers. The Carbohydrate-Active Enzyme (CAZy) database lists twenty CE families. CEs are industrially used as biocatalyst in diverse bioprocesses, making them prime research targets. Currently only a few CEs have been analysed biochemically, and even fewer three-dimensional structures have been elucidated. CE enzymes of families 1, 4, 5 and 6 deacetylate the acetyl esterification at positions 2 and 3 on mono-and di-O-acetylated D-xylopyranosyl residues releasing acetic acid and rendering the xylan main chain more accessible to xylanases and β-xylosidases (Biely, 2012; Biely et al., 2013; Neumüller et al., 2015).

1.8.2.1. Carbohydrate esterase family 4

The carbohydrate esterase family 4 (CE4) is the largest and best studied CE family. In August 2023, the CaZy database listed 44783 CE4 protein sequences, of which 51 have been biochemically characterised with 28 crystal structures.

Enzymatically, CE4s are grouped as follows:

1) Acetyl xylan esterases (AXEs) (EC 3.1.1.72) catalyse the N- or O-deacetylation of xylan (Aspinall, 1959).
2) Chitooligosaccharide deacetylases (EC 3.5.1.-) and chitin deacetylases (EC 3.5.1.41) N- or O-deacetylate chitin, a linear polysaccharide of β-(1-4)-linked N-acetylglucosamine (GlcNAc) (Tsigos, Martinou, Kafetzopoulos, & Bouriotis, 2000).
3) Peptidoglycan N-acetylmuramic acid deacetylases (EC 3.5.1.-) and peptidoglycan N-acetylglucosamine deacetylases (EC 3.5.1.104) N- or O-deacetylate peptidoglycans or murein, a polymer consisting of alternate β-(1-4)-linked N-acetylglucosamine and N-acetylmuramic acid (MurNAc) (Johannsen, 1993).

CE4 family members share a conserved NodB domain, identified in rhizobial deacetylase, the first CE4 protein characterised (Long, 1989). Rhizobial deacetylase required metal ions for catalysis, a defining property of CE4 enzymes (Blair et al., 2005). An active-site His-His-Asp triad typically coordinates a single divalent metal ion, often Zn^{2+} or Co^{2+} (Hernick & Fierke, 2005). Another active-site Asp and His function as catalytic base and catalytic acid. Together with a water molecule complete the catalytic machinery of the enzyme. All five active-site

residues form part of a conserved CE4 enzyme motif (Arnaouteli et al., 2015, Blair et al., 2005). Representative structures for all CE4 classes have been published. Most have distorted $(\alpha/\beta)_8$ or more rarely $(\alpha/\beta)_7$ barrel folds. Additional domains include β-sandwich (Arnaouteli et al., 2015, Fadouloglou et al., 2013), α-helical (Deng et al., 2009), α/β (Blair et al., 2005), or CBM domains (Andrés et al., 2014).

CE4 Enzyme Classes:

1) Enzymes of this class deacetylate poly-β-1,6-N-acetyl-D-glucosamine (PNAG), an exopolysaccharide creating the extracellular matrix of some bacterial biofilms (Branda, Vik, Friedman, & Kolter, 2005). Biofilms support bacterial infections *inter alia* by shielding the pathogens from soluble antimicrobials and by preventing their phagocytosis (Vu, Chen, Crawford, & Ivanova, 2009). Partial deacetylation of PNAG supports biofilm growth while fully acetylated PNAG impedes its formation (Itoh et al., 2008). PNAG deacetylases are thus relevant targets in preventing biofilm formation. Crystal structures for these enzymes include IcaB, a membrane-associated enzyme from *Ammonifex degensii* (PDB id: **4WCJ**; Little *et al.,* 2014); BpsB, a periplasmatic enzyme from *Bordetella bronchiseptica* (PDB id: **5BU6**; Little *et al.,* 2015); and the *E. coli* proteins PgaB (PDB ids: 3VUS, 4F9D and 4F9J; Little *et al.*, 2012, Nishiyama *et al.,* 2013).

2) CE4 enzymes of this group deacetylate peptidoglycan (murein). Peptidoglycan is a heterogeneous polymer of alternating GlcNAc and MurNac units in bacteria cell walls (Navarre & Schneewind, 1999). Bacteria partly de-N-acetylate peptidoglycan make it less susceptible to host enzymes (Boneca, 2005). The spread of infectious bacteria may be slowed or eliminated by inhibiting peptidoglycan deacetylation. This group is represented by the largest number of crystal structures, mostly sharing a $(\alpha/\beta)_8$ barrel fold. Examples include: BaCE4 (PDB **2J13**; Oberbarnscheidt, Taylor, Davies, & Gloster, 2007) and BA0150 (PDB **4M1B**; Strunk et al., 2014) from *Bacillus anthracis*; BC0361 from *Bacillus cereus* (PDB **4HD5**; Fadouloglou et al., 2013); PdaA from *Bacillus subtilis* (PDB **1W17**; Blair & van Aalten, 2004) and PgdA from *Streptococcus mutans* (PDB **2W3Z**; Deng et al., 2009). SpPgdA deacetylase from *Streptococcus pneumoniae* (PDB id: **2C1G**; Blair et al., 2005) deviates from the other enzymes by having a C-terminal NodB-like domain and additional small N-terminal and central α/β fold domains (Blair et al., 2005).

3) CE4 chitin deacetylases process, cleave, and de-N-acetylate chitin, a polysaccharide composed of GlcNAc units, producing chitosan and chitooligosaccharide derivatives (Hoell, Vaaje-Kolstad, & Eijsink, 2010). CE4 chitin deacetylases are used to make antifungals and antibacterials (Zhao, Park, & Muzzarelli, 2010), while chitin derivatives can be useful in medical and pharmaceutical applications (Park & Kim, 2010). Three CE4 chitin deacetylase structures are known: VcCDA CE4 from *Vibrio cholera* (PDB **4NY2**; Andrés et al., 2014); ClCDA from *Colletotrichum lindemuthianum* (PDB **2IW0**; Blair et al., 2006) and the ECU11_0510 from *Encephalitozoon cuniculi* (PDB **2VYO**; Urch et al., 2009).

4) The fourth group of CE4 enzymes includes acetylxylan esterases (AXE) (EC 3.1.1.72). Two crystal structures for this group include CtCE4 and SlCE4 from *Clostridium thermocellum* and *Streptomyces lividans* respectively (PDB ids: **2C71** and **2CC0**; Taylor et al., 2006). CtCE4 and SlCE4 retain the canonical $(\alpha/\beta)_8$ barrel fold of other CE4s but their conserved metal ion binding site prefers Co^{2+} over Zn^{2+}, the more common metal ion among CE4s. Four water molecules, a histidine and an aspartate residue complete the catalytic machinery of CtCE4, rather than the more classical His-His-Asp family 4 “consensus” displayed SlCE4 (Taylor et al., 2006).

To date, only 0.6 % of all available CE sequences have been biochemically characterised and only 0.2 % have been resolved experimentally. The abundance of CE sequences and their potential significance in industrial and antimicrobial drug development imply that more structural information is critical. In this study, the structure and function of the CE4 domain of the multidomain enzymes Xyl was characterised.

1.9. State of the art of multidomain CAZymes

Multidomain carbohydrate-active enzymes (CAZymes) continue to be a subject of intense research due to their pivotal roles in the breakdown, modification, and synthesis of complex carbohydrates (Thurimella et al., 2023). Recent advancements in the field have provided deeper insights into the structural and functional attributes of these enzymes (Krska et al., 2021; Chettri et al., 2020). Multidomain CAZymes typically comprise distinct catalytic domains responsible for specific enzymatic activities, often linked by flexible regions that facilitate domain movement (Cedillo, & Montanier, 2023; Sützl et al., 2018). The flexibility conferred

by these linkers is crucial for the enzymes' ability to act on diverse substrates and perform multiple functions within a single polypeptide chain (Reddy Chichili etal., 2013).

Recent studies have uncovered novel multidomain CAZymes with unique substrate specificities and catalytic efficiencies. Advances in protein engineering, aided by sophisticated techniques such as X-ray crystallography and cryo-electron microscopy, have allowed researchers to characterize these enzymes at high resolution (Li etal., 2020). Additionally, bioinformatics tools have played a key role in identifying new multidomain CAZymes and predicting their functions based on sequence analysis. This integrative approach has contributed to the design of engineered enzymes with enhanced capabilities for various biotechnological applications, including the development of more efficient biofuel production processes and the modification of plant cell walls for agricultural purposes (Jones et al., 2018; Song et al., 2023) .

The biotechnological relevance of multidomain CAZymes extends to their potential in addressing global challenges, such as sustainable energy production and environmentally friendly industrial processes (Lv, 2023; Mesbah, 2022). Ongoing efforts focus on unraveling the regulatory mechanisms governing these enzymes and harnessing their multifunctionality for practical applications (Kim et al., 2024). As researchers continue to delve into the intricacies of multidomain CAZymes, the field stands poised for further discoveries that could revolutionize bioengineering and our understanding of carbohydrate metabolism (Kim et al., 2024).

1.10. Preliminary characterization of Xylanase (Xyl)

The multidomain xylanase Xyl consists of an N-terminal GH11 domain linked to two CBM36 domains and a CE4 esterase domain. A previous MSc study of Xyl had analysed the effects of the two almost identical CBM domains on the GH11 catalytic domain (Kruger, 2017). Three variants of Xyl were prepared, encompassing 1) only the GH11 catalytic domain, 2) the GH11 domain and the first CBM domain, and 3) the GH11 domain with both CBM domains. All three variants were produced and functionally characterised using a modified DNS assay. With only one CBM present, the pH profile of the GH11 catalytic domain was observed to shift such that its activity increased at lower pH. Deleting the CBMs lowered enzyme performance to 56 % (Standard deviation = 2.9) from the 90 % enzyme activity in the presence of both CBMs.

Although, there was also no observed statistically significant change in the Michaelis constants (K_M) of the GH11 domain in the absence of both CBMs, the turnover number (k_{cat}) decreased. These observations suggested some form of cooperativity between the domains.

Here, we further study this enzyme by biochemical, X-ray crystallography and biophysical methods to *inter alia* explore the relative effect of the domains on each other. The aim was to expand current understanding of multidomain xylanases and to add structural information. Each domain and combinations of domains was produced and structurally characterised by X-ray crystallography while interdomain interaction within Xyl was investigated by isothermal titration calorimetry (ITC).

2. Aim and Objectives

2.1. Aim

The aim of this study was to investigate the multidomain xylanase Xyl in terms of structure, function and biophysical properties based on individual domains and domain combinations.

2.2. Objectives

a) Cloning

Amplify DNA fragments encoding Xyl-CBM36-1, Xyl-CBM36-2, and Xyl-CE4 and insert them into appropriate expression plasmids.

b) Protein production

Transform *E. coli* BL21 (DE3) (Table 3.3) cells with expression plasmids bearing full length *xyl*, *xyl-gh11, gh11-cbm36-1, gh11-2cmb36, xyl-cbm36-1, xyl-cbm36-2* and *xyl-ce4*. Culture the transformed cells in LB media and induce the production of the individual proteins.

c) Protein purification

Purify all protein constructs by affinity chromatography, anion exchange chromatography and/or size exclusion chromatography with a final buffer suitable for downstream experiments.

d) Protein characterization

Investigate domains Xyl-CBM36-1, Xyl-CBM36-2 and Xyl-CE4 by circular dichroism spectroscopy to assess folding and melting temperatures. For Xyl-CE4, determine the optimal temperature, pH and divalent metal ion composition for enzyme activity.

e) Protein crystallization and structure solution

Crystallize all protein constructs using commercial screens and optimizing lead conditions for optimal crystal quality. Collect diffraction data and solve individual crystal structures by molecular replacement, refine and analyse the structures.

f) Interdomain interactions within Xyl

Use isothermal titration calorimetry (ITC) to investigate possible inter-domain interactions within Xyl: I) Xyl-GH11 and Xyl-CBM36-1 II) Xyl-GH11 and Xyl-CBM36-2 III) Xyl-GH11 and Xyl-CE4 IV) Xyl-CBM36-1 and Xyl-CE4 V) Xyl-CBM36-2 and Xyl-CE4

3. Materials and Methods

3.1. Standard materials

3.1.1. General chemicals and reagents

Unless otherwise specified, all chemicals and reagents used in this thesis were of laboratory quality, supplied by Bio-Rad, Roche, Fermentas, ThermoFisher Scientific, Promega, Sigma-Aldrich, Merck, Qiagen, KPL Inc. and Hampton Research. These materials were carefully selected to ensure the reliability and consistency of experimental results. Furthermore, a range of specialized kits, enzymes, and molecular weight standards were employed to facilitate various experimental procedures and analyses. Table 3.1 provides a comprehensive overview of these tools and materials, detailing their specific applications and contributions to the study.

Table 3.1: Kits, enzymes and molecular weight standards used in this study

Purpose	Product	Supplier
Plasmid DNA isolation	GeneJET Plasmid Mini-Prep Kit	ThermoFisher Scientific, Waltham, MA, USA
PCR product purification	GeneJET PCR Purification Kit	
Nucleic acid stain	GelRed	
Restriction enzymes	XhoI-HF	New England Biolabs, Ipswich, MA, USA
	NcoI-HF	
	NdeI-HF	
	BamHI-HF	
Protein marker	Precision plus protein marker	Bio-Rad, Hercules, CA, USA
DNA marker	GeneRuler 1 kb DNA Ladder	ThermoFisher Scientific, Waltham, MA, USA

3.1.2. Buffers, stock solutions and growth media

Stock solutions, buffers, and growth are fundamental in every experimental procedure and cell culture techniques. In this study, all buffers, stock solutions, and growth media were sterilized, either through autoclaving at 121 °C for 20 minutes or filtration using 0.22 µm filters, to ensure the elimination of contaminants. Additionally, the pH of solutions and buffers was carefully adjusted to the desired levels using either hydrochloric acid or sodium hydroxide, maintaining optimal conditions for experimental protocols. Table 3.2 provides a comprehensive overview of the specific compositions and preparations of these essential solutions and media, aiding in the understanding and replication of experimental methodologies employed throughout the research.

Table 3.2: Stock solutions, buffers, and media

Stock	Composition
4 x Separating gel buffer	1.5 M Tris pH 8.8
4 x Stacking gel buffer	0.5 M Tris pH 6.8
5 x SDS running buffer	25 mM Tris pH 8.3, 0.1 % SDS and 250 mM glycine,
6 x DNA loading dye	0.25 % (w/v) bromophenol blue, 0.25 % (w/v) xylene cyanol FF and 30 % (v/v) glycerol.
10 x Phosphate-buffered saline (PBS)	80 g NaCl, 2 g KCl, 14.4 g Na_2HPO_4, 2.4 g KH_2PO_4, in deionized water and pH adjusted to 7.4
Ammonium persulphate	0.5 M stock
Ampicillin	A 100 mg/mL stock solution
Lysis-Equilibration-Wash Buffer (LEW Buffer)	50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 1 EDTA-free Complete Protease Inhibitor Cocktail Tablet (Roche, Basel, Switzerland)
Coomassie staining solution	0.25 % (w/v) Coomassie Brilliant Blue R-250, 30 % (v/v) ethanol, 10 % (v/v) acetic acid
Destaining solution	40 % (v/v) ethanol, 10 % (v/v) acetic acid
DTT	1 M stock solution in 10 mM Na acetate pH 5.2, sterile filtered
EDTA	0.5 M stock in distilled water
IPTG	1 M stock in distilled water, sterile filtered
Lysogeny broth media (LB media)	10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl. Sterilised by autoclaving
LB agar plates	7.5 g/L agar, 5 g/L NaCl, 5 g/L yeast extract, and 10 g/L tryptone, autoclaved. Supplemented with 100 µg/mL ampicillin or kanamycin. Poured into sterile plates in a laminar flow cabinet when partly cooled to set.
Lysozyme	50 mg/mL stock solution

PreScission 3C protease cleavage buffer	50 mM Tris pH 7.0, 150 mM NaCl, 5 mM DTT, 1.0 mM EDTA
10 x TAE	400 mM Tris, 400 mM glacial acetic acid, 10 mM EDTA

Table 3.3: Bacterial strains

***E. coli* strain**	**Function**	**Supplier**
BL21 (DE3)	Protein production	Stratagene, San Diego, CA, USA
BL21-CodonPlus RIL	Protein production when rare codons are present	
DH5α	Plasmid propagation	

3.1.3. Primers for polymerase chain reactions

The polymerase chain reaction (PCR) is a fundamental technique used to amplify specific DNA sequences for further analysis. Primers, short single-stranded DNA sequences complementary to the target DNA region, are essential components of PCR reactions. In Table 3.4, we present a detailed compilation of the primers utilized in this study. These primers were meticulously designed to selectively amplify target DNA fragments of interest, allowing for the precise interrogation and characterization of genetic material.

Table 3.4: Primers for polymerase chain reaction

Name	Sequence (5' to 3')
pGEX-6P-1_*cbm1* Fd (BamHI)	ACTTGGGGATCC**GGCAACACCGGTAGTCAAACG**
pGEX-6P-1_*cbm1* Rv (NotI)	AATTAGCGGCCGCTCA**CGATGATTCATTGCCACCAAC**
pGEX-6P-1_*cbm2* Fd (BamHI)	ACTTGGGGATCC**GGTGGCCAGGCTCAAAGC**
pGEX-6P-1_*cbm2* Rv (NotI)	AATTAGCGGCCGCTCA**CGTTTGAACCGGCATTGTTGC**
pET28a_*ce4* Fd (NcoI)	ACTACCCATGG**GCAATGAAAAACTGATCG**
pET28a_*ce4* Rv (XhoI)	TATACCTCGAG**TTGACCGTTAGCGTTTGAC**

Restriction enzymes and their target sequences are underlined

3.2. Molecular biological methods

3.2.1. Preparing competent bacterial cells

Cell competence refers to the ability of a cell to take up extracellular DNA and is a prerequisite for cell transformation. To generate competent bacteria, an appropriate strain of *E. coli* was plated on LB agar and incubated overnight at 37 °C. A single colony from the plate was used

to inoculate 5 mL of LB broth and incubated for 16 h, at 37 °C and 170 rpm. The 5 mL culture was diluted into 100 mL LB broth and further incubated until the OD_{600} was between 0.4 and 0.6. The cell culture was cooled on ice and centrifuged at 11 000 x *g* for 10 min at 4 °C. The cell pellet was dissolved in 35 mL ice-cold transformation buffer (10 mM PIPES pH 6.7, 55 mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl), incubated on ice for 15 min and centrifuged at 11 000 x *g* and 4 °C. The cell pellet was gently re-suspended in 2 mL ice-cold transformation buffer and centrifuged at 4 500 x *g* at 4 °C for 5 min in a bench top refrigerator centrifuge. The cell pellet was resuspended in 100 mM $CaCl_2$ solution containing 20 % (v/v) glycerol. Equal volumes (50 µL) of the cells were aliquoted into 1.5 mL Eppendorf tubes, flash cooled in liquid nitrogen and stored at -80 °C.

3.2.2. Transformation of competent bacteria cells by plasmid DNA

Here, 50 µL of competent *E. coli* cells were transformed by incubating them with 25 to 50 ng plasmid DNA. The frozen competent cells were thawed on ice, mixed with plasmid DNA, and incubated on ice for 30 min. The ice cooled cells were heat shocked at 42 °C for 45 s and returned to ice for 2 min. The transformed cells were supplemented with 800 µL of pre-warmed LB and incubated with shaking at 37 °C for 1 h. One tenth of the incubated cell culture was spread on LB agar plates with appropriate antibiotics and incubated overnight. As a control, untransformed competent cells were plated on agar plates with and without antibiotics and incubated alongside the experimental plates.

3.2.3. Polymerase chain reaction (PCR)

DNA was amplified by PCR to generate plasmid constructs, and to screen for successful plasmid construction and transformation. A standard PCR amplification generally employed a volume of 50 µL containing 10 to 100 ng template DNA, reaction buffer, 200 µM of each dNTP, 1 µM of each primer and 1 U Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA). Thermocycling in a T100 Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA), was initiated by a denaturation step at 98 °C for 30 s followed by 35 cycles of denaturation (98 °C, 10 s), primer annealing (45 to 68 °C, 10 to 30 s) and fragment extension (72 °C, 30 s/kb). The process was terminated with a final extension step at 72 °C, 7 min.

3.2.4. Plasmid DNA isolation from *Escherichia coli* cells

Plasmid DNA from *E. coli* was extracted from overnight cultures using the DNA extraction mini-prep kit (Table 3.1) according to manufacturer's instructions. The protocol is based on the conventional alkaline lysis method where plasmid DNA is isolated by alkaline denaturation of cellular proteins and chromosomal DNA. The mini-prep kit includes spin columns fitted with silica-based membranes to trap DNA through hydrogen-binding interactions between the negatively charged DNA backbone and the positively charged silica particles.

3.2.5. Restriction digest of plasmid DNA and PCR products

Restriction enzymes were used to generate and verify plasmid constructs. The desired amount of DNA, varying from 200 ng to 1 µg, was treated with the appropriate amount of restriction enzyme viz. 1 to 5 U restriction enzyme/µg DNA depending on the purpose of the digest. Restriction digest reactions were incubated at 37 °C for 30 min in restriction buffers provided by enzyme manufacturer. Double restriction digests using two restriction enzymes to cleave the DNA at unique sites was used to create overhanging or "sticky" ends, or to cut out DNA segments of interest. For restriction double digestion, a manufacturer provided buffer compatible to both enzymes was chosen.

3.2.6. Agarose gel electrophoresis

DNA fragments may be separated by size in an agarose gel by applying an electric field. Due to the negatively charged phosphate backbone and its uniform mass to charge ratio and depending on the agarose gel density, smaller DNA fragments will move through the agarose matrix more rapidly and migrate further than larger ones (Greenfield, 2006).

PCR and restriction digest products as well as isolated plasmid DNA were analysed and assessed for quality and yield by agarose gel electrophoresis. In general, 1 µL GelRed (ThermoFisher Scientific, Waltham, MA, USA) was added to 5 µL DNA sample and the mix loaded onto a 1 % agarose gel in a Mini-Sub cell GT cell (Bio-Rad Laboratories, Hercules, CA, USA) containing TAE buffer. DNA samples were separated by applying an electric potential of 90 V for 45 min. The DNA bands were visualized on a Molecular Imager Gel Doc XR+ UV transilluminator with Image lab software (Bio-Rad Laboratories, Hercules, CA, USA).

3.2.7. DNA extraction from agarose gel

DNA was extracted from agarose gels for further experiments such as between restriction digestion and ligation. The desired DNA band was physically cut from the agarose gel under UV-light with a clean, sharp scalpel, the gel slice was weighed, and DNA extracted using a gel extraction kit (Table 3.1), according to the manufacturer's instructions.

3.2.8. DNA dephosphorylation

To prevent re-ligation of vectors, linearized plasmid DNA was 5'-dephosphorylated prior to ligation using shrimp alkaline phosphatase (Roche, Basel, Switzerland). For each microgram of restricted plasmid DNA, 2.5 µL Shrimp Alkaline Phosphatase (5 U) was added and the reaction incubated for 5 min at 37 °C. The reaction was heat inactivated by incubation at 60 °C for 1 h.

3.2.9. Ligation

Vector backbones and DNA fragments were ligated using T4-ligase or T4-quick ligase (New England Biolabs, Ipswich, MA, USA). For a standard reaction, 50 ng of dephosphorylated vector DNA was incubated with a threefold molar excess of insert at room temperature overnight for T4-ligase or at 25 °C for 30 min for T4-quick ligase. After incubation, ligation mixtures were used to transform *E. coli* DH5α and cells spread on LB agar plates containing appropriate antibiotic.

3.2.10. Colony PCR

To rapidly identify positive clones after cloning procedures, five distinct colonies were randomly selected from the transformed cells plate and individually resuspended in 20 µL nuclease free distilled water. 1 µL of the mixture was used as template DNA together with primers for the specific amplification of the insert DNA in a PCR amplification reaction with cycling parameters analogous to those in section 3.2.3. A positive amplification of insert confirms successful cloning.

3.2.11. Sequencing of plasmid DNA

To verify the DNA sequence of plasmid constructs, the plasmids were sequenced by Sanger sequencing at the Inqaba sequencing facility (Inqaba Biotec, South Africa) using either the T7

forward and reverse sequencing primers, or 5' pGEX and 3' pGEX sequencing primers. Raw sequencing data was analysed using Heracle BioSoft, DNA Sequence Assembler v4 (2013).

3.3. Protein biochemical methods

3.3.1. Protein production in *Escherichia coli*

All confirmed gene expression plasmid constructs were used to transform *E. coli* BL21-CodonPlus-RIL competent cells (Stratagene, San Diego, CA, USA) and cultured on LB agar plates supplemented with appropriate antibiotic*:* 100 µg/mL ampicillin for pGEX-6P-2, 100 µg/mL ampicillin and 34 µg/mL chloramphenicol for pET20b, and 50 µg/mL kanamycin and 34 µg/mL chloramphenicol for pET28a respectively.

Single colonies of transformed cells were resuspended in 5 mL LB medium containing appropriate antibiotics and incubated overnight at 37 °C with shaking at 180 rpm. Resulting pre-cultures were diluted into larger volumes of LB medium, which were incubated at 37 °C with 180 rpm shaking. Once the OD_{600} of the medium had reached 0.4 and 0.6 as measured by NanoDrop ND1000 (ThermoFisher Scientific, Waltham, MA, USA), the culture was cooled to 25 °C and induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG) (Table 3.2). Cultures were incubated at 22 to 37 °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. Re-suspended cells were lysed by sonication in a Q500 sonicator with 15 s on/off pulses for 3 min using a ½-inch-probe tip (Qsonica LLC, Newtown, CT, 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 (ThermoFisher Scientific, Waltham, MA, USA).

3.3.2. Protein purification

3.3.2.1. Affinity chromatography

Proteins were purified with the help of His_6- or glutathione-S-transferase (GST-) tags using the corresponding binding resins: Co^{2+}-NTA agarose (Qiagen, Hilden, Germany) for His_6-tagged proteins and glutathione sepharose resins (Cytiva, Marlborough, MA, USA) for GST-tagged proteins. For each resin type, 2 mL of resin was washed with 5 column volumes (CV) of distilled water, equilibrated with 5 CV of 1 x LEW buffer pH 8.0, and mixed with the soluble cellular fraction. The mixture was agitated on a roller mixer overnight at 4 °C and transferred to a

gravity flow purification column. Unbound proteins were collected as the flow through. The matrix was washed with 5 CV of 1 x LEW buffer pH 8.0. His_6-tagged proteins bound to matrix was further progressively washed with 2 CV 1 x LEW buffer pH 8.0 supplemented with 5, 10, 25, 50, and 100 mM imidazole and eluted with 250 mM imidazole. For GST-tagged proteins, protein of interest was separated from GST by adding 1 µg 3C protease per 100 µg 1ug fusion protein bound to matrix incubating overnight at 4 °C on a roller mixer. Separated protein of interest was collected from a gravity flow column. All fractions were analysed by SDS PAGE (Table 3.2).

3.3.2.2. Anion exchange chromatography

Protein-containing fractions as identified by SDS PAGE were pooled, sterile filtered using a 0.2 µm syringe filter and further purified by anion exchange chromatography on an ÄKTA 900 chromatography system (Cytiva, Marlborough, MA, USA) using a Mono Q 10/100 GL column (Cytiva, Marlborough, MA, USA). The column was equilibrated with 10 CV of buffer A (25 mM Tris pH 8.0, 50 mM NaCl), rinsed with 1 CV buffer B (25 mM Tris pH 8.0, 1 M NaCl) to ensure all anion binding sites are occupied by Cl^-, followed by protein loading. The column was washed with buffer A until all unbound protein had eluted (OD_{280} returned to baseline). Proteins retained on the column were eluted with a linear gradient of 0 to 100 % of buffer B. Fractions within peaks of the chromatogram were analysed by SDS PAGE for target proteins.

3.3.2.3. Size-exclusion chromatography

Samples from the anion exchange peak and confirmed by SDS PAGE to contain protein of interest were pooled and concentrated to at least 25 mg/mL in 1 mL. The concentrated sample was injected onto a Superdex 200 16/60 column (Cytiva, Marlborough, MA, USA) with running buffer (25 mM Tris pH 8.0, 50 mM NaCl) and a flow rate of 1 mL/min on an ÄKTA 900 chromatography system (Cytiva, Marlborough, MA, USA). The elution was monitored from the absorbance curves at 280 nm.

3.3.3. Protein concentration

Pure proteins were concentrated by ultracentrifugation using a Vivaspin-6 concentrator (Sartorius, Göttingen, Germany) with appropriate molecular weight cut offs. The absorbance at 280 nm (A_{280}) was recorded at intervals during centrifugation. Centrifugation was terminated once the required concentration was achieved. To exchange the buffer or reduce

the salt concentration, the concentrated protein was diluted with suitable buffer and re-concentrated by centrifugation to the desired concentration.

3.3.4. Protein quantification

UV light at 280 nm is typically used to determine protein concentrations due to the absorbance of this wavelength by aromatic amino acids. A Nandrop ND-1000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) was used to measure the A280. For concentrated protein samples, the concentrator flow through was used as a blank solution. Molar extinction coefficients ε280 were calculated from protein-specific amino acid compositions.

$C = A280 / \varepsilon . D$ …… Beer-Lambert equation

Protein concentrations C were calculated for recorded absorbance values A. D, the path length is mechanically adjusted by the Nanodrop spectrophotometer. This method does not distinguish individual proteins in a mixture. Hence protein concentrations are only accurate if contaminants are negligible.

3.3.5. Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a biophysical technique that records changes in rotation of plane-polarised UV light due to the chiral properties of a solution. It can be used to assess aspects of proteins such as dominant secondary structures and indirectly to determine protein intactness and melting temperatures. Purely α-helical proteins typically yield spectra with minima at 222 and 208 nm and a maximum at 193 nm, while β-strand proteins produce a minimum at 218 nm and a maximum at 195 nm. Disordered or unfolded proteins result in spectra with low ellipticities of the entire wavelength range (Greenfield, 2006).

Far UV circular dichroism spectra were measured from 250 to 180 nm on a Chirascan Spectrophotometer (Applied Photophysics, Leatherhead, UK). The band width was 1 nm and the step resolution was 0.5 nm. Proteins were initially scanned at 20 °C in a 2 mm quartz cuvette equilibrated with 20 mM Na_3PO_4 buffer pH 7.4. The thermal stability of the protein was assessed at 10 °C temperature increments from 20 to 80 °C in the same buffer. Each

spectrum was recorded as the average of three individual scans. Protein concentrations of 3 µM were used throughout.

3.3.6. Functional characterization of Xyl-CE4.

Acetyl esterase activity was assessed spectrophotometrically by quantifying the release of *para*-nitrophenol (p-NP) from p-NP acetate (p-NPA) at an optical density of 405 nm (OD_{405}) and 30 °C. Each assay mix contained 0.5 mM p-NPA (Merck, Darmstadt, Germany), 50 mM Tris pH 8, and 0.2 µg Xyl-CE4. Temperature and pH optima were determined by repeating acetyl esterase assays over a temperature range of 20 to 80 °C and between pH 3 and 10 at 30 °C in 50 mM of the following buffers: phosphate citrate (pH 3 to 5), MES (pH 5.5-6.5), Tris (pH 7 to 9), and CHES (pH 9.5 to 10). The pH stability between pH 3 and 10 was determined by measuring residual activity after Xyl-CE4 was incubated for 72 h at 30 °C.

3.3.7. Effects of metal ions on Xyl-CE4

The effect of 10 mM divalent cations (Ca^{2+}, Fe^{2+}, Co^{2+}, Mn^{2+}, Mg^{2+}, Ni^{2+} and Cu^{2+}) on Xyl-CE4 activity was assayed at 30 °C. First, purified Xyl-CE4 was assayed. The protein was then treated with 1 mM EDTA at 4 °C for 10 min to chelate metal ions and washed repeatedly with 50 mM Tris pH 8.0 prepared with distilled deionised water. Divalent ions were then added to the EDTA treated sample and assayed.

3.4. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were carried out on an ITC-2000 calorimeter (Malvern Panalytical, Malvern, UK) at 25 °C. To study the binding reaction between protein domains, all proteins were first dialysed against the same buffer, 10 mM Tris pH 7.0 and 50 mM NaCl. The titrant (in the syringe) was generally kept at 10 x the concentration of the titrand (in the reaction vessel). As a control, the titrant was injected into the buffer buffer only. Titrant and titrand concentrations for this study are listed in

Table 3.5. Resulting data were fitted with a single binding-site model using the Origin software provided by Malvern Panalytical.

Table 3.5: Titrant and titrand concentrations for isothermal titration experiments

Titrant	Titrand
Xyl-CBM36-1 (10mg/ml) =714 µM	Xyl-GH11 (1mg/ml) =38.5 µM
Xyl-CE4 (10mg/ml)= 437.8 µM	Xyl-CBM36-1 (0.6mg/ml) = 42.8 µM
Xyl-CE4 (10mg/ml)= 437.8 µM	Xyl-GH11 (1.1mg/ml)= 42 µM

3.5. Protein crystallization

3.5.1. Sample preparation for crystallization

Proteins for crystallization were first buffer exchanged to 10 mM NaCl, 10 mM Tris pH 8.0, or 10 mM acetate pH 5 before being concentrated to 10 mg/mL. Concentrated proteins were filtered using either a 13 mm syringe filter with 0.20 µm pore size or a concentrator with a 10 kDa molecular weight cut off. The concentrated and filtered proteins were stored at 4 °C for crystallization.

3.5.2. Crystallization screening

Proteins were crystallized by the sitting-drop vapour-diffusion experiments using commercial crystallization screens that provided a range of precipitant, salt, buffer, and pH formulations. Screens used included $AmSO_4$, PEGs and PEGs II Suites (Qiagen, Hilden, Germany) and HR2-110 Crystal Screen and HR2-112 Crystal Screen II (Hampton Research, Aliso Viejo, CA, USA). Crystallization drops consisted of 1 µL protein solution and 1 µL reservoir solution equilibrated against 80 µL of reservoir solution at 18 °C. Conditions producing lead crystals were manually optimized by varying physicochemical parameters such as precipitant and protein concentrations, ionic strength, and temperature in hanging drop vapour-diffusion 24-well VDX crystallization plates (Hampton Research, Aliso Viejo, CA, USA). Further optimization was attempted using 1) Additive Screen (Hampton Research, Aliso Viejo, CA, USA), 2) micro, cross and streak seeding, 3) varying the ratio of protein to reservoir volumes, and 3) adding 0.5 to 3 % (v/v) glycerol.

3.5.3. Data collection and evaluation

Crystals are generally held in a cryo-protectant during data collection. To prepare a cryo-protectant, the reservoir fluid was supplemented with 20 to 25 % PEG 400 or glycerol. The cryo-protectant is then evaluated for clarity under a nitrogen stream on a diffractometer. A clear stream is appropriate for crystal data collection. Diffraction size crystals were picked up using nylon loops and soaked into the cryo-protectant for 10 s to keep the crystal hydrated and prevent formation of ice rings when crystal is frozen. The cryo-soaked crystals were immediately cooled in liquid nitrogen and stored in cryogenic vials in a unipuck. Unipucks with 15 samples each were stored in liquid nitrogen shippers and couriered either to the European Synchrotron Radiation Facility (ESRF), Grenoble, France, for remote synchrotron X-ray data collection or to the University of Cape Town, South Africa, for data collection on a rotating-anode X-ray generator (MicroMax-007HF, Rigaku, Japan).

3.5.3.1. Remote data collection, processing, and transfer

For remote data collection at the ESRF, unipucks with crystals were loaded into a robot Dewar by beamline staff and complete data sets automatically collected for suitable crystals on beamline ID29 at the ESRF. Each data set consisted of 3000 frames of fine-sliced data collected over 1.85 s of X-ray exposure time, with frames written every 37 ms. Diffraction data were processed using an automated script that integrates and scales with XDS (Kabsch, 2010) and converts reflection intensities to amplitudes with Aimless (Evans & Murshudov, 2013). A high-resolution cut-off was selected so that I/σ(I) in the high-resolution bin was approximately 2.0 I/σ(I). The latter is a measure of signal to noise, which is an intuitive metric that applies to most experiments. Processed reflection files were transferred by secure file transfer protocol immediately after data collection and stored in an online data base available to users for molecular replacement and model building.

3.5.3.2. Data collection on a home source diffractometer

A home-source rotating-anode X-ray generator (MicroMax-007HF, Rigaku, Japan) was ramped to an input current of 30 mA and a voltage of 40 mV resulting in a vacuum ion gauge value of 130-160. Diffraction data was collected using Cu-Kα radiation (λ = 1.54 Å) and a Saturn 944HG CCD detector (Rigaku, Japan). The crystals were X-irradiated for 10 s per image and two test images 90° apart were first collected. Full data sets consisting of 180 or 360 1°

images were collected for suitable crystals. Denzo and Scalepack of the HKL3000 suite (Minor et al., 2006; Otwinowski & Minor, 1997), and Mosflm (http://www.mrc-lmb.cam.ac.uk/imosflm/) were used for indexing, unit cell and space group determination, data integration and scaling.

3.5.3.3. Data integration, merging and scaling

Intensities of diffraction spots differ from one image to another owing to continues decrease in the diffracting ability of the crystal the longer it is irradiated and to the variability in the intensity of the X-ray beam. Integration is the measure of diffraction intensities. During data integration all diffraction images are converted into a single file containing the Miller index, error approximation and partiality for every reflection (Ravelli & Garman 2006). Data measured by integration programs are not on a common scale, and this needs to be considered when using the data in subsequent molecular structure analysis. The variation in the intensity of the diffraction spots is corrected by assigning a scale factor to which all the intensities are averaged. Scaling and merging may be thought of as an attempt to model the changes occurring during the diffraction experiment. The most useful statistics on the self-consistency of the data are only obtained after scaling and merging and not directly from integration.

Some of the parameters used to measure the quality of the diffraction data are resolution, signal-to-noise ratio, completeness, multiplicity and Rmerge. Resolution is the minimum distance between diffracting planes of a crystal with the crystal still yielding diffraction data (Wlodawer et al., 2008). The smaller the spacing the more the reflections recorded, which means more data for structure solution and refinement. A complete data set is a data set in which all the reflections have been recorded. This is not always the case as crystals are very susceptible to radiation damage. Multiplicity is the average number of measurements for equivalent reflections. Rmerge is a measure of the agreement between equivalent reflections.

3.5.4. Crystal structure solution by molecular replacement

X-ray diffraction experiments are able to determine two of three parameters required to calculate the distribution of the electron density in a crystal. Thus, they provide the intensity of individual reflections (equivalent to the amplitude of the structure factor) and its Miller

indices (related to the structure factor frequency). Information about the phase is however not easily measured. The resulting “phase problem” can be solved by various techniques of which molecular replacement (MR) is possibly the fastest. The PHENIX (http://www.phenix-online.org/) software suite for automated determination of macromolecular structures using X-ray crystallography and other methods was used to solve crystal structures by MR.

Molecular replacement is a phasing method that uses prior information in the form of a (partial) model structure as an approximation of the actual structure to calculate initial phases and hence an initial electron density distribution, which can then be improved by improving the experimental model of the protein of interest. In practice, a homologous protein structure (partial) model is positioned within the crystal unit cell of the protein of interest by solving a rotation and a translation function. The best (partial) model structures were identified by searching the amino acid sequence of the protein of interest on the Protein Data Bank (PDB) for highest sequence identity and sequence coverage. The (partial) model coordinate files were downloaded and edited to provide a single chain without water molecules and ligands. The experimental diffraction data file (mtz) was combined with the (partial) model coordinates and the amino acid sequence of the protein of interest to locate the molecules in the crystal to build a model structure of the protein of interest.

The resulting starting model was uploaded into the graphics programme “Crystallographic Object-Oriented Toolkit” (Coot) together with an mtz file providing coefficients for the electron density. The model was improved iteratively by manually adjusting coordinates within electron density in Coot, and reciprocal space refinement in PHENIX, until a satisfactory model with a low R-free was obtained. The refined model was analysed using the molecular graphics program PyMol (Schrödinger LLC & DeLano, 2022).

3.6. AlphaFold2 prediction of Xyl 3D structure

AlphaFold2 is a novel protein structure prediction tool that has greatly improved the accuracy of protein structure prediction. The programme applies neural network architectures and training procedures based on evolutionary, physical and geometric constraints of protein structures. It takes advantage of multiple sequence alignments (MSAs) and pairwise features for accurate structure prediction. The full-length sequence of Xyl was input into AlphaFold2 structure prediction interphase made available through AlphaFold2.ipynb

(https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb). The output, predicted structure of Xyl, was analysed and presented using Coot (Emsley & Cowtan, 2004) and ChimeraX (Pettersen et al., 2021).

4. Results

The Xyl protein (NCBI database protein accession number **AMO13186**) is composed of four protein domains separated by linkers of varying lengths rich in glycine and serine residues. The N-terminal 28 residues are predicted to function as a secretion signal peptide (SP), which is removed during the secretion process and hence does not form part of the mature, functional protein. The signal peptide is followed by a predicted glycosyl hydrolase family 11 (Xyl-GH11) catalytic domain, two carbohydrate binding modules identified as CBM36-like family members (Xyl-CBM36-1 and Xyl-CBM36-2), a carbohydrate esterase family 4 (Xyl-CE4) module and a C-terminal unstructured region. The Xyl-GH11 and a combination of Xyl-GH11 and the two Xyl-CBM36 domains had previously been partly investigated for an MSc project in our laboratory (Kruger, 2017). Xyl-CE4 has, by contrast, not been characterised previously. The functional and structural characterisation of Xyl-CE4 will therefore be presented first. All other stable protein domains and domain combinations were also crystallised and structurally characterised. These results will follow and will be described relative to those of the Xyl-CE4 domain.

4.1. Cloning, production, and purification of Xyl-CE4

To characterize the Xyl-CE4 domain without the other domains of Xyl, PCR primers (

Primers for polymerase chain reactions

The polymerase chain reaction (PCR) is a fundamental technique used to amplify specific DNA sequences for further analysis. Primers, short single-stranded DNA sequences complementary to the target DNA region, are essential components of PCR reactions. In Table 3.4, we present a detailed compilation of the primers utilized in this study. These primers were meticulously designed to selectively amplify target DNA fragments of interest, allowing for the precise interrogation and characterization of genetic material.

Table 3.4) were designed to amplify a gene fragment encoding Xyl-CE4 from a recombinant plasmid bearing the full-length *xyl* gene. The PCR primers included NcoI and XhoI restriction sites in the forward and reverse primer respectively. PCR amplification products were extracted from agarose gel (section 3.2.7) and digested with NcoI and XhoI. The plasmid DNA pET28a was similarly treated with NcoI and XhoI (section 3.2.5). The resulting insert and

plasmid fragments were ligated (section 3.2.9) and used to transform competent, plasmid propagating *E. coli* cells. Transformed cells were grown on LB plates supplemented with 50 µg/mL kanamycin. Single colonies were diluted in water and used as template DNA for amplification of *xyl-CE4* gene.

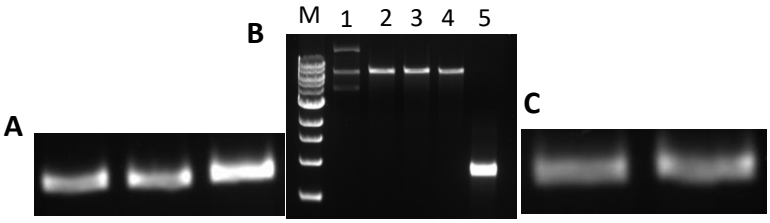


Figure 4.1: PCR Amplification, ligation, and confirmation of the pET28a/xyl-ce4 construct

A) A PCR amplified xyl-ce4 fragment for cloning in pET28a.
B) Preparation of pET28a and xyl-ce4 for ligation. Lane M: 1 kb DNA Ladder (New England BioLabs, Ipswich, MA, USA); Lane 1: Undigested pET28a; Lanes 2, 3 and 4: NcoI, XhoI and NcoI/XhoI digested pET28a; Lane 5: NcoI/Xho digested xyl-ce4 PCR product.
C) PCR amplified xyl-ce4 fragment from newly cloned pET28a/xyl-ce4 constructs.

A *xyl-ce4* fragment was successfully PCR amplified in three independent experiments from full-length *xyl* gene (Figure 4.1A). The plasmid pET28a was initially intact resulting in three visible bands corresponding to supercoiled, circular and linear plasmid DNA (Figure 4.1B, Lane 1). Cleavage of pET28a with NcoI (Lane 2), XhoI (Lane 3) and NcoI/XhoI (Lane 4) result in plasmid linearization such that the bands for supercoiled and circular plasmids are no longer visible. PCR amplified *xyl-ce4* was similarly treated with NcoI/XhoI (Lane 5). Following ligation and transformation of *E. coli* BL21 CodonPlus-RIL cells, the cells were allowed to grow on LB/kanamycin plates at 37°C. Two colonies were selected and their *xyl-ce4* insert directly amplified (Figure 4.1C) indicating the successful cloning and ligation of pET28a/*xyl-ce4.* The construct pET28a/*xyl-ce4* was additionally confirmed by Sanger sequencing.

4.1.1. Small scale Xyl-CE4 production and analysis

A single colony of the *E. coli* BL21-CodonPlus-RIL (Table 3.3) cells harbouring the pET28a/*xyl-ce4* plasmid was cultured in 100 mL LB medium with 50 µg/mL kanamycin at 37 °C and

induced at OD_{600} = 0.6 with 1 mM IPTG for 2 h for recombinant protein production. The experiment was repeated for three equivalent cultures cultured at 22, 28 and 37 °C for 2 h.

SDS-PAGE analysis of small-scale, whole-cell Xyl-CE4 production in *E. coli* BL21-CodonPlus-RIL revealed protein bands at the expected molecular weight of ~23 kDa (Figure 4.2A). Equivalent analyses of soluble (S) and insoluble (I) cell culture fractions for Xyl-CE4 production at 22, 28 and 37 °C indicated optimal production at 28 °C (Figure 4.2B).

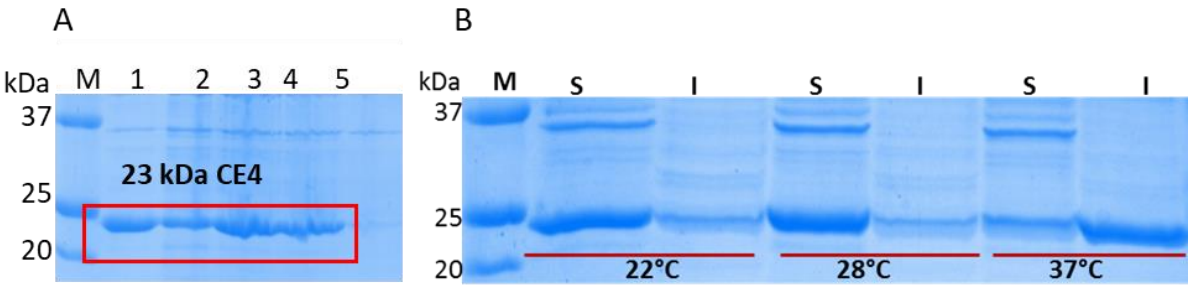
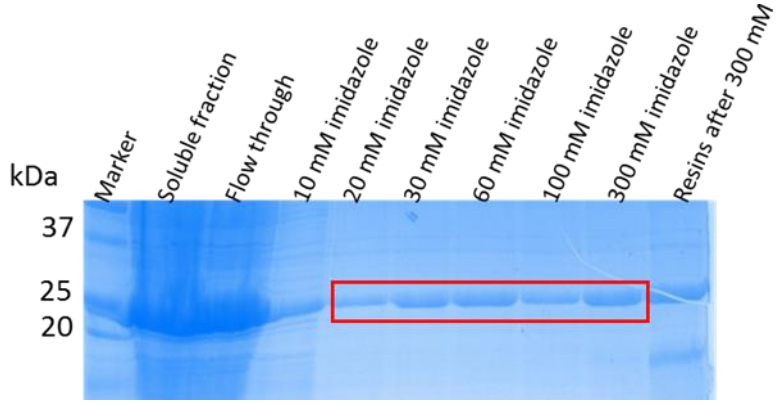


Figure 4.2: Recombinant Xyl-CE4 (23 kDa) production in *E. coli* BL21-CodonPlus-RIL

A) Whole cell analysis of Xyl-CE4 production from five colonies (Lanes 1-5) at 37 °C. The red rectangle encloses bands of overexpressed, recombinant Xyl-CE4.
B) Quantitative analysis of soluble (S) and insoluble (I) cell fractions grown at 22, 28 and 37 °C. Lanes M: Protein marker: Precision Plus Protein All Blue Standards (Bio-Rad). Soluble protein production appears to peak at 28 °C, compared to 22 and 37 °C.

4.1.2. Large scale production and purification of Xyl-CE4

For large scale protein production, 1 L LB medium with 50 µg/mL kanamycin was inoculated with pET28a/*Xyl-ce4* transformed *E. coli* BL21-CodonPlus RIL cells, cultured at 37 °C. At OD_{600} of 0.6 the culture was induced with 1 mM IPTG and grown with agitation overnight at 28 °C. Cells were collected, lysed in LEW buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl) and clarified by centrifugation at 13,000 x *g* for 5 min at 4°C. Soluble fractions were incubated with Co^{2+}-



NTA resins, washed, and eluted with LEW buffer supplemented with 10 to 300 mM imidazole (Figure 4.3).

Figure 4.3: SDS PAGE analysis of Xyl-CE4 purification using Co^{2+}-NTA

Bands corresponding to Xyl-CE4 in fractions eluted from Co^{2+}-NTA resins with increasing imidazole concentrations are emphasised by a red box.

SDS-PAGE analysis of fractions collected during Co^{2+}-NTA purification of Xyl-CE4 reveal a consistent band at 23 kDa that corresponds to expected size of Xyl-CE4. This is already visible in the soluble cellular fraction (Figure 4.3, Lane 2) but also in the "flow-through" fraction that was eluted after incubation with Co^{2+}-NTA resins (Lane 3) implying that significant amounts of Xyl-CE4 were lost in this procedure. Fractions eluted with increasing concentrations of imidazole indicate partly purified Xyl-CE4 with weaker but recognizable remaining bands especially at sizes larger then Xyl-CE4. The fractions eluted with 20 to 300 mM imidazole were pooled and concentrated to 25 mg/mL for further purification by size-exclusion chromatography (Section 3.3.2.3).

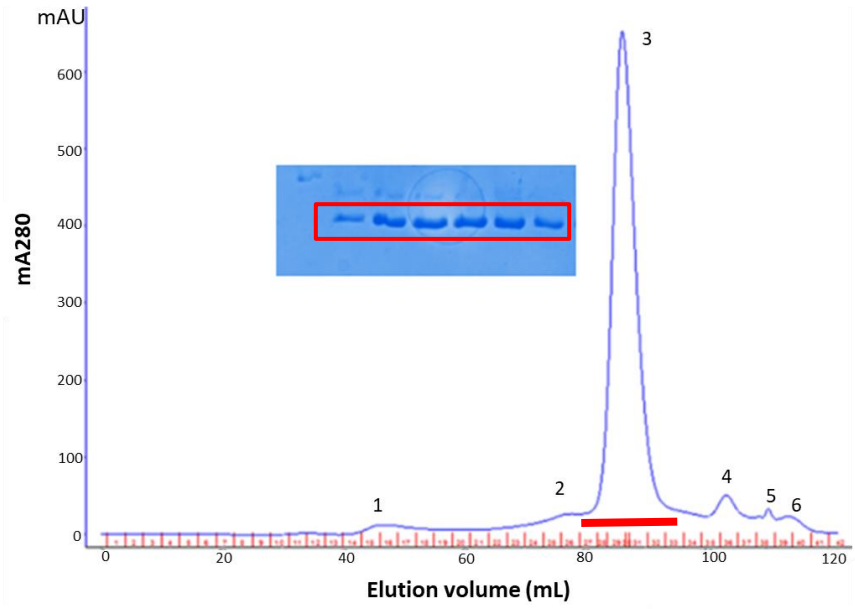


Figure 4.4: Purification of Xyl-CE4 by size-exclusion chromatography

The chromatogram for the purification of Xyl-CE4 by size-exclusion using a Superdex 200 16/60 column (Cytiva, Marlborough, MA, USA) reveals a main prominent peak (peak 3, Figure 4.4) alongside four to six minor peaks (peaks 1, 2 and 4 to 6). Samples from the main peak fractions (red bar) were analysed by SDS-PAGE (Figure 4.4 Insert) revealed a dominant protein

band for Xyl-CE4 (red rectangle) alongside at least one minor, higher-molecular weight contaminant. The protein sample obtained by combining all fractions corresponding to peak 3 was stored at 4 °C for downstream experiments.

4.1.3. Far-UV circular dichroism spectroscopy of Xyl-CE4

The stability of Xyl-CE4 was analysed by recording circular dichroism spectra for the protein at different temperatures (Figure 4.5). Purified Xyl-CE4 was transferred to a 2 mm quartz cuvette and far UV circular dichroism spectra measured from 250 to 180 nm as described (Section 3.3.5).

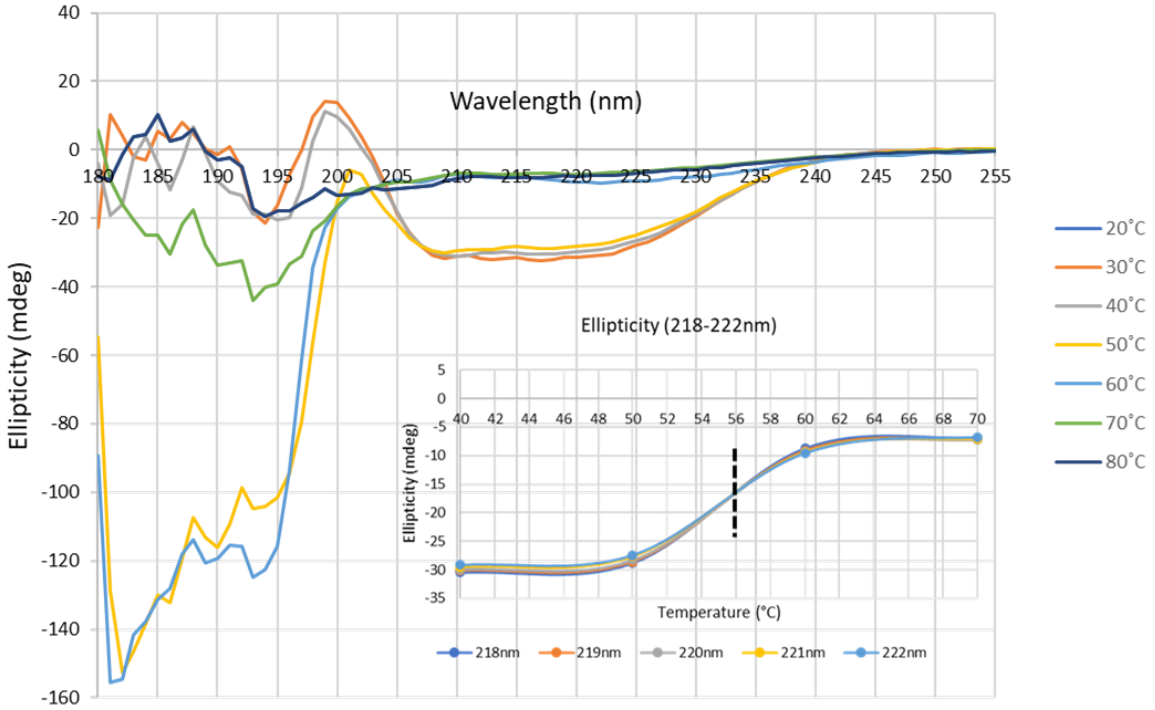


Figure 4.5: Analysis of the Xyl-CE4 domain by far-UV circular dichroism

The CD signature and thermal scan for Xyl-CE4 at temperatures between 20 and 80 °C.

The CD spectrum for Xyl-CE4 at 20 °C reveals two minima at 210 nm and 222 nm, and a maximum at 200 nm (Figure 4.5). This spectrum is typical for predominantly α-helical proteins (Miles et al., 2021). As the minimum at 222 nm is, however, not as distinct as typically seen for purely α-helical proteins, Xyl-CE4 presumably contains additional secondary structural elements. A well-defined CD spectrum moreover indicates that the protein is folded in solution. Additional CD spectra for Xyl-CE4 were recorded between 20 and 80 °C. Ellipticity versus temperature plots between 218-222 nm indicated an inflection point or melting temperature of 56 °C for Xyl-CE4 (Figure 4.5, Inset). This indicates that Xyl-CE4 domain is mesostable, retaining its folded structure up to ~50 °C. Above 60 °C the CD spectrum of Xyl-

CE4 yielded mostly low ellipticity across the entire spectrum, implying a largely unfolded protein. This observation could be critical for the industrial application of Xyl-CE4 as enzymes retain their folded state to function properly. Xyl-CE4 would thus only be useful at temperatures below 50 °C.

4.1.4. pH and temperature profiles of Xyl-CE4

Utilizing spectrophotometry, we assessed the activity of Xyl-CE4 by tracking the liberation of *para*-nitrophenol (*p*-NP) from *p*-NP acetate (*p*-NPA) within a 200 µL reaction mixture at 405 nm and 30 °C. The pH and temperature dependence of Xyl-CE4 activity are elucidated in Figures 4.6A and 4.6B, showcasing the comprehensive exploration of its enzymatic performance under varying environmental conditions.

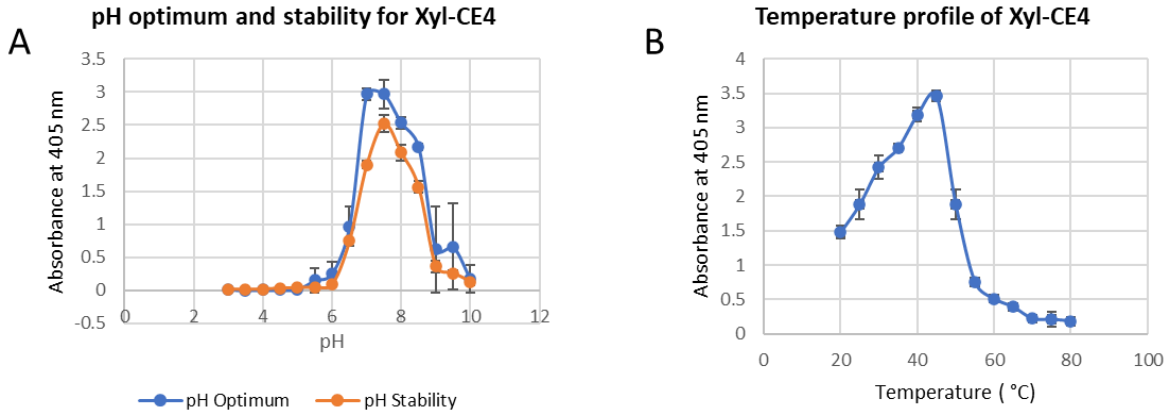


Figure 4.6: pH and temperature profiles of Xyl-CE4

A) Optimum pH and pH stability profiles for Xyl-CE4.
B) Activity of Xyl-CE4 with increasing temperature.

Xyl-CE4 was mostly active in a narrow pH range from 6.5 to 8.5. Maximum activity was observed between pH 7 and 7.5. The enzyme was stable retaining 64 % and 95 % activity at pH 7 and 7.5 after 72 h of incubation, respectively (Figure 4.6A). The activity of Xyl-CE4 was monitored within a temperature range of 20 to 80 °C. The temperature optimum was observed at 45 °C (Figure 4.6B).

4.1.5. Effects of divalent metal ions on Xyl-CE4 activity

The effect of divalent metal ions on Xyl-CE4 activity was assayed at 30 °C by first removing metal ions with 1 mM EDTA and then adding divalent ions before assessing enzyme activity. EDTA treatment reduced enzyme activity to 0 % of the untreated sample (0.2 µg Xyl-CE4 in 50 mM Tris pH 8.0). Addition of 10 mM Co^{2+}, Mg^{2+} and Cu^{2+} restored the enzyme activity to

133, 122 and 110 % of the original activity, respectively. By contrast, 10 mM Mn^{2+}, Ni^{2+}, Zn^{2+} and Ca^{2+} respectively only restored activity to 87, 76, 64 and 34 % (Figure 4.7). The effects of these metal ions on Xyl-CE4 activity may only be a result of bioavailability, not necessarily a reflection of the *in vivo* metal ion preference.

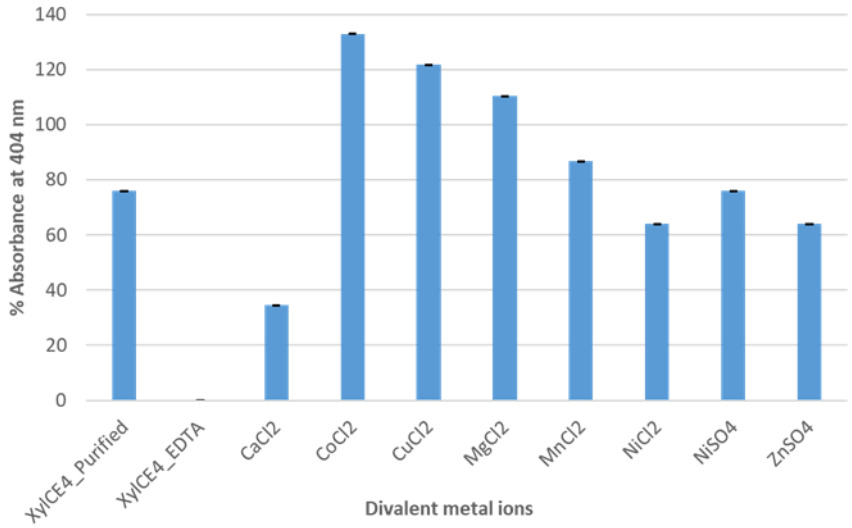


Figure 4.7: Metal ion dependence of Xyl-CE4 catalysis.

The dependence of Xyl-CE4 catalysis on divalent transition metal ions assessed by removing metal ions with 1 mM EDTA and adding divalent ions before assessing enzyme activity. EDTA treatment reduced enzyme activity to 0 % of the untreated sample (XylCE4_purified: 0.2 µg Xyl-CE4 in 50 mM Tris pH 8.0). Co^{2+} is most highly activating. All assays were performed in triplicate and standard deviation error bars indicated.

4.1.6. Crystallization of Xyl-CE4

Purified Xyl-CE4 was buffer exchanged into 20mM Tris-HCl, ph7.5 buffer with 10mM NaCl and concentrated to 10 mg/mL for crystallization. A total of 242 crystallization conditions from the Crystal Screens, PEGs and PEGs II suites were screened by mixing equal volumes of protein and reservoir solution, and equilibrating over 80 µL reservoir solution in seating drop crystallization experiments. After two weeks, hexagonal crystal grew in the following eight conditions of the PEGs II suite as shown in Figure 4.8:

1. condition 6: 0.2 M Li_2SO_4, Tris pH 8.5, 25 % (w/v) PEG 400;
2. condition 25: 0.1 M Na acetate pH 4.5, 4 % (w/v) PEG 4000;
3. condition 26: 8 % (w/v) PEG 4000;
4. condition 27: 0.1 M Na acetate pH 4.5, 8 % (w/v) PEG 4000;
5. condition 28: 0.1 M $MgCl_2$, 0.1 M MES pH 6.5, 10 % (w/v) PEG 4000;
6. condition 29: 0.1 M Na acetate, 0.1 M HEPES pH 7.5, 12 % (w/v) PEG 4000;
7. condition 30: 0.1 M Tris pH 8.0, 12 % (w/v) PEG 4000; and

8. condition 51: 0.1 M Na_2 citrate pH 5.6, 10 % (w/v) PEG 4000, 10 % (w/v) isopropanol.

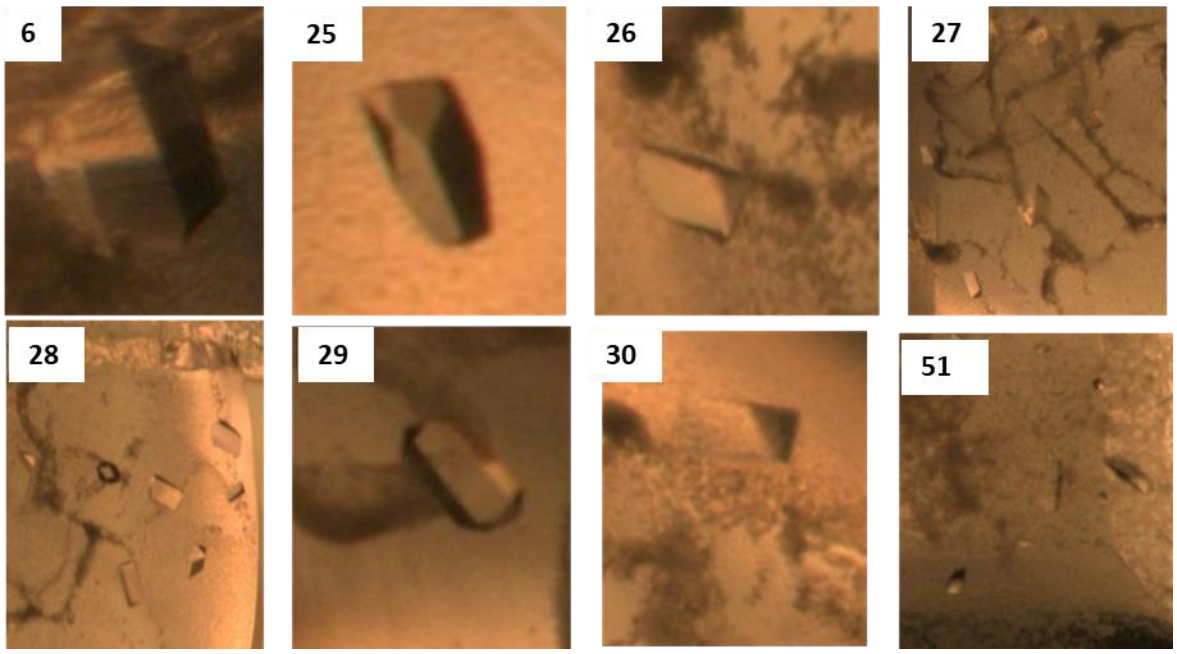


Figure 4.8: Xyl-CE4 crystals grown from PEGs II crystallization suites

Panels depict crystals respectively observed for conditions 6, 25, 26, 27, 28, 29, 30 and 51.

4.1.7. Crystal data collection and evaluation

Selected Xyl-CE4 crystals were mounted in nylon loops, soaked in 20 % PEG400 as a cryo-protectant, placed in Unipucks and loaded into a liquid nitrogen containing Dewar (Section 3.5.3). The Dewar was transported to the University of Cape Town imaging facility for data collection on a home source diffractometer. The cryo-cooled crystals were individually mounted on the goniometer of the X-ray diffractometer where their temperature was maintained at -100 °C by a stream of nitrogen gas. Two test diffraction exposures were recorded 90 ° apart and used to evaluate the crystal quality before a full dataset of 180 exposures representing consecutive 1 ° rotations was collected. Xyl-CE4 crystals diffracted X-rays to a maximum resolution of 1.9 Å, with strong and distinct reflections (Figure 4.9B). The diffraction images were indexed using HKL3000 (Minor et al., 2006; Otwinowski & Minor, 1997) identifying the Bravais lattice as orthorhombic. The space group was $P2_12_12_1$ with unit cell dimensions a, b, and c = 40.1, 59.7, and 86.1 Å while α = β = γ = 90°. The dataset was

integrated, scaled, and merged into an experimental diffraction data file (Section 3.5.4). Table 4.1 summarises experimental data and refinement statistics from Xyl-CE4 crystal.

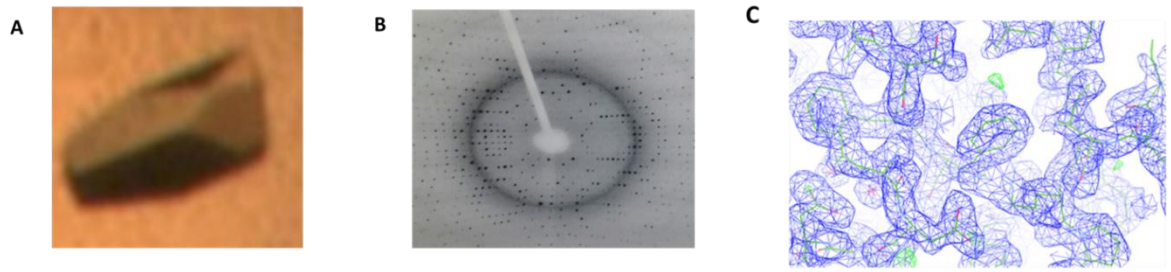
Table 4.1: Data collection and refinement statistics for Xyl-CE4 crystal structure

Data collection statistics	
Wavelength	1.54 Å
Space group	$P2_12_12_1$
Unit cell a b c (Å), α β γ (°)	40.1 59.7 86.1, 90 90 90
Resolution range* (Å)	23.07-1.93 (2.12 - 2.05)
Total reflections*	109092 (2558)
Unique reflections*	13512 (1307)
Multiplicity*	7.1 (6.6)
R_{merge}	0.029 (0.086)
Completeness* (%)	99.0 (96.9)
I/I_{sig}	41.1 (17.9)
Refinement statistics	
R-work*	0.173 (0.253)
R-free*	0.243 (0.372)
Ramachandran Favoured, additional, outliers (%)	97.5, 0.5, 2.0
Rotamer outliers (%)	1.66
Clash score	2.83
RMS bonds (Å), angles (°)	0.006, 0.708
PDB code	**7AX7**

4.1.8. Structure solution for Xyl-CE4

The crystallographic "phase problem" for Xyl-CE4 was sloved by molecular replacement (Section 3.5.4). Briefly, the experimental data file (mtz) combining the diffraction data from 180 exposures after scaling and merging all recorded reflections was combined with the structural model of a family 4 acetyl xylan esterase from *Clostridium thermocellum* (PDB ID: 2C71, sequence identity of 45%, sequence coverage of 93 %) in Phaser (McCoy et al., 2007) as implemented within PHENIX (Adams et al., 2010). After solving the rotational and translational searches for Xyl-CE4, initial phases for Xyl-CE4 were calculated and the corresponding electron density displayed in the graphical program Coot (Emsley & Cowtan,

2004). The model was optimised manually and iteratively to match the electron density (Figure 4.9C). Refinement statistics for the crystal structure of Xyl-CE4 domain are



summarised Table 4.1.

Figure 4.9: Crystal structure determination for Xyl-CE4

A) A single crystal of Xyl-CE4 used to collect diffraction data.
B) A representative image of the diffraction pattern for Xyl-CE4 crystals.
C) A $2F_O$-F_C electron density map (blue) of a section of the Xyl-CE4 domain contoured to 1.0 σ (2.0 Å resolution). A structural model is indicated by green (carbon), red (oxygen) and blue (nitrogen) lines. F_O-F_C difference electron density in green marks deviations between experimental data and the structural model.

4.1.9. Crystal structure of Xyl-CE4 domain

The crystal structure of the Xyl-CE4 domain consist of 201 amino acid residues spanning residues 556 to 757 of the full-length Xyl enzyme (Figure 4.10).

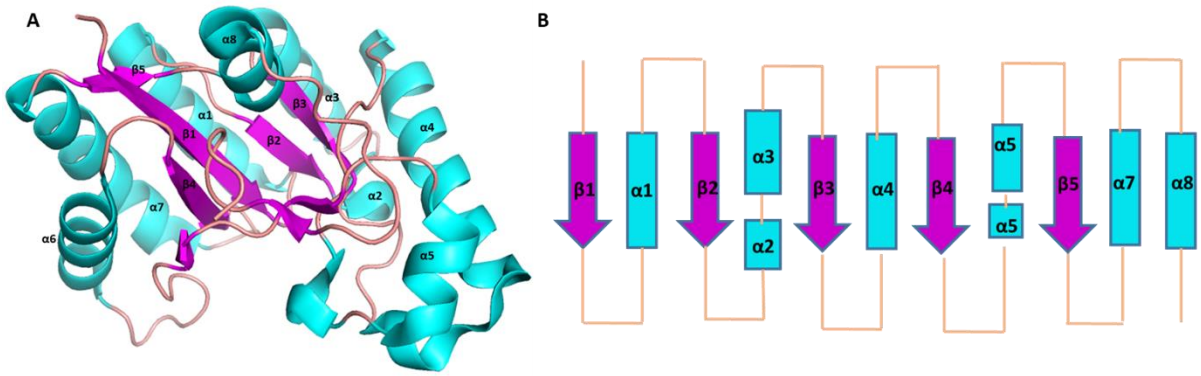


Figure 4.10: The crystal structure of Xyl-CE4

A) A cartoon representation of the Xyl-CE4 structure with α-helices in cyan and β-strands in magenta.
B) Topology diagram of Xyl-CE4. Colours as in A.

Xyl-CE4 crystal structure (Figure 4.10A) contains eight α-helices and five β-strands of varying length that together create a strongly distorted, semi-circular $(\alpha/\beta)_8$ barrel. The α-helices vary from five residues for α2 to 22 residues for α4. Val687 disrupts α5 creating two partial α-helices. The eight α-helices of Xyl-CE4 align to form a roughly semi-circular structure whose concave side is filled by a parallel β-sheet and loops. The β-sheet consists of five mostly short β-strands encompassing three residues for β5 to eight residues in β1. The secondary

structural elements are connected by loops of average six residues in length. The longest loops from residues 696 to 712 and 772 to 786 contain three distorted β-strands. Overall, the structure is related to that of other CE4 enzymes except for the three highly distorted β-strands.

4.1.10. Metal ion binding by Xyl-CE4

A metal ion coordination site was identified in the electron density map of Xyl-CE4 (Figure 4.11). The average interatomic distances between coordinating atoms and metal ion of 2.3 Å confirm metal dative bonding as they are slightly shorter than hydrogen bonds. Coordinating residues include His643 and His647, Asp593 and a free acetate ion.

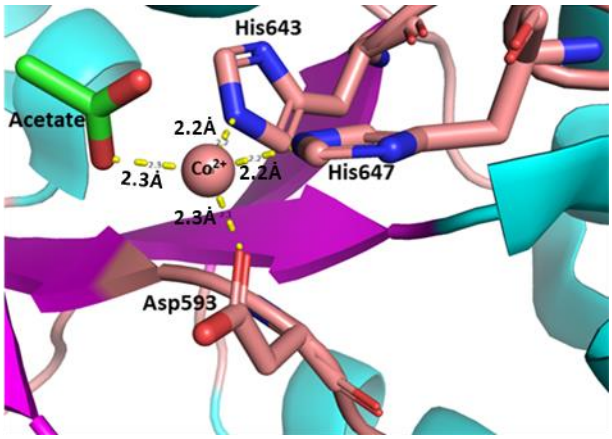


Figure 4.11: The metal ion coordination sphere of Xyl-CE4

The amino acid residues coordinating the central ion are shown as sticks, with dark salmon for carbon, red for oxygen and blue for nitrogen atoms. The Co^{2+} is rendered as a salmon-coloured sphere and the acetate ion as green sticks with read oxygen atoms.

The observed metal coordination lengths are typical for a range of divalent metal ions such as Co^{2+}, Mg^{2+} and Mn^{2+}. Each was placed into the electron density and the model refined. The best refinement statistics were obtained for Co^{2+}, which was hence retained. Metal ions are co-factors for many enzymes. By binding to proteins, they stabilize the structure supporting crystallization. As Xyl-CE4 revealed optimal enzymatic activity with bound Co^{2+} (Section 4.1.5) it was retained in the structure.

4.2. Cloning, production, and purification of Xyl-CBM36

4.2.1. PCR amplification and cloning of *cbm36* genes into pGEX-6P-2

The Xyl protein contains two family 36 carbohydrate binding modules (Xyl-CBM36). These two modules share an 84 % similarity in amino acid sequence. DNA fragments coding for Xyl-CBM36-1 and Xyl-CBM36-2 were individually amplified by PCR from a plasmid with full-length *xyl*.

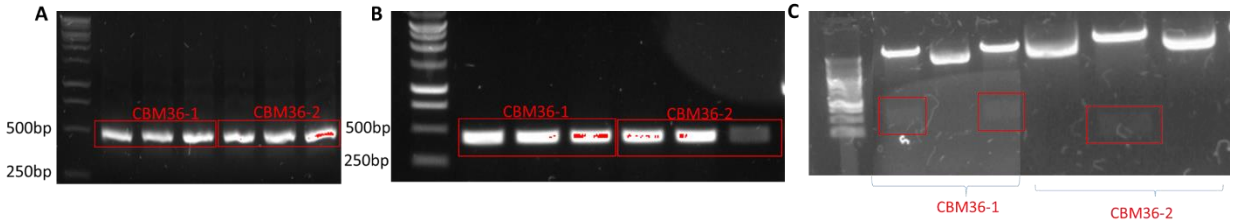


Figure 4.12: Cloning of *xyl-cbm36* genes

A) *xyl-cbm36-1* and *-2* gene fragments (red boxes) amplified by PCR from a recombinant *xyl* full-length plasmid. B) PCR product using three colonies from LB agar plates containing *E. coli* cells transformed with pGEX-6P-2/*xyl-cbm36-1* or *xyl-cbm36-2* recombinant plasmids. All colonies showed positive for *xyl-cbm36-1* and *xyl-cbm36-2* (red boxes).
C) Restriction digestion of plasmids extracted from colonies positive for colony PCR using BamHI and NotI restriction enzymes. Two of the three pGEX-6P-2/*xyl-cbm36-1* recombinant plasmids digested showed insert sizes equivalent to *xyl-cbm36-1* (red boxes) while only one of the three pGEX-6P-2/*cbm36-2* showed a *xyl-cbm36-2* insert (red box).

The PCR products were analysed by agarose gel (Figure 4.12A) before being extracted and purified (section 3.2.7). Forward and reverse primers

Primers for polymerase chain reactions

The polymerase chain reaction (PCR) is a fundamental technique used to amplify specific DNA sequences for further analysis. Primers, short single-stranded DNA sequences complementary to the target DNA region, are essential components of PCR reactions. In Table 3.4, we present a detailed compilation of the primers utilized in this study. These primers were meticulously designed to selectively amplify target DNA fragments of interest, allowing for the precise interrogation and characterization of genetic material.

Table 3.4) respectively included BamHI and NotI restriction sites for direct insertion into the correspondingly digested p-GEX-6P-2 plasmid. The resulting ligated pGEX-6P-2/*xyl-cbm36-1* or *-2* plasmids were used to transform *E. coli* DH5α cells (Table 3.3) and selected for on LB plates augmented with 100 µg/mL ampicillin. Transformed cells produced distinct colonies

while untransformed *E. coli* DH5α cells did not grow confirming the successful ligation of the recombinant plasmids and the uptake of the plasmid.

To confirm that the plasmid taken up by the cells contains the correct gene fragment, a colony PCR was performed using three colonies from the *xyl-cbm36-1* and *xyl-cbm36-2* plates. The resulting bands had the expected sizes of around 480 bp for *xyl-cbm36-1* and *xyl-cbm36-2* (Figure 4.12B).

Positive clones were grown in LB medium. Extracted plasmids were restriction digested with BamHI and Not1 (Figure 4.12C). Of three pGEX-6P-2/*xyl-cbm36-1* plasmids, two revealed inserts of the expected sizes for *xyl-cbm36-1.* Note though that the restriction digest had not proceeded to completion such that the uncleaved plasmid dominates while bands corresponding to the insert and the naked plasmid are only poorly visible. For pGEX-6P-2/*xyl-cbm36-2* plasmids only one of three revealed an insert equivalent in size to *xyl-cbm36-2.*

Plasmids with correctly sized inserts were further checked by Sanger sequencing (Inqaba Biotech, Pretoria, South Africa). All sequenced plasmids showed no mutations, thus, were stored and used for downstream protein production.

4.2.2. Xyl-CBM36-1 and 2 Production and purification

Recombinant plasmids confirmed to contain the *xyl-cbm36-1* and *2* genes were separately used to transform competent *E. coli* BL21-CodonPlus-RIL cells for production of GST-tagged proteins. The pGEX-6P-2 plasmid encodes an N-terminal GST purification tag and a 3C protease cleavage site followed by the protein of interest. Glutathione sepharose (GS) resins (Cytiva, Marlborough, MA, USA) provides covalently bound glutathione to which the fusion protein can bind with high affinity for rapid purification. Flow through fractions were collected and analysed on SDS-PAGE (Figure 4.13A). Partially pure Xyl-CBM36-2 was further purified by size exclusion chromatography (Figure 4.13B).

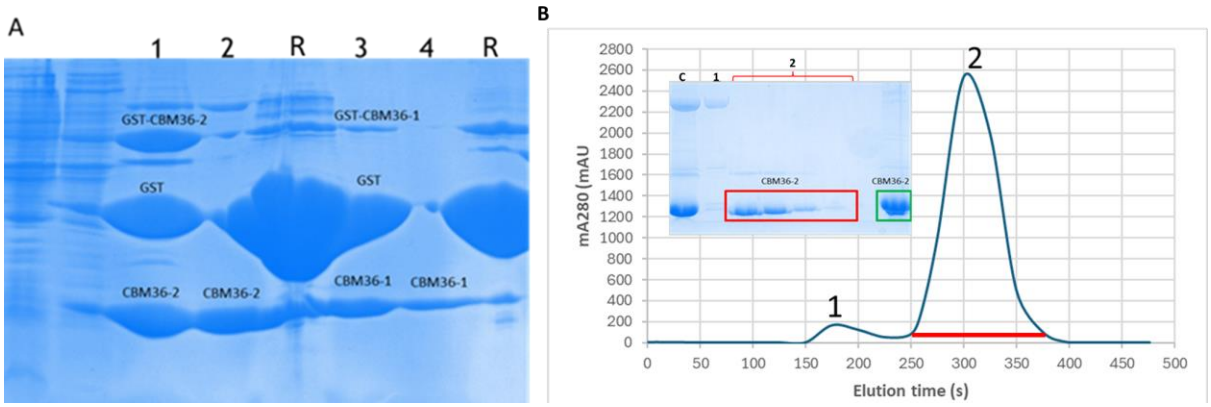


Figure 4.13: Purification of Xyl-CBM36-2 by affinity and size exclusion chromatography

A) Cleavage of GST-Xyl-CBM36-1 and GST-Xyl-CBM36-2 fusion proteins with 3C protease. Lanes 1 and 3: Fusion proteins treated with 3C protease. Lanes 2 and 4: Eluted proteins from 3C protease cleavage. Lane R: GS resins with mainly GST tag.
B) Xyl-CBM36-2 purified by size exclusion chromatography. Insert: SDS-PAGE for fractions under each peak. Lane C: Concentrated Xyl-CBM36-2 for purification. Lane 1: Pooled fractions from peak 1 indicating impurities. The red box marks fractions of peak 2 corresponding to Xyl-CBM36-2. The green box marks concentrated Xyl-CBM36-2 for crystallization.

Figure 4.13A depicts the purification of Xyl-CBM36-1 and 2. In lane 1, a sample of the GS resin, Xyl-CBM36-2, and 3C protease per mixture agitated overnight, 3 prominent protein bands are observed. The top band represent the GST- Xyl-CBM36-2 fusion protein not successfully cleaved by 3C protease. The middle band is GST after Xyl-CBM36-2 is cleaved off and the bottom band is the Xyl-CBM36-2 protein separated from GST by 3C protease. The same is true for Xyl-CBM36-1 in lane 3. Lanes 2 and 4 are flow through fractions collected from allowing the overnight mixture pass through a gravity flow column. Small bands of fusion protein and GST can still be seen in Lane 2 but lane 4 is largely pure Xyl-CBM36-1 and thus was concentrated buffered exchanged into relevant buffers for downstream experiments without further purification. Xyl-CBM36-2 shown in lane 2 was further purified by size exclusion chromatography (Figure 4.13B) before used for downstream experiments.

4.2.3. Characterization of Xyl-CBM36 domains by CD spectroscopy

Far UV circular dichroism (CD) spectra for Xyl-CBM36 domains were measured from 250 to 180 nm on a Chirascan Spectrophotometer (Applied Photophysics, Leatherhead, UK) using a bandwidth of 1 nm and a step resolution of 0.5 nm. Both Xyl-CBM36-1 and 2 were initially scanned at 20 °C in a 2 mm quartz cuvette equilibrated with 20 mM Na_3PO_4 buffer pH 7.4. The thermal stability of the protein was assessed at 10 °C increments from 20 to 80 °C in

the same buffer. Experiments were done in triplicate using protein at a concentration of 4 µM. Purified Xyl-CM36-1 and -2 were characterised by CD spectroscopy (Figure 4.14).

As seen in Figure 4.14, Xyl-CBM36-1 yielded a well-defined CD spectrum with a minimum at 218 and a maximum at 198 nm. This CD spectrum is typical for β-stranded proteins implying that the CBM26-1 domain is mainly β-stranded. CBM36-2, by contrast, produced a spectrum with undifferentiated ellipticity across the wavelength range, indicating that the protein is unfolded in solution. Additional CD spectra of Xyl-CBM36-1 were recorded between 20 and 80 °C at 10 °C intervals to assess the effect of temperature on the protein. The spectra remained largely unchanged across the temperature range implying Xyl-CBM36-1 to be thermostable up to 80 °C.

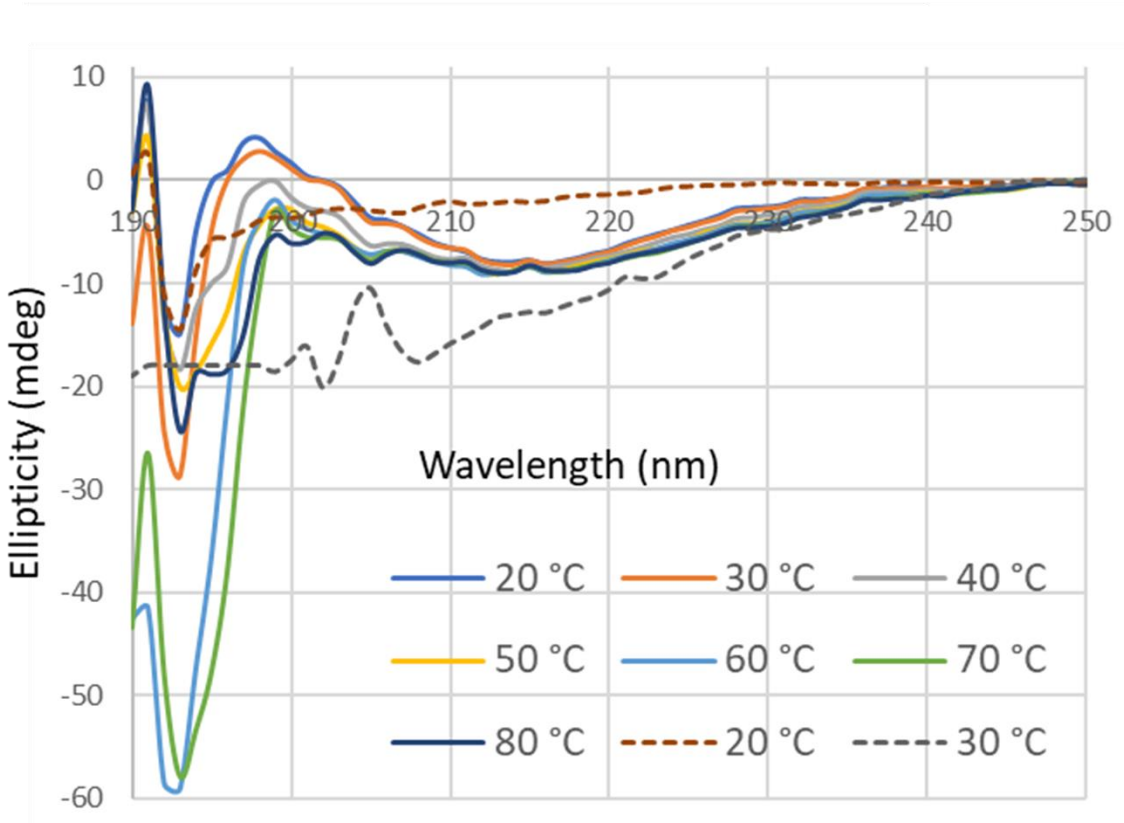


Figure 4.14: CD spectroscopy of Xyl-CBM36 domains

CD spectra of Xyl-CBM36-1 from 20 to 80 °C at 10 °C intervals (continuous lines). The Xyl-CBM36-1 structure is thermophilic remaining largely intact up to 80 °C. Xyl-CBM36-2 spectra at 20 and 30 °C (broken lines). The Xyl-CBM36-2 protein at 20 °C loses its fold at 30 °C indicating a partially folded protein.

4.2.4. Crystallization of Xyl-CBM36 domains

The buffers of purified Xyl-CBM36-1 and 2 protein samples were replaced by 20 mM Tris pH 7.5 and 10 mM NaCl via repeated dilution and concentration. The final concentration for

crystallization was 10 mg/mL. Crystallization was achieved by sitting drop vapour diffusion, with 1 µL of protein and 1 µL of reservoir solution in equilibrium with 80 µL reservoir solution. After two weeks, trigonal bipyramidal crystals of Xyl-CBM36-1 were observed for reservoir conditions of 0.1 M Hepes pH 7.5 and 25 % PEG 3000 (Figure 4.15). Consistent with the CD spectra, Xyl-CBM36-2 did not form crystals under any screened conditions.

Figure 4.15: Xyl-CBM36-1 crystals

Crystals were produced by sitting drop vapour diffusion experiments using 0.1 M Hepes pH 7.5 and 25 % PEG 3000 as precipitant.

4.2.5. Crystal data collection and evaluation

Xyl-CBM36-1 crystals were X-irradiated at the University of Cape Town imaging facility and diffraction data collected and evaluated as described for Xyl-CE4 domain (Section 3.1.7). Data collection and refinement statistics are summarised in Table 4.2

Table 4.2: Xyl-CBM36-1 crystal data collection and refinement statistics

Data collection statistics	
Wavelength	1.54 Å
Space group	$P2_12_12_1$
Unit cell a b c (Å), α β γ (°)	28.4 37.3 108.1, 90 90 90
Resolution range* (Å)	25.16-1.99 (2.06 - 1.99)
Total reflections*	51380 (2092)
Unique reflections*	8260 (521)
Multiplicity*	6.3 (4.0)

R_{merge}	0.055 (0.102)
Completeness* (%)	98.3 (87.8)
I/Isig	22.0 (9.4)
Refinement statistics	
R-work*	0.175 (0.219)
R-free*	0.249 (0.294)
Ramachandran Favoured, additional, outliers (%)	95.9, 3.3, 0.8
Rotamer outliers (%)	0.00
Clash score	1.11
RMS bonds (Å), angles (°)	0.006, 0.785
PDB code	**7AY3**

4.2.6. Xyl-CBM36-1 structure solution

The structure of Xyl-CBM36-1 was solved by molecular replacement using a family 36 carbohydrate-binding module (PDB ID: **1UX7**, Jamal-Talabani et al., 2004) as a homologous model with 49 % sequence identity to Xyl-CBM36-1 in Phaser (McCoy et al., 2007). Phaser added calculated phases to the original diffraction data file (mtz). This was uploaded into the graphical program Coot (Emsley & Cowtan, 2004) together with the coordinate PDB file to manually adjust the Xyl-CBM36-1 crystal structure to the electron density (Figure 4.16). Refinement statistics for the crystal structure of Xyl-CBM36-1 are summarised Table 4.2.

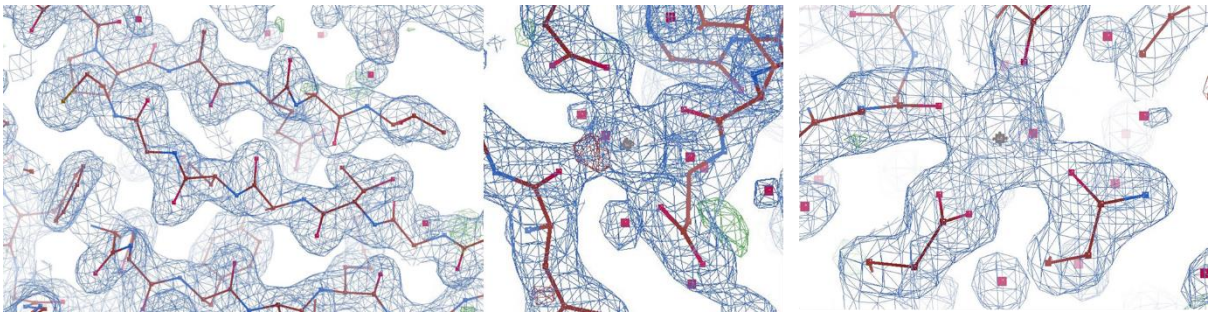


Figure 4.16: Electron density map for Xyl-CBM36-1

A 2Fo-Fc electron density map of Xyl-CBM36-1 contoured to 1.0 σ (2.0 Å resolution). Electron density is represented by blue mesh around amino acid residues shown as sticks.

4.2.7. Crystal structure of Xyl-CBM36-1

The crystal structure of Xyl-CBM36-1 was analysed using PyMol (Schrödinger LLC & DeLano, 2022).

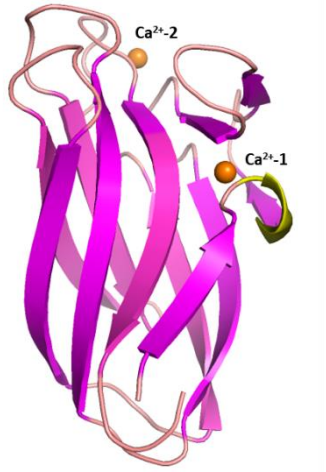


Figure 4.17: Crystal structure of Xyl-CBM36-1 at 1.99 Å resolution.

The structure of Xyl-CBM36-1 is dominated by a sandwich of two antiparallel β-sheets (magenta). It reveals coordination sites for two Ca^{2+} (orange spheres). A sole α-helix fragment is shown in yellow.

The crystal structure of Xyl-CBM36-1 (Figure 4.17) consists of 11 β-strands and one short α-helix between the first two β-strands. Eight of the β-strands form a β-jelly-roll consisting of two antiparallel β-sheets with four strands per sheet. Two metal ion binding sites were observed and modelled as Ca^{2+} in both cases. While Ca^{2+} coordination matched the electron density well, the identities of the bound ions were not cross checked by other procedures (Ravnik et al., 2023; Handing, 2018). The first Ca^{2+} is near the sole α-helix while the second is located near a cleft at one end of the domain. The location of the Ca^{2+} on the domain surface and within a prominent cleft implies that it may be involved in substrate binding.

4.3. Production and purification of Xyl-GH11 domain

The Xyl-GH11 domain, which had been cloned and partially characterised in a previous study (Kruger, 2017) was purified, further characterised and crystallised as part of this study.

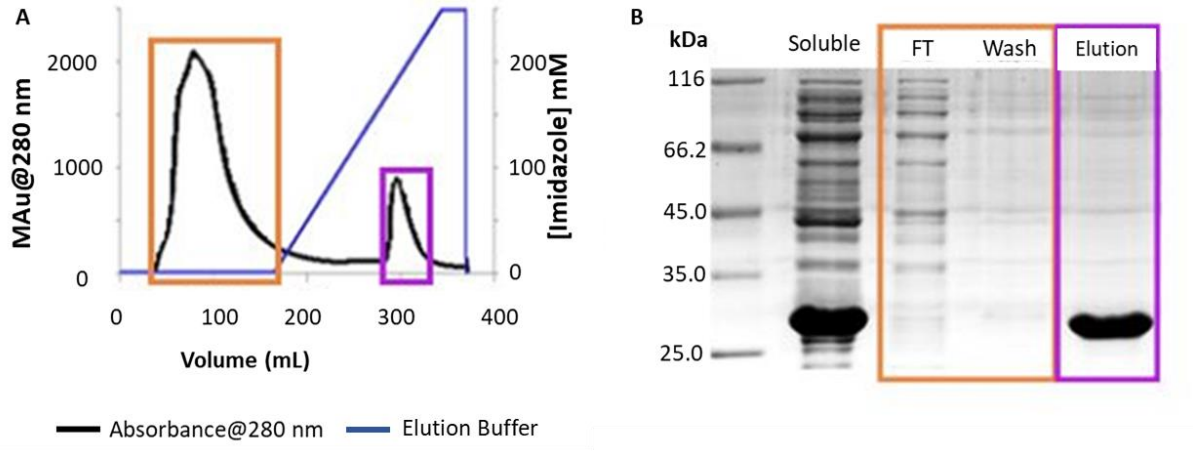


Figure 4.18*:* IMAC purification of Xyl-GH11.

A) IMAC purification of Xyl-GH11. Orange rectangle encloses the flow through (FT) fractions while purple rectangle encloses the eluted Xyl-GH11.
B) SDS PAGE analysis of the total soluble protein fraction loaded into the IMAC and the low through (FT), wash and elution fractions from the IMAC.

Figure 4.18A shows a chromatogram with two peaks and a blue imidazole concentration gradient line. The larger peak marked by an orange rectangle represents the unbound protein fraction. The small peak within a purple rectangle represents where Xyl-GH11 eluted from the column. The His_6-tag on the Xyl-GH11 bound to the Nickel on the column as the soluble proteins fraction was passed through the column. Increasing concentration of imidazole from 0 to 250 mM flowing through the column served as both wash and elution buffer. Xyl-GH11 eluted at an imidazole concentration of 100 mM. The SDS-PAGE, Figure 4.18B, represents the purification scheme. The first lane (soluble) is the entire soluble fraction that was passed through the IMAC column, the overexpressed Xyl-GH11 is seen as a prominent band at 26 kDa. Lanes FT no longer show the prominent band because it is now bound to the column material. Finally, at imidazole concentration of 100 mM, the Xyl-GH11 eluted as a pure protein. The eluted protein was concentrated, and buffer exchanged through size exclusion chromatography (Section 3.3.2.3) into 10 mM Tris pH 7.5, 20 mM NaCl and stored at 4 °C for downstream experiments.

4.3.1. Crystallization and structure solution for Xyl-GH11

Crystals were obtained using 10 mg/mL Xyl-GH11 in 50 mM Na acetate pH 4.8, 500 mM $(NH_4)_2SO_4$ at 18 °C in a hanging drop vapour diffusion crystallization experiment. The crystals were picked up using nylon loop and placed into clean mother liquor mixed with 20 % glycerol

as cryoprotectant. Five diffraction size Xyl-crystals were picked and stored in a liquid nitrogen containing Dewar for transportation to the European Synchrotron Radiation Facility (ESRF), Grenoble, France. X-ray diffraction data were collected on beamline ID23-1 and processed automatically by the Grenoble automatic data processing system (GrenADeS) in fastprocessing mode (EDNA_proc, grenades_fastproc, grenades_parallelproc and autoPROC) (Delagenière et al., 2011).

The Xyl-GH11 domain crystallised in space group P1 diffracting X-rays to a resolution of 1.39 Å but yielding a low completeness of only 87.5 % for the entire data set and 67.0 % for the shell of highest resolution. The low completeness often occurs for low symmetry P1 space groups and a non-ideal crystal orientation. The data, however, still reveals most critical structural details. The structure was solved by molecular replacement using the crystal structure of a xylanase from *Bacillus* sp. 41M-1 (PDB ID: **2DCJ,** unpublished) as a search model. Xyl-GH11 coordinates were deposited in the protein data bank (PDB ID **7ZSZ)**. Another Xyl-GH11 structure was resolved with 99.9 % completeness from a GH11-CBM36-1 construct described in section 4.4.1.

Table 4.3: Xyl-CBM36-1 crystal data collection and refinement statistics

Data collection statistics	
Wavelength	1.54 Å
Space group	P1
Unit cell a b c (Å), α β γ (°)	38.5 43.4 69.9, 100.9 96.3 116.0
Resolution range* (Å)	37.31-1.43 (1.39-1.39)
Total reflections*	107556 (4194)
Unique reflections*	68762 (2652)
Multiplicity*	1.6 (1.6)
R_{merge}	0.049 (0.278)
Completeness* (%)	87.5 (67.0)
I/Isig	13.6 (2.1)
Refinement statistics	
R-work*	0.181(0.216)
R-free*	0.217 (0.268)
Ramachandran Favoured, additional, outliers (%)	97.5, 2.5, 0.0
Rotamer outliers (%)	1.16
Clash score	3.08
RMS bonds (Å), angles (°)	0.006, 0.886
PDB code	**7ZSZ**

4.3.2. Crystal structure of Xyl-GH11

The crystal structure of Xyl-GH11 (Figure 4.19) was analysed and presented using the molecular graphics program PyMol (Schrödinger LLC & DeLano, 2022). Two GH11 monomers were found to constitute the asymmetric unit of Xyl-GH11 crystal. The monomers were named Xyl-GH11_A and Xyl-GH11_B (Figure 4.19A and B) to aid comparison.

The two monomers were overall very similar with slight differences highlighted below. Xyl-GH11_B (Figure 4.19C) was used to draw a topology diagram (Figure 4.19D). As illustrated on the topology diagram, two anti-parallel β-sheets, denoted A and B (Figure 4.19D) make up the Xyl-GH11 crystal structure. The slightly bent β-sheet A is composed of six anti-parallel β-strands whereas β-sheet B, which is more strongly twisted is composed of nine anti-parallel β-strands. The twisted β-sheet B creates a long and deep cleft holding active site residues of Xyl-GH11.

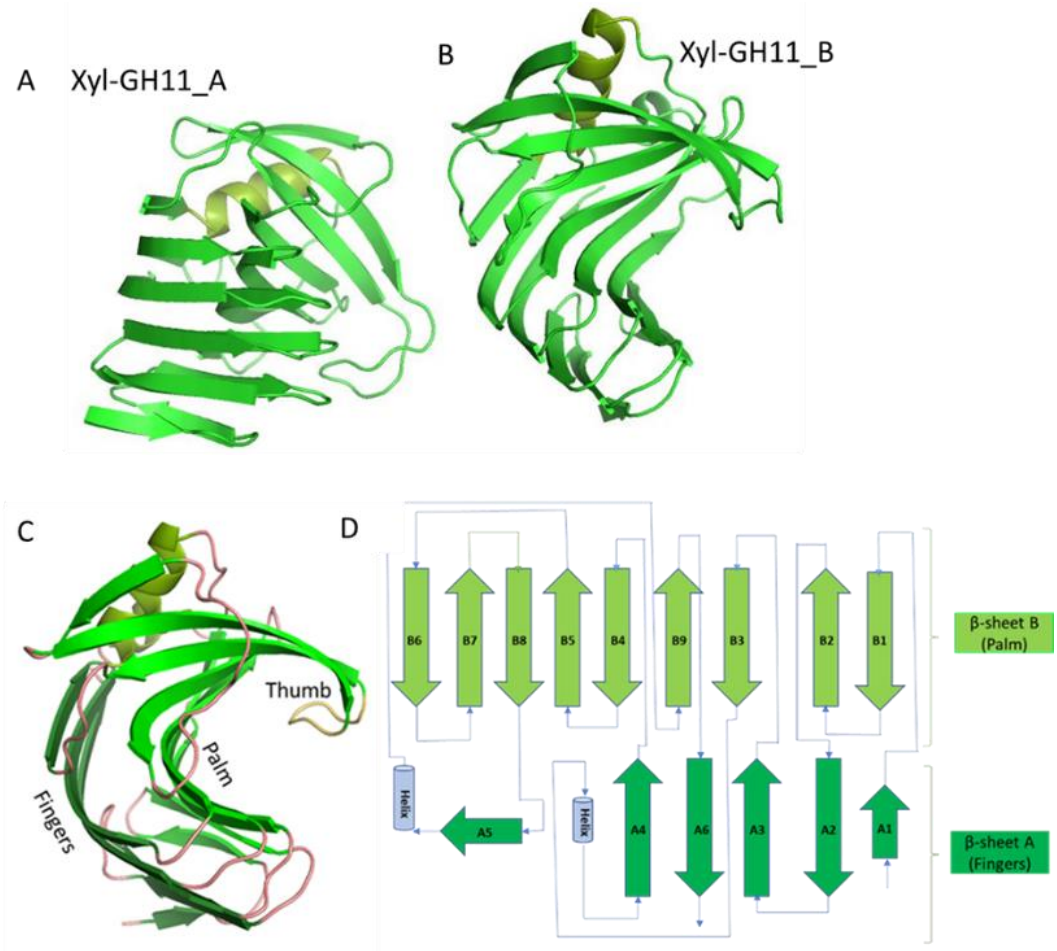


Figure 4.19: Cartoons and topology diagram of Xyl-GH11.

The β-sheets A and B are shown in matt and light green respectively.

A single α-helix and a short α-helical turn lie alongside sheet A, but similarly sandwiched onto β-sheet B. The fold may be compared to a partly closed human right hand where β-sheets A and B respectively represent the “fingers” and “palm” while the loop connecting β-strands B7 and B8 creates a “thumb”.

The two Xyl-GH11 monomers constituting the asymmetric unit were superposed to identify possible differences (Figure 4.20). Of 200 residues, 161 are structurally comparable. Monomer Xyl-GH11_A contains 202 visible amino acids residues compared to 200 in Xyl-GH11_B. Notable structural differences include the following:

1. Ala29 at the N-terminus of Xyl-GH11_A could not be modelled in Xyl-GH11_B resulting to a shorter β-strand A1 in the latter (Figure 4.20c).
2. The loop connecting β-strands B3 to A4 consists of 13 residues extending from Lys81 to Ile94. The loop is continuous in Xyl-GH11_A but is interrupted by a sing helical turn

in Xyl-GH11_B (Figure 4.20d) composed of residues Tyr88, Glu89, Glu90 and Ile91. The observation that identical loop regions adopt slightly different conformations is not unusual and may have functional implications.

3. In Xyl-GH11_A, β-strand A4 is longer by one residue than the same β-strand in monomer B, shortening the B3-A4 loop by one residue compared to monomer B (Figure 4.20d, red arrow).
4. The "thumb" loop between β-strands B7 and B8 consists of 13 residues in monomer A compared to eight residues in Xyl-GH11_B (Figure 4.20b). The thumb loop is important in the opening and closing of the active site, thus difference in the loop length may hold implications in the state of the active site in the monomers.
5. Finally, the C-terminus of Xyl-GH11_A is one residue longer than in Xyl-GH11_B resulting in a longer A6 strand in Xyl-GH11_A (Figure 4.20f).

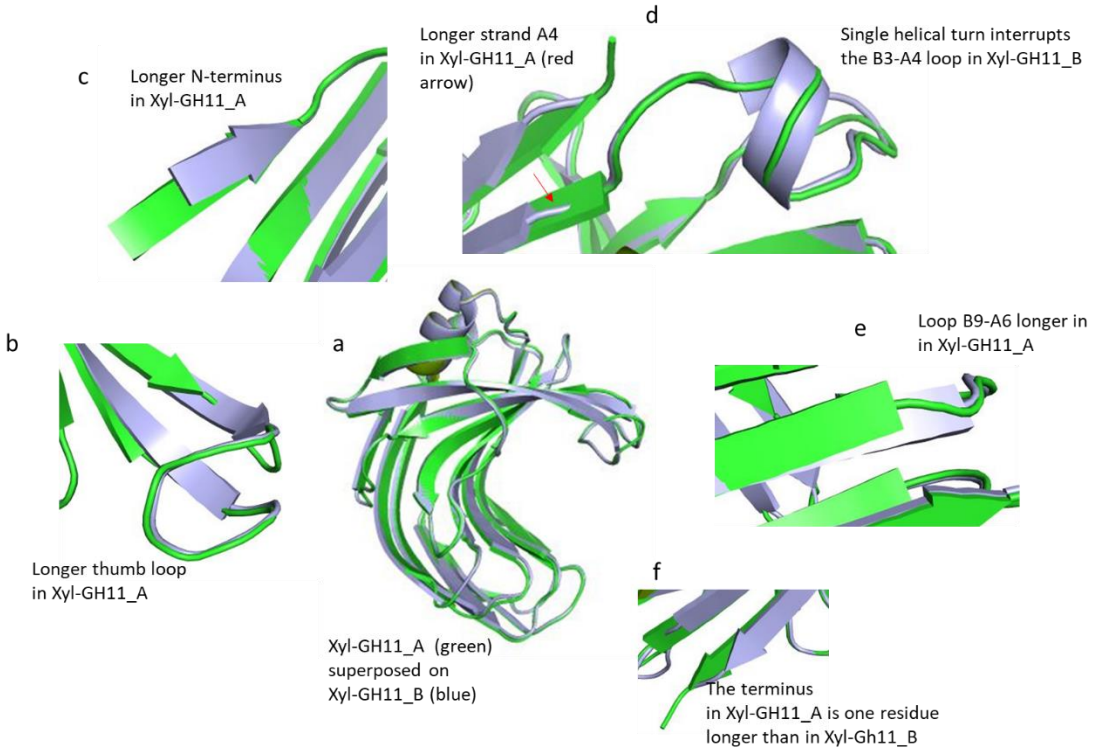


Figure 4.20: Superimposed Xyl-GH11 monomers highlighting differences.

Xyl-GH11_A (green) superimposed on Xyl-GH11_B (blue). The monomers aligned with an overall rmsd = 0.170 (161 to 161 atoms). Areas of differences are highlighted from a to f.

4.4. Purification and crystallization of GH11-CMB36-1

The two-domain protein GH11-CBM36-1 was produced and purified analogously to Xyl-GH11 (Section 4.3).

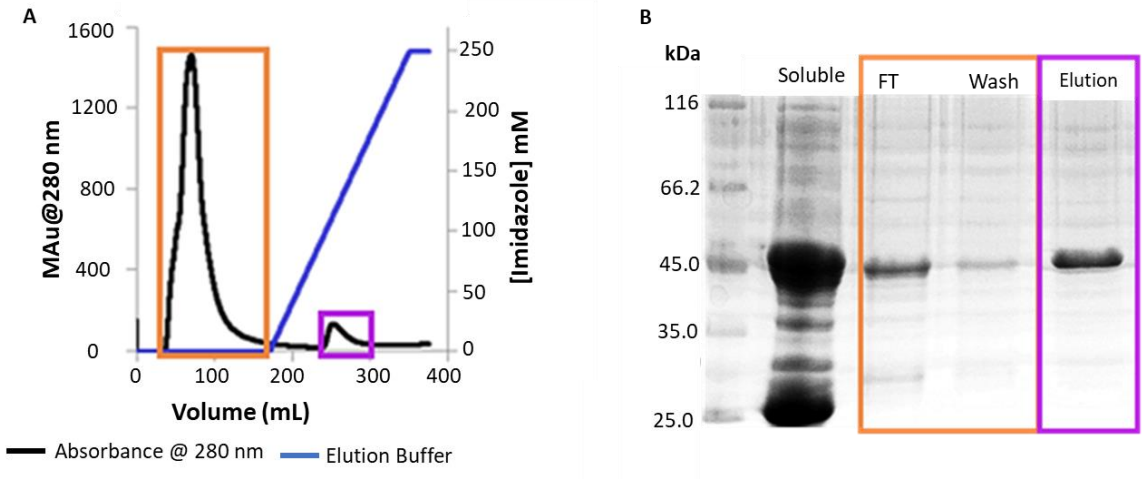


Figure 4.21: IMAC purification of GH11-CBM36-1

Protein of interest was collected from fractions under the chromatogram peak enclosed in a purple box and analysed on SDS PAGE.

The pure protein identified by a single band on an SDS-PAGE (purple rectangle) was concentrated and buffer exchanged by size exclusion chromatography into 10 mM Tris HCl pH 7.5, 20 mM NaCl and stored at 4 °C for downstream experiments. The protein was concentrated to 10 mg/mL and used for crystallization screening. Small needle clusters and microcrystals of GH11-CBM36-1 grew in a reservoir solution of 1 M Li_2SO_4 and 10 mM Tris-HCl pH 7. Crystal optimization strategies included 1) checking for additives using an Additive Screen (Hampton Research, UK), 2) micro, cross and streak seeding, 3) varying the ratio of protein to reservoir volumes, and 4) adding 0.5 to 3 % (v/v) glycerol. Crystal quality was not improved by any of these. After incubating the crystallization experiments at 18 °C for three months, produced a single large crystal in 1 M Li_2SO_4 and 10 mM Tris-HCl pH 7 supplemented with 2 % (v/v) glycerol. The crystal was harvested, and diffraction data collected at the ESRF, Grenoble, France as described for Xyl-GH11. Diffraction data for GH11-CBM36-1 (Table 4.4) revealed that protein crystallised in space group $P6_3$ with unit cell dimensions of 108.2 x 108.2 x 54.2 Å and 90 x 90 x 120 °.

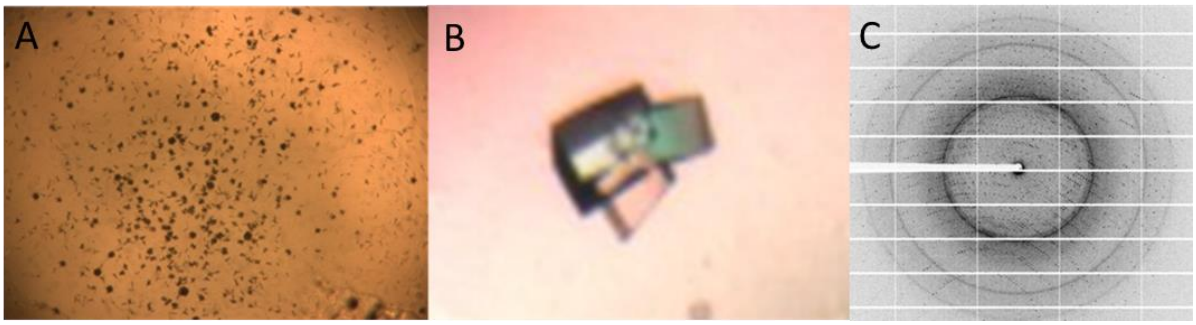


Figure 4.22: Crystals and diffraction image of GH11-CBM36-1

A) Microcrystals of GH11-CBM36-1 from 1 M Li_2SO_4 and 10 mM Tris-HCl pH 7 after 2 weeks of incubation at 18 °C.
B) Single crystal of GH11-CBM36-1 formed after 3 months of incubation of 10 mg/mL in 1 M Li_2SO_4 and 10 mM Tris-HCl pH 7 supplemented with 2 % (v/v) glycerol in a hanging drop vapour diffusion experiment at 18 °C.
C) 1 of 180 diffraction images from crystal in B recorded at the ESRF, Grenoble France.

Table 4.4: GH11-CBM36-1 crystal data collection and refinement statistics

Data collection statistics	
Wavelength	1.54 Å
Space group	$P6_3$
Unit cell a b c (Å), α β γ (°)	108.2 108.2 54.2 90 90 120
Resolution range* (Å)	93.69 - 1.47 (3.90-1.47)
Total reflections*	353537 (11941)
Unique reflections*	65364 (3225)
Multiplicity*	5.4 (3.7)
R_{merge}	0.051 (0.473)
Completeness* (%)	99.9 (99.0)
I/Isig	15.8 (2.2)
Refinement statistics	
R-work*	0.171 (0.199)
R-free*	0.217 (0.268)
Ramachandran Favoured, additional, outliers (%)	97.0, 3.0, 0.0
Rotamer outliers (%)	0.58
Clash score	1.94
RMS bonds (Å), angles (°)	0.006, 0.934
PDB code	**7AYP**

4.4.1. Crystal structure of GH11-CBM36-1

Packing analyses indicated that the crystal structure of GH11-CBM36-1 bears a single molecule in the asymmetric unit. As the protein construct consists of two domains, a molecular replacement procedure was initiated for each using the program Phaser (McCoy et al., 2007). First, an attempt was made to solve the crystal structure of GH11-CBM36-1 by combining the diffraction data file with the coordinates of a related two-domain alkaliphilic xylanase from Bacillus sp. 41M-1 (PDB ID: 2DCJ). The N-terminal Xyl-GH11 domain produced an unambiguous solution resulting in clearly defined electron density (Figure 4.23). The initial structure and the corresponding phases were improved by alternating rounds of manual refinement in Coot (Emsley & Cowtan et al, 2004) and automated refinement in Phenix (Afonine et al., 2012) yielding an R-free of 22.99 %.

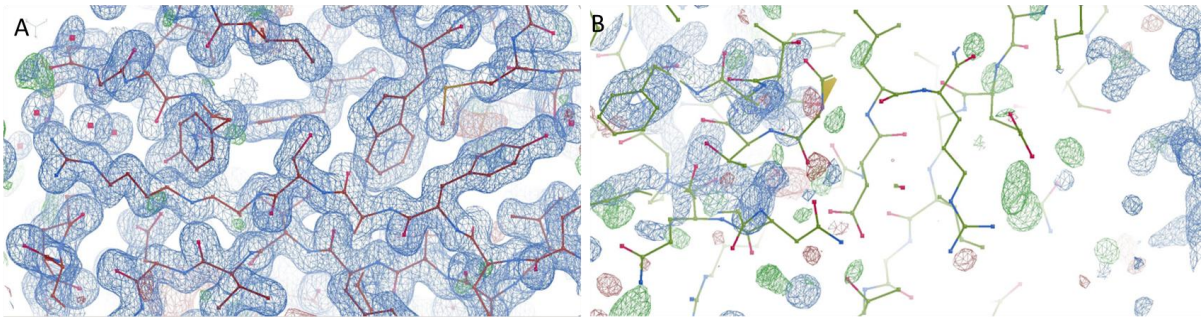


Figure 4.23: 2Fo-Fc electron density map of GH11-CBM36-1

A) Correctly placed Xyl-GH11 domain with amino acid residues (sticks) within electron density map mesh contoured to 1.0 σ (1.5 Å resolution).
B) Incorrectly placed Xyl-CBM36-1 domain lacking electron density mesh around amino acid residues.

By contrast, to the N-terminal domain, no clear solution for the C-terminal carbohydrate binding module Xyl-CBM36-1 was obtained. One solution that physically overlapped with the Xyl-GH11 domain, yielded an R-free of 52 %, revealing it as an incorrect solution. Including the solution for the N-terminal domain as a partial structure in the molecular replacement procedure also yielded no solutions, implying that the C-terminal domain does not adopt a defined position and orientation in the crystal.

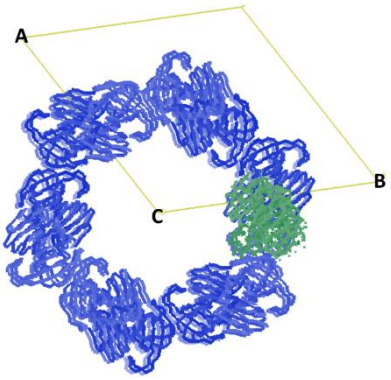


Figure 4.24: Packing diagram for GH11-CBM36-1

Symmetry generated chains of GH11-CBM36-1 represented as C-alpha traces. Only the Xyl-GH11 molecules (green) and its symmetry elements (purple) are observed, forming an empty tunnel presumably to be occupied by Xyl-CBM36-1

A packing analysis of the Xyl-GH11 in Coot (Emsley & Cowtan et al, 2004) reveals that the Xyl-GH11 domain creates a tunnel-like packing arrangement along the C-axis with substantial unoccupied space around the C-axis (). This space would be sufficient to accommodate the Xyl-CBM36-1 domain. The lack of visible electron density within the tunnel, appears to imply that the Xyl-CBM36-1 domains occupy the central void in a disordered fashion. Correspondingly, the Xyl-CBM36-1 domains would not contribute to the diffraction pattern and reciprocally the domain would not be traceable in the electron density.

The Xyl-GH11 crystal structure from GH11-CBM36-1 complements the previously solved crystal structure of Xyl-GH11 (Section 4.3.2) by increasing the data completeness from 87.5 (67.0) % to 99.9 (99.0) %, with values in parentheses indicating the shell of highest resolution. The higher data completeness can be attributed in part to the higher symmetry of the space group ($P6_3$) in which the crystals formed.

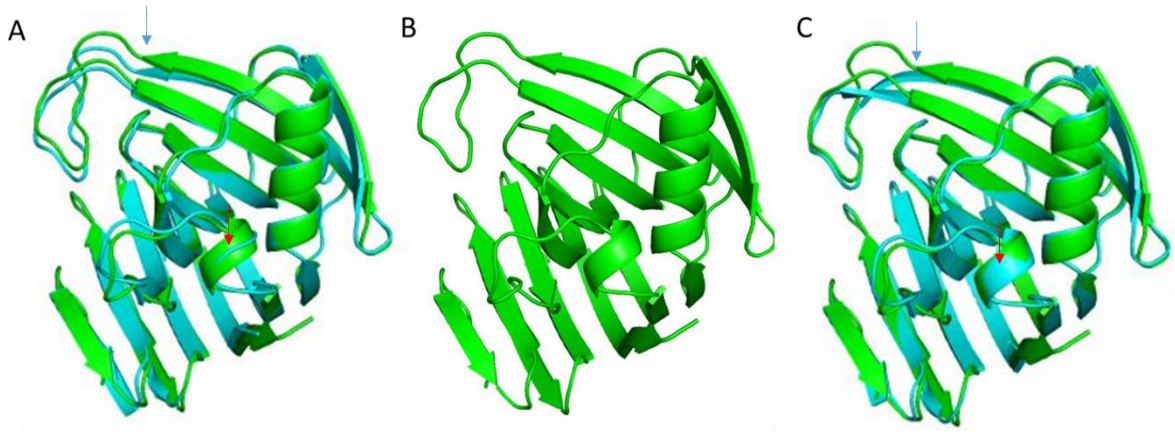


Figure 4.25: Comparison of Xyl-GH11 (green) and GH11-CBM36-1 (cyan).

A) Superimposition of Xyl-GH11 from GH11-CBM36-1 (Figure 38B) onto Xyl-GH11_A. The thumb loops (blue arrow) are of equal length whereas the loop connecting B3 to A4 (red arrow) is interrupted by a single helical turn in GH11-CBM36-1 but no helical turn in Xyl-GH11_A.
B) crystal structure of Xyl-GH11 domain from GH11-CBM36-1.
C) Xyl-GH11_B superimposed on Xyl-GH11 domain from GH11-CBM36-1. Xyl-GH11_B shows a shorter thumb loop (blue arrow) but both molecules have their B3 to A4 loops interrupted by a single helical turn (red arrow).

Superimposing the two crystal structures (Figure 4.25) reveals that the Xyl-GH11 domain from the GH11-CBM36-1 crystal structures is largely identical to the two symmetry independent molecules in the asymmetric unit of Xyl-GH11 (Section 4.3.2). Minor differences involve the thumb loop, where GH11-CBM36-1 does contain the single-turn α-helix observed for Xyl-GH11_B but not for Xyl-GH11_A.

4.5. Production and purification of Xyl and GH11-2CBM36

Previously cloned and sequenced recombinant pET20a plasmids bearing full-length *xyl* and *xyl* without the *ce4* gene (GH11-2CBM36) were separately used to transform *E. coli BL21-CodonPlus-RIL cells.* The transformed cells were cultured in LB medium supplemented with 100 µg/mL ampicillin (Section 3.3.1) and induced for protein production with 1 mM IPTG at OD_{600} = 0.6-0.8. Total protein production was analysed by SDS-PAGE (Figure 4.26A) and showed that the GH11-2CBM36 construct produced significantly more protein than the full-length Xyl construct. Both proteins were partially purified by gravity-flow chromatography using Co^{2+}-NTA resins (Qiagen, Hilden, Germany) with yields of 1.75 mg and 7 mg for the full-length Xyl and GH11-2CBM36 partially pure fractions, respectively. SDS-PAGE analysis (Figure 4.26B) of the recovered Xyl protein showed a second band almost equal in molecular weight to the size of GH11-2CBM36 which completely degraded with overnight storage at 4 °C. GH11-2CBM36 partially purified from the GH11-2CBM36 construct was analysed by size exclusion chromatography (Figure 4.26C) and showed multiple peaks suggestive of different oligomeric states of the protein. To confirm the oligomeric states, a Native PAGE analysis was performed on the samples and results showed protein bands at two different sizes possibly implying different oligomeric states (Figure 4.26C).

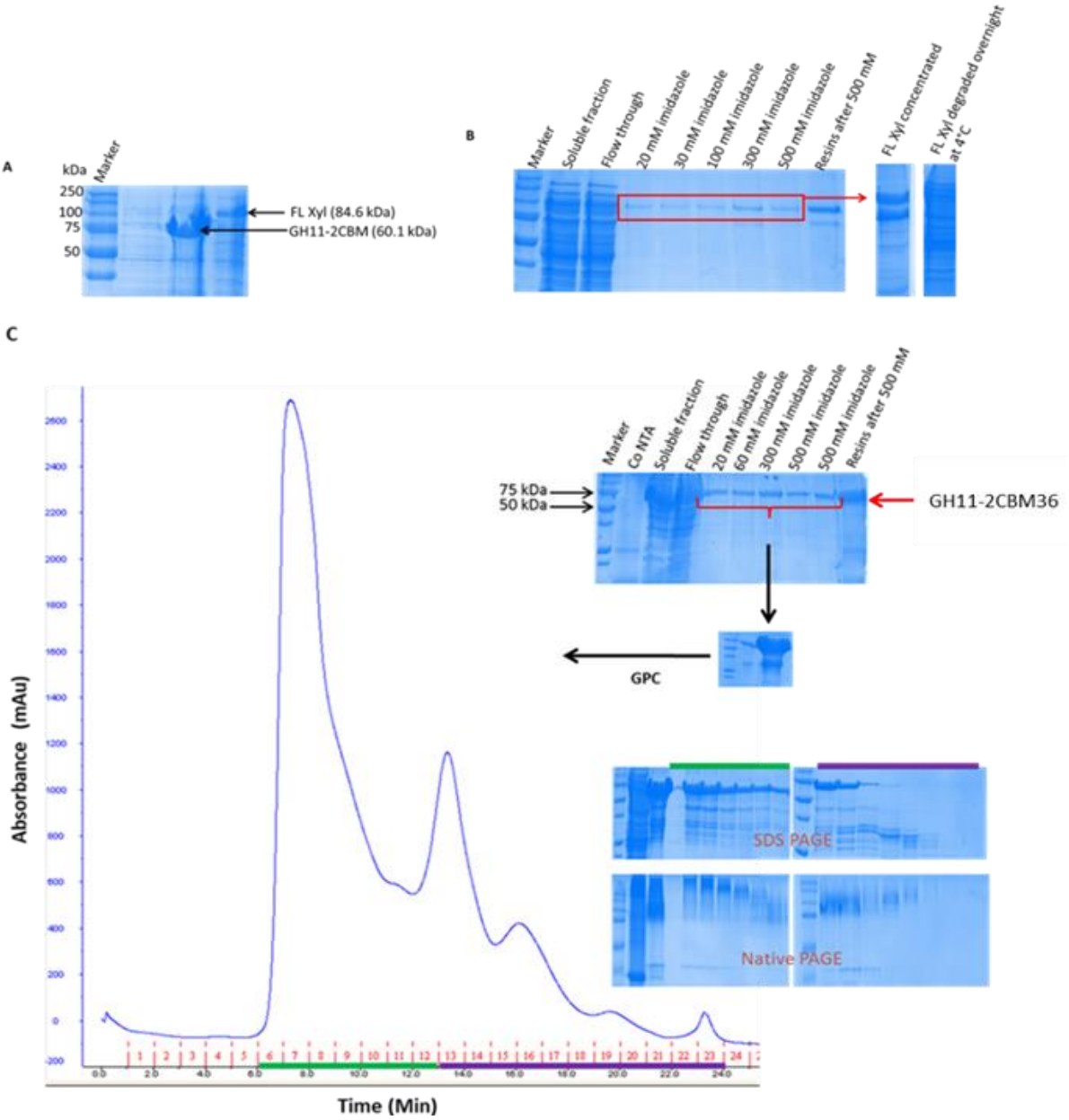


Figure 4.26: Purification of Xyl and GH11-2CBM36

A) Whole cell analysis of protein production.
B) Co^{2+}-NTA purification of Xyl. Xyl was eluted with increasing concentration of imidazole. Elution fractions were pooled and concentrated. Partly purified and concentrated Xyl revealed two major bands at 85 and 60 kDa but protein degraded completely when stored overnight at 4 °C.
C) Co^{2+}-NTA and size exclusion purification of GH11-2CBM36. Top right, SDS-PAGE analysis of protein samples after Co^{2+}-NTA. Protein samples underlined in red were pooled and concentrated for SEC. Samples under the peaks in the chromatograph were analysed by SDS PAGE and Native PAGE.

A re-chromatography (Figure 4.27) was performed on GH11-2CBM36 fractions collected from the major chromatogram peak (Figure 4.26C, green line) to further purify for crystallization. Fractions underlined in purple (Figure 4.26C, purple line) were also pooled but did not yield sufficient protein for further analysis.

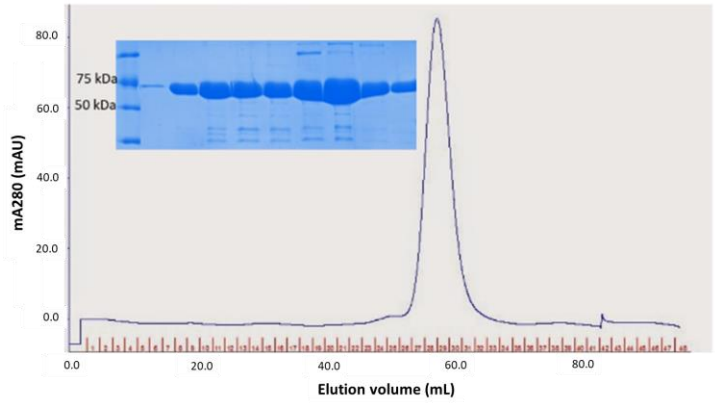


Figure 4.27: Size exclusion chromatography purification of GH11-2CBM36.

Fractions under peak were analysed by SDS-PAGE (see insert).

4.5.1. Crystallization of GH11-2CBM36

Following size exclusion chromatography purification of GH11-2CBM, elution fractions under peak of the chromatogram were pooled and concentrated for crystallization screening. Commercial crystallization screens providing over 400 different buffer conditions were used for initial screening (section 3.5.2). Microcrystals were identified in screen conditions 21 and 81 from the Qiagen NeXtal AmSO4 namely containing 0.2 or 1 M Li_2SO_4 and 2.2 or 1.6 M $(NH_4)_2SO_4$. Three screen conditions from the Qiagen PEGs II suite also yielded microcrystals. The conditions shared 2 M lithium cations and 0.1 M Tris pH 8.5 and contained different PEG sizes and concentrations. Crystal optimization strategies did not significantly improve the size of the crystals. Small, needle clusters or phase separation were observed for reservoir solution consisting of 1 M Li_2SO_4 and 10 mM Tris pH 6 or 7. Microcrystals were too tiny for diffraction experiments.

4.6. AlphaFold prediction of full-length Xyl structure

Ab initio Protein structure prediction methods remained quite limited for many decades while only homology modelling provided somewhat reliable predicted structures (Masrati et al., 2021). The development of AlphaFold (Pearce & Zhang 2021) has shifted the protein structure prediction landscape considerably. AlphaFold can predict protein structures with remarkable accuracy closely approaching experimentally determined structures (Pereira et al., 2021). It combines a deep learning approach with evolutionary information obtained from the sequence alignment of related proteins. The alignments provide information on residue co-variances that can be translated to contact maps. Full-length Xyl amino acid sequence was

submitted to AlphaFold2 structure prediction interphase made available through ColabFold on ChimeraX (Pettersen et al., 2021). The output (Figure 4.28) was analysed and presented using ChimeraX.

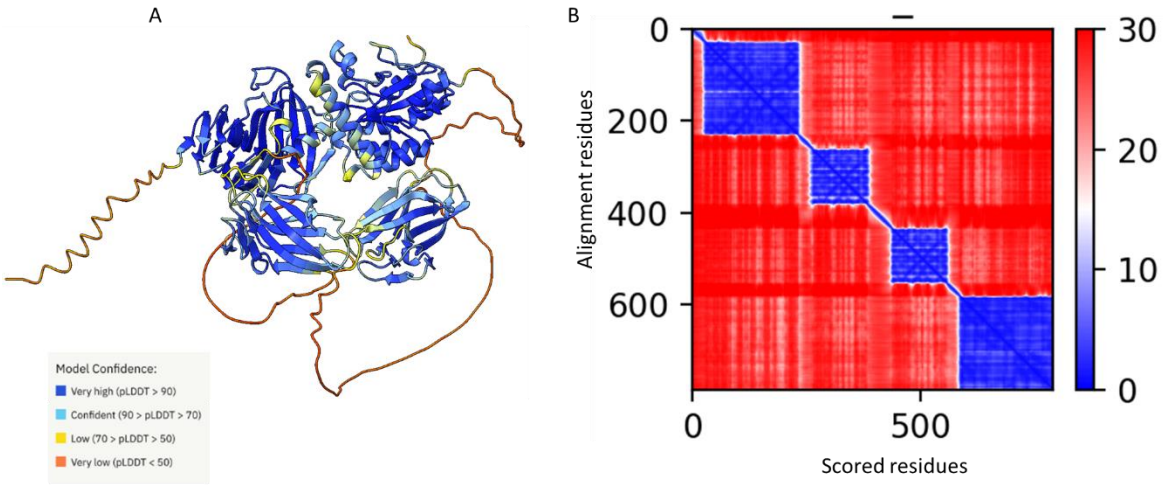


Figure 4.28: AlphaFold2 structure of full-length Xyl

A) The four structural domains of Xyl are predicted with high confidence. Colours indicate the confidence of prediction as indicated on bottom left legend. The confidence of loops prediction is low.
B) The PAE plot highlighting domain and conformational confidence in the predicted Xyl structure. High confidence regions are clustered on the diagonal and coloured blue and correspond to the individual domains. Interdomain regions are coloured red indicating low confidence for domain interactions.

In Figure 4.28, the full-length Xyl 3D model (A) and the predicted alignment error plot (PAE) (B) are presented. The colours in the model reflect the local confidence of the model, as given by the scale in the lower left side: dark blue regions have large confidence (pLDDT > 90) and red regions have low confidence (pLDDT < 50). pLDDT corresponds to the model's prediction of its score on the local distance difference test (lDDT-Cα) and is a measure of local accuracy. This metric replaces the B-Factor found in structural models built from empirical methods. Different reasons account for low confidence score including limited number of sub-sequence homologs in training databases and/or no fix conformation for the regions (intrinsically disordered regions). Individual Xyl domains are predicted with high confidence (Figure 4.28A), however, interdomain linkers have low confidence scores (pLDDT < 50). The PAE plot (Figure 4.28B), identifies in a pairwise fashion, the absolute error of relative position between residues. In the colour scale of AlphaFoldDB, blue correspond to 0 Å, whereas larger errors are coloured in red. Along the diagonal of the heat map, most elements are expected to be close to 0 Å. Well-defined blocks represent a domain of high confidence. A very good model

would display a complete blue heat map. However, this is not true for Xyl where its four domains are well predicted individually (blue squares in the diagonal of the heat map) but their relative positions are not well defined (Red area on heat map). This prediction mirrors the experimental outcomes for Xyl, where three of its four domain structures were readily determined but the structures of combined domains where not achievable. Detailed comparison of experimental versus predicted structures are discussed in section 5.

4.7. Biophysical investigation of interdomain interactions in Xyl

Isothermal titration calorimetry (ITC) was used in this study to investigate possible interdomain interaction between the Xyl domains. In a typical ITC experiment, a solution of a ligand "biomolecule" is titrated into a solution of its binding partner or target molecule(s) inside the cell. The release or absorption of energy in the form of heat is indicated by compensatory peaks following every ligand injection (Leavitt & Freire, 2001). When the target protein is saturated, the signal peaks diminish until only the heat of dilution remains – implying that no more binding is occurring. By integrating the area under each injection peak against the ratio of ligand and target protein, a binding curve is generated (Doyle, 1997). Reproducing the binding curve by fitting to a desired binding model the binding parameters can be inferred (Leavitt & Freire, 2001).

The following domain combinations were investigated in this study: 1) Xyl-GH11 and Xyl-CBM36-1, 2) Xyl-GH11 and Xyl-CE4, 3) Xyl-CE4 and Xyl-CBM36-1. In all cases, we observed no noticeable energies of interaction. Instead, only heats of dilution as exemplified by Figure 4.29 were observed. This implies that the individual domains of Xyl do not appear to bind to each other. Thus, any modulation of other domain functions is in all probability, not due to allosteric or structural interactions.

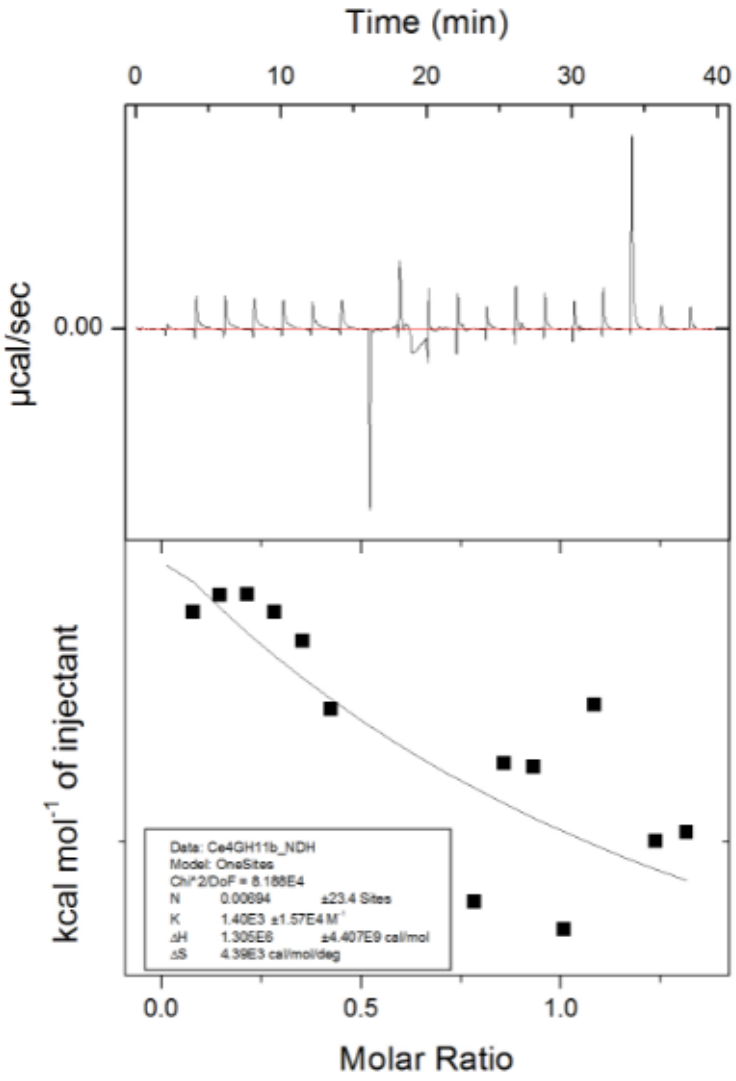


Figure 4.29: Investigation of interaction between the Xyl-CBM36-1 and Xyl-GH11.

Peaks sizes on the top section of the image are almost equal in size, implying heats of dilution than heats of interaction.

5. Discussion

Production of biofuels from lignocellulose requires the latter to be depolymerised ideally by a set of suitable enzymes (Lopes et al., 2018; Wilson, 2011). While lignocellulose may be depolymerised by single-domain proteins, the combination of domains in multidomain proteins and their interactions with each other (Verma et al., 2019), with ligands and/or with other biological structures serve to enhance the depolymerization process. This study examined the multidomain protein Xyl (NCBI GenBank accession number **AMO13186**), a xylanase protein that releases xylobiose and xylotriose from glucuronoxylan and wheat arabinoxylan. The study focused on the structure, function, and possible interactions of the four domains of Xyl. Glycoside hydrolases (GHs) from this hindgut metagenome, including GH5, GH11 and a multidomain GH11 xylanase with similar domain architectures to Xyl but lacking the C-terminal CE4 domain, were previously functionally characterised (Rashamuse et al., 2017).

The presence of homologous multidomain GH11 xylanases in *T. trinervoides* hindgut symbionts incorporating tandem CBM36 modules may imply a possible benefit of the tandem domain arrangement in the hindgut environment, which is based on the grass-derived diet of the harvester termite. To this end, a previous study in our laboratory investigated the influence of the tandem Xyl-CBM36 domains on the Xyl-GH11 enzyme. The study revealed that the two Xyl-CBM36 domains improved catalysis of the Xyl-GH11 catalytic domain to a small but significant degree (Robert Kruger, 2017). This study further investigates this enzyme as an example of lignocellulose degrading enzyme.

An aspect of the Xyl protein not previously investigated is the C-terminal CE4 domain (Xyl-CE4). A similar domain architecture has been reported for a xylanase from rumen Butyrivibrio-related bacteria (Čepeljnik, Križaj, & Logar, 2004). This domain arrangement is not unexpected, considering that many plant carbohydrates, especially from grasses, are acetylated to prevent enzyme recognition and degradation (Zhao, Zhang, & Liu, 2012a). Xyl-CE4 was delineated, produced, and characterised in this study.

5.1. Crystal structure of Xyl-CE4

Xylan esterases hydrolyse both O- and N-acetylation sites in xylan backbones of plant cell walls, alongside other enzymes. Xyl-CE4 is a de-O-acetylase. It is structurally related to de-N-acetylases of the NodB superfamily that degrade and/or remodel N-acetylglucosamine substrates (Hoell et al., 2010). Xyl-CE4 adopts a distorted $(\alpha/\beta)_8$ fold structurally related to other CE4 enzymes. It retains the five catalytically important motifs [T(F/Y)DD, H(S/T)xxH, R(P/x)PY, DxxD(W/Y) and I(V/I)LxHD] of NodB homology domains (Figure 5.1) (Blair et al., 2005; Caufrier et al., 2003).

```
|pdb|5O6Y|A        ---------------------------------WTPFSWVEKYAYAFSGPYNKA---EVAL     25
|pdb|2C1G|A        FDYKDSQIILYPSPVVENLEEIALPVSAFFDVIQSSYLLEKDAALYQSYFDKKHQKVVAL     240
|pdb|2C1I|A        FDYKDSQIILYPSPVVENLEEIALPVSAFFDVIQSSYLLEKDAALYQSYFDKKHQKVVAL     240
|pdb|2CC0|A        ------------------------------------------------AACNGYVGL      9
|pdb|5LFZ|A        ---------------------------------HHHHHHAGQPEPVATPPAVDCATTKCVAL     29
Xyl-CE4            ------------------------------------------------GNEKLIAL      8
|pdb|2C71|A        ------------------------------------------------MPANKLVAL      9
                                                                     :.*

|pdb|5O6Y|A        TFDDGPDLEFTPKILDKLKQHNVKATFFLLGENAEKF-PNIVKRIANEGHVIGNHTYSHP      84
|pdb|2C1G|A        TFDDGPNPATTPQVLETLAKYDIKATFFVLGKNVSGN-EDLVKRIKSEGHVVGNHSWSHP     299
|pdb|2C1I|A        TFNDGPNPATTPQVLETLAKYDIKATFFVLGKNVSGN-EDLVKRIKSEGHVVGNHSWSHP     299
|pdb|2CC0|A        TFDDGPS-GSTQSLLNALRQNGLRATMFNQGQYAAQN-PSLVRAQVDAGMWVANHSYTHP      67
|pdb|5LFZ|A        TFDDGPG-EYTNRLLDELSEQHTPATFFVLGKNVKKY-PKTLKRMVDEGHQIGSHTFDHK      87
Xyl-CE4            TFDDGPS-STTSDVLDILERYGVKATFFLIGQNVNSNTLSIMQRQVRMGCELASHSYTHQ      67
|pdb|2C71|A        TFDDGPDNVLTARVLDKLDKYNVKATFMVVGQRVNDSTAAIIRRMVNSGHEIGNHSWSYS      69
                   Motif 1                                                Motif 2

|pdb|5O6Y|A     NLAKVNEDEYRNQIIKTEEILNRLAGYAPKFIRPPYGEILENQLKWATEQNFMIVQWSVD     144
|pdb|2C1G|A     ILSQLSLDEAKKQITDTEDVLTKVLGSSSKLMRPPYGAITDDI-RNSLD--LSFIMWDVD     356
|pdb|2C1I|A     ILSQLSLDEAKKQITDTEDVLTKVLGSSSKLMRPPYGAITDDI-RNSLD--LSFIMWDVD     356
|pdb|2CC0|A     HMTQLGQAQMDSEISRTQQAIAGAGGGTPKLFRPPYGETNATLRSVEAKYGLTEVIWDVD     127
|pdb|5LFZ|A     DITKLTAEGIEHEVQWTDEAIEQAAGVKPQILRPPYGAHGAVYDR---LIPYPLVLWDVD     144
Xyl-CE4         DMTNMSAQNIRNEMEWTSSAIKNSVGVDVKFFRPPYIAVNNTMYQ---NIDYP-FIQGTL     123
|pdb|2C71|A     GMANMSPDQIRKSIADTNAVIQKYAGTTPKFFRPPNLETSPTLFN---NVDLV-FVGGLT     125
                                          Motif 3

|pdb|5O6Y|A     TVDWK-GVSADTITNNVLGNSFPGSVILQHSTPGGHLQGSVDALDKIIPQLKTKGARFVT     203
|pdb|2C1G|A     SLDWK-SKNEASILTEIQHQVANGSIVLMHDIHSP----TVNALPRVIEYLKNQGYTFVT     411
|pdb|2C1I|A     SLDWK-SKNEASILTEIQHQVANGSIVLMHDIHSP----TVNALPRVIEYLKNQGYTFVT     411
|pdb|2CC0|A     SQDWN-NASTDAIVQAVS-RLGNGQVILMHDWPAN----TLAAIPRIAQTLAGKGLCSGM     181
|pdb|5LFZ|A     TLDWK-HHDPQKTVRIALEEAKPGSIILMHDIHES----SVKAVPQLVSKLHDAGYTLVT     199
Xyl-CE4         INDWENSTSVQQRVNNALGAAKDGQIILLHDFQGN--SQTVQALPQIIEGLKNQGYTFVT     181
|pdb|2C71|A     ANDWIPSTTAEQRAAAVINGVRDGTIILLHDVQPEP-HPTPEALDIIIPTLKSRGYEFVT     184
                Motif 4                 Motif 5

|pdb|5O6Y|A     LPSMFQ-TSKER----------------------   214
|pdb|2C1G|A     IPEMLN-TRLKAHELYYSRDE------------   431
|pdb|2C1I|A     IPEMLN-TRLKAHELYYSRDE------------   431
|pdb|2CC0|A     ISPQTG-RAVAPDGS------------------   195
|pdb|5LFZ|A     VDQLFAGTDFKPAKAYDHRFKTNP---------   223
Xyl-CE4         VSELFEKKGVNPNV-EY-KIWSNANGQ------   206
|pdb|2C71|A     LTELFTLKGVPIDPSVK-RMYNSVPLEHHHHHH   216
```

Figure 5.1: Sequence alignment of CE4 enzymes

The five motifs that make up the active site are highlighted and numbered. The amino acid sequence of Xyl-CE4 is highlighted with a grey background. All motifs are conserved in Xyl-CE4 except a single residue change in motif 4 from a conserved Asp to a Leu.

The second aspartate of motif 1 and the two histidine residues of motif 2 coordinate a metal ion, preferably Co^{2+} (Blair et al., 2005; Taylor et al., 2006). The first aspartate of motif 1 is the probable base during catalysis, by activating a water molecule to carry out a nucleophilic

attack on the carbon in the scissile C-N bond. The histidine in motif 5, when protonated, promotes C-N bond cleavage by functioning as a catalytic acid and protonating the leaving amino-sugar. Tyrosine in the backbone of motif 3 stabilizes the oxyanion intermediate created during catalysis (Andrés et al., 2014; Blair et al., 2006), whereas motifs 3 and 4 form a shallow active site groove (Blair et al., 2005). The first Asp in motif 4 is conserved in most CE4 enzymes but is replaced by a Leu in Xyl-CE4 implying that it is not involved in the hydrolysis reaction, nevertheless, Xyl-CE4 was shown to be active in this study implying another residue(s) is involved in the hydrolysis. The positioning of Thr703 suggest that its carbonyl oxygen possibly aids in hydrolysis reactions instead.

5.1.1. Metal ion binding by Xyl-CE4

CE4 enzymes have been observed to coordinate a single divalent metal ion with a preference for Co^{2+} (Taylor et al., 2006). In Xyl-CE4, electron density for a corresponding metal ion is discernible. It is coordinated by the side chains of His643, His647 and Asp593, a single water molecule, and an acetate ion. The metal ion was modelled as Co^{2+}. After multiple cycles of refinement, its temperature factor was 16 $Å^2$. This is similar to the average temperature factor of 14 $Å^2$ for the coordinating atoms, implying this is a suitable interpretation. The metal ion coordination sphere is distortedly octahedral with a bidentate but asymmetric acetate ion creating one metal ligand with coordinating distances of 2.2 and 2.3 Å. The arrangement of two histidine and an aspartate is a "classical" coordination for bivalent metal-ion dependent hydrolases in which the metal-ion gives Lewis-acid assistance to nucleophilic attack.

The crystal structure of Xyl-CE4 and its metal-ion coordination is similar to that described for the NodB homology domain of the peptidoglycan de-N-acetylases from *Streptococcus pneumonia* (SpPgdA) (Blair et al., 2005) and a marine *Arthrobacter* species (ArCE4A) (Tuveng et al., 2017). SpPgdA and ArCE4A share amino acid sequence identities of 38 and 36 %, respectively, with Xyl-CE4. The overall structures of SpPgdA and ArCE4A show rmsd of 0.77 Å for 128 equivalent Cα positions and 0.94 Å for 146 equivalent Cα positions with the crystal structure of Xyl-CE4, respectively. Structural difference between SpPgdA and Xyl-CE4 structures are seen at the N-terminus, where SpPgdA has two additional domains.

5.1.2. Catalytic mechanism of CE4 esterase

The catalytic mechanism of CE4 enzymes generally involves a base-catalysed generation of a hydroxide ion, stabilized, in a Lewis-acid sense, by the metal ion, which then undergoes direct nucleophilic attack at the electrophilic carbon centre concomitant with leaving group departure. Departure of the leaving group may, if necessary, receive enzymatic general acid assistance. The Mg^{2+} and Co^{2+} complexes of CtCE4 and the acetate/Zn^{2+} complex of SlCE4, and by extension the acetate/Co^{2+} of Xyl-CE4 are certainly consistent with such a mechanism. Following proposals made for the peptidoglycan de-N-acetylases, SpPgdA (Blair *et al.*, 2005; Blair & van Aalten, 2004), we infer that acetyl xylan ester hydrolysis involves general base activation of water, with CtCE4 Asp488, SlCE4 Asp13, and Xyl-CE4 Asp593 playing the role of the Brønsted base with acid-catalysed assistance to the departure of the xylose-moiety provided by His539 for CtCE4, His66 for SlCE4, and His647 for Xyl-CE4. Site directed mutagenesis of the SpPgdA enzyme has confirmed that both His and Asp are essential for catalysis (Blair et al*.,* 2005). Xyl-CE4 activity at increasing temperature correlated well with secondary structure changes demonstrated by the Far-UV-CD data. At 50 where the Xyl-CE4 structure starts to change, there is a corresponding drop the in activity of the enzyme.

5.2. Xyl-GH11

Xyl-GH11 is the main catalytic domain of Xyl. Its crystal structure revealed two GH11 molecules in the asymmetric unit, providing for an internal corroboration of the structure. Overall, the two domains have a rmsd of 0.170 Å over 161 C_α-atoms, indicating that they are mostly identical. In both GH11 molecules, the β-sheet B bends to create a substrate-binding cleft. A glutamate dyad consisting of Glu122 and Glu212 within the cleft creates the active centre that cleaves the substrate (Figure 5.2A). Glu122 on β-strand B6 serves as the catalytic nucleophile while Glu212 on β-strand B4 is the general acid. The distance between the glutamate carboxylate groups is 10 Å, close to the average distance of 9.5 Å in inverting glycoside hydrolases. Xyl-GH11, is however, known to be a retaining glycoside hydrolase (McCarter & Withers, 1994; Sinnott ML, 1990). Correspondingly, the general acid in GH11 retaining enzymes can adopt two distinct conformations during the catalytic cycle (Wan, 2015). The “up” conformation is observed in a closed active site resulting in a distance of ~5.5 Å between the catalytic carboxylate groups (Figure 5.2B, C2). The “down” conformation

is seen in open conformations resulting in an inter-carboxylate distance of 10 Å (Figure 5.2 B, C1). This corresponds to the distance in Xyl-GH11.

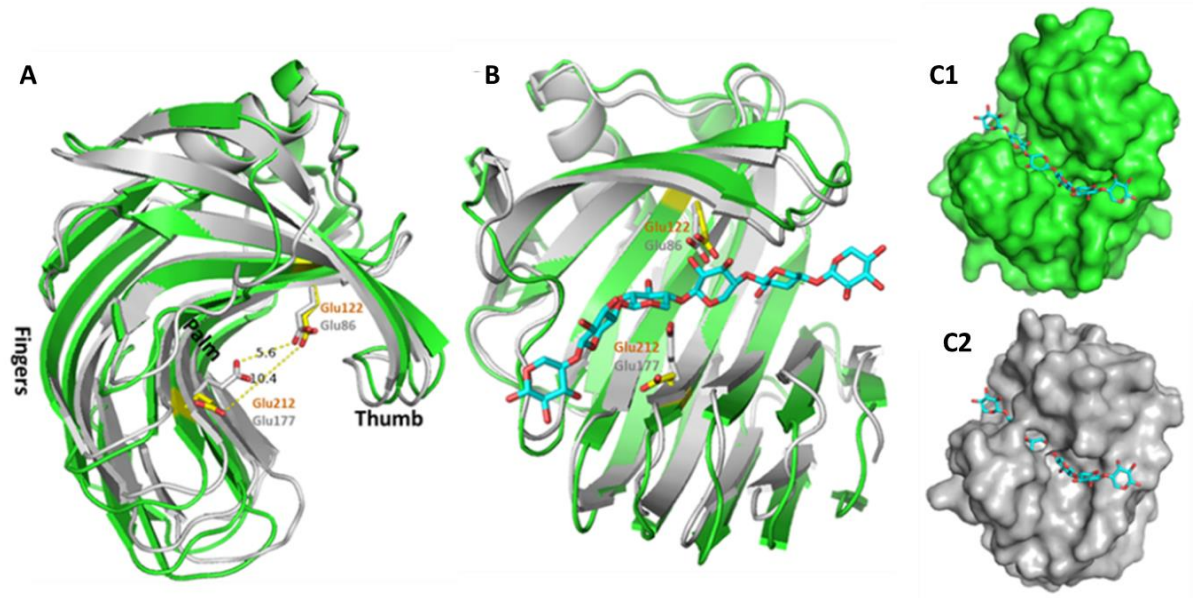


Figure 5.2: Comparison of Xyl-GH11 and TrXyn11

A) Crystal structure of Xyl-GH11 (green) superimposed on that of TrXyn11 (gray) "Fingers", "palm" (outer and inner β-sheets), and the "thumb" regions are indicated. The catalytic residues are shown as sticks. Glu177 of TrXyn11 adopts an "up" conformation (gray) whereas the corresponding residue Glu212 in Xyl-GH11 is in the "down" conformation (yellow).
B) Xylohexose modelled into the active site of Xyl-GH11 and TrXyn11. Surface view shows the modelled substrate is enclosed by the enzyme in TrXyn11 with the up conformation **(C2**) but open in Xyl-GH11 (**C1**).

In Figure 5.2, the crystal structure of the *Trichoderma reesei* enzyme Xyn11 (TrXyn11) represents the "up" conformation, while Xyl-GH11 adopts the "down" conformation. The catalytic residues in both enzymes are shown as sticks. In the closed conformation the functional groups are separated by 5.6 Å as in Xyn11 (Figure 5.2 C2), but 10 Å the open conformation as in GH11-Xyl (Figure 5.2 C1).

The glutamate dyad represents a "retaining" glycoside hydrolase working by a double displacement mechanism. The pH optima of GH11 xylanases are influenced by amino acids near the general acid catalytic residue. In enzymes with an acidic or basic pH optimum, this residue is either aspartic acid (Asp) or asparagine (Asn) (Li & Turunen 2015; Joshi et al., 2000). The residue adjacent to Glu212, the catalytic general acid of Xyl-GH11, is Asn210. However, Xyl-GH11 is active in a pH range of 5-8.5, with optimum activity at pH 6 (Kruger, 2017; Rashamuse et al., 2017), indicating a deviation from the general rules. The wide pH activity range for Xyl-GH11 makes the enzyme adaptable to different industrial applications and increases its value as an industrial enzyme candidate.

5.3. Crystal structure of the Xyl-CBM36-1

The crystal structure of the Xyl-CBM36-1 (Figure 4.17) revealed a β-jelly-roll fold. This fold is commonly observed for CBMs that bind short polysaccharides. The architecture matches that of the "X9" domain from the *Paenibacillus polymyxa* xylanase Xyn43A, the first and previously only structure of a CBM36 domain (Jamal-Talabani et al., 2004).

5.3.1. Modelled substrate binding site for Xyl-CBM36-1

In the context of multidomain proteins, CBM36 domains are critical in protein stability and substrate binding (Jamal-Talabani et al., 2004; Kitamoto et al., 2019). To model substrate binding in Xyl-CBM36-1, its crystal structure was superimposed on the substrate bound crystal structure of the "X9" domain of Xyn43A (PDB ID **1UX7**).

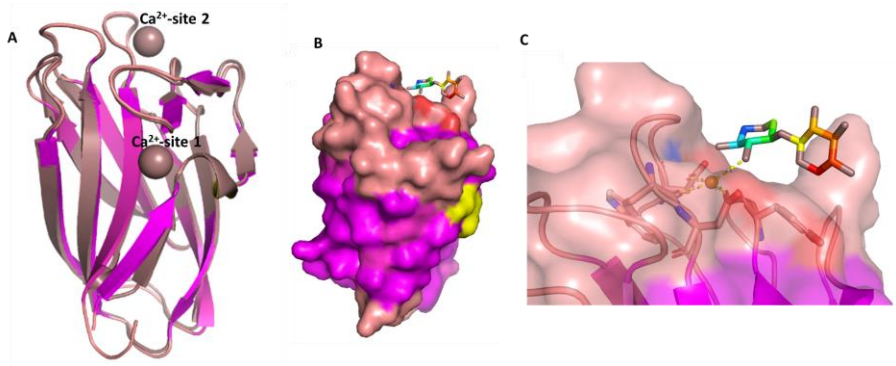


Figure 5.3: Modelled substrate binding in Xyl-CBM36-1:

A) Superimposed crystal structures of Xyl-CBM36-1 (magenta) and X-9 CBM (PDB ID: 1UX7, brown). The rmsd of 0.478 for 108 to 108 atoms.
B) Substrate modelled into the active site of Xyl-CBM36-1. Surface view (0 % transparency) highlighting the substrate binding cleft near the Ca^{2+}-site 2.
C) Surface view at 40 % transparency highlighting amino acid residues involved in protein-substrate interaction and the calcium ion properly positioned to assist in protein-substrate binding.

Surface view and substrate binding residues in the superimposed model suggest that the substrate binding site of Xyl-CBM36-1 is in the same position as that of the X9-CBM domain. This substrate binding site is analogous to "site 2" of the *Clostridium stercorarium* CBM6-3 domain (CsCBM6-3) crystallized in complex with laminaribiose (PDB ID **1OD3**). The CsCBM6-3 structure has two substrate binding sites: Site 1 is in a long open cleft on the concave face of the β sheet and is typical for type B CBMs with aromatic residues for substrate binding. Site 2 is a shallow cleft created by three loops at the edge of the domain. CBM36 domains only

have site 2 binding sites where aromatic residues recognize the substrate. The superimposed Xyl-CBM36-1 and X9 domains (Figure 5.3) identifies Trp114, Tyr35 and Tyr21 as suitably positioned for substrate binding. A flexible loop Ala109-Asp110-Asn111-Gly112 presumably contributes to substrate binding in analogy to the Xyn43A X9 domain, rather than an open groove in CBM6 domains (Jamal-Talabani et al., 2004).

5.3.2. Metal ion coordination in Xyl-CBM36-1

The crystal structure of Xyl-CBM36-1 revealed two metal ion binding sites. Both were identified as Ca^{2+} based on their coordination geometry.

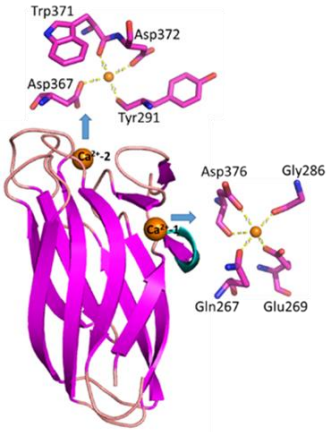


Figure 5.4: Ca^{2+} binding sites in Xyl-CBM36-1.

Ca^{2+} coordination at Site 1 (Ca^{2+}-1), is conserved in both CBM6 and CBM36 proteins. The metal ion binding at this site does not influence Xyl-CBM36-1/substrate interaction. At Site 2 (Ca^{2+}-2) metal ion coordniation is conserved for CBM36 proteins and is implicated in protein-substrate interaction.

Similar Ca^{2+} coordination sites have been identified in Xyn43A "X9" (Jamal-Talabani et al., 2004), while site 1 was also observed in the structure of CsCBM6-3 (Boraston et al., 2003). The first Ca^{2+} in Xyl-CBM36-1 is coordinated by main-chain carbonyl oxygen atoms of Asp376 and Gly286, as well as side-chain oxygen atoms of Asp376, Glu269 and Gln267. The second Ca^{2+} is coordinated by main-chain carbonyl oxygen atoms of Tyr291 and Trp371, as well as side chain oxygen atoms of Asp367 and Asp372. Site 1 in X9-CBM preferably coordinated Mg^{2+} in the absence of substrate but is replaced by Ca^{2+} in the presence of substrate and increased Ca^{2+} concentration (Jamal-Talabani et al., 2004). Ca^{2+} is also implicated in substrate recognition in site 2 in contrast to CBM6, where metal ions are not involved. Our refinement supported the presence of Ca^{2+} in both metal ion binding sites of Xyl-CBM36-1, suggesting that Ca^{2+} is the preferred ion for CBM36 domains.

5.4. GH11-2CBM36 and GH11-CBM36-1

The three-domain construct GH11-2CBM36 had previously been partly characterized to understand the evolutionary benefit of the tandem Xyl-CBM36 domains to the Xyl-GH11 catalytic domain. The study revealed that the Xyl-CBM36 domains only marginally affected the catalysis of the Xyl-GH11 domain. Nevertheless, the GH11-2CBM construct was more soluble and stable than full-length Xyl allowing rapid purification (Figure 4.26). Size exclusion chromatography purification unexpectedly revealed multiple peaks, representing at least two oligomeric states for GH11-2CBM. As this had not previously been noted, we postulated the states as different aggregation forms caused by high protein concentration. Crystallization screening experiments yielded microcrystalline material though no improvement in crystal size and quality could be achieved.

By contrast, the two-domain construct GH11-CBM36-1 did yield crystals suitable for structure determination. Only the larger Xyl-GH11 domain could, however, be located in the unit cell, while attempts for the smaller Xyl-CBM36-1 failed. Analysing the protein packing arrangement for the GH11-CBM36-1 domain, revealed that the Xyl-GH11 domains created long open voids parallel to the C-axis () that could accommodate the Xyl-CBM36-1 domain in a random orientation. This observation presumably implies that while the Xyl-GH11 domain creates a three-dimensional network required for a physical crystal, the Xyl-CBM36-1 domains occupy variable positions in the crystal, preventing their localization in the electron density map. This interpretation is consistent with a model where domains are linked by long flexible loops with little to no specific recognition between them. Comparing the packing of GH11-CBM36-1 to that of the protein Xynj (PDB ID: **2DCJ**), the only other crystal structure in the PDB combining a GH11 domain with a visible CBM36 domain, indicates a much shorter linker of only 11 residues for Xynj compared to 33-residues for GH11-CBM36-1. While XynJ also crystallised in a cylindrical arrangement, the packing is more compact with a well-defined linker. Furthermore, the packing arrangement indicates that it would essentially be impossible for the GH11 domain alone to create a three-dimensional lattice that could provide enough space to accommodate two CBM36 domains in random orientations, explaining the lack of success in resolving the GH11-2CBM36 construct by X-ray crystallography.

5.5. Full-length Xyl

The functional and structural characterization of a protein requires the protein to be pure, stably folded and resistant to proteolytic degradation. Properties that favour rapid protein degradation include a shallow folding landscape and flexible loops. Full-length Xyl was observed to degrade overnight at 4 °C – presumably aided by the shallow folding path of Xyl-CBM36-2 as well as the flexible interdomain linkers.

The rapid degradation of Xyl could also have been aided by the difference in Ca^{2+} usage by the Xyl-CBM36 and Xyl-CE4 domains. The crystal structures of the Xyl-CBM36-1 indicated a Ca^{2+}-binding site – supporting earlier data that implied the importance of Ca^{2+} for CBM36 folding, stability and function (Jamal-Talabani et al., 2004; Tanimoto et al., 2016). Interestingly, Ca^{2+} was found to partially inhibit the Xyl-CE4 domain. This apparent contradiction could be resolved by an intermediate Ca^{2+} concentration – high enough to stabilise CBM36-1 (and possibly 2) but sufficiently low not to inhibit CE4. Insights to full-length Xyl were gained from theoretical models.

5.6. Experimental versus theoretical structures of Xyl

Application of AlphaFold2 in predicting the 3D structure of Xyl affirmed the robust and efficiency of the program in predicting single domain structures. The individual Xyl domains were predicted with >90 % accuracy. Individual Xyl domains determined experimentally in this study were superimposed on the predicted 3D structure. The overall fold of each domain was correctly predicted, and very minor differences were observed between secondary structural elements. For the Xyl-GH11 domain, the experimental versus predicted structures aligned with rmsd of 0.81 Å for 190 equivalent Cα positions.

The major structural difference was observed at the thumb loop where the loop is longer in the predicted structure than in the experimental structure. The thumb loop is important in the opening and closing of the active site, thus difference in the loop length may hold implications in the state of the active site of the protein. The Xyl-CBM36-1 and Xyl-CE4 crystal structures also aligned to the AlphaFold2 predicted equivalents with rmsd of 0.43 Å for 105 and 0.338 Å for 177 equivalent Cα positions respectively. Most structural differences are observed in loop regions.

Overall, the predicted structures show long interdomain loops. This observation corroborates the crystal structure of GH11-CBM36-1 where the Xyl-CBM36-1 domains could not be seen in the crystal structure, presumably because of the long and flexible linker between the GH11 and Xyl-CBM36-1 domains. The flexible interdomain linker likely prevents the Xyl-CBM36-1 domain from parking in a fixed position. Furthermore, the longer linker between the two CBM36 domains certainly prevented the domains from packing into crystals. Even though the predicted structures hold sufficiently useful information about the 3D structure of the protein, it did not provide reliable information on how the domains are placed relative to each other and certainly did not predict the possibility of ligand or metal ion binding. The prediction program also showed the protein to be in a “closed” conformation with the catalytic dyad in the “up” position, contrary to the biological state of the protein in the experimental structure. Thus, to fully understand how protein function in nature would always need experimental methods to validate predicted structural models. Predicted structures, however, would serve as great starting points for experimental designs.

6. Conclusion

The characterization of individual and combined domains of the multidomain enzyme Xyl has significantly contributed to the field of biomass valorization. By investigating the molecular interactions between these domains, the research has shed light on the role of Xyl-CBM36 domains in enhancing or influencing the catalytic activity of Xyl-GH11. The absence of evidence for molecular interactions between domains supports the understanding that CBM36 domains do not directly interact with other protein domains. Instead, the study reveals that the impact on Xyl-GH11 catalysis is contingent upon factors such as the concentration of the Xyl-GH11 domain on the polysaccharide surface, its targeting efficiency, and the potential assistance of Xyl-CBM36 domains in disrupting the polysaccharide structure. The findings highlight the importance of linker-based co-localization in facilitating substrate-catalytic domain interactions. Therefore, Xyl enzyme exemplifies a "beads-on-a-string" protein architecture, where each domain, akin to beads, retains its individual function connected by linkers (strings). This understanding paves the way for the synthesis of novel super enzymes, offering a promising avenue for the rapid and efficient degradation of bio-waste to biofuels.

For the future, research could focus on refining the engineering of multidomain enzymes for enhanced biomass conversion. Strategies might involve optimizing linker lengths to fine-tune substrate interactions, exploring additional domains to augment catalytic efficiency, and investigating the potential synergies between different multidomain enzymes. Additionally, advancements in biotechnological approaches could be explored to scale up the production of these engineered enzymes for practical applications in biofuel production. Continuous research in this area holds promise for developing sustainable and economically viable solutions for the conversion of bio-waste into valuable biofuels.

7. References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography,* 68(Pt 4), 352–367.

Ahuja, V., Macho, M., Ewe, D., Singh, M., Saha, S., & Saurav, K. (2020). Biological and pharmacological potential of xylitol: a molecular insight of unique metabolism. *Foods*, 9(11), 1592.

Akintola, O., Ren, W., Adabala, P. J. P., Bhosale, S., Wang, Y., Ganga-Sah, Y., Britton, R., & Bennet, A. J. (2021). Intrinsic nucleophilicity of inverting and retaining glycoside hydrolases revealed using carbasugar Glyco-Tools. *ACS Catalysis*, 11(15), 9377–9389.

Al-Masaudi, S., El Kaoutari, A., Drula, E., Al-Mehdar, H., Redwan, E. M., Lombard, V., & Henrissat, B. (2017). A metagenomics investigation of carbohydrate-active enzymes along the gastrointestinal tract of saudi sheep. *Frontiers in Microbiology*, 8, 666.

Amin, F. R., Khalid, H., Zhang, H., Rahman, S. U., Zhang, R., Liu, G., & Chen, C. (2017). Pre-treatment methods of lignocellulosic biomass for anaerobic digestion. *AMB Express*, 7, 72.

Amoah, A., Kwablah, E., Korle, K., & Offei, D. (2020). Renewable energy consumption in Africa: the role of economic well-being and economic freedom. *Energy, Sustainability and Society*, 10(1).

Andrés, E., Albesa-Jové, D., Biarnés, X., Moerschbacher, B. M., Guerin, M. E., & Planas, A. (2014). Structural basis of chitin oligosaccharide deacetylation. *Angewandte Chemie*, 53, 6882–6887.

Arnaouteli, S., Giastas, P., Andreou, A., Tzanodaskalaki, M., Aldridge, C., Tzartos, S. J., Vollmer, W., Eliopoulos, E., & Bouriotis, V. (2015). Two putative polysaccharide deacetylases are required for osmotic stability and cell shape maintenance in *Bacillus anthracis*. *The Journal of Biological Chemistry*, 290, 13465–13478.

AspinalL G. O. (1959). Structural chemistry of the hemicelluloses. *Advances in carbohydrate chemistry*, 14, 429–468.

Badhan, A., Wang, Y., Gruninger, R., Patton, D., Powlowski, J., Tsang, A., & McAllister, T. (2014). Formulation of enzyme blends to maximize the hydrolysis of alkaline peroxide pretreated alfalfa hay and barley straw by rumen enzymes and commercial cellulases. *BMC Biotechnology*, 14, 31.

Bae, B., Ohene-Adjei, S., Kocherginskaya, S., Mackie, R. I., Spies, M. A., Cann, I. K., & Nair, S. K. (2008). Molecular basis for the selectivity and specificity of ligand recognition by the family 16 carbohydrate-binding modules from *Thermoanaerobacterium polysaccharolyticum* ManA. *The Journal of Biological Chemistry*, 283, 12415–12425.

Balat, M., Balat, H, & Öz, C. (2008). Progress in bioethanol processing. *Progress in Energy and Combustion Science,* 34, 551–73.

Baruah, J., Nath, B.K., Sharma, R., Kumar, S., Deka, R.C., Baruah, D.C., & Kalita, E. (2018). Recent trends in the pre-treatment of lignocellulosic biomass for value-added products. *Frontiers in Energy Research*, 6, 141.

Bayer, E. A., Chanzy, H., Lamed, R., & Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology,* 8, 548–557.

Behera, B., Parida, S., Dutta, S. K., & Thatoi, H. (2014). Isolation and identification of cellulose degrading bacteria from mangrove soil of Mahanadi River delta and their cellulase production ability. *American Journal of Microbiological Research*, 2, 41–46.

Belisle, J. T., Vissa, V. D., Sievert, T., Takayama, K., Brennan, P. J., & Besra, G. S. (1997). Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science*, 276, 1420–1422.

Benatti, A. L. T., & Polizeli, M. L. T. M. (2023). Lignocellulolytic biocatalysts: The main players involved in multiple biotechnological processes for biomass valorization. *Microorganisms*, 11(1), 162.

Bhardwaj, N. K., Kumar, B., & Verma, P. (2019). A detailed overview of xylanases: an emerging biomolecule for current and future prospective. *Bioresources and Bioprocessing*, 6(1).

Biely P. (2012). Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances*, 30, 1575–1588.

Biely, P., Cziszárová, M., Uhliariková, I., Agger, J. W., Li, X. L., Eijsink, V. G., & Westereng, B. (2013). Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase. *Biochimica et Biophysica Acta*, 1830, 5075–5086.

Binod, P., Sindhu, R., Singhania, R.R., Vikram, S., Devi, L., Nagalakshmi, S., Kurien, N., Sukumaran, R.K., & Pandey, A. (2010). Bioethanol production from rice straw: an overview. *Bioresource Technology*, 101, 4767-4774.

Bitto, E., Bingman, C. A., McCoy, J. G., Allard, S. T. M., Wesenberg, G. E., & Phillips, G. N., Jr. (2005). The structure at 1.6 Å resolution of the protein product of the *At4g34215* gene from *Arabidopsis thaliana*. *Acta Crystallographica Section D*, 61, 1655-1661.

Blair, D. E., & van Aalten, D. M. F. (2004). Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with N-acetyl-glucosamine. *FEBS Letters*, 570, 13–19.

Blair, D. E., Hekmat, O., Schüttelkopf, A. W., Shrestha, B., Tokuyasu, K., Withers, S. G., & van Aalten, D. M. F. (2006). Structure and mechanism of chitin deacetylase from the fungal pathogen *Colletotrichum lindemuthianum*. *Biochemistry*, 45, 9416–9426.

Blair, D. E., Schüttelkopf, A. W., MacRae, J. I., & van Aalten, D. M. F. (2005). Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 15429–15434.

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

Boneca, I. G. (2005). The role of peptidoglycan in pathogenesis. *Current Opinion in Microbiology*, 8, 46-53.

Boraston, A. B., Chiu, P., Warren, R. A. J., & Kilburn, D. G. (2000). Specificity and affinity of substrate binding by a family 17 carbohydrate-binding module from *Clostridium cellulovorans* Cellulase 5A. *Biochemistry*, 39, 11129–11136.

Boraston, A. B., McLean, B. W., Chen, G., Li, A., Warren, R. A., & Kilburn, D. G. (2002). Co-operative binding of triplicate carbohydrate-binding modules from a thermophilic xylanase. Molecular Microbiology, 43, 187–194.

Boraston, A. B., Notenboom, V., Warren, R. A., Kilburn, D. G., Rose, D. R., & Davies, G. (2003). Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and "galactose-binding" domains. *Journal of Molecular Biology*, 327, 659–669.

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

Branda, S. S., Vik, Å., Friedman, L., & Kolter, R. (2005). Biofilms: The matrix revisited. *Trends in Microbiology*, 13, 20-26.

Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews, Microbiology*, 12, 168–180.

Buijs, N. A., Siewers, V., & Nielsen, J. (2013). Advanced biofuel production by the yeast *Saccharomyces cerevisiae*. *Current Opinion in Chemical Biology*, 17, 480–488.

Buranov, A. U, & Mazza, G. (2008). Lignin in straw of herbaceous crops. *Industrial Crops and Products*, 28, 237-259.

Bussemaker, M.J., Zhang, D. (2013). Effect of ultrasound on lignocellulosic biomass as a pre-treatment for biorefinery and biofuel applications. *Industrial & Engineering Chemistry Research*, 52, 3563-3580.

Caufrier, F., Martinou, A., Dupont, C., & Bouriotis, V. (2003). Carbohydrate esterase family 4 enzymes: substrate specificity. *Carbohydrate Research*, 338, 687–692.

Cedillo, V. B., & Montanier, C. (2023). Effect of multimodularity and spatial organization of glycoside hydrolases on catalysis. Essays in Biochemistry, 67(3), 629–638.

Čepeljnik, T., Križaj, I., & Logar, R. (2004). Isolation and characterization of the *Pseudobutyrivibrio xylanivorans* Mz5T xylanase XynT - The first family 11 endoxylanase from rumen Butyrivibrio-related bacteria. *Enzyme and Microbial Technology*, 34, 219–227.

Chakdar, H., Kumar, M., Pandiyan, K., Singh, A., Karthikeyan, N., Kashyap, P. L., & Srivastava, A. K. (2016). Bacterial xylanases: biology to biotechnology. *3 Biotech*, 6(2).

Chang, M.C.Y. (2007). Harnessing energy from plant biomass. *Current Opinion in Chemical Biology*, 11, 677-684.

Chang, S. C., Saldivar, R. K., Liang, P. H., & Hsieh, Y. S. Y. (2021). Structures, biosynthesis, and physiological functions of (1,3;1,4)-β-d-glucans. Cells, 10(3), 510.

Charnock, S.J., Bolam, D.N., Nurizzo, D., Szabo, L., McKie, V.A., Gilbert, H.J., 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. *National Academy of Sciences*, 99, 14077–14082.

Cheah, W., Sankaran, R., Show, P., Tg. Ibrahim, T., Chew, K., Culaba, A., & Chang, J. (2020). Pre-treatment methods for lignocellulosic biofuels production: current advances, challenges and future prospects. *Biofuel Research Journal*, 7, 1115-1127.

Chen, X. Y., & Kim, J. Y. (2009). Callose synthesis in higher plants. Plant Signaling & Behavior, 4(6), 489–492.

Chen, Z., & Liu, D. (2016). Toward glycerol biorefinery: metabolic engineering for the production of biofuels and chemicals from glycerol. *Biotechnology for Biofuels*, 9(1).

Chettri, D., Verma, A. K., & Verma, A. K. (2020). Innovations in CAZyme gene diversity and its modification for biorefinery applications. Biotechnology Reports, 28, e00525.

Choi, J., Choi, D.-H., Faix, O. (2007). Characterization of lignin-carbohydrate linkages in the residual lignins isolated from chemical pulps of spruce (*Picea abies*) and beech wood (*Fagus sylvatica*). *Journal of Wood Science*, 53, 309-313.

Chong, S. L., Virkki, L., Maaheimo, H., Juvonen, M., Derba-Maceluch, M., Koutaniemi, S., Roach, M., Sundberg, B., Tuomainen, P., Mellerowicz, E. J., & Tenkanen, M. (2014). O-acetylation of glucuronoxylan in *Arabidopsis thaliana* wild type and its change in xylan biosynthesis mutants. *Glycobiology*, 24, 494–506.

Cragg, S. M., Beckham, G. T., Bruce, N. C., Bugg, T. D. H., Distel, D. L., Dupree, P., Etxabe, A. G., Goodell, B., Jellison, J., McGeehan, J., McQueen-Mason, S. J., Schnorr, K., Walton, P. H., Watts, J. E. M., & Zimmer, M. (2015). Lignocellulose degradation mechanisms across the Tree of Life. *Current Opinion in Chemical Biology*, 29, 108–119.

Cunha, J. T., Soares, P. O., Baptista, S. L., Costa, C. E., & Domingues, L. (2020). Engineered *Saccharomyces cerevisiae* for lignocellulosic valorization: a review and perspectives on bioethanol production. *Bioengineered*, 11(1), 883–903.

Curry, T. M., Peña, M. J., & Urbanowicz, B. R. (2023). An update on xylan structure, biosynthesis, and potential commercial applications. *Cell surface*, 9, 100101.

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., & 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. *Journal of Biological Chemistry*, 276, 48580–48587.

da Costa, R. R., Hu, H., Li, H., & Poulsen, M. (2019). Symbiotic plant biomass decomposition in fungus-growing termites. *Insects*, 10(4), 87.

Dahman, Y., Syed, K., Begum S., Roy P., Mohtasebi B. (2019). Biofuels: Their characteristics and analysis, Editor(s): Deepak Verma, Elena Fortunati, Siddharth Jain, Xiaolei Zhang, In woodhead publishing series in *Composites Science and Engineering, Biomass, Biopolymer-based Materials, and Bioenergy, Woodhead Publishing*, 277-325.

Dahmen, N., Lewandowski, I., Zibek, S., Weidtmann, A. (2019). Integrated lignocellulosic value chains in a growing bioeconomy: Status quo and perspectives. *GCB Bioenergy*, 11, 107–117.

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

de O Buanafina, M. M., Buanafina, M. F., Dalton, S., Morris, P., Kowalski, M., Yadav, M. K., & Capper, L. (2020). Probing the role of cell wall feruloylation during maize development by differential expression of an apoplast targeted fungal ferulic acid esterase. PloS One, 15(10), e0240369.

de Oliveira Santos, V.T., Siqueira, G., Milagres, A.M.F., Ferraz, A. (2018). Role of hemicellulose removal during dilute acid pre-treatment on the cellulose accessibility and enzymatic hydrolysis of compositionally diverse sugarcane hybrids. *Industrial Crops Products*, 111, 722-730.

de Vries, R. P., & Visser, J. (2001). Aspergillus enzymes involved in degradation of plant cell wall polysaccharides. *Microbiology and Molecular Biology Reviews*, 65, 497–522.

DeChellis, A., Nemmaru, B., Sammond, D. W., Douglas, J. T., Patil, N., Reste, O., & Chundawat, S. P. S. (2023). Supercharging carbohydrate-binding module alone enhances endocellulase thermostability, binding, and activity on cellulosic biomass. *BioRxiv* (Cold Spring Harbor Laboratory).

Delagenière, S., Brenchereau, P., Launer, L., Ashton, A. W., Leal, R., Veyrier, S., Gabadinho, J., Gordon, E. J., Jones, S. D., Levik, K. E., McSweeney, S. M., Monaco, S., Nanao, M., Spruce, D., Svensson, O., Walsh, M. A., &

Leonard, G. A. (2011). ISPyB: an information management system for synchrotron macromolecular crystallography. *Bioinformatics*, 27, 3186–3192.

Deng, D. M., Urch, J. E., ten Cate, J. M., Rao, V. A., van Aalten, D. M., & Crielaard, W. (2009). Streptococcus mutans SMU.623c codes for a functional, metal-dependent polysaccharide deacetylase that modulates interactions with salivary agglutinin. *Journal of Bacteriology*, 191, 394–402.

Devi, J., Deb, U., Barman, S., Das, S., Sundar Bhattacharya, S., Fai Tsang, Y., Lee, J. H., & Kim, K. H. (2020). Appraisal of lignocellulosic biomass degrading potential of three earthworm species using vermireactor mediated with spent mushroom substrate: Compost quality, crystallinity, and microbial community structural analysis. *The Science of the Total Environment*, 716, 135215.

Dey, P., & Roy, A. (2018). Molecular structure and catalytic mechanism of fungal family G acidophilic xylanases. *Biotech*, 8(2), 78.

Dodd, D., & Cann, I. K. (2009). Enzymatic deconstruction of xylan for biofuel production. Global change biology. *Bioenergy*, 1, 2–17.

Dou, T., Chen, J., Hao, Y., & Qi, X. (2018). Xylanosomes produced by *Cellulosimicrobium cellulans* F16 were diverse in size, but resembled in subunit composition. *Archives of Microbiology*, 201(2), 163–170.

Doyle M. L. (1997). Characterization of binding interactions by isothermal titration calorimetry. *Current Opinion in Biotechnology*, 8, 31–35.

Drula, É., Garron, M., Doğan, S., Lombard, V., Henrissat, B., & Terrapon, N. (2021). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research*, 50(D1), D571–D577.

Ebringerová, A., Hromádková, Z., & Heinze, T. (2005). Hemicellulose polysaccharides. *Springer*, 1-67.

El Achkar, J.H., Lendormi, T., Salameh, D., Louka, N., Maroun, R.G., Lanoisellé, J.L., & Hobaika, Z. (2018). Influence of pre-treatment conditions on lignocellulosic fractions and methane production from grape pomace. *Bioresource Technology*, 247, 881-889

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological crystallography*, 60, 2126–2132.

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects - diversity in structure and function. *FEMS Microbiology Reviews*, 37, 699–735.

Evans, P. R., & Murshudov, G. N. (2013). How good my data and what is the resolution? *Acta Crystallographica. Section D, Biological Crystallography*, 69, 1204–1214.

Ezeilo, U. R., Zakaria, I. I., Huyop, F., & Wahab, R. A. (2017). Enzymatic breakdown of lignocellulosic biomass: the role of glycosyl hydrolases and lytic polysaccharide monooxygenases. *Biotechnology & Biotechnological Equipment,* 1–16.

Fadouloglou, V. E., Kapanidou, M., Agiomirgianaki, A., Arnaouteli,S., Bouriotis, V., Glykos, N. M., & Kokkinidis, M. (2013). Structure determination through homology modelling and torsion-angle simulated annealing: Application to a polysaccharide deacetylase from *Bacillus cereus. Acta Crystallographica Section D*,69, 276-283.

Fradera-Soler, M., Grace, O. M., Jørgensen, B., & Mravec, J. (2022). Elastic and collapsible: current understanding of cell walls in succulent plants. Journal of Experimental Botany, 73(8), 2290–2307.

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

Fujita, A., Hojo, M., Aoyagi, T., Hayashi, Y., Arakawa, G., Tokuda, G., & Watanabe, H. (2010). Details of the digestive system in the midgut of *Coptotermes formosanus*. *Journal of Wood Science*, 56, 222–226.

Gao, D., Uppugundla, N., Chundawat, S. P. S., Yu, X., Hermanson, S., Gowda, K., Brumm, P. J., Mead, D. A., Balan, V., & Dale, B. E. (2011). Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnology for Biofuels*, 4(1), 5.

Görgens, J. F., Planas, J., van Zyl, W. H., Knoetze, J. H., & Hahn-Hägerdal, B. (2004). Comparison of three expression systems for heterologous xylanase production by *S. cerevisiae* in defined medium. *Yeast*, 21, 1205–1217.

Grabber, J. H., Ralph, J., Lapierre, C., & Barrière, Y. (2004). Genetic and molecular basis of grass cell-wall degradability. Lignin-cell wall matrix interactions. *Comptes Rendus Biologies*, 327, 455–465.

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, 1, 2876–2890.

Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W., & Kratky, C. (1998). Thermophilic xylanase from *Thermomyces lanuginosus*: high-resolution X-ray structure and modeling studies. *Biochemistry*, 37, 13475–13485.

Gruninger, R. J., Cote, C., McAllister, T. A., & Abbott, D. W. (2016). Contributions of a unique β-clamp to substrate recognition illuminates the molecular basis of exolysis in ferulic acid esterases. *The Biochemical Journal*, 473, 839–849.

Gu, B. J., Wang, J., Wolcott, M. P., & Ganjyal, G. M. (2018). Increased sugar yield from pre-milled Douglas-fir forest residuals with lower energy consumption by using planetary ball milling. *Bioresource Technology*, 251, 93–98.

Hahn-Hägerdal, B., Galbe, M., Gorwa-Grauslund, M. F., Lidén, G., & Zacchi, G. (2006). Bioethanol the fuel of tomorrow from the residues of today. *Trends in Biotechnology*, 24, 549–556.

Handing, K. B., Niedzialkowska, E., Shabalin, I. G., Kuhn, M. L., Zheng, H., & Minor, W. (2018). Characterizing metal-binding sites in proteins with X-ray crystallography. *Nature protocols*, *13*(5), 1062–1090.

Havukainen, R., Törrönen, A., Laitinen, T., & Rouvinen, J. (1996). Covalent binding of three epoxyalkyl xylosides to the active site of endo-1,4-xylanase II from *Trichoderma reesei*. *Biochemistry*, 35, 9617–9624.

He, B., Jin, S., Cao, J., Mi, L., & Wang, J. (2019). Metatranscriptomics of the Hu sheep rumen microbiome reveals novel cellulases. *Biotechnology for Biofuels*, 12(1).

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., & Gilbert, H. J. (2006). Family 6 carbohydrate binding modules in beta-agarases display exquisite selectivity for the non-reducing termini of agarose chains. *The Journal of Biological Chemistry*, 281, 17099–17107.

Hernick, M., Gennadios, H. A., Whittington, D. A., Rusche, K. M., Christianson, D. W., & Fierke, C. A. (2005). UDP-3-O-((R)-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase functions through a general acid-base catalyst pair mechanism. *The Journal of Biological Chemistry*, 280, 16969–16978.

Ho, D. P., Ngo, H. H., & Guo, W. (2014). A mini review on renewable sources for biofuel. *Bioresource Technology*, 169, 742–749.

Hoell, I. A., Vaaje-Kolstad, G., & Eijsink, V. G. H. (2010). Structure and function of enzymes acting on chitin and chitosan. *Biotechnology and Genetic Engineering Reviews*, 27, 331-366.

Horwitz, M. A., Lee, B. W., Dillon, B. J., & Harth, G. (1995). Protective immunity against tuberculosis induced by vaccination with major extracellular proteins of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 1530–1534.

Hosseini Koupaie, E., Dahadha, S., Bazyar Lakeh, A. A., Azizi, A., & Elbeshbishy, E. (2019). Enzymatic pre-treatment of lignocellulosic biomass for enhanced biomethane production-A review. *Journal of Environmental Management*, 233, 774–784.

Hou, X., Wang, Z., Sun, J., Li, M., Wang, S., Chen, K., & Gao, Z. (2019). A microwave-assisted aqueous ionic liquid pre-treatment to enhance enzymatic hydrolysis of Eucalyptus and its mechanism. *Bioresource Technology*, 272, 99–104.

Hu, X., Cheng, L., Gu, Z., Hong, Y., Li, Z., & Li, C. (2018). Effects of ionic liquid/water mixture pre-treatment on the composition, the structure and the enzymatic hydrolysis of corn stalk. *Industrial Crops Production*, 122, 142-147.

Huang, L. Z., Ma, M. G., Ji, X. X., Choi, S. E., & Si, C. (2021). Recent developments and applications of hemicellulose from wheat straw: A Review. Frontiers in Bioengineering and Biotechnology, 9, 690773.

Inward, D., Beccaloni, G., & Eggleton, P. (2007). Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biology Letters*, 3, 331–335.

Ito, K., Ikemasu, T., & Ishikawa, T. (1992). Cloning and sequencing of the xynA gene encoding xylanase A of *Aspergillus kawachii*. *Bioscience, Biotechnology, and Biochemistry*, 56, 906–912.

Itoh, Y., Rice, J. D., Goller, C., Pannuri, A., Taylor, J., Meisner, J., Beveridge, T. J., Preston, J. F., 3rd, & Romeo, T. (2008). Roles of pgaABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine. *Journal of Bacteriology*, 190, 3670–3680.

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.

Jänis, J., Hakanpää, J., Hakulinen, N., Ibatullin, F. M., Hoxha, A., Derrick, P. J., Rouvinen, J., & Vainiotalo, P. (2005). Determination of thioxylo-oligosaccharide binding to family 11 xylanases using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and X-ray crystallography. *The FEBS Journal,* 272, 2317–2333.

Jędrzejczyk, M., Soszka, E., Czapnik, M., Ruppert, A.M., Grams, J. (2019). Physical and chemical pre-treatment of lignocellulosic biomass. Second third generation feedstocks. *Elsevier*, 143-196.

Jiang, J. X., Zhu, L. W., Wang, K., and Wang, W. G. (2006). Simultaneous saccharification and fermentation of steam pretreated Lespedeza stalks for the production of ethanol. *Forestry Studies in China* 8, 30-33.

Jiang, Z., Dang, W., Yan, Q., Zhai, Q., Li, L., & Kusakabe, I. (2006). Subunit composition of a large xylanolytic complex (xylanosome) from *Streptomyces olivaceoviridis* E-86. *Journal of Biotechnology*, 126, 304–312.

Joelsson, E., Erdei, B., Galbe, M., Wallberg, O. (2016). Techno-economic evaluation of integrated first-and second-generation ethanol production from grain and straw. *Biotechnology Biofuels*, 9, 1.

Johannsen L. (1993). Biological properties of bacterial peptidoglycan. *Acta Pathologica, Microbiologica, et Immunologica Scandinavica*, 101, 337–344.

Jones, D. R., Thomas, D., Alger, N., Ghavidel, A., Inglis, G. D., & Abbott, D. W. (2018). SACCHARIS: an automated pipeline to streamline discovery of carbohydrate active enzyme activities within polyspecific families and de novo sequence datasets. Biotechnology for Biofuels, 11, 27.

Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G., & McIntosh, L. P. (2000). Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. *Journal of Molecular Biology*, 299, 255–279.

Kabsch, W. (2010). XDS. *Acta Crystallographica Section D: Biological Crystallography*, 66, 125–132.

Kari, J., Molina, G. A., Schaller, K., Schiano-di-Cola, C., Christensen, S. J., Badino, S. F., Sørensen, T. H., Røjel, N. S., Keller, M. B., Sørensen, N. R., Kołaczkowski, B. M., Olsen, J. P., Krogh, K. B. R. M., Jensen, K., Cavaleiro, A. M., Peters, G. H., Spodsberg, N., & Borch, K. (2021). Physical constraints and functional plasticity of cellulases. *Nature Communications*, 12(1).

Kariuki, E. G., Kibet, C., Paredes, J. C., Mboowa, G., Mwaura, O., Njogu, J., Masiga, D., Bugg, T. D. H., & Tanga, C. M. (2023). Metatranscriptomic analysis of the gut microbiome of black soldier fly larvae reared on lignocellulose-rich fiber diets unveils key lignocellulolytic enzymes. *Frontiers in Microbiology*, 14.

Karunanithy, C., Muthukumarappan, K., & Gibbons, W. R. (2012). Effect of extruder screw speed, temperature, and enzyme levels on sugar recovery from different biomasses. *Biotechnology*, 2013, 942810.

Katapodis, P., Vrsanská, M., Kekos, D., Nerinckx, W., Biely, P., Claeyssens, M., Macris, B. J., & Christakopoulos, P. (2003). Biochemical and catalytic properties of an endoxylanase purified from the culture filtrate of *Sporotrichum thermophile*. *Carbohydrate Research*, 338, 1881–1890.

Kim D. (2018). Physico-Chemical conversion of lignocellulose: Inhibitor effects and detoxification strategies: A mini review. *Molecules,* 23, 309.

Kim, I. J., Bornscheuer, U. T., & Nam, K. H. (2022). Biochemical and structural analysis of a glucose-tolerant β-glucosidase from the hemicellulose-degrading *Thermoanaerobacterium saccharolyticum*. *Molecules*, 27(1), 290.

Kim, I. J., Kim, S. R., Bornscheuer, U. T., & Nam, K. H. (2023). Engineering of GH11 xylanases for optimal pH shifting for industrial applications. *Catalysts*, 13(11), 1405.

Kim, J. W., Kim, K. S., Lee, J. S., Park, S. M., Cho, H. Y., Park, J. C., & Kim, J. S. (2011). Two-stage pre-treatment of rice straw using aqueous ammonia and dilute acid. *Bioresource Technology*, 102, 8992–8999.

Kim, S., Ga, S., Bae, H., Sluyter, R., Konstantinov, K., Shrestha, L. K., Kim, Y. H., Kim, J. H., & Ariga, K. (2024). Multidisciplinary approaches for enzyme biocatalysis in pharmaceuticals: protein engineering, computational biology, and nanoarchitectonics. EES Catalysis.

Kiribayeva, A., Mukanov, B., Silayev, D., Akishev, Z., Ramankulov, Y., & Khassenov, B. (2022). Cloning, expression, and characterization of a recombinant xylanase from *Bacillus sonorensis* T6. *PLOS ONE*, 17(3), e0265647.

Kitamoto, M., Tokuda, G., Watanabe, H., & Arioka, M. (2019). Characterization of CBM36-containing GH11 endoxylanase NtSymX11 from the hindgut metagenome of higher termite *Nasutitermes takasagoensis* displaying prominent catalytic activity. *Carbohydrate Research*, 474, 1–7.

Kmezik, C., Bonzom, C., Olsson, L., Mazurkewich, S., & Larsbrink, J. (2020). Multimodular fused acetyl-feruloyl esterases from soil and gut Bacteroidetes improve xylanase depolymerization of recalcitrant biomass. Biotechnology for Biofuels, 13, 60.

Kolenová, K., Vrsanská, M., & Biely, P. (2006). Mode of action of endo-beta-1,4-xylanases of families 10 and 11 on acidic xylooligosaccharides. *Journal of Biotechnology*, 121, 338–345.

Koller, M., Salerno, A., Tuffner, P., Koinigg, M., Böchzelt, H., Schober, S., Pieber, S., Schnitzer, H., Mittelbach, M., & Braunegg, G. (2012). Characteristics and potential of micro algal cultivation strategies: a review. *Journal of Cleaner Production*, 37, 377-388.

Koshijima, T. & Watanabe, T. (2003). Association between lignin and carbohydrates in Wood and Other Plant Tissues. *Springer Series in Wood Science*. Springer.

Kowalski, Z., Kulczycka, J., Verhé, R., Desender, L., De Clercq, G., Makara, A., Generowicz, N., & Harazin, P. (2022). Second-generation biofuel production from the organic fraction of municipal solid waste. *Frontiers in Energy Research*, 10.

Krska, D., Mazurkewich, S., Brown, H. A., Theibich, Y. A., Poulsen, J. N., Morris, A. L., Koropatkin, N. M., Lo Leggio, L., & Larsbrink, J. (2021). Structural and functional analysis of a multimodular hyperthermostable xylanase-glucuronoyl esterase from Caldicellulosiruptor kristjansonii. Biochemistry, 60(27), 2206–2220.

Kruger RF. "Functional and Structural characterisation of a multimodular xylanase from the hindgut metagenome of the snouted harvester termite." Master's Dissertation. University of Pretoria, 2017.

Kumar, A. K., & Sharma, S. (2017). Recent updates on different methods of pre-treatment of lignocellulosic feedstocks: a review. *Bioresources and Bioprocessing*, 4, 7.

Kumar, R., & Wyman, C. E. (2009). Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pre-treatment technologies. *Bioresource Technology*, 100, 4203–4213.

Lapierre, C., Monties, B., Rolando, C. (1986). Thioacidolysis of poplar lignins - identification of monomeric syringyl products and characterization of guaiacyl-syringyl lignin fractions. *Holzforschung*, 40, 113-118

Lara-Espinoza, C., Carvajal-Millán, E., Balandrán-Quintana, R., López-Franco, Y., & Rascón-Chu, A. (2018). Pectin and pectin-based composite materials: beyond food texture. Molecules, 23(4), 942.

Lawson, C. L., van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., de Vries, G. E., Penninga, D., Dijkhuizen, L., & Dijkstra, B. W. (1994). Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. *Journal of Molecular Biology*, 236, 590–600.

Leavitt, S., & Freire, E. (2001). Direct measurement of protein binding energetics by isothermal titration calorimetry. *Current Opinion in Structural Biology*, 11, 560–566.

Li, C., Knierim, B., Manisseri, C., Arora, R., Scheller, H. V., Auer, M., Vogel, K. P., Simmons, B. A., & Singh, S. (2010). Comparison of dilute acid and ionic liquid pre-treatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. *Bioresource Technology*, 101, 4900–4906.

Li, C., Zhang, R., Wang, J., Wilson, L. M., & Yan, Y. (2020). Protein Engineering for Improving and Diversifying Natural Product Biosynthesis. Trends in Biotechnology, 38(7), 729–744.

Li, H., & Turunen, O. (2015). Effect of acidic amino acids engineered into the active site cleft of *Thermopolyspora flexuosa* GH11 xylanase. *Biotechnology and Applied Biochemistry*, 62, 433–440.

Li, X., Dilokpimol, A., & Kabel, M. A. (2022). Fungal xylanolytic enzymes: Diversity and applications. *Bioresource Technology*, 344, 126290.

Li, Y., Zhang, X., Lu, C., Lu, P., Yin, C., Ye, Z., & Huang, Z. (2022). Identification and characterization of a novel endo-β-1,4-xylanase from Streptomyces sp. T7 and its application in xylo-oligosaccharide production. *Molecules*, 27(8), 2516

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.

Lin, L. L., & Thomson, J. A. (1991). An analysis of the extracellular xylanases and cellulases of *Butyrivibrio fibrisolvens* H17c. *FEMS Microbiology Letters*, 68, 197–203.

Lin, S., Hunt, C. J., Holck, J., Brask, J., Krogh, K. B. R. M., Meyer, A. S., Wilkens, C., & Agger, J. W. (2023). Fungal feruloyl esterases can catalyze release of diferulic acids from complex arabinoxylan. *International Journal of Biological Macromolecules*, 232, 123365.

Lin, Y., & Tanaka, S. (2006). Ethanol fermentation from biomass resources: current state and prospects. *Applied Microbiology and Biotechnology*, 69, 627–642.

Linares-Pastén, J. A., Hero, J. S., Pisa, J. H., Teixeira, C., Nyman, M., Adlercreutz, P., Martínez, M. A., & Karlsson, E. N. (2021). Novel xylan degrading enzymes from polysaccharide utilizing loci of *Prevotella copri* DSM18205. *Glycobiology*.

Little, D. J., Bamford, N. C., Pokrovskaya, V., Robinson, H., Nitz, M., & Howell, P. L. (2014). Structural basis for the de-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine in Gram-positive bacteria. *The Journal of Biological Chemistry*, 289, 35907–35917.

Little, D. J., Milek, S., Bamford, N. C., Ganguly, T., DiFrancesco, B. R., Nitz, M., Deora, R., & Howell, P. L. (2015). The protein BpsB is a poly-β-1,6-N-acetyl-D-glucosamine deacetylase required for biofilm formation in *Bordetella bronchiseptica*. *The Journal of Biological Chemistry*, 290, 22827–22840.

Little, D. J., Poloczek, J., Whitney, J. C., Robinson, H., Nitz, M., & Howell, P. L. (2012). The structure- and metal-dependent activity of *Escherichia coli* PgaB provides insight into the partial de-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine. *The Journal of Biological Chemistry*, 287, 31126–31137.

Liu, G., & Qu, Y. (2021). Integrated engineering of enzymes and microorganisms for improving the efficiency of industrial lignocellulose deconstruction. *Engineering Microbiology*, 1, 100005.

Liu, J., Sun, D., Zhu, J., Cong, L., & Liu, W. (2021). Carbohydrate-binding modules targeting branched polysaccharides: overcoming side-chain recalcitrance in a non-catalytic approach. *Bioresources and Bioprocessing*, 8(1).

Lo, N., Tokuda, G., & Watanabe, H. (2011). Evolution and function of endogenous termite cellulases. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites*: *a modern synthesis*, 51–67.

Long S. R. (1989). Rhizobium-legume nodulation: life together in the underground. *Cell*, 56, 203–214.

Lopes, A. M., Ferreira Filho, E. X., & Moreira, L. (2018). An update on enzymatic cocktails for lignocellulose breakdown. *Journal of Applied Microbiology*, 125, 632–645.

Ludwiczek, M. L., Heller, M., Kantner, T., & McIntosh, L. P. (2007). A secondary xylan-binding site enhances the catalytic activity of a single-domain family 11 glycoside hydrolase. *Journal of Molecular Biology*, 373, 337–354.

Lunin, V. V., Wang, H. T., Bharadwaj, V. S., Alahuhta, M., Peña, M. J., Yang, J. Y., Archer-Hartmann, S. A., Azadi, P., Himmel, M. E., Moremen, K. W., York, W. S., Bomble, Y. J., & Urbanowicz, B. R. (2020). Molecular mechanism of polysaccharide acetylation by the arabidopsis xylan o-acetyltransferase xoat1. The Plant *Cell*, 32, 2367–2382.

Lv, Y. (2023). Transitioning to sustainable energy: opportunities, challenges, and the potential of blockchain technology. Frontiers in Energy Research, 11.

Lynd, L., van Zyl, W., McBride, J., & Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: An update. *Current Opinion in Biotechnology*, 16, 577–583.

Machado De Castro, A., Fragoso Dos Santos, A., Kachrimanidou, V., Koutinas, A. A., & Freire, D. M. (2018). Solid-state fermentation for the production of proteases and amylases and their application in nutrient medium production. *Current Developments in Biotechnology and Bioengineering*, 185–210.

Masrati, G., Landau, M., Ben-Tal, N., Lupas, A. N., Kosloff, M., & Kosinski, J. (2021). Integrative Structural Biology in the era of accurate structure prediction. *Journal of Molecular Biology*, 433, 167127.

Maurya, D. P., Singla, A., & Negi, S. (2015). An overview of key pre-treatment processes for biological conversion of lignocellulosic biomass to bioethanol. *Biotechnology*, 5, 597–609.

McCarter, J.D., & Withers, G.S. (1994). Mechanisms of enzymatic glycoside hydrolysis, *Current Opinion in Structural Biology*, 4, 885-892.

McCarthy, A. A., Morris, D. D., Bergquist, P. L., & Baker, E. N. (2000). Structure of XynB, a highly thermostable beta-1,4-xylanase from Dictyoglomus thermophilum Rt46B.1, at 1.8 Å resolution. *Acta Crystallographica. Section D, Biological Crystallography*, 56, 1367–1375.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40, 658–674.

Mendonça, M., Barroca, M., & Collins, T. (2023). Endo-1,4-β-xylanase-containing glycoside hydrolase families: characteristics, singularities and similarities. *Biotechnology advances*, 65, 108148.

Mesbah N. M. (2022). Industrial biotechnology based on enzymes from extreme environments. Frontiers in Bioengineering and Biotechnology, 10, 870083.

Mikkelsen, M. D., Harholt, J., Ulvskov, P., Johansen, I. E., Fangel, J. U., Doblin, M. S., Bacic, A., & Willats, W. G. (2014). Evidence for land plant cell wall biosynthetic mechanisms in charophyte green algae. Annals of Botany, 114(6), 1217–1236.

Miles, A. J., Janes, R. W., & Wallace, B. (2021). Tools and methods for circular dichroism spectroscopy of proteins: a tutorial review. *Chemical Society Reviews*, 50(15), 8400–8413.

Minor, W., Cymborowski, M., Otwinowski, Z., & Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution from diffraction images to an initial model in minutes. *Acta Crystallographica. Section D, Biological crystallography,* 62, 859–866.

Mølgaard, A., Kauppinen, S., & Larsen, S. (2000). *Rhamnogalac turonan* acetylesterase elucidates the structure and functionof a new family of hydrolases. *Structure*, 8(4), 373-383.

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., & Gilbert, H. J. (2009). Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 3065–3070.

Moser, F., Irwin, D., Chen, S., & Wilson, D. B. (2008). Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnology and Bioengineering*, 100, 1066–1077.

Muilu, J., Törrönen, A., Peräkylä, M., & Rouvinen, J. (1998). Functional conformational changes of endo-1,4-xylanase II from *Trichoderma reesei*: a molecular dynamics study. *Proteins*, 31, 434–444.

Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romão, M. J., Gilbert, H. J., Bolam, D. N., & Fontes, C. M. (2006). Xyloglucan is recognized by carbohydrate binding modules that interact with beta glucan chains. *The Journal of Biological Chemistry*, 281, 8815–8828.

Navarre, W. W., & Schneewind, O. (1999). Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews*, 63, 174–229.

Navarro, D. M. D. L., Abelilla, J. J., & Stein, H. H. (2019). Structures and characteristics of carbohydrates in diets fed to pigs: a review. Journal of Animal Science and Biotechnology, 10(1).

Neumüller, K. G., de Souza, A. C., van Rijn, J. H., Streekstra, H., Gruppen, H., & Schols, H. A. (2015). Positional preferences of acetyl esterases from different CE families towards acetylated 4-O-methyl glucuronic acid-substituted xylo-oligosaccharides. *Biotechnology for Biofuels*, 8, 7.

Nishiyama, Y., Langan, P., & Chanzy, H. (2002). Crystal structure and hydrogen bonding system in cellulose Ibeta from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, 124, 9074–9082.

Nordberg Karlsson, E., Schmitz, E., Linares-Pastén, J. A., & Adlercreutz, P. (2018). Endo-xylanases as tools for production of substituted xylooligosaccharides with prebiotic properties. *Applied microbiology and biotechnology*, 102(21), 9081–9088.

Numan, M. T., & Bhosle, N. B. (2006). Alpha-L-arabinofuranosidases: the potential applications in biotechnology. *Journal of Industrial Microbiology & Biotechnology*, 33, 247–260.

Oberbarnscheidt, L., Taylor, E. J., Davies, G. J., & Gloster, T.M. (2007). Structure of a carbohydrate esterase from *Bacillus anthracis*. *Proteins: Structure, Function, and Bioinformatics*, 66, 250-252.

Okafor, C. C., Madu, C. N., Ajaero, C. C., Ibekwe, J. C., Bebenimibo, H., & Nzekwe, C. A. (2021). Moving beyond fossil fuel in an oil-exporting and emerging economy: Paradigm shift. *AIMS Energy*, 9(2), 379–413.

Ostadjoo, S., Hammerer, F., Dietrich, K., Dumont, M. J., Friščić, T., & Auclair, K. (2019). Efficient enzymatic hydrolysis of biomass hemicellulose in the absence of bulk water. *Molecules*, 24, 4206.

Østby, H., & Várnai, A. (2023). Hemicellulolytic enzymes in lignocellulose processing. *Essays in biochemistry*, 67(3), 533–550.

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 276, 307–326.

Oyewo, A. S., Sterl, S., Khalili, S., & Breyer, C. (2023). Highly renewable energy systems in Africa: Rationale, research, and recommendations. *Joule*, 7(7), 1437–1470.

Paës, G., Skov, L. K., O'Donohue, M. J., Rémond, C., Kastrup, J. S., Gajhede, M., & Mirza, O. (2008). The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry*, 47, 7441–7451.

Paës, G., Tran, V., Takahashi, M., Boukari, I., & O'Donohue, M. J. (2007). New insights into the role of the thumb-like loop in GH-11 xylanases. *Protein Engineering, Design & Selection*, 20, 15–23.

Pang, J., Zhang, Y., Tong, X., Zhong, Y., Kong, F., Li, D., Liu, X., & Qiao, Y. (2023). Recent developments in molecular characterization, bioactivity, and application of arabinoxylans from different sources. Polymers, 15(1), 225.

Park, B. K., & Kim, M. M. (2010). Applications of chitin and its derivatives in biological medicine. *International Journal of Molecular Sciences*, 11, 5152–5164.

Pauly, M., & Keegstra, K. (2016). Biosynthesis of the plant cell wall matrix polysaccharide xyloglucan. Annual Review of Plant Biology, 67(1), 235–259.

Pavarina, G. C., Lemos, E. G. M., Lima, N. S. M., & Pizauro, J. (2021). Characterization of a new bifunctional endo-1,4-β-xylanase/esterase found in the rumen metagenome. *Scientific reports*, 11(1), 10440.

Pearce, R., & Zhang, Y. (2021). Deep learning techniques have significantly impacted protein structure prediction and protein design. *Current Opinion in Structural Biology,* 68, 194–207.

Pengthaisong, S., Piniello, B., Davies, G., Rovira, C., & Cairns, J. R. K. (2023). Reaction mechanism of glycoside hydrolase family 116 utilizes perpendicular protonation. *ACS Catalysis*, 13(9), 5850–5863.

Pereira, J., Simpkin, A. J., Hartmann, M. D., Rigden, D. J., Keegan, R. M., & Lupas, A. N. (2021). High-accuracy protein structure prediction in CASP14. *Proteins*, 89, 1687–1699.

Petsko, G.A. & Ringe, D. (2004). Protein structure and function. Chapter I. From sequence to structure. New Science Press Ltd

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF chimeraX: Structure visualization for researchers, educators, and developers. Protein Science, 30, 70–82.

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., & 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. *The Journal of Biological Chemistry*, 279, 21560–21568.

Polizeli, M. L., Rizzatti, A. C., Monti, R., Terenzi, H. F., Jorge, J. A., & Amorim, D. S. (2005). Xylanases from fungi: properties and industrial applications. *Applied Microbiology and Biotechnology*, 67, 577–591.

Pollet, A., Delcour, J. A., & Courtin, C. M. (2010). Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. *Critical Reviews in Biotechnology*, 30, 176–191.

Prasad, A., Sotenko, M., Blenkinsopp, T., Coles, S.R., (2016). Life cycle assessment of lignocellulosic biomass pre-treatment methods in biofuel production. *Internal Journal of Life Cycle Assessment*, 21, 44-50.

Purmonen, M., Valjakka, J., Takkinen, K., Laitinen, T., & Rouvinen, J. (2007). Molecular dynamics studies on the thermostability of family 11 xylanases. *Protein Engineering, Design & Selection*, 20, 551–559.

Purushotham, P., Ho, R., Yu, L., Fincher, G. B., Bulone, V., & Zimmer, J. (2022). Mechanism of mixed-linkage glucan biosynthesis by barley cellulose synthase-like CslF6 (1,3;1,4)-β-glucan synthase. Science Advances, 8(45), eadd1596.

Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Jr, Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R., & Tschaplinski, T. (2006). The path forward for biofuels and biomaterials. *Science*, 311, 484–489.

Ralph, J., Grabber, J. H., & Hatfield, R. D. (1995). Lignin-ferulate cross-links in grasses: Active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydrate Research*, 275, 167-178.

Ralph, J., Hatfield R. D., Grabber J. H., Jung, H. G., Quideau, S., Helm, R. F. (1998). Cell wall cross-linking in grasses by ferulates and diferulates lignin and lignan biosynthesis. *American Chemical Society*, 209-236.

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, 174–186.

Ravelli, R. B. G., Cipriani, F., & Lavault, B. (2006). C3D: A program for the automated centring of cryocooled crystals. *Acta Crystallographica Section D*, 62, 1348-1357

Ravnik, V., Jukić, M., & Bren, U. (2023). Identifying metal binding sites in proteins using homologous structures, the MADE approach. *Journal of Chemical Information and Modeling*, *63*(16), 5204–5219.

Reddy Chichili, V. P., Kumar, V., & Sivaraman, J. (2013). Linkers in the structural biology of protein-protein interactions. Protein science. Protein Society, 22(2), 153–167.

Rodionova, M. V., Bozieva, A. M., Zharmukhamedov, S. K., Leong, Y. K., Lan, J. C., Veziroğlu, A., Везироглу, Т. Н., Tomo, T., Chang, J., & Allakhverdiev, S. I. (2022). A comprehensive review on lignocellulosic biomass biorefinery for sustainable biofuel production. *International Journal of Hydrogen Energy*, 47(3), 1481–1498.

Rohman, A., Dijkstra, B. W., & Puspaningsih, N. N. T. (2019). β-Xylosidases: Structural Diversity, Catalytic Mechanism, and Inhibition by Monosaccharides. International journal of molecular sciences, 20(22), 5524.

Rudjito, R. C., Jiménez-Quero, A., Muñoz, M. A., Kuil, T., Olsson, L., Stringer, M. A., Krogh, K. B. R. M., Eklöf, J., & Vilaplana, F. (2023). Arabinoxylan source and xylanase specificity influence the production of oligosaccharides with prebiotic potential. Carbohydrate Polymers, 320, 121233.

Saha, B. (2003). Hemicellulose bioconversion. *Journal of Industrial Microbiology & Biotechnology*, 30, 279–291.

Salazar-Cerezo, S., De Vries, R. P., & Garrigues, S. (2023). Strategies for the development of industrial fungal producing strains. *Journal of Fungi*, 9(8), 834.

Sánchez, O. J., & Cardona, C. A. (2008). Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technology*, 99, 5270–5295.

Saroj, P., Manasa, P., & Narasimhulu, K. (2018). Biofuels: Production of fungal-mediated ligninolytic enzymes and the modes of bioprocesses utilizing agro-based residues. Biocatalysis and Agricultural Biotechnology, 14, 57–71

Scheller, H. V., & Ulvskov, P. (2010). Hemicelluloses. Annual Review of Plant Biology, 61(1), 263–289.

Schröder, R., Atkinson, R. G., & Redgwell, R. J. (2009). Re-interpreting the role of endo-beta-mannanases as mannan endotransglycosylase/hydrolases in the plant cell wall. Annals of Botany, 104(2), 197–204.

Schultink, A., Liu, L., Zhu, L., & Pauly, M. (2014). Structural diversity and function of xyloglucan sidechain substituents. Plants, 3(4), 526–542.

Sharma, H. B., Sarmah, A. K., & Dubey, B. K. (2020). Hydrothermal carbonization of renewable waste biomass for solid biofuel production: A discussion on process mechanism, the influence of process parameters, environmental performance and fuel properties of hydrochar. *Renewable & Sustainable Energy Reviews,* 123, 109761.

Sharma, N., & Sharma, N. (2017). Microbial xylanases and their industrial applications as well as future perspectives: a review. *Global Journal of Biology and Agricultural Health Sciences*, 6, 5-12.

Shi, Z., Xu, J., Li, X., Li, R., & Li, Q. (2019). Links of extracellular enzyme activities, microbial metabolism, and community composition in the River-Impacted coastal waters. *Journal of Geophysical Research: Biogeosciences,* 124(11), 3507–3520.

Silva, J. C. R., Salgado, J. C. S., Vici, A. C., Ward, R. J., Polizeli, M. L. T. M., Guimarães, L. H. S., Furriel, R. P. M., & Jorge, J. A. (2020). A novel *Trichoderma reesei* mutant RP698 with enhanced cellulase production. *Brazilian journal of microbiology*, 51(2), 537–545.

Sivaramakrishnan, R., Suresh, S., Kanwal, S., Ramadoss, G., Ramprakash, B., & Incharoensakdi, A. (2022). Microalgal biorefinery concepts' developments for biofuel and bioproducts: current perspective and bottlenecks. *International Journal of Molecular Sciences*, 23(5), 2623.

Smith, P. J., Wang, H. T., York, W. S., Peña, M. J., & Urbanowicz, B. R. (2017). Designer biomass for next-generation biorefineries: leveraging recent insights into xylan structure and biosynthesis. *Biotechnology for Biofuels,* 10, 286.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., Vorwerk, S., & Youngs, H. (2004). Toward a systems approach to understanding plant cell walls. *Science*, 306, 2206–2211.

Song, Z., Zhang, Q., Wu, W., Zhang, P., & Yu, H. (2023). Rational design of enzyme activity and enantioselectivity. Frontiers in Bioengineering and Biotechnology, 11.

Srikrishnan, S., Chen, W., & Da Silva, N. A. (2013). Functional assembly and characterization of a modular xylanosome for hemicellulose hydrolysis in yeast. *Biotechnology and bioengineering*, 110(1), 275–285.

Strunk, R. J., Piemonte, K. M., Petersen, N. M., Koutsioulis, D., Bou-riotis, V., Perry, K., & Cole, K. E. (2014). Structure determination of BA0150, a putative polysaccharide deacetylase from *Bacillus anthracis*. *Acta Crystallographica Section F,* 70, 156-159.

Sützl, L., Laurent, C. V. F. P., Abrera, A. T., Schütz, G., Ludwig, R., & Haltrich, D. (2018). Multiplicity of enzymatic functions in the CAZy AA3 family. Applied Microbiology and Biotechnology, 102(6), 2477–2492.

Szydlowski, L., Boschetti, C., Crisp, A., Barbosa, E. G., & Tunnacliffe, A. (2015). Multiple horizontally acquired genes from fungal and prokaryotic donors encode cellulolytic enzymes in the bdelloid rotifer Adineta ricciae. *Gene*, 566(2), 125–137.

Tame, J.R.H. (2013). The structure of the deacetylase domain of *Escherichia coli* PgaB, an enzyme required for biofilm formation: A circularly permuted member of the carbohydrate esterase 4 family. *Acta Crystallographica Section D*, 69, 44-51.

Tanimoto, S., Higashi, M., Yoshida, N., & Nakano, H. (2016). The ion dependence of carbohydrate binding of CBM36: an MD and 3D-RISM study. *An Institute of Physics Journal*, 28, 344005.

Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Vincent, F., Brzozowski, A. M., Dupont, C., Shareck, F., Centeno, M. S., Prates, J. A., Puchart, V., Ferreira, L. M., Fontes, C. M., Biely, P., & Davies, G. J. (2006). Structure and activity of two metal ion-dependent acetylxylan esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases. *The Journal of Biological Chemistry*, 281, 10968–10975.

Taylor, M. J., Alabdrabalameer, H. A., & Skoulou, V. (2019). Choosing physical, physicochemical and chemical methods of pre-treating lignocellulosic wastes to repurpose into solid fuels. *Sustainability*, 11(13), 3604.

Tenkanen, M., Tamminen, T., & Hortling, B. (1999). Investigation of lignin-carbohydrate complexes in kraft pulps by selective enzymatic treatments. *Applied Microbiology and Biotechnology*, 51.

Terry, S. A., Badhan, A., Wang, Y., Chaves, A. V., & McAllister, T. A. (2019). Fibre digestion by rumen microbiota — a review of recent metagenomic and metatranscriptomic studies. *Canadian Journal of Animal Science,* 99(4), 678–692.

The European Parliament and the Council of the European Union, Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC 2009, 16-62.

The European Parliament and the Council of the European Union, Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources (recast), 82-209

Thompson W., Meyer S., (2013). Second generation biofuels and food crops: Co-products or competitors*? Global Food Security*, 2, 89-96.

Thurimella, K., Mohamed, A. M. T., Graham, D. B., Owens, R. M., La Rosa, S. L., Plichta, D. R., Bacallado, S., & Xavier, R. J. (2023). Protein language models uncover carbohydrate-active enzyme function in metagenomics. *bioRxiv*: the preprint *server for biology*, 2023.10.23.563620.

Trincone A. (2018). Update on marine carbohydrate hydrolyzing enzymes: Biotechnological applications. *Molecules*, 23, 901.

Tsai, S., Sun, Q., & Chen, W. (2022). Advances in consolidated bioprocessing using synthetic cellulosomes. *Current Opinion in Biotechnology*, 78, 102840.

Tsigos, I., Martinou, A., Kafetzopoulos, D., & Bouriotis, V. (2000). Chitin deacetylases: New, versatile tools in biotechnology. *Trends in Biotechnology*, 18, 305-312.

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

Tuveng, T. R., Rothweiler, U., Udatha, G., Vaaje-Kolstad, G., Smalås, A., & Eijsink, V. G. H. (2017). Structure and function of a CE4 deacetylase isolated from a marine environment. *PloS One*, 12, e0187544.

Urbanowicz, B. R., Peña, M. J., Moniz, H. A., Moremen, K. W., & York, W. S. (2014). Two Arabidopsis proteins synthesize acetylated xylan in vitro. *The Plant Journal for Cell and Molecular Biology,* 80, 197–206.

Urch, J. E., Hurtado-Guerrero, R., Brosson, D., Liu, Z., Eijsink,V. G. H., Texier, C., & van Aalten, D. M. F. (2009). Structural and functional characterization of a putative polysaccharide deacetylase of the human parasite *Encephalitozoon cuniculi*. *Protein Science*, 18, 1197-1209.

US Department of Energy (US DOE) (2006) Breaking the biological barriers to cellulosic ethanol: A Joint Research Agenda. In: Houghton, J., Weatherwax, S. and Ferrell, J., Eds., A research roadmap resulting from the biomass to biofuels workshop, Rockville, 7-9.

Vandermarliere, E., Bourgois, T. M., Rombouts, S., Van Campenhout, S., Volckaert, G., Strelkov, S. V., Delcour, J. A., Rabijns, A., & Courtin, C. M. (2008). Crystallographic analysis shows substrate binding at the -3 to +1 active-site subsites and at the surface of glycoside hydrolase family 11 endo-1,4-beta-xylanases. *The Biochemical Journal*, 410, 71–79.

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

Varelas, V. & Langton, M. (2017). Forest biomass waste as a potential innovative source for rearing edible insects for food and feed-a review. *Innovative Food Science & Emerging Technologies*, 41, 193-205.

Vasco-Correa, J., Luo, X., Li, Y.& Shah, A. (2019). Comparative study of changes in composition and structure during sequential fungal pre-treatment of non-sterile lignocellulosic feedstocks. *Industrial Crops and Products*, 133, 383-394.

Verma, D., & Satyanarayana, T. (2020). Xylanolytic extremozymes retrieved from environmental metagenomes: characteristics, genetic engineering, and applications. *Frontiers in microbiology*, 11, 551109.

Verma, R., & Pandit, S. B. (2019). Unraveling the structural landscape of intra-chain domain interfaces: Implication in the evolution of domain-domain interactions. *PLOS One*, 14(8), e0220336.

Viegas, I., Palma, M., Plagnes-Juan, É., Silva, E., Rito, J., Henriques, L. F., Tavares, L. C., Ozório, R., Panserat, S., & Magnoni, L. J. (2022). On the Utilization of dietary glycerol in carnivorous fish—Part II: insights into lipid metabolism of rainbow trout (*Oncorhynchus mykiss*) and European seabass (*Dicentrarchus labrax*). *Frontiers in Marine Science*, 9.

Vu, B., Chen, M., Crawford, J. R., & Ivanova, P. E. (2009). Bacterial extracellular polysaccharides involved in biofilm formation. *Molecules*, 14, 2535–2554.

Waghmare, P.R., Khandare, R.V., Jeon, B.H. & Govindwar, S.P. (2018). Enzymatic hydrolysis of biologically pretreated sorghum husk for bioethanol production. *Biofuel Research Journal*, 5, 846-853.

Wang, B., Andargie, M., & Fang, R. (2022). The function and biosynthesis of callose in high plants. Heliyon, 8(4), e09248.

Wang, J., Shi, T., Ding, L., Xie, J., & Zhao, P. (2023a). Multifunctional enzymes in microbial secondary metabolic processes. *Catalysts*, 13(3), 581.

Wang, L., Zhang, Y., & Gao, P. (2008). A novel function for the cellulose binding module of cellobiohydrolase I. *Science in China Series C, Life sciences*, 51, 620–629.

Wang, W., Zheng, D., Zhang, Z., Ye, H., Cao, Q., Zhang, C., Dong, Z., Feng, D., & Zuo, J. (2022). Efficacy of combination of endo-xylanase and xylan-debranching enzymes in improving cereal bran utilization in piglet diet. *Animal bioscience*, 35(11), 1733–1743.

Wang, X., Jiang, Y., Liu, H., Yuan, H., Huang, D., & Wang, T. (2023b). Research progress of multi-enzyme complexes based on the design of scaffold protein. *Bioresources and Bioprocessing*, 10(1).

Watanabe, H., & Tokuda, G. (2010). Cellulolytic systems in insects. *Annual Review of Entomology*, 55, 609–632.

Watanabe, H., Noda, H., Tokuda, G., & Lo, N. (1998). A cellulase gene of termite origin. Nature, 394, 330–331.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2 a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189–1191.

Wenzel, M., Schönig, I., Berchtold, M., Kämpfer, P., & König, H. (2002). Aerobic and facultatively anaerobic cellulolytic bacteria from the gut of the termite *Zootermopsis angusticollis*. *Journal of Applied Microbiology*, 92, 32–40.

White paper on renewable energy. (2003). Department of Minerals and Energy. Republic of South Africa.

Wilmanns, M. & Weiss, M. S. (2005). Molecular crystallography. *Encyclopedia of Condensed Matter Physics, Academic Press*, 1, 453–458.

Wilson D. B. (2011). Microbial diversity of cellulose hydrolysis. *Current Opinion in Microbiology*, 14, 259–263.

Wlodawer, A., Minor, W., Dauter, Z., & Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal*, 275, 1-21.

Xu, A., Zhang, Y., Zhao, Y., & Wang, J. (2013). Cellulose dissolution at ambient temperature: role of preferential solvation of cations of ionic liquids by a cosolvent. *Carbohydrate Polymers*, 92, 540–544.

Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., & Harvey, T. S. (1995). Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry*, 34, 6993–7009.

Xu, X., Xu, Z., Shi, S., & Lin, M. (2017). Lignocellulose degradation patterns, structural changes, and enzyme secretion by *Inonotus obliquus* on straw biomass under submerged fermentation. *Bioresource Technology*, 241, 415–423.

Yagi, H., Maehara, T., Tanaka, T., Takehara, R., Teramoto, K., Yaoi, K., & Kaneko, S. (2017). 4-O-methyl modifications of glucuronic acids in xylans are indispensable for substrate discrimination by GH67 α-glucuronidase from *Bacillus halodurans* C-125. *Journal of Applied Glycoscience*, 64(4), 115–121.

Yamada, R., Nakatani, Y., Ogino, C., & Kondo, A. (2013). Efficient direct ethanol production from cellulose by cellulase- and cellodextrin transporter-co-expressing *Saccharomyces cerevisiae*. *AMB Express*, 3, 34-42.

Yang, S., Wu, C., Yan, Q., Li, X., & Jiang, Z. (2023). Nondigestible functional oligosaccharides: enzymatic production and food applications for intestinal health. *Annual Review of Food Science and Technology*, 14(1), 297–322.

Yazdani, S. S., & Gonzalez, R. (2007). Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Current Opinion in Biotechnology*, 18, 213–219.

Ye, T. J., Huang, K. F., Ko, T. P., & Wu, S. H. (2022). Synergic action of an inserted carbohydrate-binding module in a glycoside hydrolase family 5 endoglucanase. *Acta crystallographica. Section D, Structural biology*, 78(Pt 5), 633–646.

Zakaria, M. R., Fujimoto, S., Hirata, S., & Hassan, M. A. (2014). Ball milling pre-treatment of oil palm biomass for enhancing enzymatic hydrolysis. *Applied Biochemistry and Biotechnology*, 173, 1778–1789.

Zannini, E., Bravo Núñez, Á., Sahin, A. W., & Arendt, E. K. (2022). Arabinoxylans as Functional Food Ingredients: A Review. Foods, 11(7), 1026.

Zhang, L., Zhou, Y., & Zhang, B. (2023). Xylan-Directed cell wall assembly in grasses. Plant Physiology.

Zhang, S., Zhang, L., Xu, G., Li, F., & Li, X. (2022). A review on biodiesel production from microalgae: Influencing parameters and recent advanced technologies. *Frontiers in microbiology*, 13, 970028.

Zhang, S., Zhao, S., Shang, W., Yan, Z., Wu, X., Li, Y., Chen, G., Liu, X., & Wang, L. (2021). Synergistic mechanism of GH11 xylanases with different action modes from Aspergillus niger An76. *Biotechnology for biofuels*, 14(1), 118.

Zhang, Y., Fu, X., & Chen, H. (2012). Pre-treatment based on two-step steam explosion combined with an intermediate separation of fibre cells optimization of fermentation of corn straw hydrolysates. *Bioresource Technology*, 121, 100–104.

Zhao, X., Zhang, L., & Liu, D. (2012). Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. *Biofuels, Bioproducts and Biorefining*, 6, 465–482.

Zhao, Z., Baltar, F., & Herndl, G. J. (2020). Linking extracellular enzymes to phylogeny indicates a predominantly particle-associated lifestyle of deep-sea prokaryotes. *Science Advances*, 6(16).

Zoghlami, A., & Paës, G. (2019). Lignocellulosic Biomass: Understanding recalcitrance and predicting hydrolysis. *Frontiers in Chemistry*, 7.