Structural characterization of xylanolytic enzymes and their application in the production of anti-cancer xylooligosaccharides

A Thesis

Submitted in partial fulfillment of the
requirements for the Degree of

Doctor of Philosophy

by

Kedar Sharma

Under supervision of

Professor Arun Goyal

September 2019

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI – 781039, ASSAM, INDIA

Structural characterization of xylanolytic enzymes and their application in the production of anti-cancer xylooligosaccharides

A Thesis

Submitted in partial fulfillment of the
requirements for the Degree of

Doctor of Philosophy

by

Kedar Sharma

Under supervision of

Professor Arun Goyal

September 2019

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI – 781039, ASSAM, INDIA

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

STATEMENT

I do hereby declare that the content embodied in this thesis entitled as **"Structural characterization of xylanolytic enzymes and their application in the production of anti-cancer xylooligosaccharides"** is the result of investigations carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India under the guidance of Professor Arun Goyal.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the work described is based on the findings of other investigators.

September, 2019

Kedar Sharma

(146106031)

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

CERTIFICATE

It is certified that the work described in this thesis entitled **"Structural characterization of xylanolytic enzymes and their application in the production of anti-cancer xylooligosaccharides"** by **Kedar Sharma (Roll No. 146106031)** for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision at the Department of Biosciences & Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India and this work has not been submitted elsewhere for a degree.

Dr. Arun Goyal *(MTech, PhD)*
(FAMI, FBRS, FABAP, FNABS, FNAAS, FIFIB)
Professor
(Thesis Supervisor)
Department of Biosciences & Bioengineering
Indian Institute of Technology Guwahati
Guwahati, 781 039, India

ACKNOWLEDGEMENTS

Working towards the accomplishment of this thesis has been an incredible journey of my life, which would not have been possible without the encouragement and support of many people including my supervisor, doctoral committee members, my family, my friends and colleagues.

I will always remain indebted to the Almighty God for being the guiding light at all times.

As correctly mentioned by Late President of India, Dr. A.P.J. Abdul Kalam. "Teaching is a very noble profession that shapes the character, caliber, and future of an individual." Therefore, I am incredibly grateful to my thesis supervisor, Professor Arun Goyal, Department of Biosciences & Bioengineering, IIT Guwahati, for providing me the opportunity to carry out research work under his supervision. I wish to thank and acknowledge him for his excellent guidance, caring, patience and providing me with an excellent atmosphere to think freely for designing the experiments for carrying out my research work. Without his diligent planning, valuable suggestions and constructive criticism, this work couldn't have been assumed in its current form. His dedication to science and research has inspired me a lot during the research work. I shall forever remain grateful to him for all that he has done.

I would also like to express my heartfelt appreciation to all my doctoral committee members Dr. Lalit Mohan Pandey, Dr. Soumen Kumar Maiti, Prof. Subrata Kumar Majumder and Prof. Ranjan Tamuli for their valuable suggestions and constructive criticism that has led to the successful completion of my thesis.

I am thankful to Department of Biosciences & Bioengineering and Central Instrumentation Facility (CIF), IITG for providing me instruments for my research work.

I extended my heartfelt gratitude to Dr. Sachin Kumar, IIT Guwahati and his lab members for cell culture experiments and valuable suggestion during manuscript writing.

I would also like to express my sincere gratitude to Dr. Ashish, Senior Principal Scientist, Institute of Microbial Technology, Chandigarh and his research group, Dr. Ravishankar Ramachandran, Senior Principal Scientist, Central Drug Research Institute Lucknow and his research group for providing SAXS facility. I would also like to express my sincere gratitude to Mr. Prasad Gosavi and Ms. Navjot Kaur Saini, Anton Paar India, for helping in SAXS data collection and analysis.

I would also like to thank the present and former heads of the Department of Biosciences & Bioengineering, IIT Guwahati, Prof. Latha Rangan, Prof. K. Pakshirajan, and Prof. Venkata. V. Dasu for providing me with the necessary facilities.

I would also like to express my sincere gratitude to Prof. Carlos M.G.A. Fontes CIISA-Faculdade de Medicina Veterinária, Avenida da Universidade Técnica, Lisbon, Portugal, for the necessary facilities to carry out the work. I feel lucky to be a one the beneficiary of his in-depth knowledge in the field of carbohydrate active enzymes and his critical suggestions in the manuscripts, which ultimately improved the overall thesis.

I extended my heartfelt gratitude to Dr. Shabir Najmuddin for providing his valuable suggestions for X-ray crystal diffraction and SAXS data analysis and constructive criticism during manuscript writing.

I am also thankful to my seniors Dr. Shadab Ahamed, Dr. Rishikesh Shukla, Dr. Shraddha Shukla, Dr. Arabinda Ghosh, Dr. A.K. Verma, Dr. Damini Kothari, Dr. Soumyadeep Chakraborty, Dr. Shuchi Singh, Dr. Aruna Rani, Dr. Arun Dhillon and Dr. Rwivoo Baruah for their help and suggestions. I am immensely thankful to my all the research group members, specially Abhijeet and Kaustubh for their support. I am thankful to M.Tech *students, Karthika and Aakash for the help. I am grateful to all the people with whom I have worked in the lab at the Department of Biosciences and Bioengineering for their cooperation and support.*

I want to thank my friends cum family Pukha Raj, Deepika, Swati, Devesh, Suman, Madhuri, Monika, Pooja, Nemashri, Pooja, Vikram, Neha, Ravi, Bhanupratap, Babita, Rahul, Sudhir, Balwant, Ajay, Gaurav, Debika, Vinay, Vartika, Dhana, Suraj, Anurag, Ram and Jitendra for their support.

I wish to acknowledge the support received from other teaching and non-teaching staff of the Department of Biosciences and Bioengineering, IIT Guwahati.

I wish to acknowledge MHRD, Govt. of India for providing financial assistance and also Department of Biotechnology, Govt. of India, New Delhi for providing me fellowship through its sponsored project.

I would like to express my heartfelt gratitude to all the teachers Mr. Ramavtar Sharma (Bade Papa), Mr. Karansingh Yadav, Mr. Ramesh Chand Sharma, Mr. Vikas Yadav, Mr. Madhusudan Verma, Dr. Sunil Kumar Senapati, Dr. Pradeep Shukla, Dr. Pradeep Verma and Dr. Prithvi Sukumar, who taught me and motivated me from the beginning of my life to present day,

My Ph.D. endeavor would not have been successful without the love, trust, support and blessings of Bade Papa, Badi Mummy, Papa, Mummy, Bhai, Didi and I owe my achievements to all the teachers, my family and friends.

Kedar Sharma
September, 2019

SYNOPSIS

Introduction

The sustainable growth of any country mainly relies on three pillars, *i.e.*, development at the social, economic and environmental level. The socio-economic development of any country, whether it is developing or developed depends heavily upon the utilization of non-renewable resources or fossil resources (Linear fossil-based economy). These resources play a potential role in making life easy at various levels ranging from home appliances, in transportation, in various industrial processes for the fulfillment of our daily needs. Due to the limited availability with unsustainable use of these non-renewable resources, the interest in the steady shift from a linear fossil-based economy to circular bio-based economy or bioeconomy is increasing day by day. The conversion of renewable resources into value-added products such as food, animal feed, bio-based other value-added products and bioenergy production through a biotechnological intervention is termed as bioeconomy. Agricultural practices are the main component of bioeconomy. The need for food and energy supply increasing with the rise in population and it implies a tremendous pressure on agriculture farming and its further processing industries. The processing of these crops is producing about $1.3\text{x}10^{10}$ metric tons per annum of lignocellulosic waste worldwide and that can be used for the production of bio-based value-added products and bioenergy.

The organic material derived from plants or agricultural waste is termed as lignocellulosic biomass. Lignocellulosic biomass contains cellulose, lignin and hemicellulose as major components (De Wild, 2011). The major portion of plant polysaccharides is composed of cellulose, while the second most abundant plant polysaccharide is hemicelluloses. Hemicellulose constitutes 15-35% of total lignocellulosic biomass (Limayem & Ricke, 2012). Hemicelluloses are made up of monomeric sugars such as pentose (xylose, arabinose), hexoses (mannose and glucose), acetylated sugars and uronic acids (Davison, 2013). The structure and composition of heteroxylan depend on the plant and its tissues and method used for the extraction (Petzold-Welcke et al., 2014). The backbone of heteroxylan is composed of β-(1→4)-linked xylose residues, which is often substituted with methylated α-D-glucopyranosyl uronic acid and α-L-arabinofuranosyl, ferulic acid, *p*-coumaric acid in the side chain (Ordaz-Ortiz and Saulnier, 2005). The complex xylan structure requires the synergistic action of an array of enzymes such as endo-β-xylanase, α-glucuronidase, β-xylosidase, α-arabinofuranosidase, *p*-coumaric acid esterase, acetyl xylan esterase and ferulic acid esterase for its complete hydrolysis (Sharma et al., 2018). Among all these enzymes, the endo-β-xylanase randomly (EC.3.2.1.8) hydrolyzes the xylan and produces xylooligosaccharides, which can be used as a food supplement for the improvement of probiotic bacterial growth (Goluguri et al., 2016). Currently, endo-1,4-β-xylanase are found in 16 GH families *viz.* 3, 5, 8, 9, 10, 11, 12, 16, 26, 30, 43, 44, 51, 62, 98 and 141 as single catalytic domain or as modular enzymes. β-xylosidase releases xylose residues from non-reducing ends of xylooligosaccharides. β-xylosidase are classified in family 1, 2, 3, 5, 10, 30, 39, 43, 51, 52, 54, 116 and 120 glycoside hydrolases. α-L-arabinofuranosidases catalyze the removal of α-L-arabinofuranosyl residue from the side chain of arabinoxylan. Currently, α-L-arabinofuranosidases are reported from several GH families *viz.* 2, 3, 5, 10, 43, 51, 54 and 62. However, a complete degradation requires synergistic action of acetyl esterase to remove acetyl substituents from β-1,4-linked D-xylose backbone of xylan (Wong and Saddler 1993; Coughlan and Hazlewood 1993).

Family 43 glycoside hydrolases

The family 43 glycoside hydrolase (GH43) contains mainly hemicellulolytic enzymes. As of today (September 2019), the family GH43 enzymes deposited in the CAZy database includes 19 archaeal origins, 13229 bacterial and 481 from eukaryotes, 5 from viruses and 28 are unclassified. These family 43 glycoside hydrolases are further subclassified into 37 subfamilies (http://www.cazy.org/GH43.html). Till date, 13 arabinofuranosidase from various organisms reported GH43_16 subfamily had been biochemically characterized. Out of these, the crystal structure of only 1 modular arabinofuranosidase (*Bs*AXH-m2,3) from *Bacillus subtilis* containing GH43 domain as a catalytic module and family 6 carbohydrate-binding module (CBM6) as a non-catalytic module (PDB Id: 3C7E) have been reported and available in PDB (Vandermarliere *et al*., 2009).

Family 30 glycoside hydrolase

The family 30 Glycoside hydrolases (GH30) a variety of activities which include endo-β-1,4-xylanase (EC 3.2.1.8); β-xylosidase (EC 3.2.1.37) and glucuronoarabinoxylan endo-β-1,4-xylanase (EC 3.2.1.136). Till date (September 2019), family GH30 entries in CAZy database include 2376 bacterial and 105 protein entries from eukaryotes, while only 3 entries are as unclassified. These family 30 glycoside hydrolases are further subclassified into 9 subfamilies (http://www.cazy.org/GH30.html). Under GH30_8 subfamily 596 entries are listed and out of these the crystal structure of 6 proteins are already determined, but the role of linker peptide present within the protein is not determined so far.

1.3.3 Family 10 glycoside hydrolase

The family 10 glycoside hydrolase (GH10) mainly endo-1,4-β-xylanase (EC 3.2.1.8). The family GH10 contains a total of 3851 protein sequences in the CAZy database as of September 2019 (http://www.cazy.org/GH10.html). Out of 3851 sequences, 19 belong to archaea, 3015 belong to bacteria, 461 belong to eukaryote and 356 belongs to unclassified sequences. As of now (September 2019) 360 xylanases are characterized and out of 360 xylanases, only a few mesophilic bacterial xylanases belonging to family 10 glycoside hydrolase have been identified and characterized [Fontes et al., 2000, Gallardo et al., 2010; Sharma et al., 2018). Out of these 360 characterized xylanases, the crystal structure of 45 xylanases is available in the PDB database. Majority of the crystal structures solved till date are of thermophilic xylanases, while the solution shape analysis of xylanases by Small Angle X-ray Scattering (SAXS) analysis is not reported till date.

These xylanolytic enzymes exist in nature either in a modular organization or as a single catalytic module. As, several questions remain unanswered about the conformational dynamics, molecular arrangement and protein stability of the modular and non-modular enzymes in solution. Therefore, the present study deals with the structural organization, dynamics and function characterization of modular and non-modular xylanolytic enzymes from *Clostridium thermocellum and Pseudopedobacter saltans.*

Present work

The present work entitled as “Structural characterization of xylanolytic enzymes and their application in the production of anti-cancer xylooligosaccharides” has been divided into 6 chapters.

Chapter 1 is a General Introduction, which mainly focuses on the brief review of literature dedicated to the bioeconomy, the lignocellulosic waste and the structural component of the plant cell wall with elaborated description of heteroxylan. This chapter describes different types of carbohydrate-active enzymes and their sequence-based classification. It includes a review of literature on families 43, 30 and 10 of glycoside hydrolases and their potential applications followed by significance of the work. The description of *Clostridium thermocellum* and *Pseudopedobacter saltans* has also been included in this chapter.

Chapter 2 describes the crystallization and the structure determination of α-L-arabinofuranosidase (EC 3.2.1.55) of family 43 glycoside hydrolases (*Ct*Abf43A) from *C. thermocellum.* The crystal structure of *CtAbf43A* comprises a five-bladed β-propeller fold typical of GH43 enzymes. *CtAbf43A* displays a highly compact architecture compatible with its high thermostability. The structural comparison of GH43_16 subfamily and GH43_29 subfamily enzymes reveals the presence of a highly conserved substrate-binding cleft, that reflects the topology of arabinoxylan. GH43_16 is thus unable to bind the arabinan backbone that presents a curved shape. This work presents novel information that contributes to understanding the molecular determinants of substrate binding specificity within one of the most biotechnological relevant CAZy families, GH43.

Chapter 3 mainly focuses on the analysis of conformational dynamics and modular arrangement of *Ct*Araf43 in solution. Enzymes participating in the hydrolysis of complex carbohydrates display a modular architecture. The significance of enzyme modularity in terms of flexibility and catalytic efficiency is not fully understood. α-L-arabinofuranosidase from *Clostridium thermocellum* (*Ct*Araf43) catalyzes the release of α(1→2)-, α(1→3)-, or α(1→5)- linked L-arabinose from arabinose decorated polysaccharides. In this study, the conformational dynamics and modular arrangement of *Ct*Araf43 in solution were described. The homology modeled structure of a family 43 modular α-L-arabinofuranosidase of *Clostridium thermocellum* revealed three distinct domains that are independently folded and achieve their stable conformation. The modeled structure of the full-length *Ct*Araf43 revealed that its N-terminal catalytic *Ct*Abf43A module displays a 5-fold β-propeller fold and the two CBMs display the typical jellyroll type β-sandwich folds. Ramachandran plot showed 98.5% residues in the favored region and only 1.5% residues in the disallowed region. Conformational dynamics analysis of *Ct*Araf43 by MD simulation displayed its flexible nature and no change in secondary structure was observed. SAXS analysis of *Ct*Araf43 revealed its monodisperse nature at two concentrations, 1.2 mg/mL and 4.7 mg/mL. The SAXS analysis of *Ct*Araf43 at both concentrations suggested elongated structures with monomeric and dimeric conformations, respectively. Kratky plot revealed that *Ct*Araf43 retains a flexible state at both concentrations, which corroborated with MD simulation results. The pyDockSAXS displayed that the catalytic domain of two

*Ct*Araf43 molecules interacted together and resulted in the dimerization of *Ct*Araf43. The *ab initio* derived molecular of *Ct*Araf43 showed the molecular envelope of Puppy shaped at 1.2 mg/mL and Genie lamp shape at 4.7 mg/mL, respectively, which superposed well with the MD simulated monomeric structure and dimeric structures of *Ct*Araf43 obtained from the pyDockSAXS analysis.

Chapter 4 describes the conformational flexibility, molecular organization and protein stability of *Ct*XynGH30 in solution. The secondary structure analysis of *Ct*XynGH30 by CD displayed the presence of 28.25% α helices and 40.5% β sheets. The melting analysis showed that the Ca^{2+} ions provide extra thermal stability to the *Ct*XynGH30 structure. The ITC and mass spectrometric analyses of individual modules, *Ct*Xyn30A and *Ct*CBM6, showed that even though there is no binding interaction between them, the two modules are required to be linked together for optimal function. The mixture of *Ct*Xyn30A and *Ct*CBM6 displayed lower xylanase activity as compared to the full-length *Ct*XynGH30 protein. The lower activity of *Ct*Xyn30A alone and the mixture (*Ct*Xyn30A and *Ct*CBM6) may be due to the absence of linker region, which may be functioning as a sliding hinge between the two modules, thereby facilitating the binding to the substrate and bringing about enhanced catalytic activity of the full-length enzyme. SAXS analysis of *Ct*XynGH30 revealed a boot-shaped, fully folded structure in solution. The SREFLEX refined model of *Ct*XynGH30 superposed well with the *ab initio* modeled structure. The optimal length and composition of linker peptide is crucial for the catalytic activity. These results confirmed that the two domains are required to be attached via a linker peptide, but with enough flexibility to maintain a specific spatial organization to achieve optimal biological function.

Chapter 5 gave insights about the structural and functional characterization of endo β-1,4 xylanase (*Ps*GH10A) of a family 10, glycoside hydrolase from *Pseudopedobacter saltans*. 3D structure of *Ps*GH10A modeled by comparative modeling was compact and stable. The secondary structure by CD analysis of *Ps*GH10A displayed the presence of α-helices (31.75%), β-strands (20.00%) and random coil (48.25%). The *Ps*GH10A modeled structure showed $(\beta/\alpha)_8$ TIM barrel fold which is conserved in GH10 family. The docking studies revealed that the active site of *Ps*GH10A could accommodate linear xylooligosaccharides and arabinose substituted xylooligosaccharides. The residues, Glu153 and Glu263 are involved in catalysis. *Ps*GH10A melting analysis confirmed that the enzyme does not require any metal ion for its stability. SAXS analysis displayed the monomeric nature of *Ps*GH10A in solution form. Guinier analysis gave the radius of gyration between (Rg) 2.23±0.15 and 2.29±0.16 nm for all the protein concentrations. Kratky plot analysis of *Ps*GH10A displayed fully folded state in solution form. The *ab initio* derived DAM model of *Ps*GH10A superposed well with its comparative modeling based 3D-structure.

Chapter 6 describes the extraction of xylan, characterization and effect of xylooligosaccharide on colon cancer cells. Glucuronoxylan was isolated from neem sawdust. Carbohydrate composition and structural characterization analysis of extracted xylan displayed that the main chain is composed of xylose backbone, which is substituted with a 4-O-methyl glucuronic acid side chain. The hydrolysis of extracted xylan from neem sawdust by endo-β-1,4-xylanase (*Ps*GH10A) resulted in the release of xylooligosaccharides ranging from the degree of polymerization (DP) 2-7. *In vitro* cell cytotoxicity analysis of xylooligosaccharide on mouse fibroblast (L929) cells displayed its biocompatible behavior. The HT-29 cells treated with xylooligosaccharides displayed the inhibition of cell growth. The mitochondrial membrane potential analysis of HT29 cells displayed reduction in the red/green fluorescence intensity confirming the disruption of mitochondrial membrane. FLICA, immunoblotting and nuclear fragmentation analysis revealed the caspase-3 mediated apoptotic killing mechanism of HT-29 cells. These results further contribute to the knowledge about xylooligosaccharides produced from extracted xylan gave better insights about their utilization in cancer therapy for the treatment of colon cancer.

CONTENTS

Chapter 5. SAXS and comparative modeling based structure analysis of endo-β-1,4-xylanase a family 10 glycoside hydrolase from *Pseudopedobacter saltans* 129

Chapter 6. Extraction, characterization of xylan from neem sawdust and its application in xylanase mediated production of anticancer xylooligosaccharides 159

Chapter 1

General Introduction

1.1. Bioeconomy

The sustainable growth of any country mainly relies on three pillars, *i.e.,* development at the social, economic and environmental level. The socio-economic development of any country, whether it is developing or developed depends heavily upon the utilization of non-renewable resources or fossil resources (Linear fossil-based economy) (Chiranjeevi et al., 2018). These resources play a potential role in making life easy at various levels ranging from home appliances, in transportation, in various industrial processes for the fulfillment of our daily needs. Due to the limited availability with unsustainable use of these non-renewable resources, the interest in the steady shift from a linear fossil-based economy to circular bio-based economy or bioeconomy is increasing day by day (Chiranjeevi et al., 2018). The conversion of renewable resources into value-added products such as food, animal feed, bio-based other value-added products and bioenergy production through a biotechnological intervention is termed as bioeconomy (Chiranjeevi et al., 2018). Agricultural practices are the main component of bioeconomy. The need for food and energy supply increasing with the rise in population and it implies a tremendous pressure on agriculture farming and its further processing industries (Rosenau-Tornow et al., 2009). The organic waste material produced by intensive agricultural practices and food processing is termed as lignocellulosic biomass or lignocellulosic waste. The production of lignocellulosic waste per year is about $1.3x10^{10}$ metric tons and its management is becoming burgeoning problem (Liu et al., 2019). In last few years the burning lignocellulosic biomass by farmers to clean the farming land has contributed immensely towards climate change, water and air pollution. The lignocellulosic waste is rich in carbon other valuable vital nutrient and therefore, it can be used for the production of high valued bio-based products and bioenergy (Liu et al., 2019).

1.2. Structure of lignocellulosic biomass

Lignocellulosic biomass contains cellulose, lignin and hemicellulose as major components (De Wild, 2011). Structure and composition of plant cell wall vary from different developmental stages of cell, tissues and plant species. A plant cell wall is composed of three different layers: the primary cell wall, the secondary cell wall and the middle lamella. The primary wall formed in young cells provides flexibility and basic structural support, mediating cell-cell interaction and protecting the cell (Houstan et al., 2016). When the cell is fully expanded, the more durable and thicker secondary wall is deposited between the plasma membrane and primary cell wall of the cell (Somerville et al., 2004). This secondary wall is highly ordered with parallel fibers of crystalline cellulose interlinked with lignin and hemicellulose. The secondary wall is seen as a crucial adaptation, which allows terrestrial plants to withstand and facilitate upright growth. The middle lamella is composed of pectin polysaccharide, which forms a composite, continuous layer and acts as a cementing material between adjacent cells. As the plant is growing, the pectin layer is lignified to reinforce the cell structure (Sjöström, 1993). These polysaccharides provide mechanical support and maintain the structural organization to the plants, but they also assume transporting functions of water, nutrients and products of photosynthesis within the plant (Showalter, 1993; McQueen-Mason *et al.*, 1994). Figure 1.1 highlights the different polysaccharides present in the plant cell wall and its ordered interconnected network. To better understand these arrangements and their functions, it is necessary to understand the role of different plant polysaccharides.

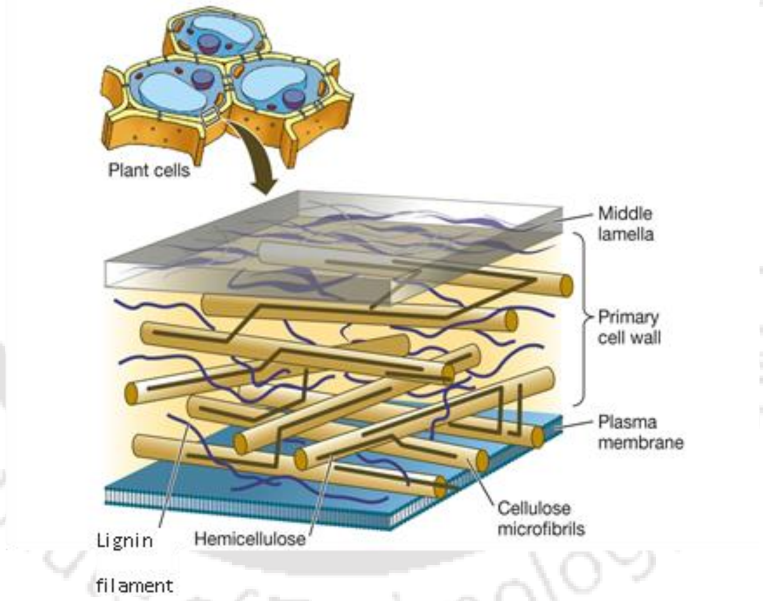


Fig. 1.1. Model of the primary cell wall in higher plants (from Sadava et al., 2009).

1.2.1. Cellulose

Cellulose is the most abundant polysaccharide present in the plant cell wall of lignocellulosic biomass, which constitutes 30–50% of the total biomass with a global production estimated to be 120-140 billion tons annually (Wang et al., 2016). Cellulose is primarily present in plant cell walls, but it is also produced by other organisms, such as bacteria, algae and some marine animals. The composition of cellulose is dependent on the plant species as well as age (Limayem and Ricke, 2012). The primary cell walls of all gymnosperms and angiosperms predominantly contain cellulosic polysaccharide. The cottonseed hairs contain 91% of cellulose, which represents its pure form in comparison to wood cellulose. The cellulosic polymer is a long unbranched homopolymer, which is composed of a repeated monomeric unit of β-D-glucopyranose residues joined by β-(1→4)-glycosidic linkages (Fig. 1.2) (Klemm *et al.*, 2005; Mussatto and Teixeira, 2010). Each repeating β-D-glucopyranose unit is rotated 180° to its neighbouring β-D-glucopyranose unit (Ten and Vermerris, 2013), which resulted in the formation of cellulose. These cellulose chains further crosslink or aggregate by forming intra- and intermolecular hydrogen bonds and van der Waals interactions to produce the cellulose microfibrils. The average degree of polymerization in the primary cell wall is approximate, 6000 glucose units and up to 14000 glucose units in the secondary cell walls. The cellulose microfibrils are highly crystalline and hydrophobic, that contributes greatly to the recalcitrance of biomass (Harris *et al.*, 2009).

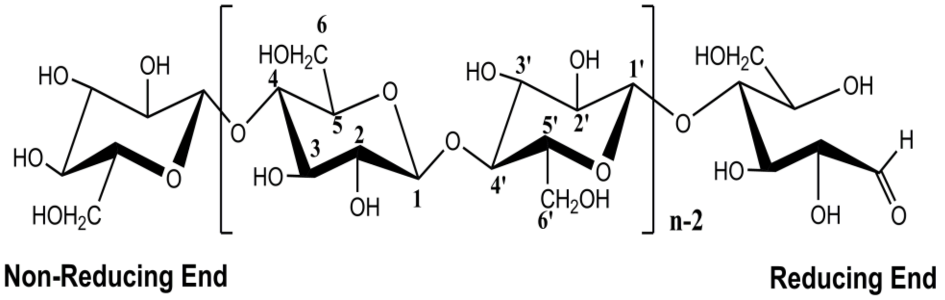


Fig. 1.2 Chemical structure of cellulose showing β-(1→4) glycosidic linkages, reducing end and non-reducing end (Klemm *et al.*, 2005).

1.2.2. Lignin

Lignin is an amorphous, highly branched and heterogeneous aromatic polymer that lacks in primary structure and aromatic functionality. Lignin is derived from three basic phenylpropane monomers as repeating units *viz*. *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol (Fig. 1.3) which are collectively termed as “monolignols.” The softwood lignin is mainly composed of coniferyl alcohol, the hardwood lignin is composed of coniferyl and sinapyl alcohols, while the lignin present in grasses contains all the three above mentioned moieties (Duval and Lawoko, 2014). Lignin is mainly present in the secondary plant cell walls and constitutes one of the major components of the plant cell wall. It contributes greatly to the transportation of nutrients and water to the xylem in vascular plants by providing mechanical support to the cell. Its recalcitrance nature also imparts the resistance against microbial pathogens (Boerjan *et al*., 2003; Ralph *et al*., 2004).

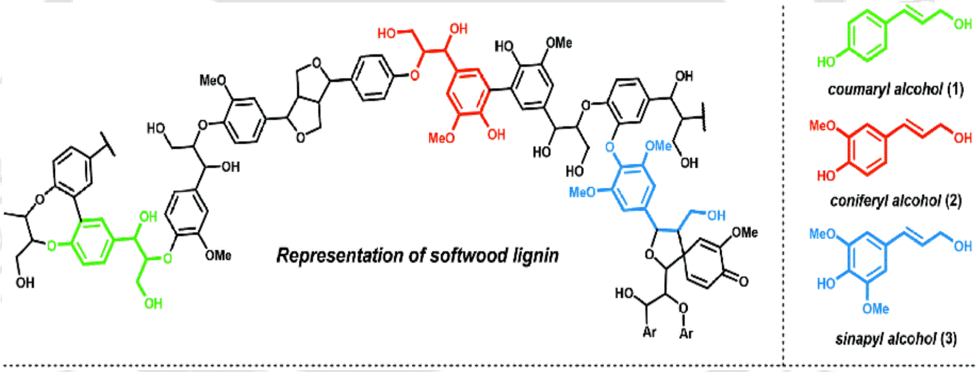


Fig. 1.3. Structure of Lignin, the side panel showing three monomers *p*-coumaryl (H), sinapyl (S) and coniferyl (G) alcohols involved in the formation of lignin (Kärkäs et al., 2016)

1.2.3. Hemicelluloses

Hemicellulose constitutes 15-35% of total lignocellulosic biomass (Limayem and Ricke, 2012). Hemicelluloses are an amorphous, single chain and branched heteropolysaccharides in nature. Hemicelluloses are made up of monomeric sugars such as pentose (xylose, arabinose), hexoses (mannose and glucose), acetylated sugars and uronic acids with an average degree of polymerization between 500-3000 monomeric units (Davison, 2013). These D- pyranosyl residues such as glucose, mannose and xylose are linked through β-1,4 glycosidic linkage at O4 position in equatorial orientation (Rose, 2003). The branches of hemicelluloses are shorter in length, which makes its non-crystalline nature (Cunha and Gandini, 2010). The presence of different functional groups in branches and their amount varies depending on the source of the lignocellulosic biomass (Limayem and Ricke, 2012). Hemicelluloses are subclassified into four main sub-groups depending on the presence of 1-4 linked monomeric sugars present in the backbone of polysaccharides; namely mannans, xyloglucans, arabinans and xylans (Scheller and Ulvskov, 2010). The most abundant type of hemicellulose is xylan and its composition greatly varies among the species (Limayem and Ricke, 2012). The most important biological role of hemicelluloses is to contribute to the strengthening of the cell wall by interacting with lignin and cellulose (Scheller and Ulvskov, 2010).

1.2.3.1. Mannans

Mannans are hetero-polysaccharides and a major component of the angiosperms cell wall and Charophytes (obligate water plants) (Popper and Fry, 2003; Popper, 2008). It is composed of D-mannopyranose residues connected through β-(1→4) mannosidic linkage in the backbone. These homo-mannan and heteromannan serves as storage compounds in seeds (Meier, 1982). The β-(1→4)-linked mannose backbone is usually substituted with glucose or galactose through β-(1→6) and β-(1→4) linkages (Sharma et al., 2018).

1.2.3.1.1. Galactomannan

Galactomannan is a major constituent of the endospermic cell wall of leguminous seeds and stored as a food component, which constitutes more than 30% of dry seed weight (Buckeridge, 2010). The backbone of the galactomannans is composed of β-(1→4)-linked D-mannopyranose residues and it is randomly substituted with a single molecule of D-galactopyranose connected through α-1-6-glycosidic linkage (Fig. 1.4). The common source of galactomannans includes locust bean gum, guar gum, fenugreek and alfalfa plants.

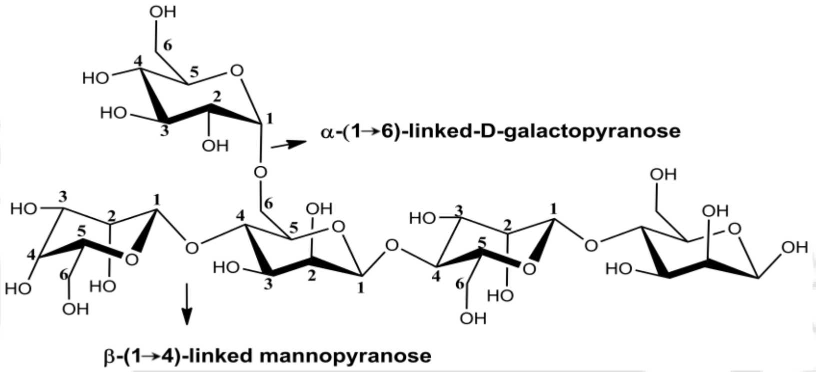


Fig. 1.4. Chemical structure of galactomannan exhibiting the repeating units of β-(1→4)-D-mannopyranose making the backbone substituted with α-(1→6)-linked-D-galactopyranose side chain. (Guo et al., 2017).

1.2.3.1.2. Glucomannan

Glucomannans are present in the seeds of family Liliaceae and Iridaceae as storage polysaccharides (Meier and Reid, 1982). Apart from the seeds, glucomannans are also found in the roots, bulbs and tubers of several other types of plants such as konjak plant (*Amorphophallus konjac*). Glucomannans are the linear polymer of β-(1→4) linked D-mannopyranose (~70%) and D-glucopyranosyl (~30%) units (Fig. 1.5). Glucomannans are soluble dietary fiber and used in the food industry as thickener, emulsifier and as therapeutic agents in for the treatment of obesity, high cholesterol, constipation, acne vulgaris and type 2 diabetes (Walsh *et al*., 1984; Passaretti *et al*., 1991; Vuksan *et al*., 1999).

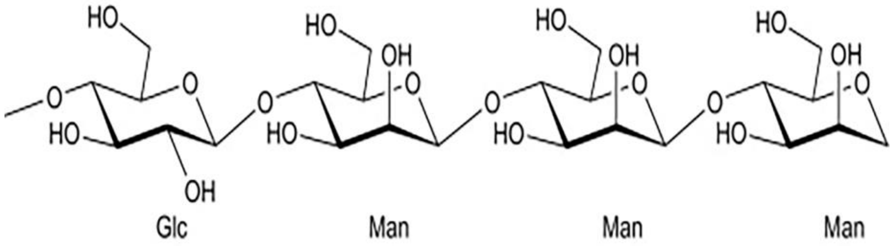


Fig. 1.5. Chemical structure of glucomannan displaying randomly linked β-(1→4)-D-mannopyranose (Man) and D-glucopyranose (Glc). (Adopted from Salinas et al., 2018).

1.2.3.2. Xyloglucans

Xyloglucan is a linear polysaccharide, which contains a backbone of β-(1→4)-linked glucose units and randomly substituted with the side chain of β-(1→6)-D-xylose linked units (Fig 1.6). Xyloglucan mainly occurs in the primary cell wall of most vascular plants. Xyloglucan polymer also exists as a storage polysaccharide in some of the seed plants such as Tamarindus, Impatiens, Annona, Tropaeolum, Hymenaea, and Detarium (Nishinari et al., 2009). In the primary cell wall, xyloglucan interacts with cellulose microfibrils and forms network structure, which acts as a major load-bearing structure and play a key role in providing the growth to cell wall (Cosgrove, 2005).

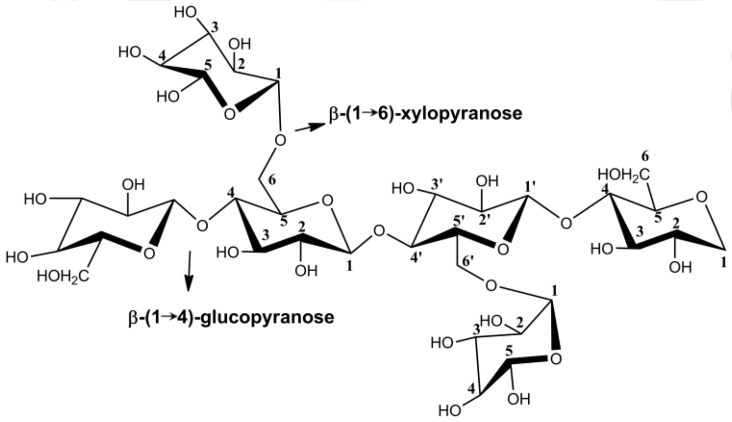


Fig. 1.6. Chemical structure of xyloglucan showing main chain β-(1→4)-glucose units substituted at β-(1→6) position with D-xylopyranose sugars side chains (Brennan and Harris, 2011).

1.2.3.3. Arabinans

Arabinans are widely distributed in the various plant tissues and important structural components of the cell wall. Arabinans are branched homopolymer, composed of α-(1→5)-linked-L-arabinofuranosyl main chain residues to which other L-arabinofuranosyl residues are α-1,3 and α-1,2 linked in either a comb-like or a ramified arrangement (Fig. 1.7) (Pitson *et al*., 1997). Arabinan degrading enzymes are helpful in the food technology, mobilization of plant biomass, nutritional, medical research, plant biochemistry and organic synthesis (Leal and Sá-Nogueira, 2004).

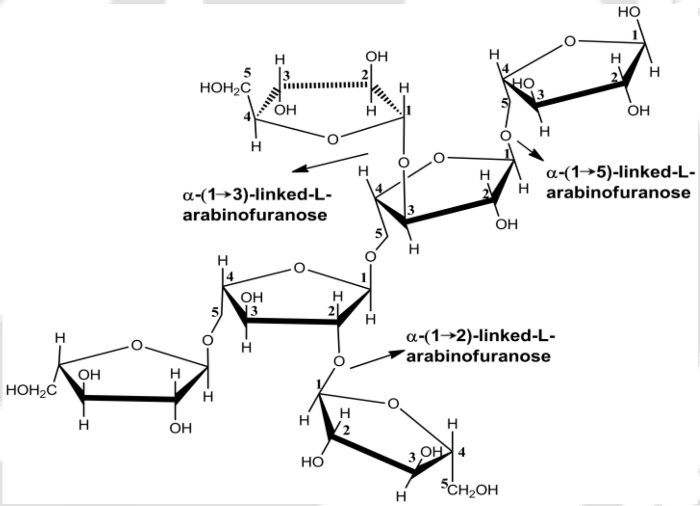


Fig. 1.7. Chemical structure of arabinan showing linear main chain of α-(1→5)-linked L-arabinofuranosyl with substituted with sidechain L-arabinofuranosyl residues with α-(1→2) or α-(1→3) linkages. (Adopted from (Pitson *et al*., 1997).

1.2.3.4. Xylans

Xylan belongs to the second most abundant structural component of the secondary walls of dicot plants. Xylan has a backbone of β-(1→4)-linked xylose residues. In general, xylans are found with numerous degrees of substitution based on the plant tissue types and species. The common side-chain substitutions on the backbone are arabinose, acetic acid, glucuronic acid, ferulic acid, *p*-coumaric acid, or

4-*O*-methyl glucuronic acid (Ordaz-Ortiz and Saulnier, 2005). Arabinose and uronic acid chains stabilize the xylan structure against alkali-catalyzed deterioration. Hemicellulose contains various more prominent substitutions in the chains as an alternative for hydroxyl groups present at positions C2, C3 and C6 (Sjostrom, 2013). Depending on the relative richness of their substitutions within the xylan, it is further categorized into the following types.

1.2.3.4.1. Glucuronoxylan

Glucuronoxylan (GX_n) is polymer of linear β-(1→4) xylosidic bond linked D-xylopyranosyl residues and having a side chain substitution of α-D-glucuronic acid (GlcA) or 4-*O*-methyl α-D-glucuronic acid (MeGlcA) at O2 or O3 position linked by α-(1→2) or α-(1→3) glycosidic bond (Fig. 1.8). The significant amount of GX_n is present in the secondary cell wall of crop residues and hardwood plant. Therefore, it constitutes a major part of plant biomass (Hayashi *et al*., 1996). The xylan backbone of glucuronoxylans at C2 or C3 position in some of the xylose residues is also partially acetylated (Gröndahl et al., 2003). The number of substitutions *viz.* GlcA or 4-OMe-GlcA or acetylation varies depending upon the species and different tissues of plant biomass. The molar ratio of xylose to 4-OMe-GlcA or GlcA may vary between 5 to 20. In most of the glucuronoxylan, the ratio of the β-(1→4)-linked D-xylopyranose and glucuronic acid units is reported to be 10:1, which means on an average every tenth xylose residue of the backbone is substituted with 4-*O*-methyl α-D-glucuronic acid (MeGlcA) residue at C-2 or C-3 position (Hurlbert and Preston, 2001).

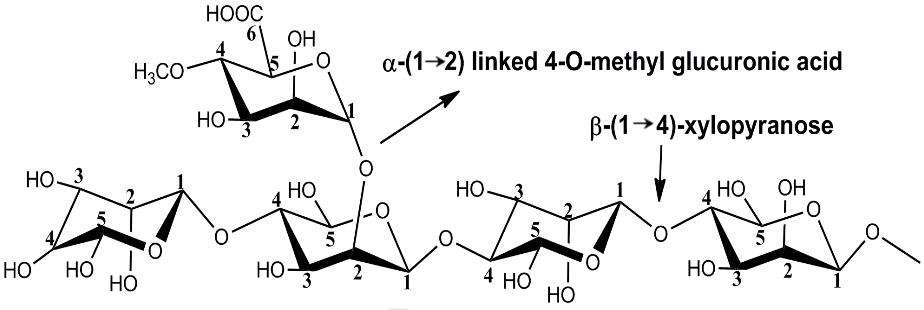


Fig. 1.8. Chemical structure of glucuronoxylan displaying a linear chain of β-(1→4)-linked D-xylopyranose substituted with α-D-glucuronic acid (GlcA) or 4-*O*-methyl α-D-glucuronic acid (MeGlcA) residue at O2 position. (Adopted from Hurlbert and Preston, 2001).

1.2.3.4.2. Arabinoxylans

Arabinoxylan (AX) is an elemental part of the plant cell wall, which is the non-starch polysaccharide and is non-digestible. Arabinoxylans have the β-(1→4)-linked xylose backbone with the substitution of one or more L-arabinofuranosyl units (Fig. 1.9), at O2 or O3 position. AX is located in primary cell walls of grasses and monocot plants cereals like wheat, rye, barley, oat, rice, corn and sorghum (Brett and Waldron, 1990). L-arabinosyl residues are extensively found in these polymers as side chains. AX exhibits various important characteristics like viscosity improvement, the formation of a gel, stabilization of foam, absorption of water, restoration of fat, and prebiotic properties; these properties make it an essential component of the food industry (Izydorczyk and Biliaderis, 1995).

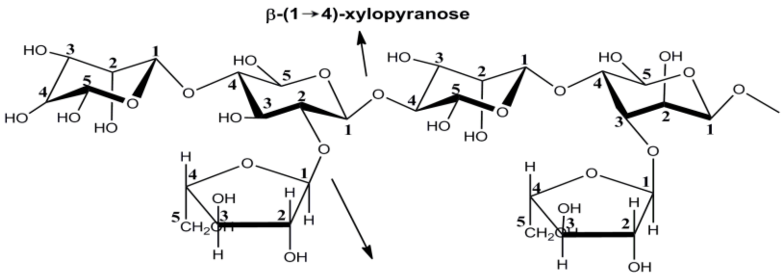


Fig. 1.9. Chemical structure of arabinoxylan displaying the β-(1→4)-linked xylose main chain substituted at O-2 and O-3 with L-arabinofuranosyl residues. (Adopted from Brett and Waldron, 1990).

1.3. Enzymatic degradation of lignocellulosic biomass

The structural complexity and heterogeneity of the polysaccharides present in the cell wall of lignocellulosic biomass make them seemingly recalcitrant and inaccessible for enzyme degradation. The yearly production of the lignocellulosic biomass in nature is increasing year by year, which is raising the environmental concern. The complete degradation of lignocellulosic biomass will make a pool of nutrients available for the recycling. Nature has gifted a diverse group of microorganisms with enzymes to disrupt the plant cell wall polysaccharides (Ochiai et al., 2007). Principally, fungi and bacteria are responsible for the cell wall polysaccharides deterioration. Few insects and mollusks also release enzymes which act on the plant cell wall (Lynd et al., 2002). As mentioned earlier that the major portion of lignocellulosic biomass is composed of complex polysaccharides such as cellulose and hemicelluloses. Their synthesis and complete breakdown into mono sugars and other biochemicals require cumulative action of enzymes and accessory proteins, which are collectively termed carbohydrate-active enzymes (CAZymes). These CAZymes are deposited into the CAZy database (www.cazy.org) and further classified into different families depending on amino acid sequence similarities that reflect structural similarities and conserved catalytic mechanisms, rather than on enzyme specificity (Cantarel et al., 2009). As a consequence of classification system, the enzyme having same enzyme commission (EC) number can be found in multiple families of CAZy database and one family may contain multiple enzymes having different EC numbers (Lombard et al., 2013). These CAZymes have been classified in the CAZy database as Glycosyl Transferases (GTs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs), Auxiliary Activities (AA) and Glycoside Hydrolases (GH). Currently

(September 2019), the CAZy database contains a total 341 families of enzymes (regularly updated), which includes 165 families of GHs, 107 families of GTs, 37 families of PLs, 16 families of CEs and 16 families of AA. Among these five enzyme classes, glycosyltransferases are the enzymes involved in the synthesis of oligosaccharides and polysaccharides, while the rest four-class, *i.e.*, Polysaccharide lyases, carbohydrate esterases, auxiliary activity enzymes and glycoside hydrolases are involved in either the breakdown of polysaccharides or the modification of the polysaccharides.

Glycosyltransferase (EC 2.4.x.y) is the only class, which catalyze the transfer of sugar moieties of activated donor molecules to specific acceptor molecules, forming glycosidic bonds. GTs contributed approx. 42.35% of total CAZymes at present (September 2019) and is virtually present in every single organism.

Polysaccharide lyase (EC: 4.2.2.x) enzymes catalyze non-hydrolytic cleavage of glycosidic bonds of uronic acid-containing polysaccharides using a β-elimination catalytic mechanism. Only 1.5% of total CAZymes are reported as PLs in the CAZy database, but these enzymes have various biotechnological and biomedical applications (Lombard *et al.*, 2010).

The carbohydrate esterase enzymes catalyze the de-O-acylation or de-N-acylation of substituted saccharides by removing the ester-based modifications present in mono-, oligo- and polysaccharides to make it accessible for other hydrolytic enzymes such as PLs and GHs. CEs accounts for 5.10% of total CAZymes.

The auxiliary activities redox enzymes family contains lignin-degrading and lytic polysaccharide mono-oxygenase enzymes. These enzymes catalyze the degradation of lignin and cleavage of mono-oxygen bond from a crystalline polysaccharide. AAs can be used in combination with other CAZymes for the complete degradation of lignocellulosic biomass. These AAs constitute only 0.92% of total CAZymes, but in the last one decade, the utilization of these enzymes in different applications have gained interest.

1.3.1. Glycoside hydrolase

Glycoside hydrolase (GHs), (EC: 3.2.1.x) represent the highest number of enzymes classified and contributed approximately 50.15% of total CAZymes. GHs includes glycosidases which can hydrolyze the glycosidic bond as well as transglycosidases which can transglycosylate the glycosidic bond between two or more monosaccharide moieties or between a monosaccharide and a non-carbohydrate moiety (Henrissat, 1991; Henrissat and Bairoch, 1993, 1996). Till date (September 2019) 664285 enzyme entries are classified into 165 glycoside hydrolase families in the CAZy database (http://www.cazy.org/Glycoside-Hydrolases.html). Some of the families out of 165 families based on conservation of protein structure folds were further categorized into 14 ‘clans or superfamilies.’ GH acts on polysaccharides or oligosaccharides substrates and hydrolyze the glycosidic bonds either within the substrate randomly or at reducing end, thus being termed as endo- or exo-acting enzymes, respectively. The shape or groove of active site of the endo-acting enzymes is present in the more open conformation, which allows random binding of non-crystalline polysaccharide resulted in the production of oligosaccharides (Davies and Henrissat, 1995). While the active site of exo-acting enzymes most often has a pocket or tunnel conformation to bind the reducing end of substrate for the release of disaccharides or monosaccharides (Rouvinen et al., 1990; Divne et al., 1994). Over the years, different types of GHs have been identified, biochemically and structurally characterized. The common feature of several GHs is the modular nature, i.e. they also contain one or more ancillary modules that are often but not always carbohydrate-binding modules (CBMs) (Henrissat and Davies, 2000). Consideration of different modules of particular GHs is the importance of correct open reading frame (ORF) annotation and functional prediction (Bourne and Henrissat, 2001). The relative abundance of the polysaccharides, a group of enzymes, based on their substrate specificities involved in the degradation of these polysaccharides are classified into cellulolytic and hemicellulolytic enzymes.

1.3.1.1. *Cellulolytic enzymes*

Glycoside hydrolases that specifically target cellulosic polysaccharide present in the lignocellulosic biomass are known as cellulolytic enzymes. These cellulolytic enzymes hydrolyze the β-(1,4)-glycosidic linkages in cellulose and are generally named as endoglucanases (EGs), cellobiohydrolases (CBHs) or β-glucosidases. Endoglucanases (EC 3.2.1.4) randomly catalyze the β-(1,4)-glycosidic linkages (internal bonds) in the amorphous regions of the polysaccharides and produces oligosaccharides. Currently, endoglucanases are found in 15 GH families *viz.* 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74, 124 and 148 as single catalytic domain or as modular enzymes. Cellulose hydrolysis by endoglucanases produces one reducing and one non-reducing end, that may be utilized by cellobiohydrolases for further hydrolysis (Rouvinen et al., 1990; Divne et al., 1994). These CBHs are mainly found in family GH5, GH6, GH7, GH9 and GH48. The cellobiose released from CBH mediated hydrolysis acts as a substrate for β-Glucosidases (EC 3.2.1.21) for the conversion of cellobiose into glucose sugar. β-Glucosidases are mainly present in family GH1, GH2, GH3, GH5, GH9, GH16, GH30, GH39 and GH116 (Henrissat et al., 1985).

1.3.1.2. *Hemicellulolytic enzymes*

Diverse and complex hemicellulosic component interacts with the cellulosic component and forms a tight association. Inefficient removal of hemicellulose from the lignocellulosic biomass may limit the efficient and complete degradation of cellulose (Saha et al., 2013). Therefore, the enzymatic saccharification of hemicellulosic components enhances the overall product yield and now attracting more attention. (Hemsworth et al., 2016). Hemicellulolytic enzymes differ in their mode of action on distinct substrates based on their substrate specificity are maybe endo- or exo-acting enzymes. These hemicellulolytic enzymes are classified in different GH families. Endo-acting hemicellulolytic enzymes include endo-β-xylanase and endo-β-mannanases. However, for complete conversion of hemicellulose into monosaccharide and other valuable products, several exo-acting enzymes and some of which primarily act on substitutions as well as on oligosaccharides are required. This susbtitutuon removing enzymes are arabinofuranosidases, arabinanases, galactosidases, glucuronidases, mannosidases and xylosidases. The hemicellulolytic enzymes involved in the complete conversion of heteroxylans are termed as xylanolytic enzymes.

1.4. *Xylanolytic enzymes*

Heteroxylans are highly substituted, heterogeneous and complex in nature. Therefore, its complete conversion requires a group of enzymes with diverse specificities and modes of action and the group of these enzymes are termed as xylanolytic enzymes (Beg et al. 2001). The xylanolytic enzyme system that orchestrates hydrolysis of heteroxylan comprises a collection of hydrolytic enzymes that includes endo-1,4-β-xylanase (E.C.3.2.1.8), xylan-1,4-β-xylosidase (E.C.3.2.1.37), α-glucosiduronase (E.C.3.2.1.139), α-L-arabinofuranosidase (E.C.3.2.1.55) and acetyl xylan esterase (E.C.3.1.1.72) (Juturu and Wu 2012). This xylanolytic enzyme system acts synergistically to convert xylan into its constituent sugars (Belancic et al. 1995). Among all the xylanases, endoxylanases are most important, as they are directly involved in the cleavage of glycosidic bonds, liberating short xylooligosaccharides (Verma and Satayanarayana 2012). Currently, endo-1,4-β-xylanase are found in 16 GH families *viz.* 3, 5, 8, 9, 10, 11, 12, 16, 26, 30, 43, 44, 51, 62, 98 and 141 as single catalytic domain or as modular enzymes. β-Xylosidases releases xylose residues from non-reducing ends of xylooligosaccharides. β-Xylosidases are classified in family 1, 2, 3, 5, 10, 30, 39, 43, 51, 52, 54, 116 and 120 glycoside hydrolases. α-L-arabinofuranosidases catalyze the removal of α-L-arabinofuranosyl residue from the side chain of arabinoxylan. Currently, α-L-arabinofuranosidases are reported from several GH families *viz.* 2, 3, 5, 10, 43, 51, 54 and 62. However, a complete degradation requires synergistic action of acetyl esterase to remove acetyl substituents from β-1,4-linked D-xylose backbone of xylan (Coughlan and Hazlewood 1993).

1.4.1. Family 43 glycoside hydrolases

The family 43 glycoside hydrolase (GH43) includes enzymes with different activities like xylanase (EC 3.2.1.8), β-1,3-xylosidase (EC 3.2.1.-), β-xylosidase (EC 3.2.1.37), exo-β-1,3-galactanase (EC 3.2.1.145), inverting type β-D-galactofuranosidase (EC 3.2.1.146), endo-α-1,5-L-arabinanase (EC 3.2.1.99), exo-α-1,5-L-arabinofuranosidase (EC 3.2.1.-), inverting type exo-α-1,5-L-arabinanase (EC 3.2.1.-), α-1,2-L-arabinofuranosidase (EC 3.2.1.-) and α-L-arabinofuranosidase (EC 3.2.1.55). The enzymes classified in family 43 GHs belong to clan GH-F possessing 5-fold β-propeller structure at its core (http://www.cazy.org/GH43.html). The enzymes reported in this family hydrolyze the polysaccharide *via* inverting type mechanism.

These enzymes require a catalytic triad formed by Aspartic acid, Glutamic acid and Aspartic acid for their activity (Vandermarliere *et al.*, 2009). As of September 2019, the family GH43 enzymes deposited in the CAZy database are, 19 from archaeal origin, 13229 from bacterial and 481 from eukaryotes, 5 from viruses and 28 are unclassified. These family 43 glycoside hydrolases are further subclassified into 37 subfamilies (http://www.cazy.org/GH43.html). Till date, 13 arabinofuranosidases from various organisms reported in GH43_16 subfamily have been biochemically characterized. Out of these, the crystal structure of only 1 modular arabinofuranosidase (*Bs*AXH-m2,3) from *Bacillus subtilis* containing a catalytic module of family GH43 and family 6 carbohydrate-binding module (CBM6) as a non-catalytic module (PDB Id: 3C7E) have been reported and available in PDB (Vandermarliere *et al.*, 2009). The crystal structures of *Bs*AXH-m2,3 displayed a 5 bladed β-propeller fold and the structure is conserved throughout the family. The analysis of ligand-bound crystal structure confirmed that, the binding of the substrate with *Bs*AXH-m2,3 is based on hydrophobic stacking interaction and provided flexibility to cleave the arabinose substitution at O2 or O3 position of arabinoxylan.

1.4.2. *Family 30 glycoside hydrolases*

The family 30 Glycoside hydrolase (GH30) contains a variety of activities which include endo-β-1,4-xylanase (EC 3.2.1.8); β-glucosidase (3.2.1.21); β-glucuronidase (EC 3.2.1.31); β-xylosidase (EC 3.2.1.37); β-fucosidase (EC 3.2.1.38); glucosylceramidase (EC 3.2.1.45); β-1,6-glucanase (EC 3.2.1.75); glucuronoarabinoxylan endo-β-1,4-xylanase (EC 3.2.1.136); endo-β-1,6-galactanase (EC:3.2.1.164); [reducing end] β-xylosidase (EC 3.2.1.-). The family 30 glycoside hydrolases (GH30) belong to the GH-A clan having $(\beta/\alpha)_8$ topology of the catalytic core region and a β-strand rich side β-domain (Urbániková et al., 2011). Family GH30 enzymes also follow the retaining type mechanism, where two glutamic acid residues are involved in the catalysis, one of these residue acts as acid/base and another residue acts as a nucleophile. Till September 2019, the family GH30 entries in CAZy database includes 2376 bacterial and 105 from eukaryotic protein, while only 3 entries are as unclassified. These family 30 glycoside hydrolases are further subclassified into 9 subfamilies (http://www.cazy.org/GH30.html).

1.4.3. Family 10 glycoside hydrolase

The family 10 glycoside hydrolase (GH10) includes enzymes with different activities like endo-1,4-β-xylanase (EC 3.2.1.8), endo-1,3-β-xylanase (EC 3.2.1.32), tomatinase (EC 3.2.1.-), Xylan endotransglycosylase (EC 2.4.2.-) and endo-β-1,4-glucanase (EC 3.2.1.4). The family GH10 contains a total of 3851 protein sequences in the CAZy database as of September 2019 (http://www.cazy.org/GH10.html). Out of 3851 sequences, 19 belong to archaea, 3015 belong to bacteria, 461 belong to eukaryote and 356 are unclassified sequences. The enzymes classified under GH10 family have a higher molecular mass of approximately 40 kDa and lower pI than those reported in other GH families (Beaugrand et al., 2004; Meng et al., 2015). Moreover, GH10 xylanases are capable of hydrolyzing the xylosidic linkage of the highly substituted xylan main chain, near the site, where substitution is present, whereas, GH11 xylanases cannot hydrolyze highly substituted xylan (Valenzuela et al., 2014). The catalytic modules of GH10 family belong to clan GH-A and they possess $(\beta/\alpha)_8$-TIM-barrel structure fold resulting in a 'salad bowl' shape (Pell et al., 2004). As of now (September 2019), 360 xylanases are characterized and out of 360 xylanases, only a few mesophilic bacterial xylanases belonging to family 10 glycoside hydrolase have been identified and characterized [Fontes et al., 2000, Gallardo et al., 2003; Sharma et al., 2018). Out of these 360 characterized xylanases, the crystal structure of 45 xylanases is available in the PDB database. Majority of the crystal structures solved till date are of thermophilic xylanases. The two acidic amino acid, *i.e.* glutamic acid acts as acid/base and nucleophile are involved in the catalysis. The solution structure of xylanases by Small Angle X-ray Scattering (SAXS) analysis has been not reported till date.

1.4.4. Mechanism of action of xylanolytic enzymes

Xylanolytic enzymes usually hydrolyze glycosidic bonds by adding water across the glycosidic bond. Asp or Glu is the most commonly found catalytically active residues, although a few exceptions do exist (Davies *et al*., 2002). There are two fundamental mechanisms of action by which these enzymes hydrolyze the substrates (www.cazy.org). These include i) Retaining mechanism and ii) Inverting mechanism and are explained in the subsections 1.3.4.1 and 1.3.4.2.

1.4.4.1. Retaining mechanism

Xylanolytic enzymes acting by retaining mechanism generally have a distance of approximately 6Å between the nucleophile and catalytic acid/base residues (Zechel and Withers, 2000). The retaining mechanism proceeds *via* a double displacement mechanism (Fig. 1.10). In the first step, the nucleophilic attack occurs at the anomeric center and the glycosidic oxygen immediately protonated by the general acid (Fig 1.10). These occur through a transition state that has oxocarbenium ion character and results in the formation of a glycosyl-enzyme intermediate and results in retention of stereochemistry at the anomeric carbon as shown in Fig. 1.10 (Zechel and Withers, 2000). Hydrolysis of the intermediate occurs and simultaneously, the general base deprotonates an incoming water molecule, which further attacks the anomeric center causing hydrolysis of the glycosyl-enzyme intermediate (Fig 1.10).

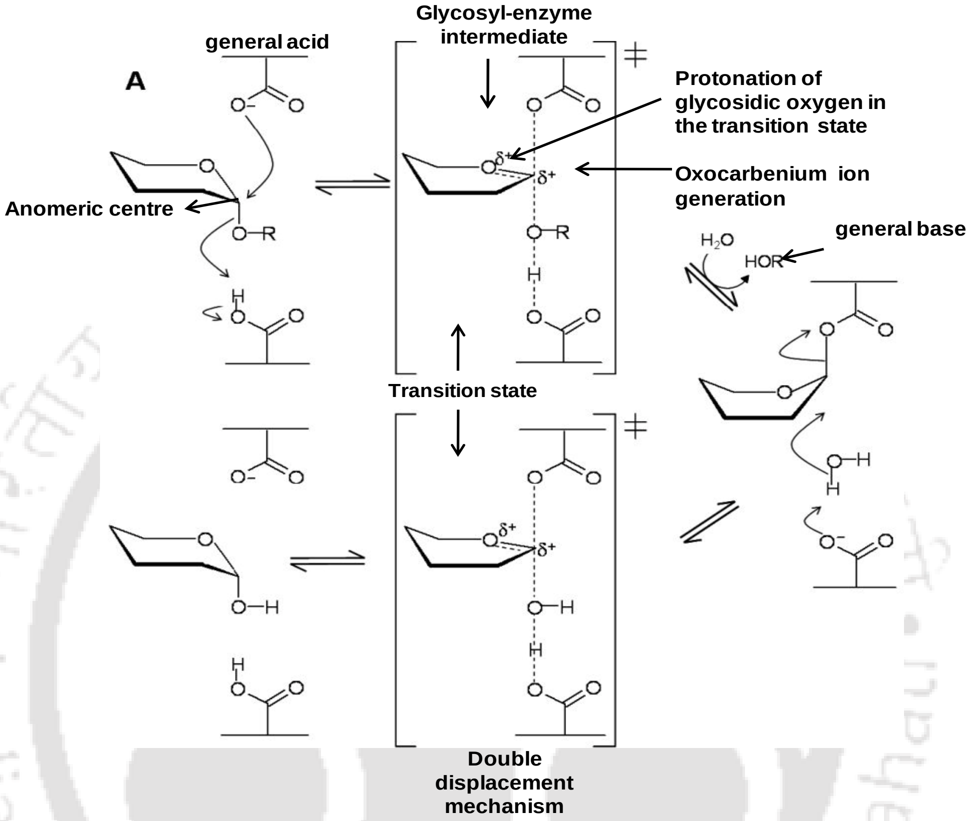


Fig. 1.10 Retaining type mechanisms of action by glycoside hydrolases.

1.4.4.4.2. Inverting mechanism

In case of inverting mechanism, glycoside hydrolases require a larger distance between the general base and general acid residues in order to accommodate the direct attack of a water molecule and the substrate (Zechel and Withers, 2000). This reaction proceeds *via* a single displacement mechanism (Fig. 1.11). Water is deprotonated by the general base and attacks the anomeric center while the general acid concomitantly protonates the leaving group (Fig. 1.11). The inverting mechanism proceeds *via* an oxacarbenium ion-like transition state and results in overall inversion of stereochemistry at the anomeric carbon (Zechel and Withers, 2000).

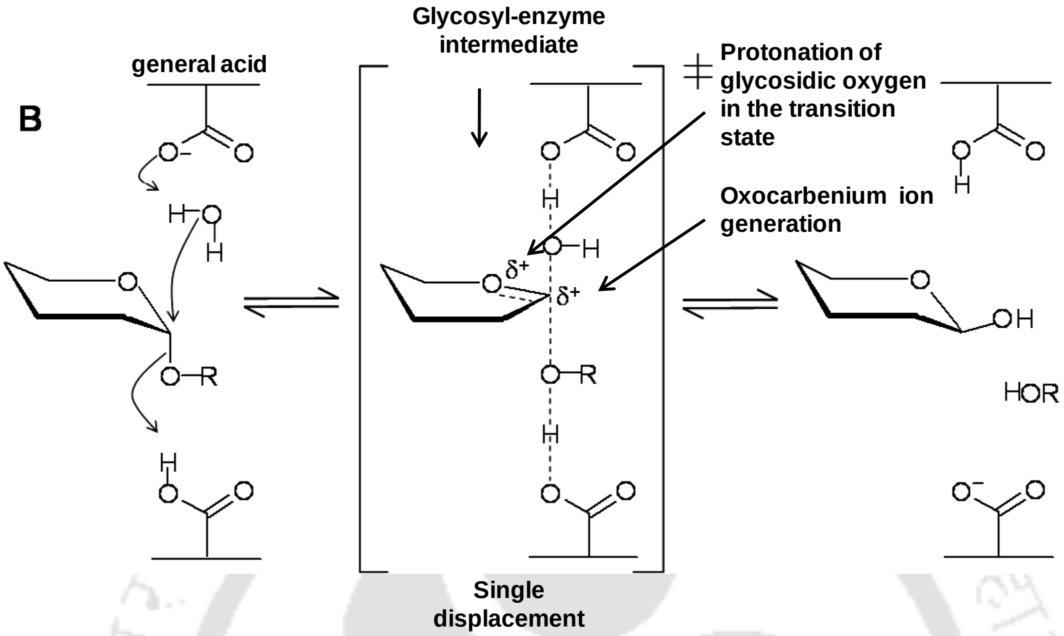


Fig. 1.11 Inverting mechanism of action by glycoside hydrolases.

1.4.5. Industrial production and application of xylanolytic enzymes

According to the Global Market Insights, Inc, the global market for food enzymes is increasing and will achieve USD 3.6 Billion by 2024. Carbohydrate-active enzymes will be on the top and acquire approximately, 50% market with USD 1.8 Billion by 2024. The global xylanolytic enzyme market is anticipated to grow at a compound annual growth rate of 6.6%. The industrial market for xylanolytic enzymes sales are dominated by Alltech Inc. USA, Adisseo France, BASF Enzymes LLC. USA, Novozymes Denmark, Enzyme Development Corporation, New York, USA, Genencor USA, Danisco Denmark, DSM Netherland, Associated British Foods plc UK and Takabio France.

There has been a remarkable increase in the utilization of xylanolytic enzymes in the conversion of lignocellulosic biomass recent years. These Xylanases find applications in various industrial processes such as food and feed, fruit juice clarification, bio-bleaching, textile, detergent and pharmaceutical preparations (Sharma et al., 2019). Similarly, arabinofuranosidases alone or in combination with other enzymes also plays a key role in various applications such as efficient conversion of hemicellulosic biomass to fuels and chemicals, delignification of paper pulp, digestibility enhancement of animal feedstock, production of important medicinal compounds production and the synthesis of oligosaccharides, fruit juice clarification and beer consistency improvement (Thakur et al., 2019). Some of the examples of these enzyme utilization in different processes are listed in Table 1.1.

Table 1.1. Various applications of xylanase and arabinofuranosidase.

Application	Xylanases	Arabinofuranosidase
Food Industry	Dornez et al., 2011.	Gobbetti et al., 2000
i. Bread Making	Guo et al., 2018,	Zhou et al., 2019
ii. Papad Manufacturing	Awalgaonkar et al. 2015,	-
iii. Fruit Juice Clarification	Adiguzel et al., 2019	İlgü et al., 2018
Wine industry	Mihnea et al., 2001	Zietsman et al., 2011
Aroma enhancement	Wang et al., 2016	Ravanal et al., 2012
Feed industry	Café et al., 2002 Zhang et al., 2014	De La Mare et al., 2013 Ravn *et al*., 2018
Paper and Pulp industry	Shrinivas et al., 2010	Numan and Bhosle, 2006
Pharmaceutical Industry	Kapilan, 2015 Ng et al., 2018	Xie et al., 2016 Choi et al., 2018
Bioethanol	Tabka et al., 2006 Thomas et al., 2016	Sørensen et al., 2003 Alvira et al., 2011

1.5. The microorganism

Clostridium thermocellum is an anaerobic, thermophilic, cellulolytic and ethanologenic, Gram-positive bacterium capable of directly converting cellulosic biomass into ethanol (Bayer *et al*., 1983). The TEM image of *C. thermocellum* clearly showed rod-shaped cells having fibrous and protuberant structures on the cell surface (Fig. 1.12)

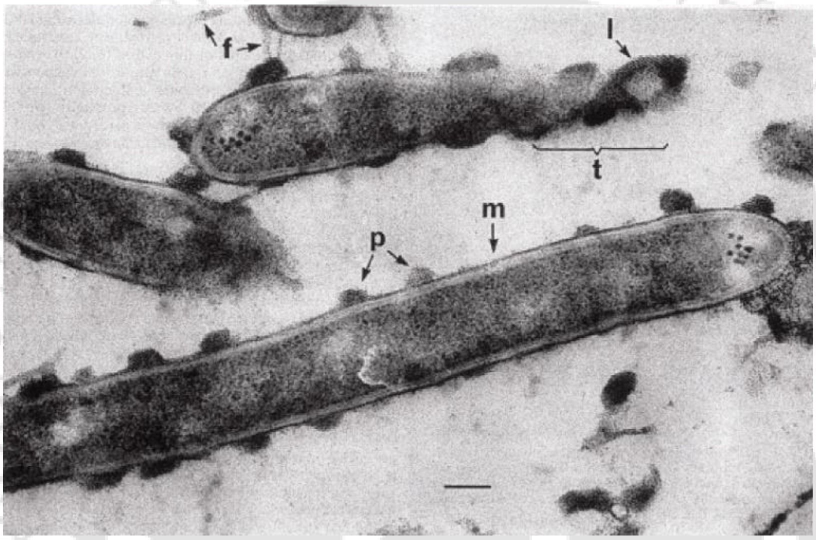


Fig. 1.12. Transmission electron microscope (TEM) image of cationized ferritin (CF) stained *Clostridium thermocellum* YS grew on cellobiose. The three major types of labeling: the monolayer (m) of CF particles enveloping the entire cell surface, the fibrous structures (f) which connecting two adjacent cells and the nodulous protuberances (p) over the entire cell surface. In tangentially sectioned areas of the cell surface (t), there are indications that the protuberances may be interconnected secondarily by low-lying structures (1). Treatment with CF resulted in labeling of the protuberances, and the ferritin particles appeared to be packed very tightly over the entire structure (adapted from Bayer and Lamed, 1986).

Pseudopedobacter saltans is a Gram-negative soil-associated bacterium and one of the currently known thirty-two species in the genus *Pedobacter* within the family *Sphingobacteriaceae*. Like other members of this genus, *P. saltans* is also an heparinolytic organism, which utilizes heparin as their sole carbon and nitrogen source. Cells of *P. saltans* show a peculiar gliding (Fig. 1.13), dancing motility and can be distinguished from other *Pedobacter* strains by their ability to utilize glycerol and the inability to assimilate D-cellobiose (Gallego *et al.*, 2006 and Liolios *et al.*, 2011).

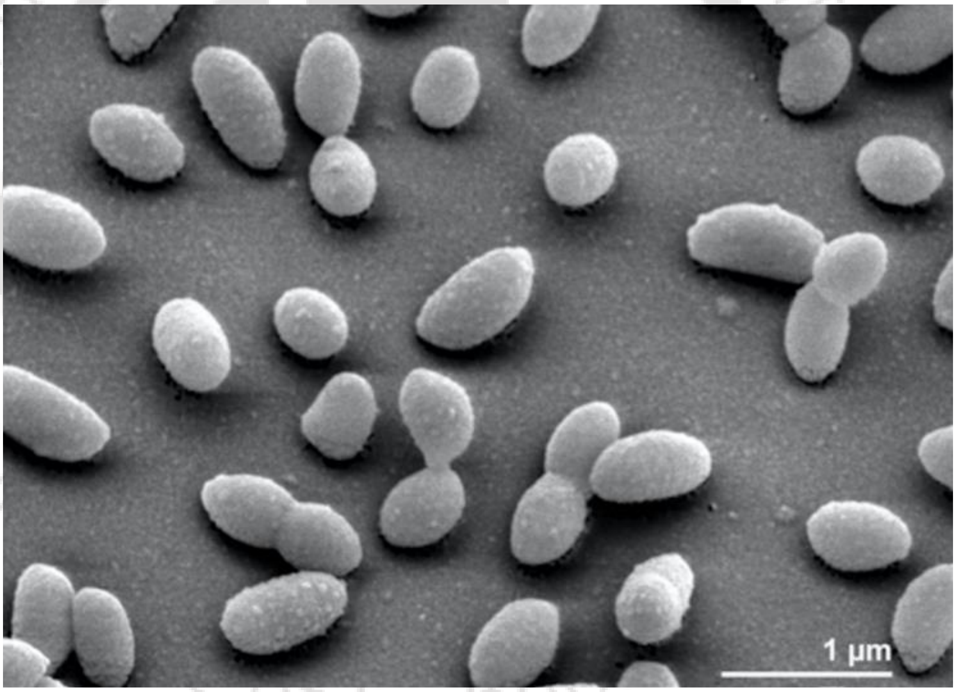


Fig. 1.13. Scanning electron micrograph of *P. saltans* strain 113T (Liolios *et al.*, 2011).

1.6. Significance of the investigation

The increase in the human population, the rapid expansion of civilizations and the depletion of fossil fuels may cause serious food and energy shortage in the near future. Therefore, introducing new technologies for effectively utilizing these abundant, sustainable resources, like plant biomass, must be developed urgently to provide sufficient amounts of animal feed, food, fuel and chemicals. The enzymes, including cellulases and xylanases, capable of degrading plant polysaccharides will play essential roles as catalysts in this process. Therefore, studies on the cellulolytic and xylanolytic enzymes have markedly increased in the last decade. The genome sequence of several plant cell wall degrading organisms and the advancement of available genetic engineering tools have revolutionized the overall approach for identification, screening and characterization of the efficient enzyme candidates. *Clostridium thermocellum* genome contains genes encoding the complex multienzyme complex termed as cellulosome, which is capable of degrading cellulosic and hemicellulosic polysaccharides effectively. Furthermore, the cellulosomal enzyme complex is 50 times more efficient in degrading polysaccharides as compared with non-cellulosomal enzymes, as reported by Fontes and Gilbert (2010). *Pseudopedobacter saltans* is an heparinolytic bacteria but, in its, genome several xylanolytic enzymes such as endo-β,1-4 xylanases encoding genes belonging to families GH10, GH11, GH43 are present. Therefore, in the proposed work, the structure, functional and potential application of xylanolytic enzymes such as endo-β,1-4 xylanase of family 10 glycoside hydrolase from *Pseudopedobacter saltans and* glucuronoxylan hydrolase of family 30 glycoside hydrolase and α-L arabinofuranosidase of family 43 glycoside hydrolase from *Clostridium thermocellum* will be studied. The biochemical characterization of these enzymes was already performed, which confirmed that these enzymes play a key role in the efficient hydrolysis of heteroxylan present in lignocellulosic biomass. However, several questions remain unanswered about the conformational dynamics, molecular arrangement and protein stability of these modular and non-modular enzymes in solution. In this regard, the present study deals with the structural organization, dynamics and function characterization and application of modular and non-modular xylanolytic enzymes from *Clostridium thermocellum* and *Pseudopedobacter saltans.*

1.7. Objectives of the present study

The full-length and truncated modules of xylanolytic enzymes *viz.* α-L arabinofuranosidase (*Ct*Araf43) of family GH43 (GenBank: Acc No. ABN52503), glucuronoxylan hydrolases (*Ct*XynGH30) (GenBank: Acc No. ABN54208) of family 30 glycoside hydrolase from *Clostridium thermocellum* and endo-β,1-4 xylanase (*Ps*GH10A) (GeneBank: Acc No. ADY53246.1) belonging to GH10 family from *Pseudopedobacter saltans* were chosen for the study. The first objective deals with the 3-Dimensional structure determination of *Ct*Abf43, a catalytic module of the full-length *Ct*Araf43 by employing X-ray crystallography. The active site residues of *Ct*Abf43 were identified and confirmed by site-directed mutagenesis. Its substrate specificity towards arabinoxylan was determined by structure comparison with other homologous protein structures. The molecular arrangement and conformational dynamics of full-length modular arabinofuranosidase (*Ct*Araf43) in solution is not known. Therefore, the second objective of the thesis deals with the computational modeling and dynamics in combination with SAXS analysis of the modular *Ct*Araf43 for better understanding about its flexibility, conformational arrangement and the overall molecular organization in solution.

The full-length *Ct*XynGH30 is a modular protein comprising a catalytic module *Ct*Xyn30A and a family 6 carbohydrate binding module, *Ct*CBM6 connected through a linker peptide. The third objective of the thesis deals with molecular organization analysis of *Ct*XynGH30 by protein-protein interaction analysis between truncated derivatives *Ct*Xyn30A and *Ct*CBM6. The conformational dynamics, molecular arrangement and protein stability in solution will be determined by Small Angle X-ray Scattering (SAXS) analysis and computational modeling of the full-length *Ct*XynGH30.

The structural and functional investigation of *Ps*GH10A is important in order to understand the differences as well as similarities among the proteins characterized so far from the same family as well as across the families. Therefore, in the fourth objective of the thesis the structure modeling and docking analysis of *Ps*GH10A with xylo-oligosaccharide ligands will be performed to obtain the necessary information regarding the key amino acids involved in substrate catalysis. Conformational behavior and the thermal stability of the *Ps*GH10A will be determined by SAXS analysis and melting curve analysis.

Extraction and characterization of xylan from neem sawdust will be explored in the last objective. The extracted xylan from neem sawdust will be used for *Ps*GH10A assisted xylooligosaccharides (XOS) production. These XOS will be identified by Thin layer chromatography and their anticancer potential and mode of cell death on human colorectal cancer cell lines will be examined.

1.7.1 Specific objectives

1. Molecular determinants of substrate specificity revealed by x-ray crystallographic structure of *Clostridium thermocellum* arabinofuranosidase 43A from family 43 subfamily glycoside hydrolase.

2. Small-angle X-ray scattering based structure, modeling and molecular dynamics analyses of family 43 glycoside hydrolase α-L-arabinofuranosidase from *Clostridium thermocellum*

3. Molecular organization of glucuronoxylan endo β-1,4-xylanase of family 30 glycoside hydrolase from *Clostridium thermocellum*.

4. SAXS and comparative modeling based structure analysis of endo-β-1,4-xylanase a family 10 glycoside hydrolase from *Pseudopedobacter saltans*.

5. Extraction and characterization of xylan from neem sawdust (*Azadirachta indica)* for the production of anti-cancer xylooligosaccharides.

1.8. References

Adiguzel, G., Faiz, O., Sisecioglu, M., Sari, B., Baltaci, O., Akbulut, S., Genc, B., Adiguzel, A., 2019. A novel endo-β-1,4-xylanase from *Pediococcus acidilactici* GC25; purification, characterization and application in clarification of fruit juices. *Int. J. Biol. Macromol.*, 129, 571-578.

Alvira, P., Negro, M.J., Ballesteros, M., 2011. Effect of endoxylanase and α-L-arabinofuranosidase supplementation on the enzymatic hydrolysis of steam exploded wheat straw. *Bioresour. Technol.*, 102, 4552-4558.

Awalgaonkar, G., Sarkar, S., Bankar, S., Singhal, R.S., 2015. Xylanase as a processing aid for papads, an Indian traditional food based on black gram. *LWT - Food Sci. Technol.*, 62, 1148-1153.

Bayer, E.A., Kenig, R., Lamed, R., 1983. Adherence of *Clostridium thermocellum* to cellulose. *J. Bacteriol.*, 156, 818-827.

Bayer, E.A., Lamed, R., 1986. Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *J. Bacteriol.*, 167, 828-836.

Beaugrand, J., Chambat, G., Wong, V.W.K., Goubet, F., Rémond, C., Paës, G., Benamrouche, S., Debeire, P., O’Donohue, M., Chabbert, B., 2004. Impact and efficiency of GH10 and GH11 thermostable endoxylanases on wheat bran and alkali-extractable arabinoxylans. *Carbohydr. Res.*, 339, 2529-2540.

Beg, Q.K., Kapoor, M., Mahajan, L., Hoondal, G.S., 2001. Microbial xylanases and their industrial application: a review. *Appl. Microbiol. Biotechnol.*, 56, 326-38.

Belancic, A., Scarpa, J., Peirano, A., Díaz, R., Steiner, J., Eyzaguirre, J., 1995. *Penicillium purpurogenum* produces several xylanases: Purification and properties of two of the enzymes. *J. Biotechnol.*, 41, 71-79.

Bourne, Y., Henrissat, B., 2001. Glycoside hydrolases and glycosyltransferases: Families and functional modules. *Curr. Opin. Struct. Biol.*, 11, 593-600.

Brennan, M., Harris, P.J., 2011. Distribution of fucosylated xyloglucans among the walls of different cell types in monocotyledons determined by immunofluorescence microscopy. *Mol. Plant.* 4, 144-156. https://doi.org/10.1093/mp/ssq067

Buckeridge, M.S., 2010. Seed cell wall storage polysaccharides: Models to understand cell wall biosynthesis and degradation. *Plant Physiol.*, 154, 1017-1023.

Café, M.B., Borges, C.A., Fritts, C.A., Waldroup, P.W., 2002. Avizyme improves performance of broilers fed corn-soybean meal-based diets. *J. Appl. Poult. Res.*, 11, 29-33.

Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.*, 37, d233-238.

Choi, J.H., Shin, K.C., Oh, D.K., 2018. An L213A variant of β-glycosidase from *Sulfolobus solfataricus* with increased α-L-arabinofuranosidase activity converts ginsenoside Rc to compound K. *PLoS One.*, 13, e0191018.

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

Coughlan, M., Hazlewood, G., 1993. beta-1,4-D-xylan-degrading enzyme systems: biochemistry, molecular biology and applications. *Biotechnol. Appl. Biochem.*, 17, 259-289.

Cunha, A.G., Gandini, A., 2010. Turning polysaccharides into hydrophobic materials: A critical review. Part 1. Cellulose. *Cellulose.*, 17, 875-889.

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

Davison, B.H., Parks, J., Davis, M.F., Donohoe, B.S., 2013. Plant cell walls: Basics of structure, chemistry, accessibility and the influence on conversion, in: Aqueous pretreatment of plant biomass for biological and chemical conversion to fuels and chemicals. https://doi.org/10.1002/9780470975831.ch3

De La Mare, M., Guais, O., Bonnin, E., Weber, J.Ô., Francois, J.M., 2013. Molecular and biochemical characterization of three GH62 α-l-arabinofuranosidases from the soil deuteromycete *Penicillium funiculosum*. *Enzyme Microb. Technol.*, 53, 351-358.

De Wild, P., Reith, H., Heeres, E., 2011. Biomass pyrolysis for chemicals. *Biofuels.*, 2, 185-208.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K.C., Teeri, T.T., Jones, T.A., 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science.,* 265, 524-528.

Dornez, E., Verjans, P., Arnaut, F., Delcour, J.A., Courtin, C.M., 2011. Use of psychrophilic xylanases provides insight into the xylanase functionality in bread making. *J. Agric. Food Chem.,* 59, 9553-9662.

Duval, A., Lawoko, M., 2014. A review on lignin-based polymeric, micro- and nano-structured materials. *React. Funct. Polym*., 85, 78-96.

Fontes, C.M.G.A., Gilbert, H.J., 2010. Cellulosomes: Highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.,* 79, 655-681.

Fontes, C.M.G.A., Gilbert, H.J., Hazlewood, G.P., Clarke, J.H., Prates, J.A.M., McKie, V.A., Nagy, T., Fernandes, T.H., Ferreira, L.M.A., 2000. A novel *Cellvibrio mixtus* family 10 xylanase that is both intracellular and expressed under non-inducing conditions. *Microbiology.,* 146, 1959-1967.

Gallardo, O., Diaz, P., Pastor, F.I.J., 2003. Characterization of a *Paenibacillus* cell-associated xylanase with high activity on aryl-xylosides: A new subclass of family 10 xylanases. *Appl. Microbiol. Biotechnol.,* 61, 226-233.

Gallego, V., García, M.T., Ventosa, A., 2006. *Pedobacter aquatilis* sp. nov., isolated from drinking water, and emended description of the genus *Pedobacter. Int. J. Syst. Evol. Microbiol.,* 56, 1853-1858.

Gobbetti, M., Lavermicocca, P., Minervini, F., De Angelis, M., Corsetti, A., 2000. Arabinose fermentation by *Lactobacillus plantarum* in sourdough with added pentosans and α-L-arabinofuranosidase: A tool to increase the production of acetic acid. *J. Appl. Microbiol*., 88, 317-324.

Gröndahl, M., Teleman, A., Gatenholm, P., 2003. Effect of acetylation on the material properties of glucuronoxylan from aspen wood. *Carbohydr. Polym*., 52, 359-366.

Guo, Y., Gao, Z., Xu, J., Chang, S., Wu, B., He, B., 2018. A family 30 glucurono-xylanase from *Bacillus subtilis* LC9: Expression, characterization and its application in Chinese bread making. *Int. J. Biol. Macromol*., 117, 377-384.

Harris, D., Stork, J., Debolt, S., 2009. Genetic modification in cellulose-synthase reduces crystallinity and improves biochemical conversion to fermentable sugar. *GCB Bioenergy*., 1, 51-61.

Hayashi, K., Inouhe, M., Aoyagi, C., Nevins, D.J., 1996. Characterization of the isozymes of glucuronoxylan xylanohydrolase in the presence of a native cell wall substrate. *Lett. Appl. Microbiol*., 22, 293-298.

Hemsworth, G.R., Déjean, G., Davies, G.J., Brumer, H., 2016. Learning from microbial strategies for polysaccharide degradation. *Biochem. Soc. Trans*., 44, 94-108.

Henrissat, B., 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280, 309-316.

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

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

Henrissat, B., Davies, G.J., 2000. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol*., 124, 1515-1519.

Henrissat, B., Driguez, H., Viet, C., Schülein, M., 1985. Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Bio/Technology*., 3, 722.

Houston, K., Tucker, M.R., Chowdhury, J., Shirley, N., Little, A., 2016. The plant cell wall: A complex and dynamic structure as revealed by the responses of genes under stress conditions. *Front. Plant Sci*., 7, 984.

Hurlbert, J.C., Preston, J.F., 2001. Functional characterization of a novel xylanase from a corn strain of *Erwinia chrysanthemi. J. Bacteriol*., 183, 2091-2100.

İlgü, H., Sürmeli, Y., Şanlı-Mohamed, G., 2018. A thermophilic α-l-Arabinofuranosidase from *Geobacillus vulcani* GS90: heterologous expression, biochemical characterization, and its synergistic action in fruit juice enrichment. *Eur. Food Res. Technol.*, 244, 1627-1636.

Izydorczyk, M.S., Biliaderis, C.G., 1995. Cereal arabinoxylans: advances in structure and physicochemical properties. *Carbohydr. Polym.*, 28, 33-48.

Kärkäs, M.D., Matsuura, B.S., Monos, T.M., Magallanes, G., Stephenson, C.R.J., 2016. Transition-metal catalyzed valorization of lignin: The key to a sustainable carbon-neutral future. *Org. Biomol. Chem.*, 14, 1853-1914.

Klemm, D., Heublein, B., Fink, H.P., Bohn, A., 2005. Cellulose: Fascinating biopolymer and sustainable raw material. *Angew. Chemie - Int. Ed.*, 44, 3358-3393.

Leal, T.F., Sá-Nogueira, I. De, 2004. Purification, characterization and functional analysis of an endo-arabinanase (AbnA) from *Bacillus subtilis. FEMS Microbiol. Lett.*, 241, 41-48.

Limayem, A., Ricke, S.C., 2012. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Prog. Energy Combust. Sci.*, 38, 449-467.

Liolios, K., Sikorski, J., Lu, M., Nolan, M., Lapidus, A., Lucas, S., Hammon, N., Deshpande, S., Cheng, J.F., Tapia, R., Han, C., Goodwin, L., Pitluck, S., Huntemann, M., Ivanova, N., Pagani, I., Mavromatis, K., Ovchinikova, G., Pati, A., Chen, A., Palaniappan, K., Land, M., Hauser, L., Brambilla, E.M., Kotsyurbenko, O., Rohde, M., Tindall, B.J., Abt, B., Göker, M., Detter, J.C., Woyke, T., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Klenk, H.P., Kyrpides, N.C., 2011. Complete genome sequence of the gliding, heparinolytic *Pedobacter saltans* type strain (113 T). *Stand. Genomic Sci.*, 5, 30.

Liu, Y., Nie, Y., Lu, X., Zhang, X., He, H., Pan, F., Zhou, L., Liu, X., Ji, X., Zhang, S., 2019. Cascade utilization of lignocellulosic biomass to high-value products. *Green Chem.*, 21, 3499-3535.

Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M., Henrissat, B., 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J.*, 432, 437-444.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, 42, D490-D495.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S., 2002. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol. Mol. Biol. Rev.*, 66, 506-577.

Mcqueen-Mason, S., Cosgrove, D.J., 1994. Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. *Proc. Natl. Acad. Sci. U. S. A.*, 91, 6574-6578.

Meier, H., Reid, J.S.G., 1982. Reserve polysaccharides other than starch in higher plants, in: Plant Carbohydrates I. https://doi.org/10.1007/978-3-642-68275-9_11

Meng, D.D., Ying, Y., Chen, X.H., Lu, M., Ning, K., Wang, L.S., Li, F.L., 2015. Distinct roles for carbohydrate-binding modules of glycoside hydrolase 10 (GH10) and GH11 xylanases from *Caldicellulosiruptor sp*. strain F32 in thermostability and catalytic efficiency. *Appl. Environ. Microbiol.*, 81, 2006-2014.

Mihnea, M., José, M.L.G.-S., Velasco-López, M.T., Rivero-Pérez, M.D., Ortega-Heras, M., Pérez-Magarino, S., 2016. Effect of pre-fermentative strategies on the composition of *Prieto Picudo* (Vitis vinífera) Red Wines. *OALib*., 3, 1.

Mussatto, S.I., Dragone, G., Guimarães, P.M.R., Silva, J.P.A., Carneiro, L.M., Roberto, I.C., Vicente, A., Domingues, L., Teixeira, J.A., 2010. Technological trends, global market, and challenges of bio-ethanol production. *Biotechnol. Adv*., 28, 817-830.

Ng, H.S., Chai, C.X.Y., Chow, Y.H., Loh, W.L.C., Yim, H.S., Tan, J.S., Lan, J.C.W., 2018. Direct recovery of *Bacillus subtilis* xylanase from fermentation broth with an alcohol/salt aqueous biphasic system. *J. Biosci. Bioeng*., 125, 585-589.

Nishinari, K., Takemasa, M., Yamatoya, K., Shirakawa, M., 2009. 19 - Xyloglucan BT-Handbook of Hydrocolloids (Second edition), in: Woodhead Publishing Series in Food Science, Technology and Nutrition., 535-566.

Numan, M.T., Bhosle, N.B., 2006. α-L-arabinofuranosidases: The potential applications in biotechnology. *J. Ind. Microbiol. Biotechnol*., 33, 247-260.

Ochiai, A., Itoh, T., Kawamata, A., Hashimoto, W., Murata, K., 2007. Plant cell wall degradation by saprophytic *Bacillus subtilis* strains: Gene clusters responsible for rhamnogalacturonan depolymerization. *Appl. Environ. Microbiol*., 73, 3803-3813.

Ordaz-Ortiz, J.J., Saulnier, L., 2005. Structural variability of arabinoxylans from wheat flour. Comparison of water-extractable and xylanase-extractable arabinoxylans. *J. Cereal Sci.*, 41, 119-125.

Passaretti, S., Franzoni, M., Comin, U., Donzelli, R., Rocca, F., Colombo, E., Ferrara, A., Dinelli, M., Prada, A., Curzio, M., Tittobello, A., 1991. Action of glucomannans on complaints in patients affected with chronic constipation: A multicentric clinical evaluation. *Ital. J. Gastroenterol.*, 23, 421-425.

Pell, G., Taylor, E.J., Gloster, T.M., Turkenburg, J.P., Fontes, C.M.G.A., Ferreira, L.M.A., Nagy, T., Clark, S.J., Davies, G.J., Gilbert, H.J., 2004. The Mechanisms by Which Family 10 Glycoside Hydrolases Bind Decorated Substrates. *J. Biol. Chem.*, 279, 9597-9605.

Pitson, S.M., Voragen, A.G.J., Vincken, J.P., Beldman, G., 1997. Action patterns and mapping of the substrate-binding regions of endo-(1→5)-α-L-arabinanases from *Aspergillus niger* and *Aspergillus aculeatus*. *Carbohydr. Res.*, 303, 207-218.

Popper, Z.A., 2008. Evolution and diversity of green plant cell walls. *Curr. Opin. Plant Biol.*, 11, 286-292.

Popper, Z.A., Fry, S.C., 2003. Primary cell wall composition of bryophytes and charophytes. *Ann. Bot.*, 91, 1-12.

Ravanal, M.C., Rosa, L., Eyzaguirre, J., 2012. α-l-Arabinofuranosidase 3 from *Penicillium purpurogenum* (ABF3): Potential application in the enhancement of wine flavour and heterologous expression of the enzyme. *Food Chem.*, 134, 888-893.

Ravn, J.L., Glitsø, V., Pettersson, D., Ducatelle, R., Van Immerseel, F., Pedersen, N.R., 2018. Combined endo-β-1,4-xylanase and α-L-arabinofuranosidase increases butyrate concentration during broiler cecal fermentation of maize glucurono-arabinoxylan. *Anim. Feed Sci. Technol.*, 236, 159-169.

Rosenau-Tornow, D., Buchholz, P., Riemann, A., Wagner, M., 2009. Assessing the long-term supply risks for mineral raw materials-a combined evaluation of past and future trends. *Resour. Policy.*, 34, 161-175.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C., Jones, T.A., 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science*., 249, 380-386.

S., V.M., Chiranjeevi, P., Dahiya, S., A., N.K., 2018. Waste derived bioeconomy in India: A perspective. *N. Biotechnol.*, 25, 60-69.

Saha, B.C., 2003. Hemicellulose bioconversion, in: *Journal of Industrial Microbiology and Biotechnology*., 30, 279-291.

Salinas, P., Salinas, C., Contreras, R.A., Zuñiga, G.E., Dupree, P., Cardemil, L., 2019. Water deficit and abscisic acid treatments increase the expression of a glucomannan mannosyltransferase gene (GMMT) in *Aloe vera* Burm. *F. Phytochemistry*., 1 306829.

Scheller, H.V., Ulvskov, P., 2010. Hemicelluloses. *Annu. Rev. Plant Biol*., 61, 263-289.

Schopfer, P., 1997. Physiology and biochemistry of plant cell walls. *Plant Sci.*, https://doi.org/10.1016/s0168-9452(96)04565-7

Sharma, K., Antunes, I.L., Rajulapati, V., Goyal, A., 2018. Molecular characterization of a first endo-acting β-1,4-xylanase of family 10 glycoside hydrolase (PsGH10A) from *Pseudopedobacter saltans* comb. nov. *Process Biochem*., 70, 79-89.

Sharma, K., Dhillon, A., Goyal, A., 2016. Insights into Structure and Reaction Mechanism of β-Mannanases. *Curr. Protein Pept. Sci*., 19, 34-47.

Sharma, K., Thakur, A. and Goyal, A., 2019. Xylanases for Food Applications. In Green Bio-processes (pp. 99-118). Springer, Singapore.

Showalter, A.M., 1993. Structure and Function of Plant Cell Wall Proteins. *Plant Cell*., 5, 9. https://doi.org/10.2307/3869424

Shrinivas, D., Savitha, G., Raviranjan, K., Naik, G.R., 2010. A highly thermostable alkaline cellulase-free xylanase from thermoalkalophilic *Bacillus sp*. JB 99 suitable for paper and pulp industry: Purification and characterization. *Appl. Biochem. Biotechnol*., 162, 2049-2057.

Sjöström, E., 1998. Analytical methods in wood chemistry, pulping, and papermaking, Springer Verlag. https://doi.org/10.1007/978-3-662-03898-7

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.

Sørensen, H.R., Meyer, A.S., Pedersen, S., 2003. Enzymatic hydrolysis of water-soluble wheat arabinoxylan. 1. Synergy between α-L-arabinofuranosidases, endo-1,4-β-xylanases, and β-xylosidase activities. *Biotechnol. Bioeng*., 81, 726-731.

Tabka, M.G., Herpoël-Gimbert, I., Monod, F., Asther, M., Sigoillot, J.C., 2006. Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. *Enzyme Microb. Technol*., 39, 897-902.

Ten, E., Vermerris, W., 2013. Functionalized polymers from lignocellulosic biomass: State of the art. *Polymers* (Basel)., 5, 600-642.

Thakur, A., Sharma, K. and Goyal, A., 2019. α-l-Arabinofuranosidase: A Potential Enzyme for the Food Industry. In Green Bio-processes (pp. 229-244). Springer, Singapore.

Thomas, L., Parameswaran, B., Pandey, A., 2016. Hydrolysis of pretreated rice straw by an enzyme cocktail comprising acidic xylanase from *Aspergillus sp*. for bioethanol production. *Renew. Energy*., 1, 9-15.

Urbániková, Ä., Vršanská, M., Mãrkeberg Krogh, K.B.R., Hoff, T., Biely, P., 2011. Structural basis for substrate recognition by *Erwinia chrysanthemi* GH30 glucuronoxylanase. *FEBS J*., 278, 2105-2116.

Valenzuela, S. V., Diaz, P., Pastor, F.I.J., 2014. Xyn11E from *Paenibacillus barcinonensis* BP-23: A LppX-chaperone-dependent xylanase with potential for upgrading paper pulps. *Appl. Microbiol. Biotechnol*., 98, 5949-5957.

Vandermarliere, E., Bourgois, T.M., Winn, M.D., Van Campenhout, S., Volckaert, G., Delcour, J.A., Strelkov, S. V., Rabijns, A., Courtin, C.M., 2009. Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. *Biochem. J*., 418, 39-47.

Vasella, A., Davies, G.J., Böhm, M., 2002. Glycosidase mechanisms. *Curr. Opin. Chem. Biol*., 6, 619-629.

Verma, D., Satyanarayana, T., 2012. Cloning, expression and applicability of thermo-alkali-stable xylanase of *Geobacillus thermoleovorans* in generating xylooligosaccharides from agro-residues. *Bioresour. Technol*., 107, 333-338.

Vuksan, V., Jenkins, D.J.A., Spadafora, P., Sievenpiper, J.L., Owen, R., Vidgen, E., Brighenti, F., Josse, R., Leiter, L.A., Bruce-Thompson, C., 1999. Konjac-mannan (glucomannan) improves glycemia and other associated risk factors for coronary heart disease in type 2 diabetes: A randomized controlled metabolic trial. *Diabetes Care*., 22, 913-919.

Walsh, D.E., Yaghoubian, V., Behforooz, A., 1984. Effect of glucomannan on obese patients: A clinical study. *Int. J. Obes*., 8, 289-293.

Wang, X., Luo, H., Yu, W., Ma, R., You, S., Liu, W., Hou, L., Zheng, F., Xie, X., Yao, B., 2016. A thermostable *Gloeophyllum trabeum* xylanase with potential for the brewing industry. *Food Chem*., 199, 516-523.

Xie, J., Zhao, D., Zhao, L., Pei, J., Xiao, W., Ding, G., Wang, Z., Xu, J., 2016. Characterization of a novel arabinose-tolerant α-L-arabinofuranosidase with high ginsenoside Rc to ginsenoside Rd bioconversion productivity. *J. Appl. Microbiol*., 120, 647-660.

Zechel, D.L., Withers, S.G., 2000. Glycosidase mechanisms: Anatomy of a finely tuned catalyst. *Acc. Chem. Res*., 33, 11-18.

Zhang, L., Xu, J., Lei, L., Jiang, Y., Gao, F., Zhou, G.H., 2014. Effects of xylanase supplementation on growth performance, nutrient digestibility and non-starch polysaccharide degradation in different sections of the gastrointestinal tract of broilers fed wheat-Based diets. *Asian-Australasian J. Anim. Sci*., 27, 855.

Zhou, T., Xue, Y., Zhang, Z., Dong, Y., Gao, R., Li, Y., 2019. Improvement of the Characteristics of Steamed Bread by Supplementation of Recombinant alpha-L-arabinofuranosidase Containing xylan-binding domain. *Food Biotechnol*., 33, 34-53.

Zietsman, A.J.J., De Klerk, D., Van Rensburg, P., 2011. Coexpression of α-l-arabinofuranosidase and β-glucosidase in *Saccharomyces cerevisiae*. *FEMS Yeast Res*., 11, 88-103.

Chapter 2

Molecular determinants of substrate specificity revealed by X-ray crystallographic structure of *Clostridium thermocellum* arabinofuranosidase 43A from glycosyl hydrolase family 43 subfamily 16

2.1. Introduction

Plant cell walls, one of the major reservoirs of organic carbon in nature, comprise a diverse repertoire of chemically complex inter-locking polysaccharides and lignin (Burton *et al.*, 2010). Within an industrial context, this recalcitrant macromolecule is regarded as possessing the significant potential for the generation of value-added chemicals and biofuels (Burton and Fincher, 2014a). In-plant cell walls, cellulose fibers are embedded in a highly entangled cross-linking matrix comprising hemicellulose and pectin (Burton *et al.*, 2010; Ding *et al.*, 2012). In contrast to cellulose, a highly homogeneous polysaccharide composed of β-1,4 linked glucose molecules, hemicellulose and pectin are predominantly heterogeneous polysaccharides in which the recalcitrant backbone is very often decorated with arabinose containing side-chains

(Burton and Fincher, 2014*b*). Arabinoxylans are ubiquitous in the primary and secondary cell walls of monocots and consist of a backbone of β-1,4 linked xylose residues decorated at O-2 and/or O-3 with L-arabinofuranosyl units (Marcotuli *et al.*, 2016). In contrast, the pectic polysaccharide rhamnogalacturonan I, contains a backbone of the repeating disaccharide 4-α-D-galacturonic acid-1,2-α-L-rhamnose-1 extensively decorated with a variety of oligosaccharides, of which arabinan is the most abundant (Park *et al.*, 2015). Arabinan side-chains are found exclusively in pectins and consist of an α-1,5-L-arabinofuranose backbone that is heavily substituted with α-1,2 and/or α-1,3-L-arabinofuranose residues.

The full realization of the biotechnological potential of plant cell walls requires a clear understanding of the biological mechanisms involved in the recycling of structural polysaccharides (Bayer *et al.*, 2007; Ding *et al.*, 2012; Gilbert *et al.*, 2013). Microbial deconstruction of structural polysaccharides involves the production of a highly diverse repertoire of modular enzymes, predominantly glycoside hydrolases (GH) but also carbohydrate esterases (CEs), polysaccharide lyases (PLs) and lytic polysaccharide monooxygenases (LPMOs). These enzymes act cooperatively in the extracellular milieu to deconstruct complex carbohydrates, thus generating simple sugars that are used by microbes as an energy source (Gilbert *et al.*, 2008). Based on primary sequence similarities, GHs has been grouped into structurally uniform families in the regularly updated CAZy database (www.cazy.org) (Cantarel *et al.*, 2009). Enzymes that are active on arabinose-linked glycosidic bonds are predominantly located in families GH43, GH51, GH54 and GH62 (Cantarel *et al.*, 2009). Of these, α-L-arabinofuranosidases (EC 3.2.1.55) are debranching enzymes that cleave decorating arabinose residues from hetero-polysaccharides, such as arabinoxylan and arabinan

(Pitson *et al.*, 1996; Kormelink *et al.*, 1993). These enzymes can either release arabinose exclusively from mono-substituted O-2 or O-3 residues of the polysaccharide backbone or double substituted backbone xylose or arabinose units of arabinoxylans or arabinans, respectively (Vandermarliere *et al.*, 2009).

α-L-arabinofuranosidases are predominantly found in families GH43, GH51, GH54 and GH62. The family GH43, in addition to arabinofuranosidases, also contains arabinanases (EC 3.2.1.99), β-xylosidases (EC 3.2.1.37) and exo-1,3-galactanases (EC 3.2.1.145). A recent seminal study provides the first classification of more than 4500 members of GH43 into 37 subfamilies based on conserved sequence residues and supported by biochemical assays and structural studies (Mewis *et al.*, 2016). Introduction of a subfamily classification in GH43 provides a solid framework to develop informed structure-function studies aiming at elucidating the molecular determinants of substrate specificity in this expanded GH family. In the present study, the structure of *Clostridium thermocellum* arabinofuranosidase *Ct*Abf43A was determined. *Ct*Abf43A is the catalytic domain of a modular protein, termed *Ct*43Araf, that was previously shown to specifically release arabinose side-chains from arabinoxylan, but not from arabinan (Ahmed *et al.*, 2013). This modular protein contains, in addition to the N-terminal *Ct*Abf43A catalytic domain, two C-terminal family 6, Carbohydrate Binding Modules (CBM6s). *Ct*Abf43A belongs to GH43 subfamily 16 (GH43_16), whose members have specialized in the deconstruction of arabinoxylans. Structural analysis of *Ct*Abf43A and its comparison with the members of other related GH43 subfamilies that are specific for removal of arabinose from arabinans allowed the identification of the structural determinants of specificity within GH43_16.

2.2. Material and methods

2.2.1. Material

Commercial screens such as Crystal Screen-1, Crystal Screen-2 and PEG/Ion were purchased from Hampton Research (California, USA), JCSG screen was procured from Molecular Dimensions (UK). All the buffer components and *p*-Nitrophenyl α-L-arabinofuranoside were purchased from Sigma Chemical Co., USA and rye arabinoxylan was acquired from Megazyme International, Ireland.

2.2.2. Sequence retrieval and analysis of modular *Ct*Araf43

The amino acid sequence of full-length α-arabinofuranosidase (*Ct*Araf43) was retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/ABN52503.1/) having a GenBank Accession ID ABN52503.1 (Ahmed et al., 2013a). *Ct*Araf43 domain boundaries were assigned based on results obtained from InterProScan (http://www.ebi.ac.uk/interpro/) server analysis (Zdobnov et al., 2001) and PSI-Blast (http://blast.ncbi.nlm.nih.gov/) (Altschul et al., 1990). The molecular architecture of *Ct*Araf43 was prepared by using IBS (Illustrator for Biological Sequences) Software v2.0 after the comprehensive analysis of sequence by multiple sequence alignment (MSA) and conserved domain database search. The subcellular localization of *Ct*Araf43 was analyzed by using the SignalP v4.0 server (Petersen et al., 2011).

2.2.3. Expression and purification of *Ct*Abf43A

Clostridium thermocellum genome possesses a gene at locus tag Cthe_1271, which encodes a modular family 43 glycoside hydrolase (GH43). The enzyme (accession number ABN52503) contains an N-terminal signal peptide, a GH43 catalytic domain, two tandemly repeated family 6 Carbohydrate Binding Modules (CBM6-1 and

CBM6-2) and a C-terminal dockerin, which enables the incorporation of the enzyme into the high molecular mass multi-enzyme complex produced by the bacterium, termed the cellulosome. The DNA sequence encoding the GH43 catalytic domain of ABN52503 (residues 31-335), termed *Ct*Abf43A, was amplified from the bacterium genomic DNA and cloned into pET28a(+) (Novagen) as described previously (Ahmed *et al.*, 2013), thus generating pAG43. Recombinant *Ct*Abf43A contains an N-terminal His-6 tag. *Escherichia coli* BL21 (DE3) pLysS cells harboring pAG43 were cultured in Luria Bertani medium containing 35 µg/mL chloramphenicol and 50 µg/mL kanamycin at 37ºC to mid-exponential phase (A_{600} 0.5) at which point isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and the cultures were incubated for a further 16 hours at 24ºC. Cells were collected by centrifugation and the cell pellet was resuspended in a 50 mM HEPES buffer, pH 7.5, containing 1 M NaCl, 5 mM $CaCl_2$ and 10 mM imidazole. Recombinant *Ct*Abf43A was purified by Immobilized Metal Affinity Chromatography (IMAC) as described previously (Goyal *et al.*, 2014). To achieve a high degree of protein purity gel filtration was included as a further purification step. The enzyme was buffer exchanged, using a PD-10 desalting column containing Sephadex G-25M resin (GE Healthcare, USA), into 50 mM Hepes buffer, pH 7.5, containing 200 mM NaCl (Buffer A), concentrated to 20 mg/mL with Amicon 10 kDa molecular-weight cutoff centrifugation membranes (Merck Millipore), and subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Healthcare) with protein eluted at 1 mL/min in buffer A. The purified enzyme was concentrated, as described before, washed three times with 0.5 mM $CaCl_2$ using the same centrifugal membranes, and the final protein concentration was adjusted to 60 or 120 mg/mL in 0.5 mM $CaCl_2$. The purified His_6-tagged *Ct*Abf43A enzyme from *C. thermocellum* showed a molecular size of approximately, 36.0 kDa, which was similar to the theoretical molecular size, 35.86 kDa.

2.2.4. Site-directed mutagenesis of *Ct*Abf43A

*Ct*Abf43A mutant derivatives were generated by using the PCR-based NZY Change site-directed mutagenesis kit (NZYtech Ltd, Portugal) using plasmid pAG43 as a template and following the manufacturer’s instructions. The primers used to generate these mutants are listed in Table 2.1. The generated nucleic acids were sequenced to ensure that only the appropriate mutations had been incorporated.

Table 2.1. Details of Primers used for site-directed mutagenesis.

Primer	Primer Sequence
Asp43Ala_F	5´-CAGCGTTTTACTGCAGCCCCAGCCGCAGTCGTG-3´
Asp43Ala_R	5´-CACGACTGCGGCTGGGGCTGCAGTAAAACGCTG-3´
Glu215Ala_F	5´-GCTCCGAGATTTTTCGCGGCAGCATATATGCAT-3´
Glu215Ala_R	5´-ATGCATATATGCTGCCGCGAAAAATCTCGGAGC-3´

2.2.5. Enzyme activity determination

Enzyme activity assays for *Ct*Abf43A, *Ct*Abf43A_Asp43Ala and *Ct*Abf43A_Glu215Ala were performed using *p*-Nitrophenyl α-L-arabinofuranoside as substrate performed as described previously (Ahmed *et al.*, 2013). The specific activity for release of arabinose from arabinoxylan was determined using the Lactose/Galactose Rapid Assay kit (Megazyme~~s~~ International, Ireland) following the manufacturer's instructions. The kit contains β-galactose dehydrogenase, which oxidizes β-D-galactose (using NAD^+) to D-galactonic acid; the amount of NADH formed in this reaction is measured, which is stoichiometric with the amount of D-galactonic acid. The NADH is measured by the increase in absorbance at 340 nm. The kit can also be used to determine the amount of L-arabinose released in a system through the same chemistry.

2.2.6. Crystallization, data collection, structure determination and refinement

Commercial screens such as Crystal Screen, Crystal Screen 2 and PEG/Ion from Hampton Research (California, USA), JCSG from Molecular Dimensions and an in-house-prepared sparse-matrix screen of 80 conditions were screened by the sitting-drop vapor-diffusion method for crystallization of *Ct*Abf43A using an Oryx8 nanodrop robotic dispensing system (Douglas Instruments). Drops consisting of 0.8 µL of either 60 or 120 mg mL^{-1} *Ct*Abf43A solution and 0.8 µL reservoir solution were prepared at 19°C (Goyal *et al.*, 2014). Two forms of crystals of *Ct*Abf43A were obtained: the cube-shaped in 0.1 M sodium acetate, pH 4.5, 2 M ammonium sulfate and the brick-shaped ones in 0.1 M Tris pH 8.5, 20% PEG2000 Monomethyl ether, 0.2 M Trimethyl N-oxide. The data from *Ct*Abf43A crystals were collected on beamline IO4 at the Diamond Light Source (Harwell, England). The data were processed by fastdp and xia2 (Winter, 2010), using the programs MOSFLM (Battye *et al.*, 2011) and AIMLESS (Evans, 2011) from the CCP4 suite (Winn *et al.*, 2011) or XDS (Kabsch, 2010). The resolution for the cube-shaped crystal ($P4_32$) extended to 1.65 Å and that for the brick-shaped crystal ($P2_1$) to 0.97 Å. The structure of the cubic form was solved by molecular replacement using BALBES (Long *et al.*, 2008). The best solution was found using the GH43 arabinoxylan arabinofuranohydrolase from *Bacillus subtilis* (PDB ID: 3C7E, Vandermarliere *et al.*, 2009). This gave a monomer in the space group $P4_32$, with final R_{factor}/R_{free} of 35.9/37.4 and Q-factor of 0.767 after REFMAC5 (Murshudov *et al.*, 2011). PHASER (McCoy *et al.*, 2007) was used to fit the PDB model output from BALBES into the monoclinic 0.97 Å data. Structure refinement was performed with REFMAC5 and model building with COOT (Emsley *et al.*, 2010) using the validation tools within. The statistics for data collection and structure refinement are shown in Table 2.1. Visualization and analysis of structures, as well as the generation of figures, were performed using UCSF Chimera (Pettersen *et al.*, 2004).

2.3. Results and Discussion

2.3.1. Sequence analysis of *Ct*Araf43

*Ct*Araf43 is a modular protein (Fig. 2.1). The N-terminal catalytic module belongs to family 43 glycoside hydrolase, termed as *Ct*Abf43A, and spans from 31-330 amino acids residues. Amino acid residues from 331-463 form a domain classified as family 6 carbohydrate binding modules, termed as *Ct*CBM6A, is tightly attached with the catalytic module by a very short linker peptide. The second CBM6, termed as *Ct*CBM6B, spans from 478-610, is connected to *Ct*CBM6A via a 14 amino acid residues long linker peptide. The SecretomeP 2.0 server analysis of *Ct*Araf43 displayed that the locus tag Cthe_1271, the encoded protein, is an extracellular enzyme. The presence of a type I dockerin in *Ct*Araf43 suggests that the modular enzyme is integrated as a component of the multi-enzyme complex "cellulosome" of *C. thermocellum*.

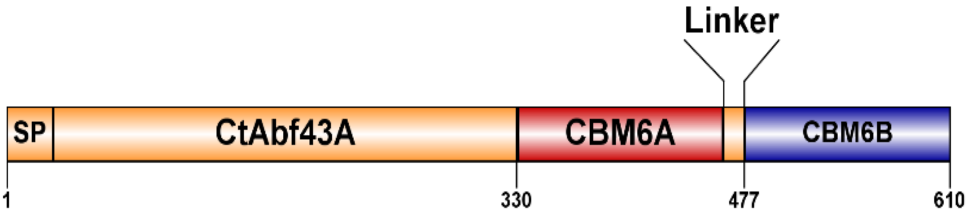


Fig. 2.1. Molecular architecture of *Ct*Araf43 based on the amino acid sequence.

2.3.2. Overall structure of *Ct*Abf43A

The crystal structures of *Ct*Abf43A were solved by molecular replacement to a resolution of 0.97 Å for the monoclinic (PDB ID 5A8C) and 1.65 Å for the cubic form (PDB ID 5A8D). Both structures consist of one molecule of *Ct*Abf43A in the asymmetric unit equating residues 31-329 of the full-length enzyme. The monoclinic *Ct*Abf43A (PDB ID 5A8C) structure contains 564 water molecules, a molecule of the additive, Tris (2-amino-2-hydroxymethyl-propane-1,3-diol, Trs) and a calcium ion. The cubic *Ct*Abf43A (PDB ID 5A8D) structure contains 232 water molecules, five sulfate ions, 2 acetate ions and 4 glycerol molecules. Full crystallographic statistics are provided in Table 2.2. *Ct*Abf43A displays a five-bladed β-propeller fold for both forms (Fig. 2.2A), typical of GH43 and other GH-F clan enzymes (Cantarel *et al.*, 2009). This fold is also observed in two other different GH families, GH families 32 and 68 of the GH-J clan (Cantarel *et al.*, 2009). The topology analysis of *Ct*Abf43A depicts the typical secondary structural elements observed in GH43, which comprise 18/19-β strands with 2 α-helices (Fig. 2.2A). Similar results have also been obtained previously by circular dichroism analysis of *Ct*GH43 (Ahmed *et al.*, 2013). *Ct*Abf43A lacks the C-terminal β-jelly roll domain seen in several GH43 enzymes, that is believed to function as stabilizing module (Fujimoto *et al.*, 2010; Vandermarliere *et al.*, 2009).

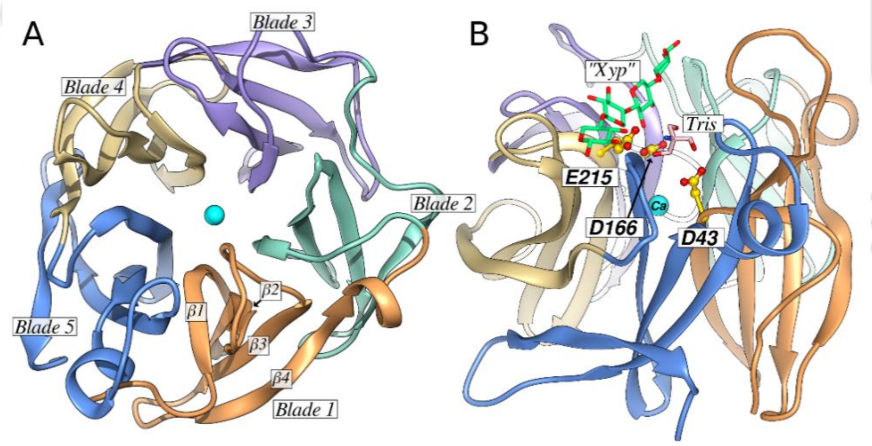


Fig. 2.2. The overall structure of *C. thermocellum* arabinofuranosidase *Ct*Abf43A. A) The secondary structure of the five-bladed β-propeller fold is differentially colored for each blade. The four β-strands in each blade are identified in blade 1. In the central axial cavity, the cyan-colored calcium atom is shown. B) Lateral view highlighting the active site cleft position where a putative tetra xylopyranose is represented (“Xyp” shown in green and based on PDB ID 3C7G) along with a bound Tris molecule (Tris, pink-colored). The three catalytic related residues are shown in a ball & stick representation and yellow-colored.

Table 2.2. X-ray crystallography data collection and refinement statistics for *Ct*Abf43A.

Dataset	*Cubic CtAbf43A*	*Monoclinic CtAbf43A*
Beamline		
Space Group	$P\,4_32$	$P\,2_1$
Wavelength (Å)	*0.9795*	*0.9795*
Unit-cell parameters		
a, b c (Å)	*140.0 140.0 140.0*	*45.4 67.7 51.0*
α, β, γ *(°)*	*90 90 90*	*90 113.1 90*
Resolution limits (Å)	*140 -1.65 (1.709 - 1.65)*	*39.95-0.97 (1.005-0.97)*
No. of observations	*1175673 (31832)*	*236832 (2499)*
No. of unique observations	*56678 (5554)*	*119271 (1368)*
Multiplicity	*20.7 (11.7)*	*2.0 (1.8)*
Completeness (%)	*99.9 (100.0)*	*71.3 (8.2)*
$<I/\sigma(I)>$	*20.0 (1.8)*	*20.5 (2.4)*
CC1/2†	*0.99 (0.67)*	*0.99 (0.90)*
Wilson B-factor	*21.3*	*8.5*
R-merge ‡	*0.09 (1.15)*	*0.02 (0.20)*
R-pim γ	*0.02 (0.36)*	*0.03 (0.28)*
R-work § *, R-free* ¥	*0.17 (0.26); 0.19 (0.28)*	*0.10 (0.21); 0.12 (0.18)*
Ccwork; Ccfree	*0.964; 9.955*	*0.984; 0.979*
No. of Non-H atoms	*2754*	*2986*
Macromolecules	*2465*	*2413*
Ligands	57	*9*
Water	*232*	*564*
Protein residues	*302*	*299*
RMS(bonds)	0.015	0.016
RMS(angles)	*1.81*	*1.80*
Ramachandran favored (%)	*89*	*95*
Ramachandran outliers (%)	*4*	*0*
Clash score	*6.32*	*2.96*
Average B-factor	*34.20*	*12.20*
Macromolecules	*33.30*	*9.70*
Ligands	*57.10*	*6.20*
Solvent	*38*	*23.10*
PDB accession ID	*5A8D*	*5A8C*

† $CC_{1/2}$ = the correlation between intensities from random half-dataset.

‡ $R_{merge} = \Sigma_{hkl}\, \Sigma_i\, |I_i(hkl) - <I(hkl)>| / \Sigma_{hkl}\, \Sigma_i\, I_i(hkl)$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl*, including symmetry-related reflections, and $I(hkl)>$ is its average.

$R_{p.i.m.}=(\sum_{hkl} (\sqrt{(1/(n-1))}\ \sum^{n}_{(j=1)} |I_{hkl,j}-<I_{hkl}>|)) \,/\, (\sum_{hkl}\sum_j I_{hkl,j})$ where $\langle I_{hkl}\rangle$$<I_{hkl}>$ is the average of symmetry-related observations of a unique reflection.

§$Rwork = \Sigma_{hkl}||F_{obs}|-|F_{calc}||/\Sigma_{hkl}|F_{obs}|$.

¥R_{free} as R_{work}, but summed over a 5% test set of reflections

The five blades of the propeller form a cylindrical shape and are organized around a central axial cavity (Fig. 2.2A). This central water-filled tunnel is approximately 33 Å long with a minimum diameter of 9 Å and contains a calcium ion (Fig. 2.3). The calcium ion is located at the middle of the cavity while the catalytic apparatus is positioned at the surface of the protein. The propeller is based upon a 5-fold repeat of blades arranged radially from the center of the propeller and formed by four antiparallel β-strands connected by hairpin turns (Fig. 2.2A). Within each blade, the loops connecting the first and second and the third and fourth strands form short hairpins. In contrast, the loops between the second and third strands and the fourth strand of one blade and the first strand of the next one are considerably longer (>10 residues long) and form the carbohydrate-binding cleft that modulates the substrate specificity of the enzyme (Fig. 2.2B). The four strands that make up each blade twist through approximately 60° from the inside to the outside of the structure. In some GH43 enzymes the fifth blade of the five-bladed β-propeller fold consists of strands from both the N- and C- terminal, providing a "Molecular Velcro" that supports the closure of the β-propeller, thus contributing to a considerable stabilization of the structure (Vandermarliere *et al.*, 2009; Cartmell *et al.*, 2011). As observed in few GH43 enzymes, in *Ct*Abf43A, the fifth strand comprises only β-strands from the C-terminal of the protein and thus the propeller fold is not "closed" by contacts with the first blade (Vandermarliere *et al.*, 2009; Cartmell *et al.*, 2011). However, in 5A8C, the N- and C-terminal β-strands are in close association providing additional stability to the structure.

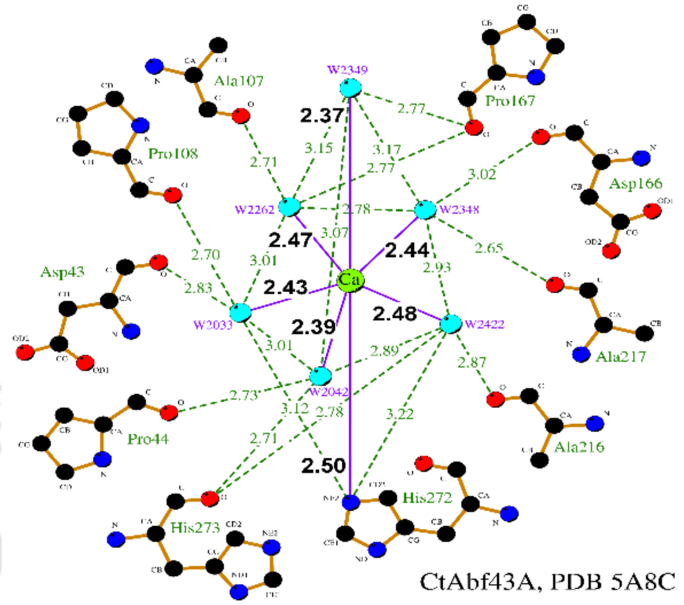


Fig. 2.3. Coordination and interactions of the central calcium ion from *Ct*Abf43A. The pentagonal bi-pyramidal coordination between the calcium (green sphere) with its seven ligands is depicted as solid purple lines (Å distance numbers in black font) while the H-bonds are shown as dashed green lines (Å distance numbers in green font). Water molecules are shown as cyan-colored spheres and labeled with their PDB numbering.

In addition, in some GH43s, such as *Bacillus subtilis* arabinofuranosidase (BsAXH-m2,3), the CBM6 modules located at the C-terminal of GH43 intimately contact the external region of the first blade and the N-terminal region, thus providing an efficient closing of the propeller and significant stabilization to the structure. The full-length *Ct*Abf43A enzyme contains two CBM6 at the C-terminal of the catalytic domain and it is likely that the first CBM6 contributes to the stabilization of the structure through a similar mechanism. A comparison of 5A8C and 5A8D by structural superpositioning revealed that although the two structures have an almost identical core, there is a significant deviation at the N- and C-terminal and in the loops that contribute to the carbohydrate binding. In the absence of the CBM6 modules in *Ct*Abf43A, the N-terminus region from residues 29-42 in 5A8D makes a 180° turn in direction compared to 5A8C, thus occupying the catalytic cleft instead of coming close to the C-terminus. This leads to significant and compensatory movement of some of the loops around the carbohydrate binding cleft. These include the loop between the second and third strand in blade 1, the loop between the fourth strand of blade 4 and first strand of blade 5, and the second helix and the following loop in blade 5. Although 5A8C is likely to be the conformation expected in the full-length protein in which the *Ct*Abf43A is attached to two CBM6 modules and will be used for all further analysis, the 5A8D structure highlights the flexible nature of protein structures, particularly around loop regions and how it may be feasible to engineer different substrate-binding properties around a central β-propeller fold by altering the loop structure. An analysis of structural homologs of *Ct*Abf43A using the SSM site (http://www.ebi.ac.uk/msd-srv/ssm/) identified a large number of GH43 enzymes that displayed significant structural similarity to *Ct*Abf43A, especially in the core five-bladed β-propeller fold. The *B. subtilis Bs*AXH-m2,3 arabinofuranosidase (PDB IDs 3C7E/F/G/H/O; Vandermarliere *et al.*, 2009), with a root mean square deviation (rmsd) of 1.31 Å over 286 Cα atoms and a Z-score of 17.3, the *Cellvibrio japonicus* highly specific α-1,2-L-arabinofuranosidase (PDB IDs 3QED/E/F; Cartmell *et al.*, 2011), with a root mean square deviation (rmsd) of 1.31 Å over 275 Cα atoms and a Z-score of 16.7, and the ruminal GH43 arabinofuranosidase of *Butyrivibrio proteoclasticus* (PDB ID 4NOV; Till *et al.*, 2014), with an rmsd of 1.36 Å over 268 Cα atoms and a Z-score of 16.5, are the closest structural homologs to *Ct*Abf43A. Together with the structure reported here, the *Bacillus* enzyme is the only structure known so far in subfamily GH43_16 while the other two enzymes belong to GH43_29, a subfamily closely related to GH43_16.

2.3.3. Substrate recognition by *Ct*Abf43A

The structure of the closest homolog to *Ct*Abf43A, the subfamily GH43_16 *B. subtilis* arabinofuranosidase, (*Bs*AXH-m2,3) was previously solved in complex with a xylotetraose molecule bound to the substrate-binding cleft (PDB ID 3C7G, Vandermarliere *et al.*, 2009). An alignment of GH43_16 enzymes, including *B. subtilis* arabinofuranosidase, reveals a close homology of enzyme belonging to this subfamily (Fig. 2.4). More information on the mechanism used by *Ct*Abf43A and GH43_16 enzymes for substrate recognition was derived from the superposition of the two structures (Fig. 2.5). A narrow and deep V-shaped cleft, of approximately 35 Ǻ length, extends along the entire dimension of the top of *Ct*Abf43A propeller. The β-1,4 tetrasaccharide is located in this cleft that likely accommodates the arabinoxylan backbone. Following the terminology proposed previously to label the subsites of arabinofuranosidases (Cartmell *et al.*, 2011), the bound ligand occupies subsites +3R, +2R, +1, and +2NR. The arabinose decoration, which binds the backbone xylose at +1, is accommodated at -1 subsite (the active site). However, the side chain sugar of arabinoxylan is not observed in *Bs*AXH-m2,3 (PDB ID 3C7G) structure. *Ct*Abf43A cleft dimensions and the position of the catalytic apparatus suggest that the substrate-binding cleft is unlikely to extend distal to the +3R subsite, although the enzyme may contain at least one additional non-reducing subsite (+3NR).

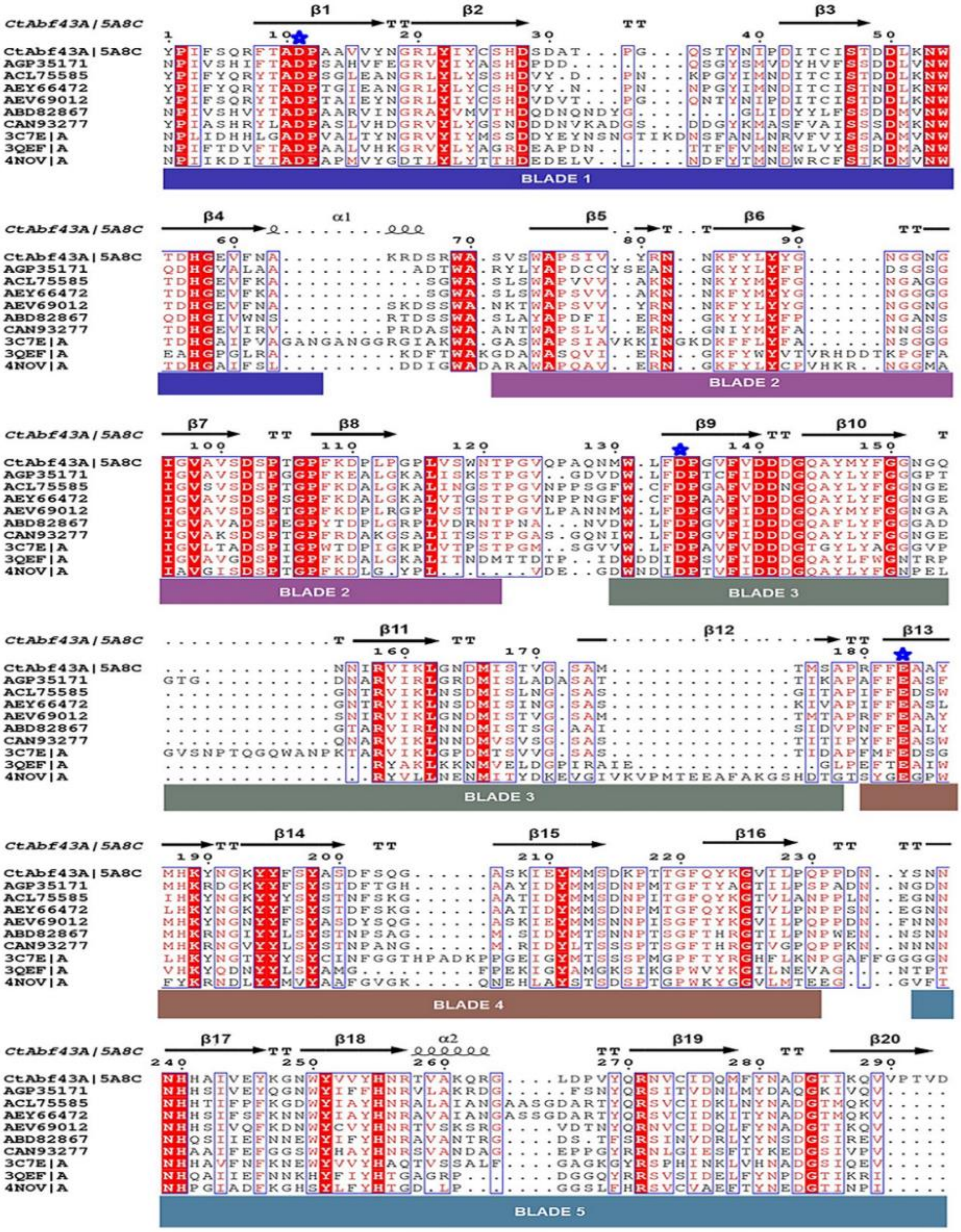


Fig 2.4. Structure-based sequence alignment of *Ct*Abf43A (PDB ID 5A8C) from *C. thermocellum*, with other members of GH43_16 sub-family named by accession numbers and *B. subtilis Bs*AXH-m2,3 arabinofuranosidase (PDB ID 3C7E). The alignment includes the sequences of *Cellvibrio japonicus* highly specific α-1,2-L-arabinofuranosidase (PDB ID 3QEF) and the ruminal GH43 arabinofuranosidase of *Butyrivibrio proteoclasticus* (PDB ID 4NOV), which are members of GH43_29 sub-family. Blue filled stars represent conserved key catalytic residues. The regions contributing to the five structural blades of *Ct*Abf43A are represented.

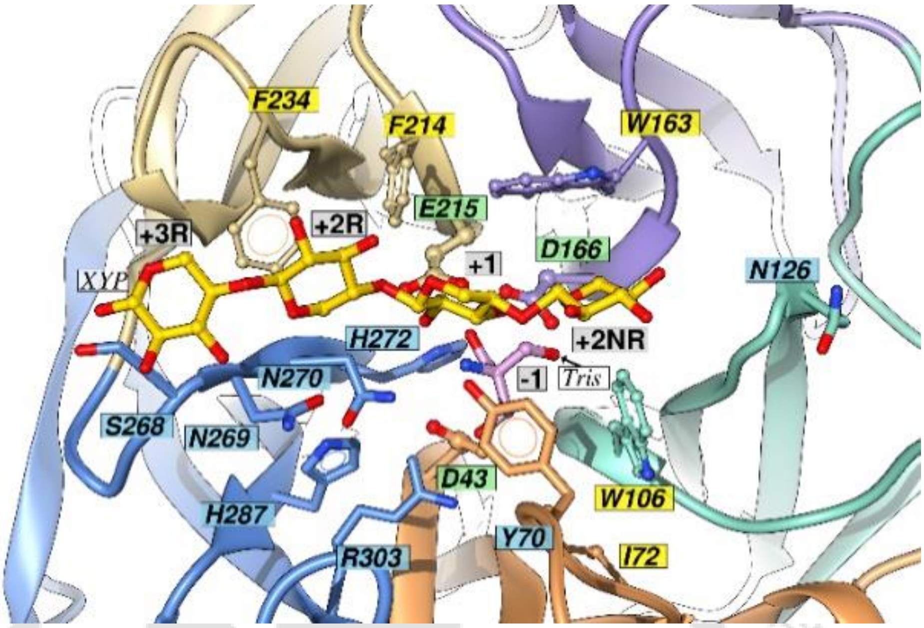


Fig. 2.5. The active site of *Ct*Abf43A with a putative superposed xylotetraose ligand from *B. subtilis Bs*AXH-m2,3 complex (PDB ID 3C7G). The active site subsites are labeled. The β-1,4 tetrasaccharide is depicted in yellow in a ball & stick representation while the Tris molecule in purple was found at subsite -1 in *Ct*Abf43A. The side chains of relevant residues are shown: related catalytic residues in ball & stick representation with green labels; polar contact residues in stick representation with blue labels and hydrophobic/stacking residues in wire representation with yellow labels. XYP: xylotetraose.

Superposition of *Ct*Abf43A with BsAXH-m2,3 in complex with xylotetraose (Vandermarliere *et al.*, 2009) revealed the structural properties of the various subsites of GH43_16 enzymes that modulate arabinoxylan specificity (Fig. 2.5). Thus, at the +3R subsite Ser-268 is predicted to make several hydrogen bonds with xylose O-2 and O-3. At subsite, +2R Phe-234 makes strong hydrophobic interactions with xylose and Asn-269, replaced by Gly-289 in the *Bacillus* enzyme, could make polar contacts with the sugar. At subsite +1, Glu-215, the putative catalytic acid, makes two strong hydrogen bonds with xylose O-2 and O-3, which may comprise the scissile glycosidic bond, and the N$_{\delta 2}$ of Asn-270 is within hydrogen-bonding distance of O-2 of the sugar moiety. Phe-214 provides a hydrophobic environment to stack xylose at subsite +1 and may also create an apolar context that contributes to elevate the pK_a of the acidic residue. Significantly, at subsite +1, *Ct*Abf43A contains a tyrosine residue, Tyr-70, which is not present in *Bs*AXH-m2,3. In the *Bacillus* enzyme this residue is replaced by phenylalanine. The O_η of Tyr-70 is at hydrogen bond distance to O-2 and O-3 of the xylose residue accommodated at subsite +1, the two putative scissile bonds. This suggests an important role of Tyr-70 in the stabilization of the substrate in the proximity of the catalytic site.

Similarly, to subsite +2R, in subsite +2NR binding is solely based in hydrophobic stacking interactions as the xylose residue is sandwiched between the aromatic residues Trp-163 and Trp-106. As described above, *Ct*Abf43A may contain a +3NR subsite that may accommodate the xylose backbone or eventually make productive interactions with it. Although this subsite could be solvent-exposed, $O_{\delta1}$ and $N_{\delta2}$ of Asn-126 could turn the +3NR subsite functional through the establishment of polar contacts. With the exception of the replacement of Tyr-70 in *Ct*Abf43A by Phe-56 in *B. subtilis* (*Bs*AXH-m2,3) at the +1 subsite, and Asn-269 by Gly-289 at the +2R subsite, the residues that participate in recognition of the arabinoxylan backbone are remarkably conserved in the two GH43_16 enzymes. Inspection of the primary sequence of GH43_16 members, suggest that this conservation spans the entire sub-family.

The catalytic pocket that houses the -1 subsite reveals several conserved residues in GH43 enzymes which may contribute to the fine-tuning of substrate binding and transition state stabilization. Thus, Asp-43, the general base, Glu-215, the general acid and Asp-166, the third catalytic residue that is believed to modulate the pK_a of the catalytic acid as well as orientating the catalytic acid and the substrate, are conserved in GH43 enzymes (Vandermarliere *et al.*, 2009; Cartmell *et al.*, 2011). These residues are located on the most internal strands of blades 1, 4 and 3, respectively and are conserved in GH43 family (Fig. 2.4). As the catalytic base activates a water molecule, which attacks the anomeric C atom from one side, the catalytic acid donates a proton to the leaving group on the other side, cleaving the glycosidic bond between the two sugar moieties and leading to an inversion of the configuration of the anomeric C atom. The catalytic role of Asp-43 and Glu-215 was confirmed by the observation that *Ct*Abf43A mutant derivatives Asp43Ala and Glu215Ala are inactive. The inability of the two mutant derivatives to perform catalysis was confirmed by assaying the enzymes with both natural and artificial substrates and by measuring product release by spectrophotometrically. Trp-106 and Ile-72 are highly conserved in several GH43 enzymes and provide the hydrophobic platform that stacks the sugar rings at the -1 subsite. In addition, Arg-303, His-287 and His-272 are mostly unchanged in GH43 enzymes and may also contribute polar contacts with the arabinose housed at the enzyme active site. Also, owing to their basic nature, these three residues may contribute to decrease the pK_a of the catalytic base, Asp-43. The structure of *Ct*Abf43A revealed that the -1 subsite houses a molecule of 2-amino-2-hydroxymethyl-propane-1,3-diol that is most likely occupying the position of arabinose during catalysis. A close inspection of *Ct*Abf43A -1 subsite revealed that this molecule establishes several hydrogen bonds with the catalytic residues Asp-43, Asp-166 and Glu-215, but also with Arg-303, Tyr-70 and His-272, confirming the putative involvement of these later amino acids in substrate recognition (Fig. 2.6).

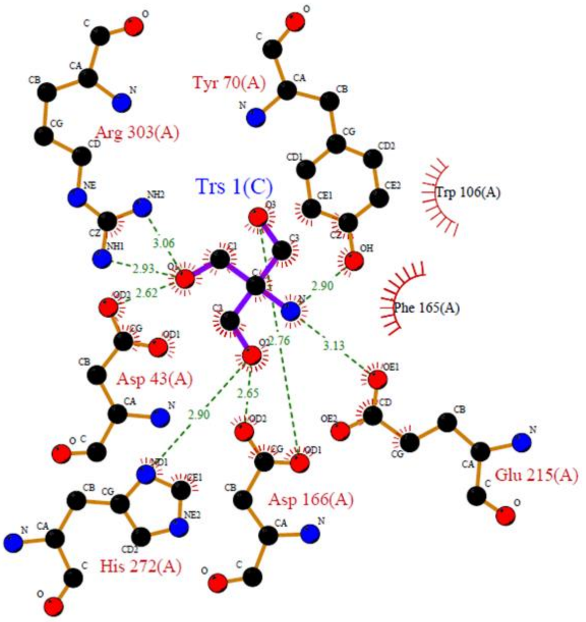


Fig. 2.6. Interactions established between a molecule of Tris, identified at subsite -1, and *Ct*Abf43A.

Recently, it was suggested that His-272 is involved in catalysis via an electron-withdrawing pathway between the divalent calcium ion and the catalytic acid Glu215, thus helping to modulate the pK_a of the acidic residue (Till *et al.*, 2014). The N∈2 of His-272 of *Ct*Abf43A participates in the coordination of the central calcium ion (Fig. 2.3). In *Ct*Abf43A, $N_{\delta 1}$ of His-272 is hydrogen-bonded to the Tris molecule found at the -1 subsite which subsequently binds to $O_{\delta 1}$ of Glu-215. This electron-withdrawing pathway observed in *Ct*Abf43A structure supports the suggestion that the calcium ion and His-272 are positioned in a way that could exert an effect on the pK_a of Glu-215 (Till *et al.*, 2014). *Ct*Abf43A calcium ion interacts with seven ligands in pentagonal bi-pyramidal coordination previously observed in other GH43 enzymes. The axial positions are occupied by $N_{\epsilon 2}$ of His-272 described above and one water molecule. The five equatorial ligands are all water molecules and are positioned midway between two of the blades of the propeller allowing each water molecule to hydrogen-bond to either proline or the preceding residue located in the loops. In *Ct*Abf43A these prolines are located near the catalytic residues. One of these prolines, Pro-108, is located at the end of the loop that terminates at the first β-strand of the first blade located next to Asp-43, the general base. Pro-108 participates in hydrogen bonding to one of the equatorial waters and belongs to the highly conserved motif "WAP" of GH43 domains (Vandermarliere *et al.*, 2009), of which the tryptophan, Trp-106, was suggested to form hydrophobic stacking interactions with the substrate.

2.3.4. Structural determinants of specificity in GH43_16

At present, the structure of two members of GH43_16 subfamily are known; the structure of *Ct*Abf43A reported here and the structure of *Bs*AXH-m2,3 that was previously described (Vandermarliere *et al.*, 2007, 2009). The two enzymes act specifically on the removal of arabinose side-chains from arabinoxylans, but not from arabinans (Bourgois *et al.*, 2007; Ahmed *et al.*, 2013). Biochemical analysis of two other GH43_16 enzymes also revealed a strict specificity to arabinoxylans and no debranching activity to arabinans (Sakka *et al.*, 2012; Valls *et al.*, 2016), suggesting that GH43_16 subfamily comprises enzymes that specialize in arabinoxylan degradation. The superposition of *Ct*Abf43A and *Bs*AXH-m2,3 structures (Fig. 2.7A), revealed that loops connecting β-strands 2 and 3 of blades 1, 3 and 4 are considerably longer in *Bs*AXH-m2,3. Similarly, the loop connecting β-strand 4 of blade 1 to β-strand 1 of blade 2 is also longer in the *Bacillus* enzyme. The functional significance of the extended loops observed in *Bs*AXH-m2,3 remains unknown. However, it is possible that the higher thermostability revealed by *C. thermocellum* enzyme requires a more compact protein structure that is afforded by shorter hairpins. In contrast, the topology of the substrate-binding cleft is similar in the two enzymes.

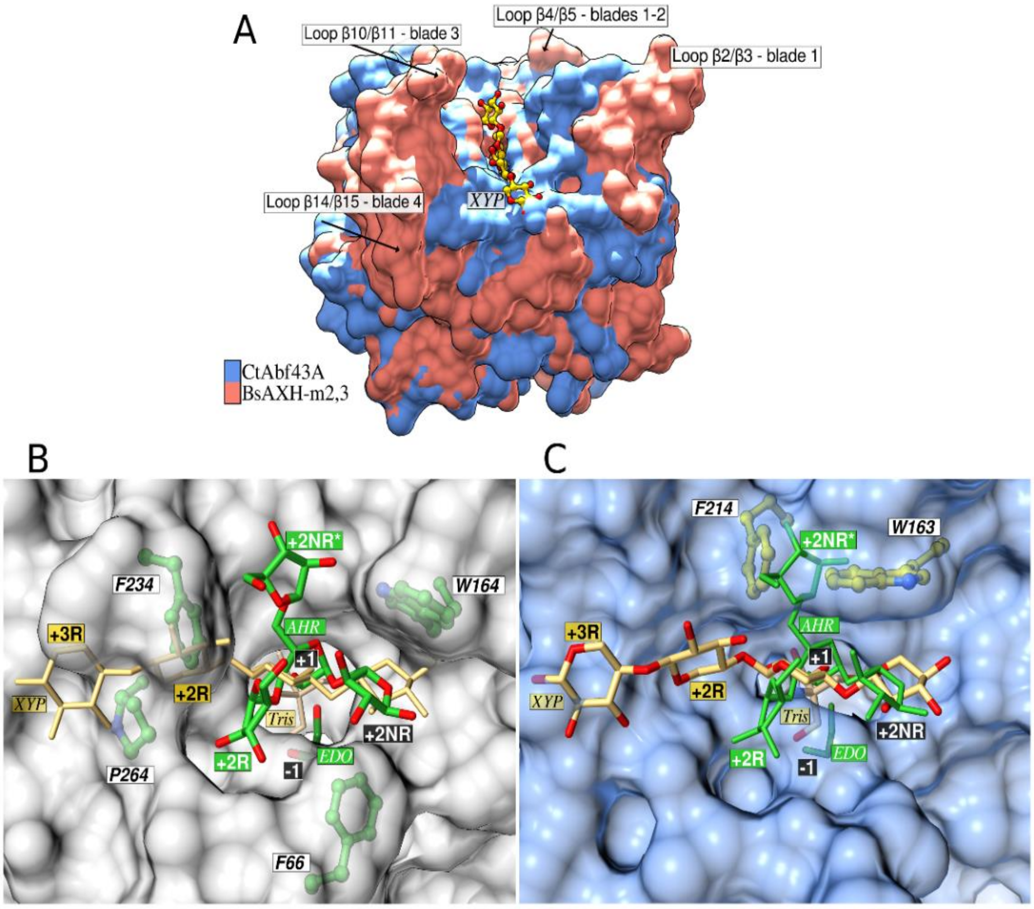


Fig 2.7. Structural determinants of specificity in GH43_16. A) Landscape topography loop differences between superposed *Ct*Abf43A (PDB ID 5A8C) and *Bs*AXH-m2,3 (PDB ID 3C7G). The molecular surface overviewing the substrate-binding cleft of both structures, blue and salmon-colored respectively, exposes significant loop differences (labeled for *Bs*AXH-m2,3) in both members of subfamily GH43_16. In panels B) and C) GH43_16 *Ct*Abf43A (PDB ID 5A8C) is superposed with GH43_29 *Cj*Abf43A (PDB ID 3QEF). B) The molecular surface of *Cj*Abf43A in transparent grey reveals the location of the green ball & stick relevant residues that contribute to its curved shaped cleft, overlayed with a green stick representation of an α-1,5-L-arabinofuranose arabinan backbone (AHR) and a dark green stick ethylene glycol (EDO). C) The molecular surface of *Ct*Abf43A is shown in transparent blue showing in yellow the ball & stick relevant residues for ligand specificity. The location of the GH43_16 ligands is overlayed in yellow stick representation: a molecule of Tris and a putative xylotetraose ligand (XYP) from *Bs*AXH-m2,3 (PDB ID 3C7G). Subsites are labeled with a green background for GH43_29, yellow background for GH43_16 and a black background for similarly placed subsites on both subfamilies. Occlusion of subsites +2R onwards for linear arabinoxylan is evident on GH43_29 (panel B) and conversely for curved arabinan on GH43_16 (panel C). XYP: xylotetraose, AHR: α-1,5-L-arabinofuranose, EDO: ethylene glycol.

The GH43 subfamily that is phylogenetically more proximate to GH43_16 is GH43_29. One member of subfamily GH43_29, *Cellvibrio japonicus* arabinofuranosidase 43A, *Cj*Abf43A, was shown to be strictly specific for the removal of arabinose side-chains of sugar beet arabinan while not attacking arabinoxylan decorations (Cartmell *et al.*, 2011). To identify the structural determinants of specificity within GH43_16 and GH43_29 subfamilies, the structure of GH43_16 arabinoxylan specific enzymes were overlayed with the arabinan specific *Cj*Abf43A (PDB ID 3QEF) (Fig. 2.7B and 7C). As described previously (Cartmell *et al.*, 2011), *Cj*Abf43A Phe-66, Trp-164, and Phe-234 contribute to the formation of a curved surface cleft at the -1 subsite, a topology that reflects the extended helical structure of the α-1,5-L-arabinofuranose backbone of arabinan (Fig. 2.7B and 7C). The GH43_29 enzyme is unable to interact with arabinoxylan, which presents a more linear and less curved structure than arabinan, as *Cj*Abf43A Phe-234 and Pro-264 will not allow the accommodation of the arabinoxylan backbone at subsite +2R as observed in GH43_16. The side chains of these two residues of *Cj*Abf43A occlude the +2R subsite, which results in the inactivation of the remaining subsites located in the reducing sugar direction. Thus, reducing-end positive subsites of *Cj*Abf43A are shifted by 90º concerning the GH43_16 binding cleft. In GH43_16 enzymes the +2NR* subsite of *Cj*Abf43A is occupied by the side-chains of *Ct*Abf43ATrp-163 and Phe-214, explaining why GH43_16 enzymes are unable to attack the arabinan backbone. As proposed before, the architecture of the surface substrate-binding cleft, curved in *Cj*Abf43A and linear in GH43_16 enzymes, modulate the substrate specificity of these enzymes (Cartmell et al., 2011). Although GH43_16 family members may present a strict specificity to arabinoxylan, recent structure-function studies in a second member of the

GH43_29 subfamily, Xsa43A from *Butyrivibrio proteoclasticus* (Till *et al*., 2014), suggest that occlusion of the +2R, which in *Cj*Abf43A binds the arabinan backbone, limits the capacity of the enzyme to interact with this substrate, thus suggesting that GH43_29 subfamily is polyspecific. Interestingly, inspection of Xsa43A structure explains why the enzyme preferentially binds arabinoxylooligosaccharides rather than arabinoxylans. The +3R subsite that is functional in GH43_16 enzymes is occluded by the extended loop that connects β-strands 2 and 3 at blade 5 of Xsa43A, in particular by the side-chains of phenylalanine and leucine.

2.4. Conclusion

With more than 4500 members, predominantly of bacterial origin, GH43 is one of the largest CAZy families (www.cazy.org). The major reported activities of GH43 are β-D-xylosidase (EC 3.2.1.37), α-L-arabinofuranosidase (EC 3.2.1.55), endo-α-L-arabinanase (EC3.2.1.99), and 1,3-β-galactosidase (EC 3.2.1.145). Recently GH43 was divided into subfamilies based on primary sequence homology and informed by biochemical data, providing an opportunity to develop insightful structure-function studies within GH43. GH43 subfamily 16 enzymes (GH43_16) are α-L-arabinofuranosidases that specifically act on arabinoxylans but not on arabinan. The structure of *C. thermocellum Ct*Abf43A reported here together with the structure of a second GH43_16 member reveals the presence of a highly conserved substrate binding cleft in subfamily GH43_16 that reflects the topology of arabinoxylan. GH43_16 is thus unable to bind the arabinan backbone that presents a curved shape. This work presents novel information that contributes to understanding the molecular determinants of substrate binding specificity within one of the most biotechnological relevant CAZy families, GH43.

2.5. References

Ahmed, S., Luis, A.S., Bras, J.L.A., Ghosh, A., Gautam, S., Gupta, M.N., Fontes, C.M.G.A., Goyal, A., 2013. A Novel α-L-Arabinofuranosidase of Family 43 Glycoside Hydrolase (*Ct*43Araf) from *Clostridium thermocellum*. *PLoS One* 8, 1–10.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W., 2011. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 271–281

Bayer, E.A., Lamed, R., Himmel, M.E., 2007. The potential of cellulases and cellulosomes for cellulosic waste management. *Curr. Opin. Biotechnol.* 18, 237–245.

Bourgois, T.M., Van Craeyveld, V., Van Campenhout, S., Courtin, C.M., Delcour, J.A., Robben, J., Volckaert, G., 2007. Recombinant expression and characterization of XynD from *Bacillus subtilis* subsp. subtilis ATCC 6051: A GH 43 arabinoxylan arabinofuranohydrolase. *Appl. Microbiol. Biotechnol.* 75, 1309–1317.

Burton, R.A., Fincher, G.B., 2014a. Evolution and development of cell walls in cereal grains. *Front. Plant Sci.* 5, 456.

Burton, R.A., Fincher, G.B., 2014b. Plant cell wall engineering: Applications in biofuel production and improved human health. *Curr. Opin. Biotechnol.* 26, 79–84.

Burton, R.A., Gidley, M.J., Fincher, G.B., 2010. Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat. Chem. Biol.* 6, 724–732.

Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233-238.

Cartmell, A., McKee, L.S., Pena, M.J., Larsbrink, J., Brumer, H., Kaneko, S., Ichinose, H., Lewis, R.J., Viksø-Nielsen, A., Gilbert, H.J., Marles-Wright, J., 2011. The structure and function of an arabinan-specific α-1,2- arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *J. Biol. Chem.* 286, 15483–15495.

Ding, S.Y., Liu, Y.S., Zeng, Y., Himmel, M.E., Baker, J.O., Bayer, E.A., 2012. How does plant cell wall nanoscale architecture correlate with enzymatic digestibility? *Science* 338, 1055–1060.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 486–501.

Evans, P.R., 2011. An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 282–292.

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

Gilbert, H.J., Knox, J.P., Boraston, A.B., 2013. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* 23, 669–677.

Gilbert, H.J., Stålbrand, H., Brumer, H., 2008. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr. Opin. Plant Biol.* 11, 338–348.

Goyal, A., Ahmed, S., Fontes, C.M.G.A., Najmudin, S., 2014. Crystallization and preliminary X-ray crystallographic analysis of a novel α-l-arabinofuranosidase (CtGH43) from *Clostridium thermocellum* ATCC 27405. *Acta Crystallogr. Sect. F Structural Biol. Commun.* 70, 616–618

Kabsch, W., 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr*. 66, 125–132.

Kormelink, F.J.M., Gruppen, H., Voragen, A.G.J., 1993. Mode of action of (1→4)-β-d-arabinoxylan arabinofuranohydrolase (AXH) and α-l-arabinofuranosidases on alkali-extractable wheat-flour arabinoxylan. *Carbohydr. Res.* 249, 345–353.

Long, F., Vagin, A.A., Young, P., Murshudov, G.N., 2007. BALBES: A molecular-replacement pipeline, in: *Acta Crystallogr. Sect. D Biol. Crystallogr*. 64, 125–132.

Marcotuli, I., Hsieh, Y.S.Y., Lahnstein, J., Yap, K., Burton, R.A., Blanco, A., Fincher, G.B., Gadaleta, A., 2016. Structural Variation and Content of Arabinoxylans in Endosperm and Bran of Durum Wheat (*Triticum turgidum* L.). *J. Agric. Food Chem.* 64, 2883–2892.

Mewis, K., Lenfant, N., Lombard, V., Henrissat, B., 2016. Dividing the large glycoside hydrolase family 43 into subfamilies: A motivation for detailed enzyme characterization. *Appl. Environ. Microbiol.* 82, 1686–1692.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 355–367.

Park, J.M., Jang, M.U., Oh, G.W., Lee, E.H., Kang, J.H., Song, Y.B., Han, N.S., Kim, T.J., 2014. Synergistic action modes of arabinan degradation by exo-and endo-arabinosyl hydrolases. *J. Microbiol. Biotechnol.* 25, 227–233.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.

Pitson, S.M., Voragen, A.G.J., Beldman, G., 1996. Stereochemical course of hydrolysis catalyzed by arabinofuranosyl hydrolases. *FEBS Lett.* 398, 7–11.

Sakka, M., Tachino, S., Katsuzaki, H., van Dyk, J.S., Pletschke, B.I., Kimura, T., Sakka, K., 2012. Characterization of Xyn30A and Axh43A of *Bacillus licheniformis* SVD1 identified by its genomic analysis. *Enzyme Microb. Technol.* 51, 193–199.

Till, M., Goldstone, D., Card, G., Attwood, G.T., Moon, C.D., Arcus, V.L., 2014. Structural analysis of the GH43 enzyme Xsa43E from *Butyrivibrio proteoclasticus*. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* 70, 1193–1198.

Valls, A., Diaz, P., Pastor, F.I.J., Valenzuela, S. V., 2016. A newly discovered arabinoxylan-specific arabinofuranohydrolase. Synergistic action with xylanases from different glycosyl hydrolase families. *Appl. Microbiol. Biotechnol.* 100, 1743–1751.

Vandermarliere, E., Bourgois, T.M., Van Campenhout, S., Strelkov, S. V., Volckaert, G., Delcour, J.A., Courtin, C.M., Rabijns, A., 2007. Crystallization and preliminary X-ray analysis of an arabinoxylan arabinofuranohydrolase from *Bacillus subtilis*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 63, 692–694.

Vandermarliere, E., Bourgois, T.M., Winn, M.D., van Campenhout, S., Volckaert, G., Delcour, J.A., Strelkov, S.V., Rabijns, A., Courtin, C.M., 2009. Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. *Biochem. J*. 418, 39-47.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr*. 67, 235–242.

Winter, G., 2010. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr*. 43, 186–190.s

Chapter 3

Small angle X-ray scattering based structure, modeling and molecular dynamics analyses of family 43 glycoside hydrolase α-L-arabinofuranosidase from *Clostridium thermocellum*

3.1. Introduction

Large quantities of agricultural by-products such as stover, husk, straw and bagasse are produced after harvesting and processing of cereals (Saleh et al., 2017). These agro-waste materials can be utilized as a sustainable source for biorefineries for the production of fine biochemicals, biofuels and prebiotic oligosaccharides. The agro-wastes are economically cheaper, easily accessible and available in large quantities, which make them suitable for the utilization in various industrial applications (Harris et al., 2010). The agro-wastes primarily comprise three primary polymers: cellulose, hemicellulose and lignin, which form a complex structure and are available as a lignocellulosic matrix. Hemicellulosic components of agro-waste account for 20-40% of the total biomass and are the second most abundant renewable organic carbon on earth (Thakur et al., 2019). Heteroxylans are one of the most significant components of hemicellulose, which is mainly composed of β-1,4 linked xylose residues in the backbone and often substituted at O-2 by 4-O-Methyl-D-glucuronic acid and at O-2 or O-3 by L-arabinose sugar (Zhong et al., 2014). The complete deconstruction of heteroxylan requires an array of plant cell wall degrading enzymes with different substrate specificities such as endo-β-1,4-xylanases, β-1,4-xylosidases, acetyl xylan esterases, α-glucuronidases and α-L-arabinofuranosidases (Lagaert et al., 2014). α-L-Arabinofuranosidase (EC 3.2.1.55) catalyzes the release of α-1,2-, α-1,3- or α-1,5-linked L-arabinose at the non-reducing end of arabinose containing oligosaccharides or polysaccharides (Saha, 2000) through an exolytic mode of action. α-L-arabinofuranosidases display different types of substrate specificities and are subclassified into three types (Beldman et al., 1997). Type A preferentially hydrolyzes *p*-nitrophenyl-α-L-arabinofuranoside or arabinose-oligosaccharides (Beldman et al., 1997). Type B enzymes attack branched polymers to release L-arabinose. Type C enzymes are active only against arabinosidic linkages present in arabinoxylans and are inactive on synthetic *p*NP glycosides.

α-L-arabinofuranosidases are classified into families 3, 43, 51, 54, 62, 93 and 155 of glycoside hydrolases based on primary sequence similarities (http://www.cazy.org/). The family 43 and 62 glycoside hydrolases (GH43 and GH62) are classified in Clan-F and both share a five-bladed β-propeller structural fold (Contesini et al., 2017). GH43 family contains enzymes with various substrate specificities such as α-L-arabinofuranosidase (EC 3.2.1.55), α-L-arabinanase (EC 3.2.1.99), endo-β-1,4-xylanase (EC 3.2.1.8), β-D-xylosidase (EC 3.2.1.37) and β-1,3-galactanase (EC 3.2.1.145). Several GH43 enzymes from different origins and displaying different substrate specificities have been biochemically and structurally characterized

(Vandermarliere et al., 2009; Hemsworth et al., 2016; Goyal et al., 2016; Farro et al., 2018). GHs are often multi-modular and can contain one or more catalytic modules linked to one or more auxiliary carbohydrate-binding module(s) (CBMs) by flexible linker peptides (Boraston et. al., 2002). CBMs direct the associated catalytic module(s), in the correct orientation, to the target substrate thereby enhancing the catalytic efficacy (Shoseyov, et al., 2006). The modular enzymes containing catalytic modules such as cellulases, glucosidases, xylanases, xylosidases and arabinofuranosidase interact through protein-protein interaction between cohesin and dockerin modules to form a high molecular mass multienzyme complex termed as cellulosome (Fontes et al., 2010). This cellulosomal complex coordinates the degradation of recalcitrant carbohydrates and releases the mono sugar for the metabolic conversion into ethanol or various other value-added chemicals through anaerobic fermentation.

The cellulosome of *Clostridium thermocellum* possesses a set of modular heteroxylan degrading enzymes belonging to families GH10, GH11, GH30, GH39 and GH43. α-L-arabinofuranosidase of family 43 glycoside hydrolases from *C. thermocellum* (*Ct*Araf43) is a complex modular enzyme consisting of an N-terminal catalytic module, termed *Ct*Abf43A, along with two C-terminal family 6 carbohydrate-binding modules, *Ct*CBM6A and *Ct*CBM6B and a C-terminal dockerin module (Ahmed et al., 2013a). *Ct*Araf43 hydrolyzes arabinose substituted xylan based polysaccharides such as rye and wheat arabinoxylans and oat spelt xylan (Ahmed et al., 2013a). Therefore, it can be used alone or in combination with other xylan degrading enzymes for different biotechnological applications such as production of bioethanol and arabinose free xylooligosaccharides. The structural and functional characterization of *Ct*Abf43A (Goyal et al., 2016), *Ct*CBM6A (Ahmed et al., 2013b) and *Ct*CBM6B

(Ahmed et al., 2013c) have been reported. The catalytic module *Ct*Abf43A displays a five-bladed β-propeller structural fold topology starting at Ala15 and extending to the residue Leu299 (Goyal et al., 2016). The structures of *Ct*CBM6A and *Ct*CBM6B revealed a jelly roll β-sandwich fold and their biochemical characterization *displayed higher affinity towards insoluble oat spelt xylan followed by* arabinose containing wheat arabinoxylans and rye arabinoxylans (Ahmed et al., 2013b; Ahmed et al., 2013c; Ahmed et al., 2013d). *Ct*Araf43 displayed activity against rye arabinoxylan, wheat arabinoxylan and also on synthetic substrates such as *p*-nitrophenol-*α*-L-arabinofuranoside and *p*-nitrophenol-*α*-L-arabinopyranoside (Ahmed et al., 2013a). The thermal stability of *Ct*Araf43 by melting curve analysis revealed *Ct*Abf43A and noncatalytic CBMs (*Ct*CBM6A-*Ct*CBM6B) are melting independently (Ahmed et al., 2013a). The molecular arrangement and conformational dynamics of full-length modular arabinofuranosidase (*Ct*Araf43) in solution is not known. Therefore, to understand the conformational flexibility, dynamics and the overall molecular organization of *Ct*Araf43, the computational modeling, molecular dynamics and SAXS analysis of the modular enzyme were performed.

3.2. Material and methods

3.2.1. Sequence retrieval and analysis of modular *Ct*Araf43

The amino acid sequence of full-length α-arabinofuranosidase (*Ct*Araf43) was retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/ABN52503.1/) having a GenBank Accession ID ABN52503.1 (Ahmed et al., 2013a). *Ct*Araf43 domain boundaries were assigned based on results obtained from InterProScan (http://www.ebi.ac.uk/interpro/) server analysis (Zdobnov et al., 2001) and PSI-Blast (http://blast.ncbi.nlm.nih.gov/) (Altschul et al., 1990). The molecular architecture of *Ct*Araf43 was prepared by using IBS (Illustrator for Biological Sequences) Software v2.0 after the comprehensive analysis of sequence by multiple sequence alignment (MSA) and conserved domain database search. The subcellular localization of *Ct*Araf43 was analyzed by using the SignalP v4.0 server (Petersen et al., 2011).

3.2.2. Molecular modeling and validation of *Ct*Araf43

A 3-dimensional model for CtAraf43 was generated by using comparative modeling approach. PSI-BLAST search of the amino acid sequence of CtAraf43 (580 aa long out of 671 aa as the N-terminal Signal peptide and C-terminal dockerin is not included in this study) was performed using default parameters against a Protein Data Bank (PDB) database. The modular Arabinoxylan arabinofuranohydrolase from Bacillus subtilis (PDB: 3C7G, Vandermarliere et al., 2009) was found closest to CtAraf43 with 94% query coverage and 42% sequence identity so was used as a reference. The modular structure of 3C7G is arranged as GH43-CBM6. The crystal structure of the catalytic module, CtAbf43A from Clostridium thermocellum of a modular CtAraf43 is available at high resolution (PDB: 5A8C at 0.97 Å) (Goyal et al.

2016). In order to enable the modeling of the complete structure, the homologous 3-dimensional structures of the CtCBM6A and CtCBM6B were identified by PSI-BLAST analysis. The full-length CtAraf43 structure with C-terminal His_6-tag was modeled by MODELLER v9.18 using the crystal structures of CtAbf43A (PDB: 5A8C, Goyal et al., 2016), *Clostridium cellulolyticum* CBM6 (PDB: 2V4V, sequence identity 57%, Abbott et al., 2009) for CtCBM6A and CtCBM6 of xylanase U (PDB: 1GMM, sequence identity 69%, Czjzek et al., 2001) for CtCBM6B as the templates (Webb and Sali, 2016) (Fig. *3*.1). The modeled structure of CtAraf43 was validated using Programs available online at SAVES Server v5.0 (**http://servicesn.mbi.ucla.edu/SAVES/**) and Ramachandran plot that was analyzed by using Rampage server (**http://mordred.bioc.cam.ac.uk/~rapper/rampage.php**).

3.2.3. Molecular dynamics simulation of the *Ct*Araf43 structure

The conformational dynamics and stability of *Ct*Araf43 modeled structure were analyzed by performing molecular dynamic (MD) simulation for 100 nanoseconds. GROMOS96 43a1 forcefield compiled with GROMACS v5.1.4 software program on a supercomputer (Param-Ishan) available at Indian Institute of Technology Guwahati, India was used. *Ct*Araf43 modeled structure was placed in a cubic box with dimensions of 6.50x7.70x9.74 nm and volume 968.99 nm^3 and the box was solvated with single point charge (SPC) waters. The MD simulation system containing *Ct*Araf43was neutralized by the addition of 11 Na^+ ions as counter ions. The *Ct*Araf43 modeled structure was energy minimized by using the Steepest descent method using 20,000 iteration steps and cut-off up to 1000 kJ mol^{-1}. The MD simulation system contained 586 amino acid residues involved in *Ct*Araf43 tertiary structure formation, water molecule and 11 Na^+ ions. The non-protein molecules were equilibrated by performing of NVT ensemble followed by NPT ensemble equilibration steps for 500 ps each. The final MD simulation run was performed using equilibrated MD simulation system for 100 ns and the trajectories at 10 ps interval were written. The root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) for the MD simulated system were calculated by using gmx rms and gmx rmsf command. The solvent accessible surface area (SASA), intramolecular hydrogen bonds and radius of gyration (R_g) were determined by using gmx sasa, gmx hbond and gmx gyrate respectively. Change in the secondary structure of *Ct*Araf43 was performed by DSSP v3.0 (Touw et al., 2014) compiled with Gromacs v5.14. All plots were analyzed using XMGRACE software (http://plasma-gate.weizmann.ac.il/Grace/).

3.2.4. Expression and purification of *Ct*Araf43

The recombinant plasmids pET21a-*Ct*Araf43 containing hexa-histidine tag at C-terminal was expressed in *Escherichia coli* BL21 (DE3), as reported in the previous study (Ahmed et al., 2013). The BL21 (DE3) cells containing pET21a-*Ct*Araf43 plasmid were aerobically grown in 100 µg/mL ampicillin supplemented 400 mL LB medium by incubating at 37°C and 180 rpm for 3 hours (Ahmed et al., 2013a). The cell growth was checked for its mid-exponential phase by using UV-visible spectrophotometer (A_{600} = 0.6) and further cooled to 24°C, after that the culture was induced by adding IPTG (isopropyl-β-D-thiogalactopyranoside) to a final concentration of 1 mM and further incubated at 24°C and 180 rpm for 16-18 hours for enzyme production. The cells were harvested by centrifugation at 6000*g* at 4°C for 10 min. The cell pellet was washed by deionized autoclaved water and then resuspended in 5 mL of 20 mM sodium phosphate buffer (pH 7.4), containing 250 mM NaCl, 50 mM Imidazole and serine protease inhibitor, phenylmethylsulfonyl fluoride at a final concentration of

1 mM. The resuspended cells were lysed by ultra-sonication (Sonics, Vibra cell) for 8 min (5s on and 15s off pulse; 33% amplitude) for cell lysis. The cell-free extract was obtained by centrifuging the cell lysate at 18,000g, at 4°C for 45 min followed by filtration through 0.45 µm membrane. *Ct*Araf43 purification was performed by using immobilized metal ion affinity chromatography (IMAC) using 5 mL Sepharose column (HisTrap Chelating, GE Healthcare, USA) as reported by Ahmed et al., (2013). In the second step of purification by gel filtration 2 mL fraction containing 25 mg/mL of *Ct*Araf43 was loaded on Superdex 75 column (HiLoad 16/60, GE Healthcare) and eluted by using 20 mM sodium phosphate buffer (pH 6.0) at a flow rate of 1.0 mL/min connected to FPLC (AKTA purifier chromatographic system (GE Healthcare). The fractions 50-55 mL were collected and analyzed by SDS-PAGE using (10%, w/v) gel for purity followed by pooling the fractions containing pure enzyme. The retention time of *Ct*Araf43 matched well with the retention time of albumin protein from bovine serum. The pooled fractions were concentrated using the centrifugal filter (Amicon Ultra, 3 kDa molecular weight cut-off) by centrifuging at 4000g and 4°C for 30 min to obtain the final concentration of 15 mg/mL.

3.2.5. Small Angle X-ray Scattering (SAXS) of *Ct*Araf43

Scattering profiles for *Ct*Araf43 were collected at the BM29-BioSAXS beamline at ESRF (Grenoble, France) as reported by Sharma et al., (2019). The scattering pattern was recorded at a wavelength of 0.992 Å with the sample to detector distance of 2.81 meters by using Pilatus 1M detector (Dectris), to cover the scattering vector range q of 0.03 to 4.48 nm^{-1}. The *Ct*Araf43 samples and matched buffer were centrifuged at 20,000 g for 30 min, at 4°C followed by filtration using 0.2 µm syringe filters before the measurement to prevent the interference by any aggregation and precipitation. The sample temperature was maintained at 4ºC by a Peltier system. 10 successive frames of 10s each were collected for each sample to check for beam stability and radiation damage. The scattering analysis of *Ct*Araf43 was recorded at several protein concentrations between 0.15 and 9.0 mg/mL to confirm any inter-particle interaction or aggregation effect, which may have been present, could be accounted for and to rule out their influence on the analysis. The scattering pattern of matched buffer (20 mM sodium phosphate pH 6.0) was also recorded in the same sample holder. The scattering pattern of matched buffer was subtracted from all protein samples' scattering profile to obtain the absolute scattering profile of the protein.

3.2.6. SAXS data analysis

The data treatment, intensity normalization to absolute scale and averaging procedures were performed by the use of the ATSAS package v2.8 (Franke et al., 2017). The final scattering patterns were obtained by merging multiple frames of *Ct*Araf43 at each concentrations using the Primus package of ATSAS 2.8 suite (Franke et al., 2017). The radius of gyration (Rg) for *Ct*Araf43 was estimated by using 2 independent procedures i) by the indirect Fourier transform method using the GNOM Software (Svergun et al., 1992) and ii) by Guinier approximation. The Kratky analysis was also performed to determine the folding and compactness of *Ct*Araf43. GNOM software was used to obtain the distance distribution function P(R) and the maximum dimension (D_{max}) for *Ct*Araf43. The molecular mass of *Ct*Araf43 was estimated from its scattering pattern by SAXS Mow Web server (Fischer et al., 2010). Molecular envelope of *Ct*Araf43 was reconstructed by the *ab-initio* modeling method implementing in DAMMIF by using SAXS I(Q) profile and the results obtained from its pairwise distribution of interatomic vectors (Svergun et al., 2001). A total of 20 independent molecular shapes were modeled, which were further averaged and refined by using DAMAVER (Volkov et al., 2003) and DAMMIN, respectively (Svergun et al., 1999). The simulated *Ct*Araf43 structure was superposed with the final molecular shape obtained from DAMMIN refinement by SUPCOMB from ATSAS suite (Kozin et al., 2001). The pyDockSAXS webserver (https://life.bsc.es/pid/pydocksaxs) (Jiménez-García et al., 2015) was used for protein-protein docking analysis using a *Ct*Araf43 as a protein and as a receptor to confirm the dimerization possibility. Visual analysis and figure generation were done by Open source version of Pymol.

3.3. Results and Discussion

3.3.1. Sequence analysis of *Ct*Araf43

*Ct*Araf43 is a modular protein (Fig. 3.1). The N-terminal catalytic module belongs to family 43 glycoside hydrolase, termed as *Ct*Abf43A, and spans from 31-330 amino acids residues. Amino acid residues from 331-463 form a domain classified as family 6 carbohydrate binding modules, termed as *Ct*CBM6A, tightly attached to the catalytic module by a very short linker peptide. The second CBM6, termed as *Ct*CBM6B, spans from 478-610 is connected to *Ct*CBM6A via a 14 amino acid residues long linker peptide. The SecretomeP 2.0 server analysis of *Ct*Araf43 displayed that the locus tag Cthe_1271, the encoded protein, is an extracellular enzyme. The presence of a type I dockerin in *Ct*Araf43 suggests that the modular enzyme is integrated as a component of the multi-enzyme complex “cellulosome” of *C. thermocellum*.

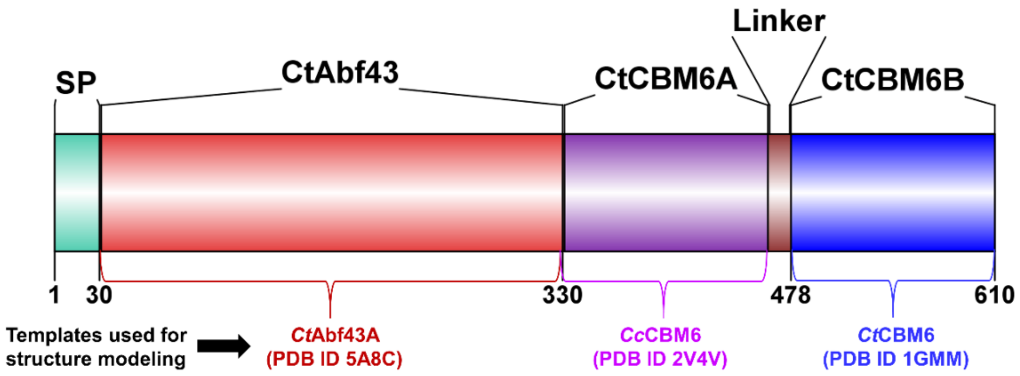


Fig. 3.1. Molecular architecture of *Ct*Araf43 based on amino acid sequence. The templates used for *Ct*Araf43 structure modeling is shown below the molecular architecture.

3.3.2. Structure modeling and validation of *Ct*Araf43

The modeled structure of *Ct*Araf43 consists of three domains: an N-terminal catalytic domain and independently folded *Ct*CBM6A and *Ct*CBM6B domains at its C-terminal. *Ct*CBM6A domain is located in the same plane as the catalytic domain, while CBM6B is attached with *Ct*Abf43A-*Ct*CBM6A at an angle of 45° through a linker peptide (Fig. 3.2A). The catalytic center displayed a five-fold β-propeller topology starting at Ala15 and extending to residue Lys299. A tightly associated carbohydrate binding module (*Ct*CBM6A) is connected with the catalytic module through 7-residue long linker peptide, QLYVDPY. The linker sequence is similar in length and sequence to the homologous arabinofuranosidase from *Bacillus subtilis* having a sequence as QISNLNP. An extended, non-conserved 14-residue long fragment links the *Ct*CBM6B to *Ct*CBM6A domain of *Ct*Araf43. The associated non-catalytic domains, *Ct*CBM6A (residues 313-432) and *Ct*CBM6B (residues 455-586) displayed a jellyroll type β-sandwich fold of two antiparallel sheets and each sheet is composed of five β-strands, in which four β-strands from each sheet are opposite to each other (Fig. 3.2A). The remaining two β-strands from each sheet extend away from the β-sandwich core β-3 and β-4 strands, making a finger-like structure, similar to *Cs*CBM6 of *Clostridium stercorarium* as reported earlier (Boraston et al., 2003). The PDBe Fold web server (Krissinel & Henrick, 2004) was used to find structural homologs of *Ct*Araf43. The result showed that the modeled *Ct*Araf43 structure is most similar to *Bs*AXH-CBM6 from *Bacillus subtilis* and gave rmsd value of 0.3 Å by superposing 488 Cα atoms. The structural topology analysis of *Ct*Araf43 from PDBSum webserver also corroborated with the modeled structure (Fig. 3.2B). The quality of modeled structure after energy minimization was assessed by the tools available at SAVES server. Ramachandran plot developed by Rampage server displayed that 98.5% residues are in the favored region, 1.5% in the disallowed region (Fig. 3.2C). This implied that dihedral angles, phi (ϕ) and psi (ψ) of the modeled *Ct*Araf43 structure had occupied favorable positions. ProSA results indicated that the modeled protein is error-free and reside in the x-ray zone with a score of -7.33 (Fig. 3.2D). Verify3D results showed that 91% of the residues had an average score of ≥ 0.2 in the 3D/1D profile (Fig. 3.2E).

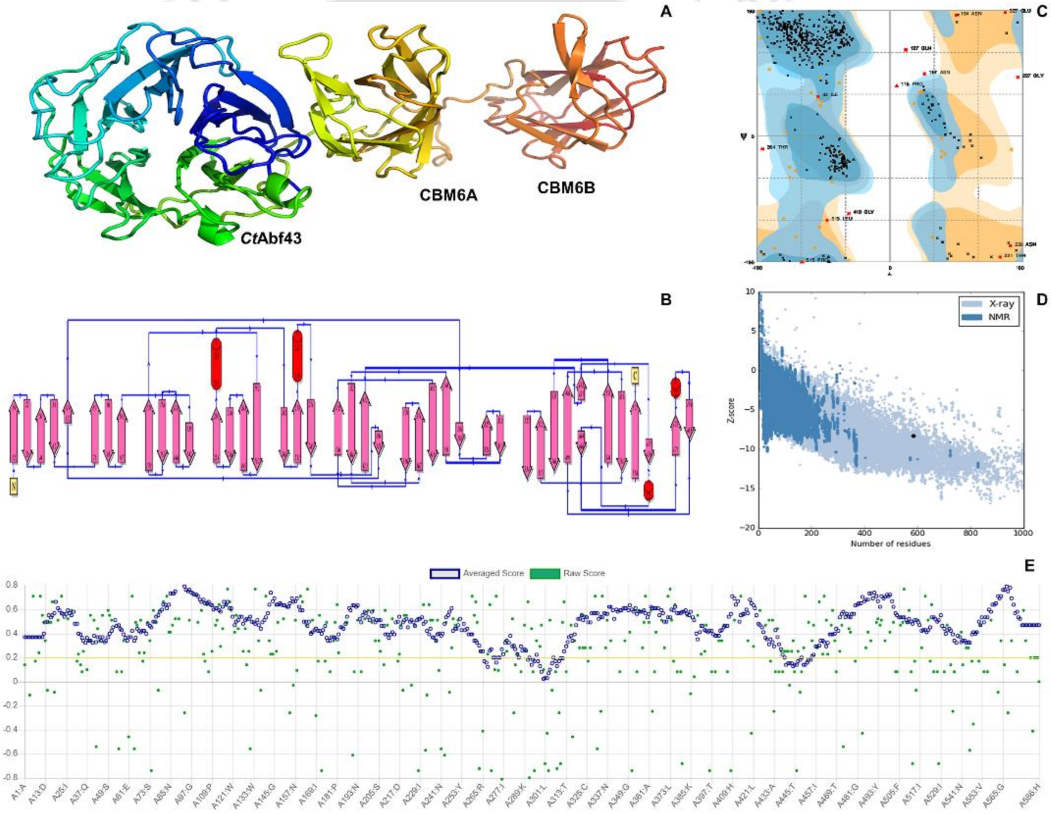


Fig. 3.2. *Ct*Araf43 structure modeling and validations. A) 3-D structure of *Ct*Araf43 B) Topology diagram *Ct*Araf43 C) Ramachandran plot, D) Prosa Plot analysis and E) Verify 3D Plot.

3.3.3. Molecular dynamics simulation of the *Ct*Araf43 structure

The structural dynamics and global compactness of *Ct*Araf43 were analyzed by performing molecular dynamics (MD) simulation. RMSD values were estimated for the entire trajectory to analyze the deviation from the original structure. It displayed a huge difference until 5 ns and then it became stable up to 100 ns with an average RMSD of 0.94 nm (Fig. 3.3A). The *Ct*Araf43 flexibility and dynamics after simulation were analyzed by estimating the RMSF value of the Cα atoms based on B-factor value. The RMSF analysis displayed that the overall structure of *Ct*Araf43 possesses flexibility. *Ct*Abf43A and *Ct*CBM6A both showed limited flexibility in the non-structural region and no flexibility was observed in secondary structure component. However, *Ct*CBM6B is more flexible in terms of structure and orientation (Fig. 3.3B). The sequence analysis of *Ct*Araf43 revealed that the protein C-terminal region contains a dockerin-I that is connected with *Ct*CBM6B via a linker peptide. The enzyme linked dockerin binds to cohesin modules of a scaffoldin providing the molecular mechanism for the integration of the enzyme into the cellulosome, which is well known to potentiate the efficient hydrolysis of plant cell wall polysaccharides. The flexible nature of *Ct*CBM6B may help in the positioning of *Ct*Abf43A-*Ct*CBM6A molecules for efficient catalysis.

The radius of gyration was determined for the MD simulated structure to evaluate the global compactness of *Ct*Araf43 during the MD simulation. The fluctuation in R_g values (3.85 nm to 3.65 nm) of *Ct*Araf43 for the initial 6 ns indicated flexibility but after the initial phase the average R_g value was found to be constant at 3.62 nm till 100 ns, suggesting that *Ct*Araf43 acquired a global compactness and achieved its stable conformation (Fig. 3.3C). The intramolecular hydrogen bonds formed in the main chain and main chain-side chain of *Ct*Araf43 were analyzed by gmx hbond program. The main chain of *Ct*Araf43 structure contains an average 414 intramolecular hydrogen bonds (Fig. 3.3D), while the main chain and the side chain of amino acid residues forms an average 57 hydrogen bonds up to 100 ns and these hydrogen bonds assisted the structure in achieving the stable state (Fig. 3.3E).

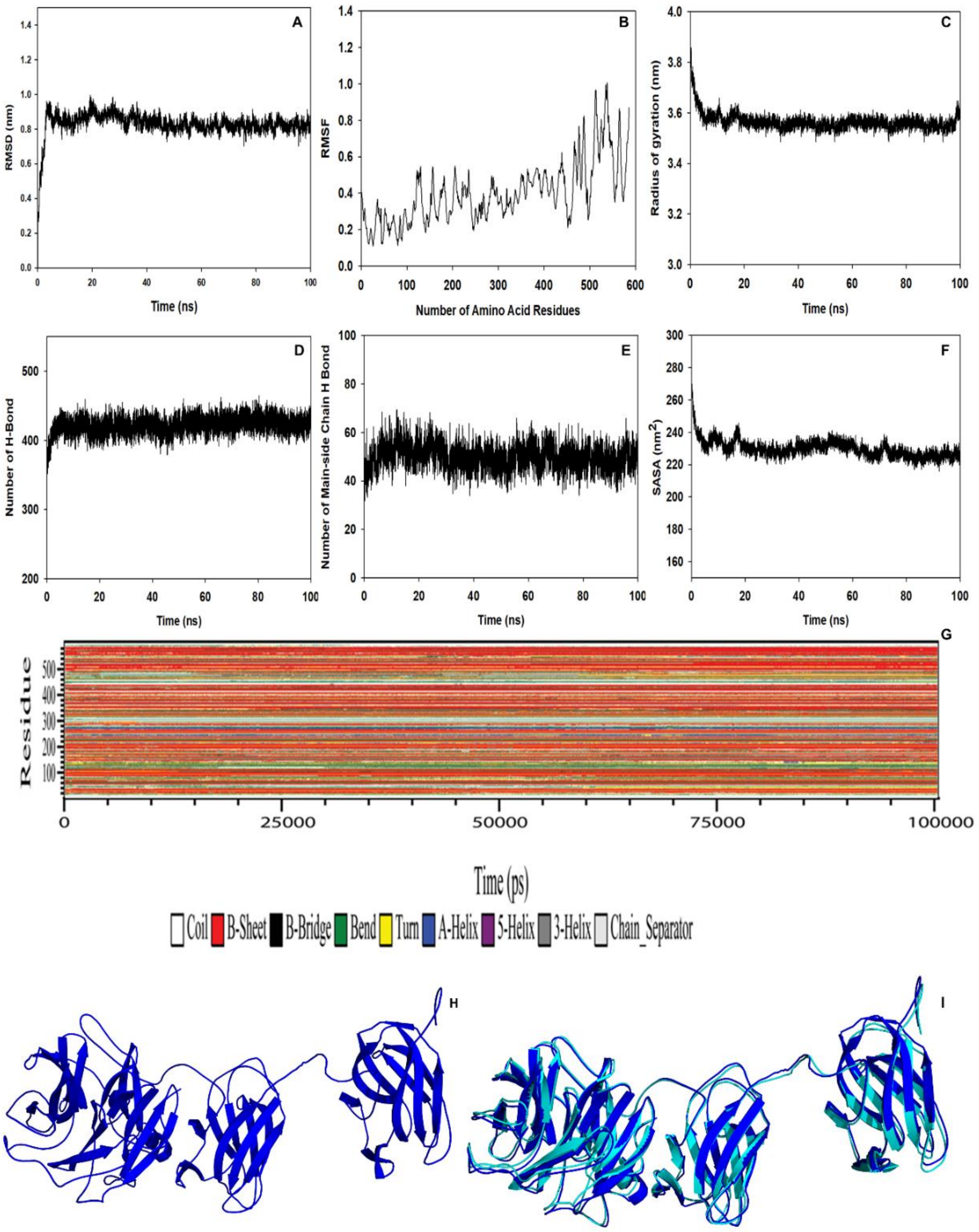


Fig. 3.3. Molecular dynamic simulation of *Ct*Araf43 (A) RMSD Plot (B) RMSF Plot. (C) Radius of gyration Plot (D) Intramolecular hydrogen bond analysis between main chain (E) Intramolecular hydrogen bond analysis between main chain and side chain (F) SASA calculation, (G) Secondary structure analysis by DSSP (H) Simulated *Ct*Araf43 structure (I) Superposition of simulated *Ct*Araf43 structure (blue) with Modeled *Ct*Araf43 structure (Cyan).

Solvent Accessible Surface Area (SASA) for *Ct*Araf43 was analyzed by gmx sas program. The average SASA up to 100 ns was 238 nm^2 and the exposure to the solvent area remains relatively unchanged (Fig. 3.3F). The DSSP analysis of *Ct*Araf43 displayed no change in secondary structure and maintained the integrity of the entire structural component involved in tertiary structure formation as shown in Fig. 3.3G. The MD simulated *Ct*Araf43 displayed stable structure conformation (Fig. 3.3H). The modeled *Ct*Araf43 structure and the MD simulated structures were aligned in PyMOL v2.30 that gave RMSD value of 0.045 nm (all Cα atom aligned) (Fig. 3.3I). The superposition of the original *Ct*Araf43 structure with the MD simulated *Ct*Araf43 structure showed that the overall structure possesses the flexibility only in the unstructured region (Fig. 3.3I). This low RMSD value suggested that *Ct*Araf43 is flexible in nature, but did not show variation in the conformation based on spatial orientation of different domains. Furthermore, the alternative views of the superposed simulated *Ct*Araf43 structure with native *Ct*Araf43 structure are shown in supplementary figure (Fig. 3.4). The *Ct*Araf43 structure obtained by MD simulation was stable and used for further analysis.

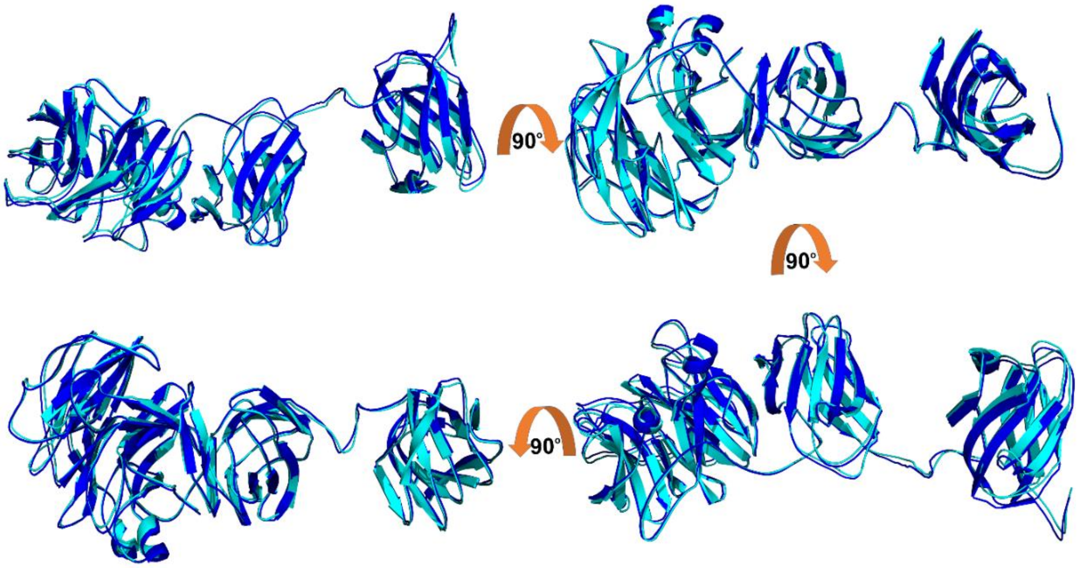


Fig. 3.4. Alternative views of superposed simulated *Ct*Araf43 structure with native modelled *Ct*Araf43 structure at different angle.

3.3.4. Solution structure analysis of *Ct*Araf43 by Small Angle X-ray Scattering

The molecular shape of *Ct*Araf43 in solution was determined by SAXS analysis at different concentrations. *Ct*Araf43 concentration-dependent SAXS pattern showed variation in structure and aggregation at higher concentration. *Ct*Araf43 showed inter-particle interaction at lower concentrations *i.e.* 0.3 and 0.6 mg/mL, while at a higher concentration, 9 mg/mL, it displayed aggregation (Fig. 3.5A). The SAXS pattern of *Ct*Araf43 at concentrations 1.2 mg/mL and 4.7 mg/mL displayed no inter-particle interaction and aggregation (Fig. 3.5B). The SAXS data of *Ct*Araf43 at all concentrations were evaluated by Guinier plot approximation and plotted as R_g vs *Ct*Araf43 concentrations (Fig. 3.5C). Based on the Guinier plot analysis the concentration 1.2 mg/mL and 4.7 mg/mL were found suitable for further analysis and the data were presented in Table 3.1.

Table 3.1. SAXS data collection parameters and derived parameters of *Ct*Araf43.

Data-collection parameters	*Ct*Araf43	
Beamline	BM29-BioSAXS, ESRF	
Wavelength (Å)	0.99 Å	
Q range (nm^{-1})	0.03-4.48	
Exposure time (Seconds)	1x10	
Temperature (°C)	4	4
Protein Concentration mg/mL	1.2	4.7
Structural parameter		
Q range (nm^{-1}) used for Rg analysis	0.17-0.41	0.15-0.38
I(0) au from Guinier	91.23±0.09	109.45±0.39
Rg nm from Guinier	3.43±0.03	4.55±0.02
Rg nm from P(R)	3.38	4.63
D_{max} (nm)	12	17
Porod volume estimate (nm^{3})	113	128
Molecular Mass Determination		
Theoretical molecular mass (kDa)	63	
Molecular Mass from SAXSMOW (kDa)	68	110
Modeling parameters		
NSD	0.537	0.539
Resolution (nm)	8.62	8.84
Q range (nm^{-1}) used for Structure Superposition by Crysol	0.18-0.45	0.16-0.43
Software employed		
Data processing	Primus	Primus
P(R) function calculation	GNOM	GNOM
Ab initio modelling	DAMMIF	DAMMIF
Validation and averaging	DAMAVER	DAMAVER
Structure superposition	SUPCOMB	SUPCOMB
3-D graphical representation	PyMOL	PyMOL

The Guinier plot analysis of *Ct*Araf43 indicated a linear behavior in low q region, which confirms that the protein sample preparation is in a monodisperse state (Fig. 3.5D). The radius of gyration for globular shape (R_g) and rod shape (R_c) determined at 1.2 mg/mL were found to be of 3.43 ± 0.03 nm and 1.22 nm, respectively and for 4.7 mg/mL the R_g and R_c value were 4.55± 0.02 nm and 1.37 nm, respectively (Fig. 3.5D). The estimated persistence length (L) of the *Ct*Araf43 at 1.2 mg/mL was found to be 11.9 nm and at 4.7 mg/mL was 15.0 nm, respectively. The flexibility and global compactness of *Ct*Araf43 in solution were evaluated by Kratky plot analysis using SAXS experimental data (Fig. 3.5E) The Kratky plot showed a Porod–Debye plateau with a Porod exponent in the range of 3.5-3.75 for *Ct*Araf43 at 1.2 mg/mL and 4.7 mg/mL, respectively, confirming the fully folded molecule with limited flexibility or inherent disorder present in the structure. The pair-distance distribution function P(R) curve obtained by indirect Fourier transform of the scattering curve showed an asymmetric profile consistent with elongated macromolecules (Fig. 3.5F). The maximum intra particle distance (D_{max}) and R_g obtained from P(R) curves of *Ct*Araf43 were 12 and 3.38 nm for 1.2 mg/mL and 17 nm and 4.63 nm for 4.7 mg/mL, respectively. The D_{max} is four times greater than the R_g value, which indicates that the *Ct*Araf43 molecule is elongated and not globular. The molecular weight (M_w) of *Ct*Araf43 estimated by SAXS MoW web server to be 68 kDa and 110 kDa at 1.2 mg/mL and 4.7 mg/mL, respectively. The calculated molecular mass of *Ct*Araf43 is equal to the monomeric protein at 1.2 mg/mL while the molecular mass at 4.7 mg/mL is almost double compared with theoretical molecular mass (63 kDa). As the protein concentration increases, the interaction between the molecules is promoted, which leads to the formation of higher molecular weight protein complexes, that further leads to the oligomerization of *Ct*Araf43.

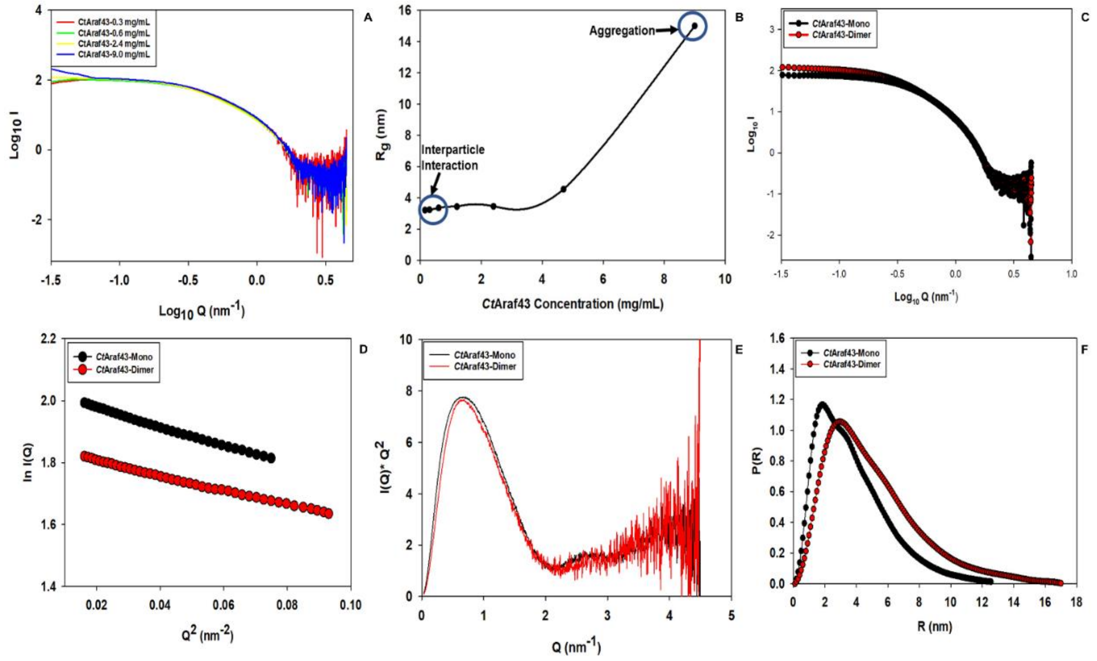


Fig. 3.5. Small angle X-ray scattering analysis of *Ct*Araf43, (A) Scattering pattern of *Ct*Araf43 at different concentrations, (B) Scattering pattern of *Ct*Araf43 at 1.2 mg/mL and 4.7 mg/mL concentrations (C) Guinier plot of R_g vs *Ct*Araf43 concentrations (D) Guinier plot analysis of *Ct*Araf43 at 1.2 mg/mL and 4.7 mg/mL concentrations, (C) Kratky analysis of *Ct*Araf43 at 1.2 mg/mL and 4.7 mg/mL concentrations and (D) P(R) Plot analysis of *Ct*Araf43 at 1.2 mg/mL and 4.7 mg/mL concentrations.

The overall molecular shape of *Ct*Araf43 was constructed by rapid *ab initio* modeling approach. The normalized spatial discrepancy (NSD) value of a set of models reconstructed by DAMMIF for monomeric and dimeric *Ct*Araf43 was 0.537. The averaged resolution of *ab initio* models at both concentrations are in the range of 8.6-8.8 nm. The *ab initio* approach derived monomeric and dimeric molecular envelope displayed the multi-modular and elongated structure (Fig. 3.6A & B). The molecular envelope of *Ct*Araf43 at 1.2 mg/mL displayed Miniature Dashund shape and at 4.7 mg/mL shows 2-fold rotational symmetry like envelope. The superposition of MD simulated *Ct*Araf43 structure on monomeric molecular envelop displayed an excellent fit (Fig. 3.6C).

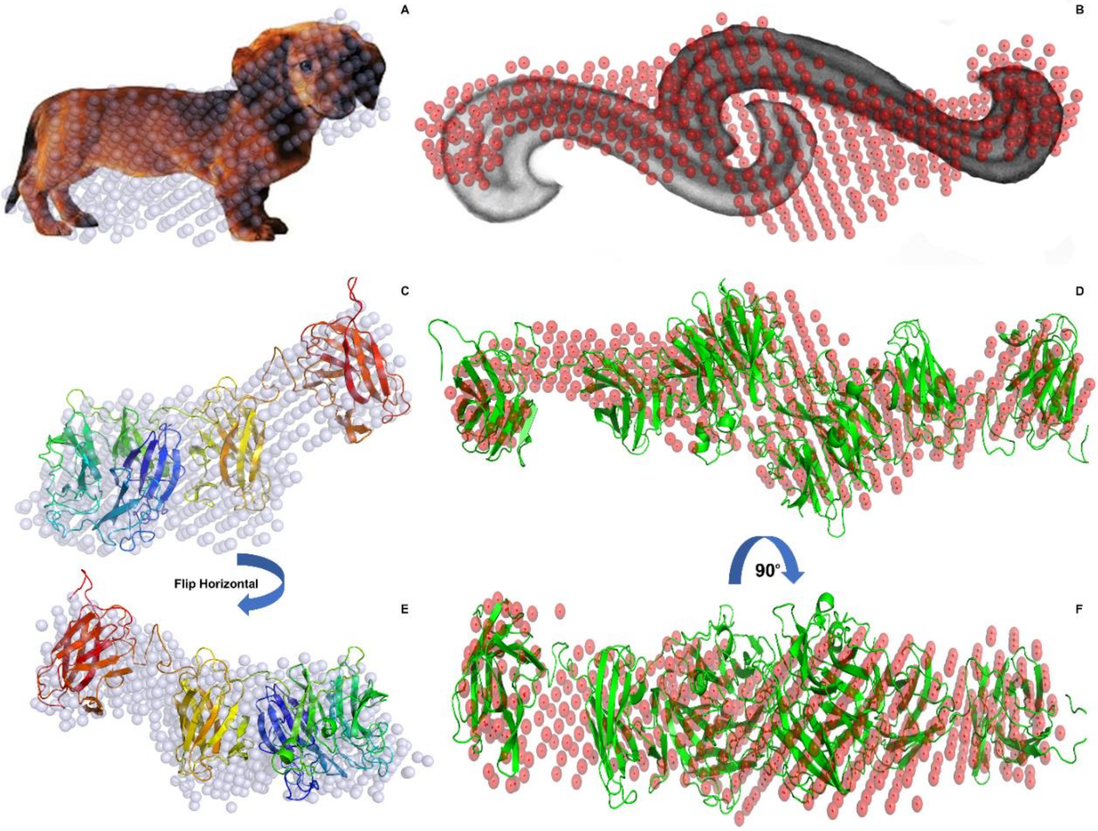


Fig. 3.6. (A & B) *Ab initio* model of *Ct*Araf43 at both concentrations. (C) Superposition of *ab initio Ct*Araf43 molecular shape and modeled structure (D). Superposition of *ab initio Ct*Araf43 molecular shape and pyDockSAXS generated dimer structure. (E & F) Alternative view of superposed monomer and dimeric *Ct*Araf43. The images were generated using PyMol (the PyMOL molecular graphics system, v2.3; Schrödinger, LLC).

However, the dimeric molecular envelope displayed the superposition of MD simulated *Ct*Araf43 structure covers only half of the *ab initio* modeled structure. Therefore, the two molecules of MD simulated *Ct*Araf structure were used as protein and receptor for protein-protein docking against the scattering data obtained at 4.7 mg/mL by using pyDockSAXS webserver. The top 10 models obtained from pyDockSAXS webserver analysis results were superposed using the SUPCOMB program and the best superposition results are shown here (Fig. 3.6D). *Ct*Araf43 dimer obtained from pyDockSAXS analysis revealed that the *Ct*Abf43A molecule of both the Chain A and Chain B interacts with each other and forms a dimer molecule. *Ct*CBM6A and *Ct*CBM6B of each chain are not interacting with the other and oriented in the opposite direction. The structure of *Ct*Abf43A was determined earlier in two different space group, which is monoclinic and cubic. One molecule of *Ct*Abf43A and 4 homotrimers of *Ct*Abf43A were observed in the asymmetric unit of monoclinic and cubic shape respectively. The significance of these interactions at higher protein concentrations remains unknown. However, in the context of the incorporation of *Ct*Abf43A in the cellulosome, the micro-concentration of the protein maybe relatively high and should promote protein-protein interactions through the catalytic modules of the enzyme that might help positioning *Ct*Abf43A in close proximity within the complex.

3.4. Conclusion

The conformational dynamics and modular arrangement of a family 43 modular α-L-arabinofuranosidase (*Ct*Araf43) from *Clostridium thermocellum* in solution were investigated. The modeled structure revealed three distinct domains that are independently folded and achieve their stable conformation. Conformational dynamics analysis of *Ct*Araf43 by MD simulation displayed the flexible nature of the molecule at its position, but no variation in its conformational organization or secondary structure was observed. SAXS analysis of *Ct*Araf43 revealed its monodisperse nature at both concentrations (1.2 mg/mL and 4.7 mg/mL). SAXS analysis of *Ct*Araf43 suggested elongated structures with monomeric conformation at 1.2 mg/mL and dimeric conformation at 4.7 mg/mL. Kratky plot revealed that *Ct*Araf43 retains a bulged-tubular shape with limited flexibility or inherently disorder at both the concentrations, which was corroborated by MD simulation results. The pyDockSAXS displayed that the catalytic module of the two *Ct*Araf43 molecules interact together resulting in the dimerization of *Ct*Araf43. The *ab initio* derived molecular shape of *Ct*Araf43 showed the molecular envelope with an of Miniature Dashund puppy shape at 1.2 mg/mL that is dimerised by a 2-fold rotational symmetry at 4.7 mg/mL. The *ab initio* dummy atom model superposed well with the MD simulated monomeric structure and dimeric structures of *Ct*Araf43 obtained from the pyDockSAXS analysis. The flexible nature of the modular enzyme upon its integration on a multi-enzymatic complex, the cellulosome, remains to be elucidated.

3.5. References

Abbott, D.W., Ficko-Blean, E., van Bueren, A.L., Rogowski, A., Cartmell, A., Coutinho, P.M., Henrissat, B., Gilbert, H.J., Boraston, A.B., 2009. Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules. *Biochemistry*, 48(43), 10395-10404.

Ahmed, S., Luis, A.S., Bras, J.L., Ghosh, A., Gautam, S., Gupta, M.N., Fontes, C.M., Goyal, A., 2013a. A novel α-l-arabinofuranosidase of family 43 glycoside hydrolase (*Ct*43Araf) from *Clostridium thermocellum*. *PloS One*, *8*(9), e73575.

Ahmed, S., Luís, A. S., Brás, J. L. A., Fontes, C. M. G. A., Goyal, A., 2013b. Functional and structural characterization of family 6 carbohydrate-binding module (*Ct*CBM6A) of *Clostridium thermocellum* α-L-arabinofuranosidase. *Biochem. (Moscow)*, *78*(11), 1272-1279.

Ahmed, S., Luis, A. S., Brás, J. L., Fontes, C. M., Goyal, A., 2013c. The family 6 carbohydrate-binding module (*Ct*CBM6B) of *Clostridium thermocellum* alpha-L-arabinofuranosidase binds xylans and thermally stabilized by Ca^{2+} ions. *Biocatal. Biotransformatio*, *31*(4), 217-225.

Ahmed, S., Gautam, S., Gupta, M. N., Goyal, A., 2013d. Analysis of structural element of family 6 carbohydrate-binding module (*Ct*CBM6B) of alpha-L-arabinofuranosidase from *Clostridium thermocellum*. *J. Proteins Proteom., 4(1) 27-34.*

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., 1990. Basic local alignment search tool. *J. Mol. Biol., 215,* 403-410.

Beldman, G., Schols, H. A., Pitson, S. M., Searle-van Leeuwen, M. J. F., Voragen, A. G. J., 1997. Arabinans and arabinan degrading enzymes. *Adv. Macromol. Carbohydr.,* 1, 1-64.

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

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

Contesini, F.J., Liberato, M.V., Rubio, M.V., Calzado, F., Zubieta, M.P., Riaño-Pachón, D.M., Squina, F.M., Bracht, F., Skaf, M.S., Damasio, A.R., 2017. Structural and functional characterization of a highly secreted α-L-arabinofuranosidase (GH62) from *Aspergillus nidulans* grown on sugarcane bagasse. *Biochim. Biophys. Acta-Proteins Proteomics*, 1865, 1758-1769.

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 2001. The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. *J. Biol. Chem.,* 276, 48580-48587.

Farro, E. G. S., Leite, A. E. T., Silva, I. A., Filgueiras, J. G., de Azevedo, E. R., Polikarpov, I., Nascimento, A. S., 2018. GH43 endo-arabinanase from *Bacillus licheniformis*: Structure, activity and unexpected synergistic effect on cellulose enzymatic hydrolysis. *Int. J. Biol. Macromol., 117,* 7-16.

Fischer, H., de Oliveira Neto, M., Napolitano, H. B., Polikarpov, I., Craievich, A. F., 2010. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr*., *43*, 101-109.

Fontes, C. M. G. A., Gilbert, H. J., 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem*., 79, 655-681.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., Svergun, D.I., 2017. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr*., 50, 1212-1225.

Goyal, A., Ahmed, S., Sharma, K., Gupta, V., Bule, P., Alves, V.D., Fontes, C.M., Najmudin, S. 2016. Molecular determinants of substrate specificity revealed by the structure of *Clostridium thermocellum* arabinofuranosidase 43A from glycosyl hydrolase family 43 subfamily 16. *Acta Crystallogr. Sect. D Struct. Biol*., *72*, 1281-1289.

Harris, D., DeBolt, S., 2010. Synthesis, regulation and utilization of lignocellulosic biomass. *Plant Biotechnol. J.*, 8, 244–262.

Hemsworth, G.R., Thompson, A.J., Stepper, J., Sobala, Ł.F., Coyle, T., Larsbrink, J., Spadiut, O., Goddard-Borger, E.D., Stubbs, K.A., Brumer, H., Davies, G.J., 2016. Structural dissection of a complex *Bacteroides ovatus* gene locus conferring xyloglucan metabolism in the human gut. *Open Biol*., *6*(7), 160142.

Jiménez-García, B., Pons, C., Svergun, D. I., Bernadó, P., Fernández-Recio, J., 2015. *Py*DockSAXS: Protein-protein complex structure by SAXS and computational docking. *Nucleic Acids Res*., 43(W1), W356-W361.

Kozin, M. B., Svergun, D. I., 2001. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr*., 34, 33–41.

Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Struct. Biol*., *60*, 2256-2268.

Lagaert, S., Pollet, A., Courtin, C. M., Volckaert, G., 2014. β-Xylosidases and α-L-arabinofuranosidases: Accessory enzymes for arabinoxylan degradation. *Biotechnol. Adv., 32*, 316-332.

Petersen, T. N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8, 785.

Saha, B. C., 2000. α-L-Arabinofuranosidases: Biochemistry, molecular biology and application in biotechnology. *Biotechnol. Adv., 18,* 403-423.

Saleh, M. A., Han, W. J., Lu, M., Wang, B., Li, H., Kelly, R. M., Li, F.L., 2017. Two distinct α-l-Arabinofuranosidases in *Caldicellulosiruptor* species drive degradation of arabinose-based polysaccharides. *Appl. Environ. Microbiol., 83,* e00574-17.

Webb, B., Sali, A., 2017. Protein structure modeling using MODELLER. *Methods Mol. Biol., 1654, 39-54.*

Sharma, K., Fontes, C. M. G. A., Najmudin, S., Goyal, A., 2019. Molecular organization and protein stability of the *Clostridium thermocellum* glucuronoxylan endo-β-1,4-xylanase of family 30 glycoside hydrolase in solution. *J. Struct. Biol., 206.* 335-344.

Shoseyov, O., Shani, Z., Levy, I., 2006. Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. *Microbiol. Mol. Biol. Rev.*, 70, 283–295.

Svergun, D. I., 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr*., 25, 495-503.

Svergun, D. I., 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J., 76*, 2879-2886.

Svergun, D. I., Petoukhov, M. V., Koch, M. H. J., 2001. Determination of domain structure of proteins from x-ray solution scattering. *Biophys. J.*, 80, 2946–2953.

Thakur, A., Sharma, K., Goyal, A., 2019. α-l-Arabinofuranosidase: A Potential Enzyme for the Food Industry. In *Green Bio-processes*, 229-244. Springer, Singapore.

Touw, W. G., Baakman, C., Black, J., Te Beek, T. A. H., Krieger, E., Joosten, R. P., Vriend, G., 2015. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res*., *43*(D1), D364-D368.

Vandermarliere, E., Bourgois, T.M., Winn, M.D., Van Campenhout, S., Volckaert, G., Delcour, J.A., Strelkov, S.V., Rabijns, A., Courtin, C.M., 2008. Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. *Biochem. J., 418*, 39-47.

Volkov, V. V., Svergun, D. I., 2003. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr*., 36, 860-864.

Zdobnov, E. M., Apweiler, R., 2001. InterProScan - An integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, 17, 847–848.

Zhong, R., Teng, Q., Lee, C., Ye, Z. H., 2014. Identification of a disaccharide side chain 2-O-α-D-galactopyranosyl-α-D-glucuronic acid in *Arabidopsis* xylan. *Plant Signal. Behav.*, 9, 2-6.

Chapter 4

Molecular organization of glucuronoxylan endo-β-1,4-xylanase of family 30 glycoside hydrolase from *Clostridium thermocellum* in solution

4.1. Introduction

The use of lignocellulosic biomass as a renewable source for the production of biochemicals such as organic acids, biofuels and value-added products such as xylitol and xylooligosaccharides, has become increasingly relevant for economic and environmental reasons (Jönsson et al., 2016). The low cost, abundance and accessibility of lignocellulosic biomass make it suitable for the mentioned industrial applications (Harris et al., 2010). Lignocellulosic biomass predominantly contains lignin and polysaccharides existing as a complex lignocellulosic matrix (Saini et al., 2015). In this context, heteroxylan accounts for approximately one-third of the total cell wall polysaccharides available as renewable organic carbon on earth (Collins et al., 2005). Heteroxylans are mainly composed of a backbone of β-1,4 linked xylose residues often with decorations at O-2 or O-3 by arabinofuranose residues and at O-2 by 4-O-Methyl-D-glucuronic acid (Zhong et al., 2014). The composition of xylan varies widely depending on the source. The breakdown of heteroxylan requires the cumulative action of several xylanases with different specificities for its efficient utilization and conversion to xylose for production of various value-added products (Padilha et al., 2014). The majority of xylanases studied so far are grouped into family 10 and 11 glycoside hydrolase (GH10 and GH11). They can hydrolyze the β-1,4- and β-1,3-linked algal xylan, cereal arabinoxylans and hardwood glucuronoxylans (Biely et al., 1997). Other xylanases belonging to family 5 and 8 glycoside hydrolase specifically hydrolyze arabinose substituted xylans (Correia et al., 2011). A few bacterial xylanases belonging to family 30 glycoside hydrolase that have specificity towards glucuronoxylan, but no activity against arabinoxylans, have been identified and characterized (John et al., 2006; Vrsanská et al., 2007; Gallardo et al., 2010). These xylanases require 4-O-methyl-D-glucuronosyl decorations on xylan for their activity and are classified into subgroup H of the family 30 glycoside hydrolase (GH30) and are considered as glucuronoxylanases (Verma et al., 2016).

Glycoside hydrolases are often modular and contain a catalytic module linked by a flexible linker peptide to one or more ancillary carbohydrate-binding module(s) (CBMs) (Boraston et. Al., 2002). These CBMs target the appended catalytic modules to the enzyme substrate, thus enhancing the breakdown of the complex plant polysaccharides (Shoseyov, et al., 2006). *Ct*XynGH30 belongs to the cellulosomal complex of *Clostridium thermocellum* (Fontes et al., 2010). *Clostridium thermocellum* produces a set of modular xylan degrading enzymes including GH10, GH11, GH30, GH39 and GH43. Unlike other GH30 glucuronoxylanase, *Ct*XynGH30 is a complex modular enzyme consisting of an N-terminal catalytic module, termed *Ct*Xyn30A, along with a C-terminal family 6 carbohydrate-binding module, *Ct*CBM6, and a dockerin. *Ct*XynGH30 hydrolyzes 4-O-methyl-D-glucuronosyl decorated substrates (Verma et al., 2014). *Ct*XynGH30 plays a crucial role in the depolymerization of heteroxylans (Verma et al., 2016). Therefore, it can be used in various biological processes. The structural and functional characterization of *Ct*Xyn30A (Freire et al., 2016) and *Ct*CBM6 (Verma et al., 2015) have been reported earlier. The catalytic domain, *Ct*Xyn30A contains $(\beta/\alpha)_8$ TIM barrel topology starting at Arg13 and extending to residue Pro295 and a tightly associated, side β-domain structure connected with this TIM-barrel. The side β-domain structure is composed of eight β-strands oriented in an anti-parallel manner and forms immunoglobulin like fold (Freire et al., 2016).

The modeled structure of *Ct*CBM6 revealed a jelly roll β-sandwich fold and its biochemical characterization *displayed affinity towards undecorated (*beechwood- and birchwood-) *and* decorated (wheat-, rye-arabino- and oat spelt-) xylans (Verma et al., 2015). However, several questions still remain unanswered about the conformational dynamics, molecular arrangement and protein stability of the modular glucuronoxylanase *Ct*XynGH30 in solution. In order to comprehend the conformational flexibility, molecular organization and protein stability of *Ct*XynGH30 in solution, a low-resolution model was obtained using a combination of Small Angle X-ray Scattering (SAXS) analysis and computational modelling. The biophysical data collected in this study was corroborated by various biochemical techniques.

4.2. Material and methods

4.2.1. Retrieval of the protein sequence of modular *Ct*XynGH30

The full-length amino acid sequence of *Ct*XynGH30 was retrieved from the UniProtKB database (http://www.uniprot.org) having a UniProt ID A3DJS9 (Verma et al., 2014). The protein domain boundaries of *Ct*XynGH30 were assigned by analyzing the results of sequence analysis from PSI-BLAST tool (http://blast.ncbi.nlm.nih.gov/) (Altschul et al., 1990) and InterProScan (http://www.ebi.ac.uk/interpro/) servers (Zdobnov et al., 2001). The *Ct*XynGH30 domain architecture was prepared by using Illustrator for Biological Sequences (IBS) tool after the comprehensive analysis of sequence by Conserved domain database search and multiple sequence alignment (MSA). The subcellular localization of *Ct*XynGH30 was analyzed by using the SecretomeP 2.0 server (Bendtsen et al., 2005).

4.2.2. Expression and purification of *Ct*XynGH30 and its truncated derivatives

The recombinant plasmids containing the genes encoding *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 with a hexahistidine tag at N-terminal, obtained from previous study (Verma et al., 2016), were expressed in *Escherichia coli* BL21 (DE3) separately. The *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 expressing BL21 (DE3) cells were aerobically grown separately in a 2 L flask containing 400 mL Luria Bertani medium supplemented with 50 µg/mL kanamycin by incubating at 37°C and 180 rpm for 3 h (Verma et al., 2016). The cells were cooled to 24°C after cell growth had reached mid-exponential phase (A_{600} = 0.6) and isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM was added and further incubated at 24°C and 180 rpm for 16-18 h. The cells were centrifuged at 6000*g* at 4°C for 10 min. The cell pellet was washed and resuspended in 10 mL of 50 mM sodium phosphate buffer (pH 6.0) containing 1 mM phenylmethylsulfonyl fluoride, 150 mM NaCl and 50 mM Imidazole. The resuspended cells wsere subjected to ultra-sonication (Sonics, Vibra cell) for 20 min (2s on and 10s off pulse; 33% amplitude) for cell lysis. The cell lysate was centrifuged at 18,000g, at 4°C for 45 min. The cell-free extract was filtered through 0.45 µm membrane and purified by immobilized metal ion affinity chromatography (IMAC) using 5 mL Sepharose column (HisTrap Chelating, GE Healthcare, USA) (Verma et al., 2016). *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were further purified by gel filtration by loading 3 mL of 5 mg/mL of each separately to Superdex 75 column (HiLoad 16/60, GE Healthcare) using 20 mM sodium phosphate buffer (pH 6.0) at a flow rate of 1.0 mL/min on FPLC (AKTA purifier chromatographic system (GE Healthcare). The fractions were analyzed by SDS-PAGE (12%, w/v gel) for purity and pure fractions were pooled. The pooled fractions of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6, were concentrated up sto 10 mg/mL by using the centrifugal filter (Amicon Ultra, 10 kDa molecular weight cut-off) by centrifuging at 4000g and 4°C for 15 min.

4.2.3. Secondary structure analysis of *Ct*XynGH30

The secondary structure of *Ct*XynGH30 was predicted by the PSIPRED web server (http://bioinf.cs.ucl.ac.uk/psipred/). PSIPRED server predicted secondary structure using two-feed position specificity based forward neural network and scoring matrices generated by PSI-BLAST. The secondary structure was analyzed by Circular Dichroism (CD) by using 50 µM of *Ct*XynGH30 in 20 mM sodium phosphate, pH 6. The CD spectrum of *Ct*XynGH30 was recorded under Far-UV range (190-250 nm) on a spectropolarimeter (Jasco Corporation, J-815, Tokyo, Japan) at 25°C using a sample cell with a path length of 0.1 cm. The CD spectrum was collected at a scanning rate of 50 nm/min and 1 nm bandwidth with an average of five scans. The CD spectrum was corrected by buffer subtraction. The molar residual ellipticity (MRE) was calculated from the ellipticity (θ) values at each wavelength. The MRE values of 190 to 240 nm were analyzed by the K2D3 server to get the spectral deconvolutions for *Ct*XynGH30 (i.e., secondary structure composition) (HTTP:// k2d3.ogic.ca/) (Perez-Iratxeta et al., 2008). The results of the secondary structure of *Ct*XynGH30 from CD analysis and PSIPRED were compared to finalize the result.

4.2.4. Molecular modelling

A comparative modelling approach was used for obtaining a 3-dimensional model for *Ct*XynGH30, as at this moment, there is no three-dimensional structure available for the entire *Ct*XynGH30. The retrieved amino acid sequence of *Ct*XynGH30 (559 aa long out of 630 aa as the C-terminal dockerin is not included in this study) was subjected to PSI-BLAST search using default parameters against a Protein Data Bank (PDB). The modular Xyn30d from *Paenibacillus barcinonensis* (PDB: 4QAW, Sainz-Polo et al., 2014) was found closest to *Ct*XynGH30 with 59% sequence identity and thus, it was used as a reference. The modular structure of Xyn30d is arranged as Xyn30-Linker-CBM35, which is similar in length and organisation to the *Ct*XynGH30. As the crystal structure of *Ct*Xyn30A catalytic domain at high resolution is available (PDB: 5A6M at 1.17 Å) (Freire et al. 2016) it was superposed onto the Xyn30d catalytic module. SSM (from the server: PDBe Fold v2.59. (src3) 14 Apr 2014) gave a rmsd of ~0.7 Å between 389 aa on superpositioning, with the sequence identity between the two being 71%. In order to enable the modelling of the entire structure, the three-dimensional structure of the *Ct*CBM6 was modelled using bioinformatics tools. PSI-BLAST search using the *Ct*CBM6 sequence against a PDB database gave 11 structures with sequence identities ranging from 28-63%. A three-dimensional model for *Ct*CBM6 using the structure of the *Clostridium thermocellum* CBM6 from xylanase U (PDB: 1GMM, sequence identity 63%, Czjzek et al., 2001) as a template was built using MODELLER 9v17 (Šali et al., 1995). It was superposed onto the CBM35 module of the Xyn30d. SSM (from the server: PDBe Fold v2.59. (src3) 14 Apr 2014) gave a rmsd of ~2.0 Å between 110 aa on superpositioning, with the sequence identity between the two being 22%. A sequence comparison of the linker region between *Ct*XynGH30 and other homologous modular proteins shows a great diversity in both the length (from as little as 12 aa to over 20 aa) and in composition (Fig. 4.1). The *Ct*XynGH30 linker was taken as 14 aa long.

4.2.5. Protein melting studies of *Ct*XynGH30

The protein melting point of *Ct*XynGH30 was determined by using 40 µg of purified *Ct*XynGH30 in 1 mL of 20 mM sodium phosphate buffer, pH 6.0. The absorbance at 280 nm (A_{280}) was measured at a temperature from 25°C to 95°C by increasing the temperature at 5°C per min using a UV-Visible spectrophotometer (Varian, Cary 100-Bio) equipped with the peltier, temperature controller. The protein solution was kept at a particular temperature for 5 min to attain the equilibrium and then the absorbance was recorded and the melting curve of *Ct*XynGH30 was acquired. The melting curve of *Ct*XynGH30 was also studied in the presence of 1 mM Li^+ ion, 1 mM Ca^{2+} ion or 1 mM EDTA. The melting curve of absorbance at 280 nm versus temperature was plotted to generate the melting profile of *Ct*XynGH30 using Sigma plot v12.5.

4.2.6. Analysis of the contribution of modules in *Ct*XynGH30

*4.2.6.1. Ct*Xyn30A *and Ct*CBM6 *binding assay by ITC*

The role of *Ct*CBM6 in the full-length module, *Ct*XynGH30 was determined by studying the interaction between catalytic module, *Ct*Xyn30A and the carbohydrate binding module, *Ct*CBM6 by Isothermal titration calorimetry (ITC, MicroCal iTC200, GE Healthcare, USA) at 25°C. 200 µL of 70 µM of *Ct*Xyn30A filled in the cell was stirred at 150 rpm and titrated against *Ct*CBM6 (350 µM) through 20 successive injections of 1.2 µL made at regular intervals of 120 s. The non-linear regression and single binding site model in the MicroCal Origin software were used to determine the integrated heat effects.

*4.2.6.2. Ct*Xyn30A *and Ct*CBM6 *interaction by MALDI-TOF*

The proteins *Ct*Xyn30A (100 µM) and *Ct*CBM6 (100 µM) were mixed in a 1:1 molar ratio and incubated at 25ºC for 6 h. The individual proteins *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were used as control. The molecular mass of *Ct*Xyn30A:*Ct*CBM6 mixture, *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were determined by mass spectrometric analysis to confirm the interaction between *Ct*Xyn30A and *Ct*CBM6. The mass spectrometric analysis was performed on an Ultraflex workstation (Bruker Daltonics, Bremen, Germany). The matrix for MALDI analysis was prepared by dissolving 4 mg/mL of sinapinic acid (trans-3,5-dimethoxy-4-hydroxycinnamic acid) in (95%) ethanol and termed as solution A. The solution B was prepared by mixing 4 mg/mL sinapinic acid in 0.1% trifluoroacetic acid (TFA) and acetonitrile in 30:70 (v/v) ratio. 0.5 µL each of *Ct*Xyn30A:*Ct*CBM6 mixture, *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were mixed with 0.5 µL of solution B and then an equal volume of solution A was added and mixed gently by micropipette. The 2 µL *Ct*Xyn30A:*Ct*CBM6 mixture, *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were spotted on stainless steel sample target plate and dried for 1 h and used for mass spectrometric analysis.

4.2.6.3. Assay of ***Ct*****Xyn30A:*****Ct*****CBM6 mixture,** ***Ct*****XynGH30 and** ***Ct*****Xyn30A** *activity*

The effect of *Ct*CBM6 on enzyme activity was determined by estimating the activities of *Ct*Xyn30A: *Ct*CBM6 mixture, *Ct*XynGH30 and *Ct*Xyn30A against beechwood xylan, xylan (Mw. 20000-30000), birchwood xylan and 4-O-methylglucuronoxylan under optimized conditions as reported earlier (Verma et al., 2016). Ninety microliters of 1% (w/v) substrate with 10 µL of individual enzymes (17 µg) in 20 mM sodium phosphate buffer (pH 6.0) was incubated at 70ºC for 10 min. The xylanase activity was determined by quantifying the released reducing sugar as described earlier (Nelson, 1944; Somogyi, 1945). All the enzyme assays were performed in triplicates.

4.2.7. Small Angle X-ray Scattering (SAXS) of *Ct*XynGH30

SAXS data for *Ct*XynGH30 were collected at the BM29-BioSAXS beamline at The European Synchrotron Radiation Facility (ESRF) (Grenoble, France). Scattering pattern was recorded using Pilatus 1M detector (Dectris, Switzerland) at a wavelength of 0.992 Å for the sample to detector distance of 2.81 meters, covering the scattering vector range q of 0.023 to 4.5 nm^{-1}. To prevent the interference by any aggregation and precipitation, the samples were filtered using 0.2 µm syringe filters and centrifuged at 20,000 g for 30 min, at 25°C before the measurement. Scattering analysis of *Ct*XynGH30 was performed at different protein concentrations, 1.5, 2.5 and 5 mg/mL to confirm that inter-particle interaction can be ruled out and that their influence on the analysis is negligible. The scattering of buffer was also measured using the same sample holders and subtracted for the normalization of data to the absolute scale. The sample temperature was controlled by a peltier system. 10 successive frames of 10s each were collected for each sample to check for beam stability and radiation damage.

4.2.8. SAXS data analysis

The data treatment, normalization to absolute scale and averaging procedures were performed by the use of the ATSAS package (Konarev et al., 2006). Multiple frames of *Ct*XynGH30 scattering patterns were merged using the Primus package of ATSAS 2.8 suite (Franke et al., 2017). The radius of gyration (Rg) for *Ct*XynGH30 was calculated by using two independent procedures i) by Guinier approximation ii) by the indirect Fourier transform method using the GNOM package (Svergun et al., 1992). The distance distribution function P(r) and the maximum diameter (D_{max}) were also obtained with the GNOM software. The molecular weight of *Ct*XynGH30 was calculated from the scattering pattern by SAXS Mow Web server (Fischer et al., 2010). Scattering shape of *Ct*XynGH30 was reconstructed using the *ab initio* modeling based chain like ensemble modeling program GASBOR v2.3 in reciprocal space using SAXS I(Q) profile and results obtained from its pairwise distribution of interatomic vectors (Svergun et al., 2001). Dummy aspartic acid residues equal to the molecular mass of the *Ct*XynGH30 were taken into consideration to model the shape. 50 independent dummy atom models were generated and averaged by using DAMAVER suite of programs (Volkov et al., 2003). CORAL was employed to reconstruct the missing linker peptide region (14 residues long) between *Ct*Xyn30A (PDB: 5A6M) and the modelled structure of *Ct*CBM6 against the SAXS data. SREFLEX was used for the refinement and normal mode analysis of the CORAL derived structure. The refined *Ct*XynGH30 structure from SREFLEX was superposed with the averaged model from DAMAVER using SUPCOMB from ATSAS suite (Kozin et al., 2001). The CORAL derived structure was also used for flexibility analysis using conformation-sampling approach by BILBOMD web server (Pelikan et al., 2009). The conformational sampling was performed by keeping *Ct*Xyn30A and *Ct*CBM6 as rigid bodies, while keeping the linker region between the two modules as flexible. Ensemble of model conformations were generated by restraining the Rg value within 3-4 nm range with 800 conformations generation at each Rg value. For comparison and validation purposes, the above procedure was repeated with Xyn30d (PDB: 4QAW_a) against the SAXS data. Visual analysis and figure generation were done by Open source version of PYMOL v2.2.0.

4.3. Results and Discussion

4.3.1. Sequence analysis of *Ct*XynGH30

*Ct*XynGH30 is a modular protein (Fig. 4.1). The N-terminal, catalytic domain belongs to family 30 of glycoside hydrolases, termed *Ct*Xyn30A, and spans from 30-419 amino acids residues. *Ct*CBM6 is linked with the catalytic domain via a 14 amino acid residues long linker peptide. The presence of a typical signal peptide suggests that *Ct*XynGH30 is an extracellular enzyme consistent with its target activity against a large and insoluble extracellular substrate. The presence of a type I dockerin in *Ct*XynGH30 suggests that the enzyme is also integrated as a component of *C. thermocellum* multi-enzyme complex termed the cellulosome.

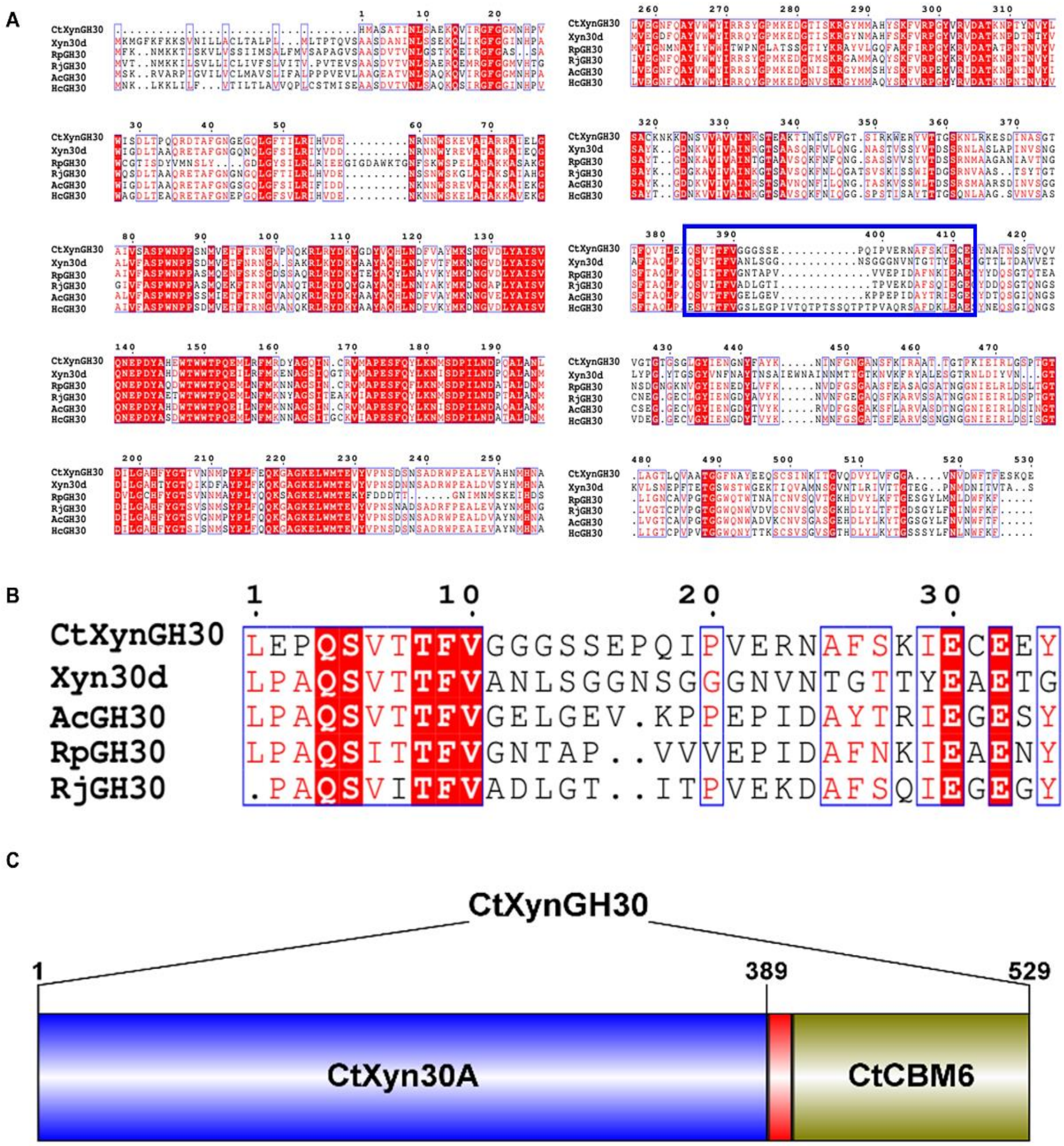


Fig. 4.1. A) Multiple sequence alignment (MSA) of *Ct*XynGH30 with other homologous modular GH30 glucuronoxylan hydrolases *viz.* Xyn30d from *Paenibacillus barcinonensis* (% sequence identities 70 & 22), *Rp*GH30 from *Ruminiclostridium papyrosolvens* (49% & 48%), *Rj*GH30 from *Ruminiclostridium josui* (69% & 46%), *Ac*GH30 from *Anaerobacterium chartisolvens* (76% & 54%) and *Hc*GH30 from *Hungateiclostridium cellulolyticum* (73% & 59%). (% sequence identities given in parenthesis correspond to the catalytic domain and carbohydrate binding modules, respectively, in comparison with *Ct*XynGH30), B) MSA analysis of linker region displaying variation in sequence composition, however for better clarity, the *Hc*GH30 sequence was removed and C) Molecular architecture of *Ct*XynGH30 based on amino acid sequence.

4.3.2. Secondary structure analysis of *Ct*XynGH30 by PSIPRED and CD

The secondary structure prediction analysis of *Ct*XynGH30 carried out by the PSIPRED software, showed 28.25% α-helices, 35.50% β-sheets and 39% random coils (Table 4.1). The secondary structure analysis of *Ct*XynGH30 by Circular Dichroism (CD) using the K2D3 server revealed that it contains 28.25% α-helices, 40.5% β-sheets and 31.25% random coils (Fig. 4.2, Table 4.1). The results of PSIPRED prediction for the secondary structural element of *Ct*XynGH30 corroborated with those of CD analysis.

Table 4.1. Secondary structure elements of *Ct*XynGH30.

Secondary structure element	Percentage predicted by CD spectrum	Percentage predicted by PSIPRED
α-helix	28.25	25.50
β-strand	40.50	35.50
Random coils	31.75	39.0

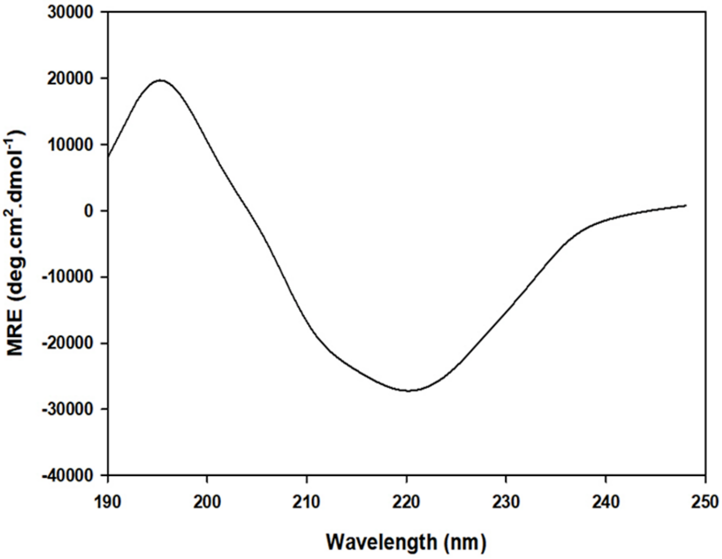


Fig. 4.2. Far UV CD spectrum of *Ct*XynGH30.

4.3.3. The protein melting analysis of *Ct*XynGH30

The melting curve of *Ct*XynGH30 displayed a single peak at 80°C. As the temperature was increased the protein started unfolding at 72ºC and completely melted at 80ºC (Fig. 4.3). In our previous report, it was observed that 1 mM Li^+ ion enhances the *Ct*XynGH30 activity by 25% (Verma et al., 2016). However, the effect of 1 mM Li^+ on *Ct*XynGH30 melting indicated that Li^+ ions are involved only in activity enhancement and not for providing stability (Fig. 4.3). The presence of 1 mM Ca^{2+} ions shifted the melting peak of *Ct*XynGH30 towards a higher temperature of 85ºC. Thus, the presence of Ca^{2+} ions increased the stability of *Ct*XynGH30 (Fig. 4.3). The treatment of *Ct*XynGH30 with 1 mM EDTA, which removes inherent divalent metal ions, did not show any enhanced effect on the melting temperature and further supported the biochemical studies of Ca^{2+} ions effect on *Ct*XynGH30 reported previously (Verma et al., 2016). The melting curves of full length modular *Ct*Man26F (Ghosh et al., 2013) and *Ct*43Ara*f* (Ahmed et al., 2013) showed that the catalytic domains have a melting peak at 80°C, whereas, carbohydrate binding module CBM35 of *Ct*Man26F has a peak at 50°C and CBM6s of *Ct*43Ara*f* at 55°C. The melting curve analysis of these proteins revealed that the two modules, the catalytic and carbohydrate binding modules, melt independently. However, the melting curve analysis of *Ct*CBM6 showed a single melting temperature at 82°C confirming it to be more thermostable (Verma et al., 2015). The presence of *Ct*CBM6 domain may thus provide greater stability to *Ct*XynGH30 resulting in a single melting peak.

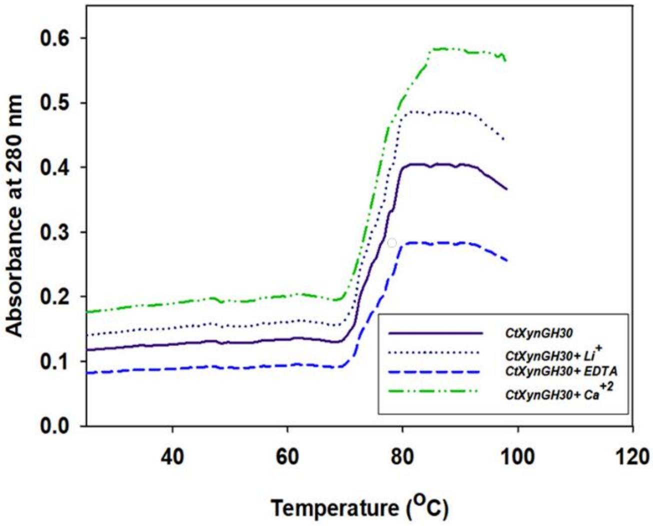


Fig. 4.3. Melting curve analysis of *Ct*XynGH30 displaying melting curve of control with no additive (—), in presence of 1 mM Li^+ ions (······), in presence of 1 mM Ca^{2+} ions (—··—) of 1 mM EDTA (---).

4.3.4. Role of *Ct*CBM6 in *Ct*XynGH30

Isothermal titration calorimetric (ITC) analysis of *Ct*Xyn30A and *Ct*CBM6 revealed no binding (Fig. 4.4A), which confirmed that within *Ct*XynGH30 there are no significant intermodular surface:surface interactions between the two domains. The mass spectrometric analysis of *Ct*Xyn30A:*Ct*CBM6 (1:1 molar ratio) displayed two separate peaks indicating the two individual molecular masses of monomeric *Ct*Xyn30A (45.15 kDa) and *Ct*CBM6 (15.5 kDa) (Fig. 4.4 B, C & D). If the two modules had interacted *in trans*, then there would have been a single peak corresponding to the combined molecular mass of the two. In contrast, the full-length protein, *Ct*XynGH30 provided a single peak with the molecular mass of 61.5 kDa (Fig. 4.4 E) corresponding to the combined molecular mass of the two modules. The mass spectrometry results corroborated the ITC results confirming that *in trans*, there are no significant surface to surface interactions between the enzyme catalytic domain and CBM.

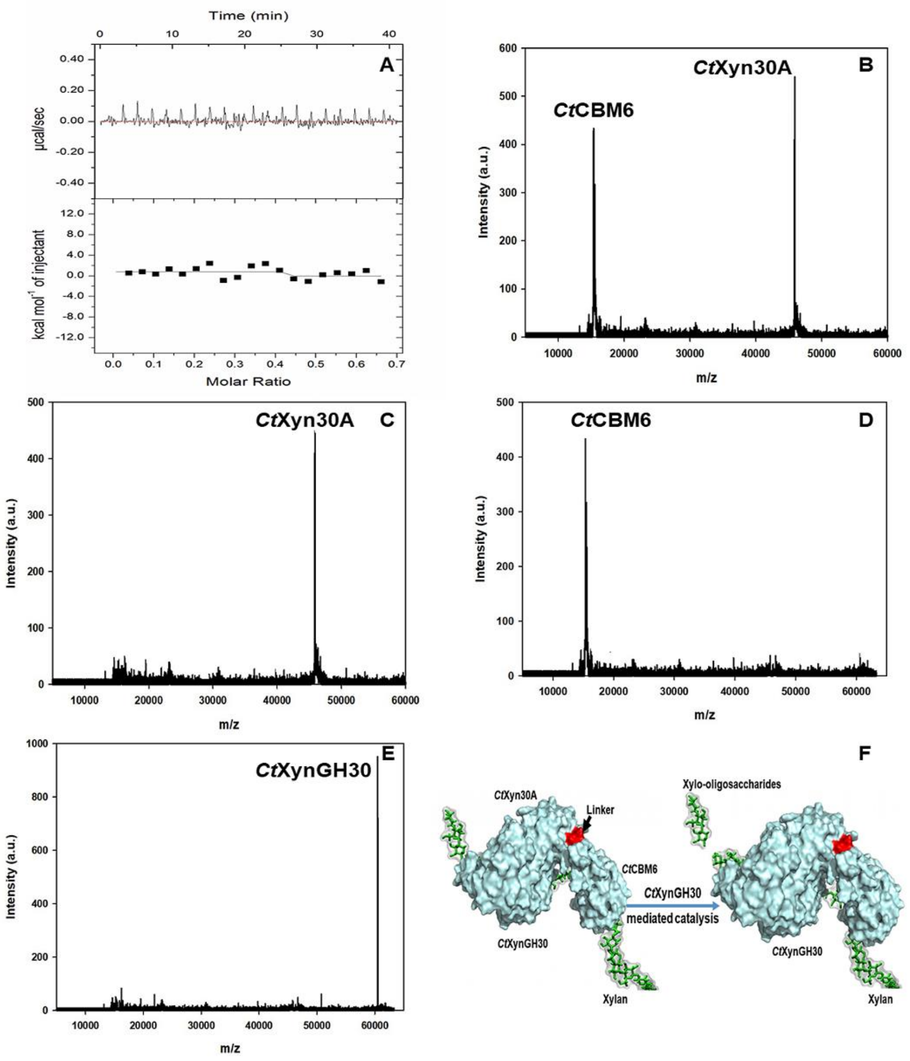


Fig. 4.4. A) Isothermal titration calorimetric analysis of *Ct*Xyn30A and *Ct*CBM6, B) mass spectrometric analysis of *Ct*Xyn30A:*Ct*CBM6, C) Mass spectrometric analysis of *Ct*Xyn30A, D) mass spectrometric analysis of *Ct*CBM6 E) mass spectrometric analysis of *Ct*XynGH30 and F) Probable mechanism of *Ct*XynGH30 catalysis.

The effect of connecting the two *Ct*XynGH30 functional domains was evaluated by determining the specific activities of *Ct*XynGH30, *Ct*Xyn30A and *Ct*Xyn30A: *Ct*CBM6 as displayed in Table 4.2 The full length enzyme, *Ct*XynGH30, displayed maximum activity against beechwood xylan (35.5±0.4 U/mg), followed by xylan (Mw., 20000-30000) (33±0.6 U/mg), birchwood xylan (32.7±0.3 U/mg) and 4-O-methylglucuronoxylan (30.2±0.5 U/mg). In contrast, the mixture of *Ct*Xyn30A and

*Ct*CBM6 *in trans*, displayed activity against beechwood xylan (28.8±0.2 U/mg), followed by xylan (Mw., 20000-30000) (27.3±0.2 U/mg), birchwood xylan (26.3±0.3 U/mg) and 4-O-methylglucuronoxylan (24.2±0.3 U/mg) which replicates the activity profile of *Ct*XynGH30. However, levels of activity of full-length *Ct*XynGH30 were higher than, those observed for the single domain *Ct*Xyn30A. Therefore, it can be concluded that the linker peptide is necessary to connect the *Ct*CBM6 module with the *Ct*Xyn30A module giving rise to the full-length *Ct*XynGH30 module orienting the two domains against the substrate resulting in enhancement of enzyme activity of *Ct*XynGH30. In the modelled structure of *Ct*XynGH30, the substrate binding sites on both *Ct*Xyn30A and CBM face the same direction (Fig. 4.4F). Thus, all these results confirm that the carbohydrate binding module helps in binding the substrate and probably functions as a sliding clamp that provides the support to the catalytic module for efficient hydrolysis of the substrate by proximity (Beckham et al., 2014).

Table 4.2. Activity analysis of *Ct*XynGH30, *Ct*Xyn30A: *Ct*CBM6 and *Ct*Xyn30A.

Substrates	**Enzyme Activity (U/mg**		
	***Ct*XynGH30**	***Ct*Xyn30A:*Ct*CBM6**	***Ct*Xyn30A**
Beechwood xylan	35.5±0.4	28.9±0.2	29.9±0.4
Xylan (Mw 20000-30000)	33.6±0.3	27.3±0.2	28.4±0.3
Birchwood xylan	32.8±0.3	26.3±0.3	26.8±0.6
4-O-methyl glucuronoxylan	30.3±0.5	24.2±0.3	25.1±0.5

.3.5. Solution structure analysis of *Ct*XynGH30 by Small Angle X-ray Scattering

The Small Angle X-ray Scattering (SAXS) analysis of *Ct*XynGH30 was carried out in order to determine its size and molecular shape in solution. The comparative SAXS pattern of *Ct*XynGH30 at different concentration showed that the concentration effects were negligible. Small aggregation effect was observed in SAXS data collected at 2.5 mg/mL and 5 mg/mL concentrations. Therefore, initial 3 data points at each concentration were removed by reducing the S_{min} value to 0.03 nm^{-1} for the removal of small aggregation effect. The SAXS pattern of *Ct*XynGH30 at a concentration of 1.5 mg/mL is shown in Fig. 4.5A and Table 4.3. The Guinier plot analysis of *Ct*XynGH30 indicated a linear behaviour in low q region, which confirms that the protein sample preparation is in a monodisperse state. The radius of gyration (Rg) of 3.4 ± 0.01 nm was computed from the slope of the Guinier area (defined by criteria S_{max}<Rg/1.3) (Fig. 4.5B). The pair-distance distribution function P(R) curve obtained by indirect Fourier transform of the scattering curve showed asymmetric profile typical for globular macromolecules (Fig. 4.5C). The maximum intra particle distance (Dmax) and Rg obtained from P(R) curves of *Ct*XynGH30 are 18 and 3.6 nm, respectively. The Dmax is five times greater than the Rg, which confirms that the *Ct*XynGH30 is extended and not fully globular.

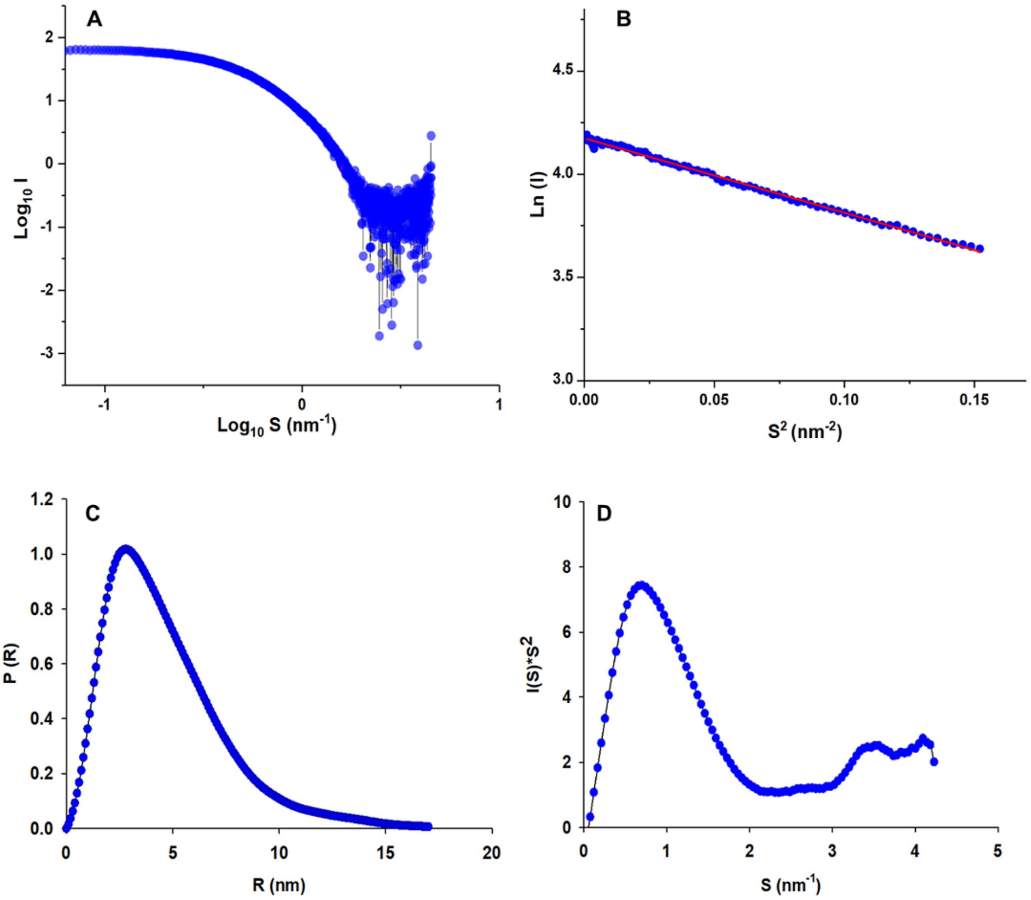


Fig. 4.5. Small angle X-ray scattering analysis of *Ct*XynGH30, A) Scattering pattern of *Ct*XynGH30, B) Guinier analysis of *Ct*XynGH30, C) P(r) plot analysis of *Ct*XynGH30 and D) Kratky analysis of *Ct*XynGH30.

The flexibility and global compactness of *Ct*XynGH30 in solution were evaluated by Kratky plot analysis using SAXS experimental data. The Kratky plot showed a bell-shaped peak which did not decay to zero at higher S values, confirming some degree of flexibility in *Ct*XynGH30 between the two modules (Fig. 4.5D). The molecular weight (M_w) of *Ct*XynGH30 was estimated to be 63 kDa by SAXS MoW web tool which is in close agreement to the theoretical M_w 60 kDa, computed by its amino acid sequence using ExPASY Protparam tool.

Table 4.3. SAXS data collection parameters and derived parameters of *Ct*XynGH30.

Data-collection parameters	*Ct*XynGH30		
Beamline	BM29-BioSAXS, ESRF		
Wavelength (Å)	0.99 Å		
Q range (nm^{-1})	0.023-4.5		
Exposure time (Seconds)	10x10		
Temperature (°C)	4		
Protein Concentration mg/mL	1.5	2.5	5
Structural parameter			
Q range (nm^{-1}) used for Rg analysis	0.11-0.33	0.16-0.37	0.13-0.35
I(0) au from Guinier	68.53±0.07	73.52±0.08	71.93±0.05
Rg nm from Guinier	3.4±0.01	3.6±0.01	3.62±0.07
Rg nm from P(r)	3.6	3.8	3.8
D_{max} (nm)	18	17	17.5
Porod volume estimate (nm^3)	90.6	98.41	98.65
Molecular Mass Determination			
Theoretical molecular mass (kDa)	63		
Molecular Mass from SAXSMOW (kDa)	60	60.5	60.8
Modeling parameters			
NSD	1.57	1.43	1.18
Resolution (Å)	24.5	26.3	39
Software employed			
Data processing	Primus	Primus	Primus
P(r) function calculation	GNOM	GNOM	GNOM
Ab initio modeling	GASBOR	GASBOR	GASBOR
Validation and averaging	DAMAVER	DAMAVER	DAMAVER
Flexible region modeling	CORAL	CORAL	CORAL
Refinement of Flexible model	SREFLEX	SREFLEX	SREFLEX
Ensemble Conformation modelling	BILBOMD	BILBOMD	BILBOMD
Structure superposition	SUPCOMB	SUPCOMB	SUPCOMB
3-D graphical representation	PyMOL	PyMOL	PyMOL

The overall molecular shape of *Ct*XynGH30 was constructed by *ab initio* method from its scattering profile using the GASBOR program (Svergun et al., 2001). The total reconstructed 50 models were averaged to obtain the most typical model by DAMAVER (Volkov et al., 2003). The normalized spatial discrepancy (NSD) value of a set of models reconstructed by GASBOR for *Ct*XynGH30 was 1.57. The molecular shape constructed by GASBOR showed a single, boot-shaped envelop (Fig. 4.6 A &

B). The reconstructed molecular shape of *Ct*XynGH30 was overlapped with its derived SREFLEX refined structure. The overall appearance of the GASBOR generated molecular shape of *Ct*XynGH30 matched well with its SREFLEX refined structure (Fig. 4.6 C & D). The experimental SAXS data were also fitted by the scattering profile calculated from the SREFLEX refined *Ct*XynGH30 structure using CRYSOL (Svergun et al., 1995). The *Ct*XynGH30 structure yielded Rg = 3.1 nm which is smaller than the experimental value. The fit showed a discrepancy of $\chi^2 = 3.8$ confirming that the overall shape of *Ct*XynGH30 is reasonably close to that of the monomeric form. The Xyn30d (Chain A) structure yielded Rg = 2.8 nm and $\chi^2 = 5.8$. The analysis by BILBOMD displayed the Rg values between 3.3-3.6 nm, even after taking a large range (3.0-4.0 nm) of Rg value. Thus, the BILBOMD analysis further confirmed that the *Ct*XynGH30 structure possesses limited flexibility. The Rg and Dmax value for best fit model was found to be 3.38 nm and 17.5 nm, respectively.

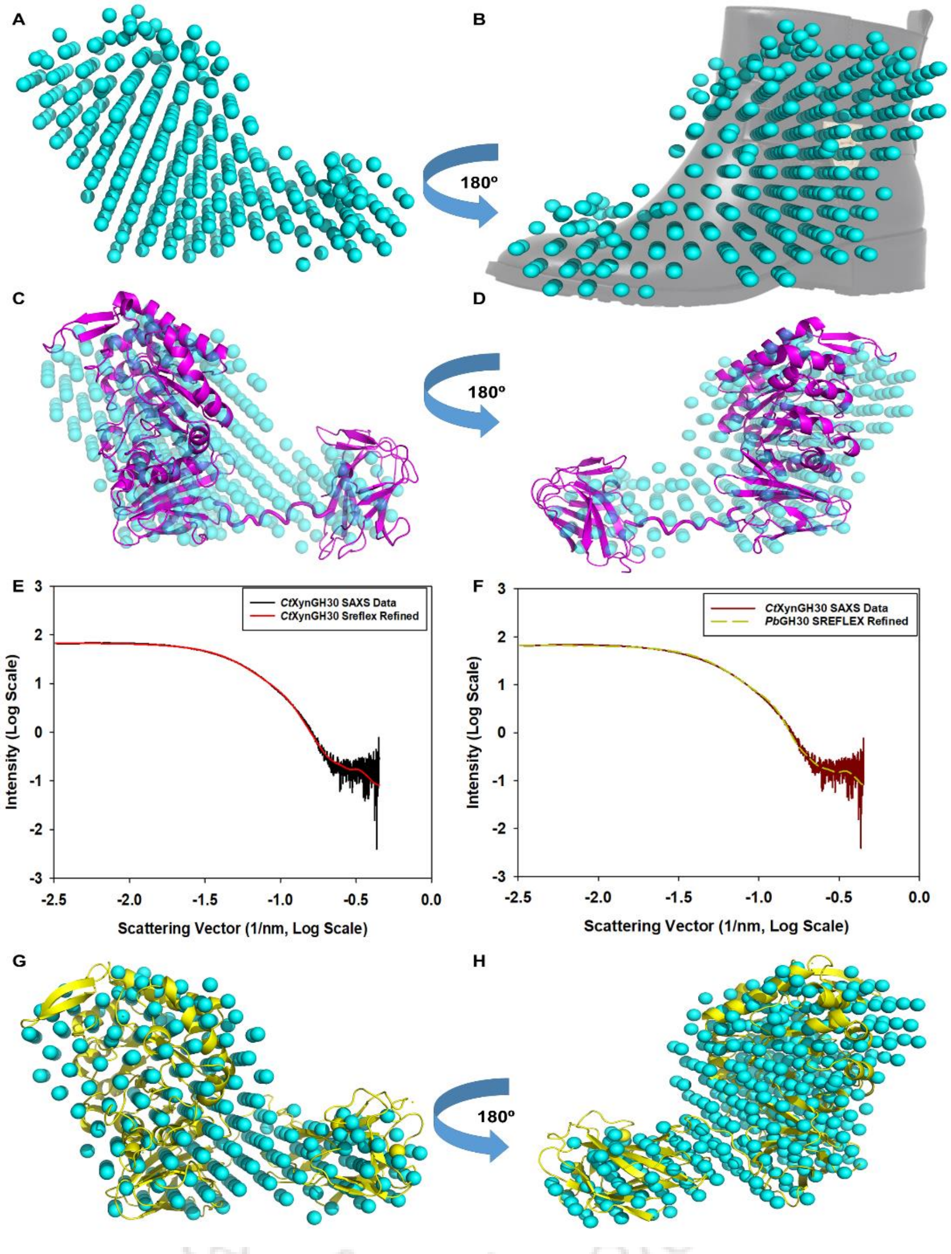


Fig. 4.6. *ab initio* model of *Ct*XynGH30 generated by constructing independent dummy atom residue models using GASBOR derived and averaging with DAMAVER at 90° and 180° rotations (A & B). Superposition of *ab initio* and SREFLEX refined *Ct*XynGH30 at 90° and 180° rotations (C & D). CRYSOL fitting of experimentally collected SAXS data on *Ct*XynGH30 (E) & Xyn30d (F) and Superposition of *ab initio* and SREFLEX refined Xyn30d at 90° and 180° rotations (G & H). The images were generated using PYMOL (the PyMOL molecular graphics system, version 2.20; Schrödinger, LLC).

The theoretical scattering profile of the best model obtained from BILBOMD analysis displayed that it is in good agreement with the experimental scattering data of *Ct*XynGH30 giving $\chi2$ = 1.59. The BILBOMD simulated model displayed a significant improvement over the CORAL derived model fit that gave $\chi2$ = 3.58. This shows that, although the solution structure of *Ct*XynGH30 has a similar molecular organisation to the homologous Xyn30d crystal structure, it is more flexible and extended, but not fully extended and completely flexible, as seen in some proteins such as cellobiohydrolase from *Trichoderma harzianum* (Lima et al., 2013). This is despite the fact that there are no significant inter-modular interactions in *Ct*XynGH30. Even though *Ct*XynGH30 and Xyn30d have equal linker length, their composition is very different (Fig. 4.1B). In Xyn30d structure, the first four residues (ANLS) are part of the catalytic domain and then there is an almost fully extended stretch GGNSGGG, with a distance between the Cα-atoms of the first and last Gly of ~21 Å. This distance corresponds roughly to the length of two extended cellobiose units. Thereafter, a highly flexible loop (NVN), which is highly disordered in 5 out of 8 Xyn30d molecules in the asymmetric unit (Sainz-Polo, et al., 2014). In *Ct*XynGH30, the linker is expected to be more flexible on the *Ct*Xyn30A side with the sequence (GGGSSE) than in Xyn30d and the second part of the linker (PQIPVERN) to be less flexible and probably forming a loop or coil, especially with the presence of two prolines. This is also reflected in the SREFLEX refined model (Fig. 4.6D). The difference in linker sequence of the two enzymes could be due to the fact that *Ct*XynGH30 is a cellulosomal protein and will have a linker at the CBM6 C-terminal connecting it to a dockerin module which in turn binds a cohesin module making it part of the cellulosome, whereas, Xyn30d is a free enzyme.

The comparison of the linker region with other homologous modular proteins shows their linkers to be of similar or double in length (Fig. 4.1A). This comparison corroborates well with the results of Ruiz et al. (2016) on the role of linkers in endo-glucanase GH5 (*Bs*Cel5A) from *Bacillus subtilis*, where they show a fine balance between flexibility (needed for substrate recognition and binding) and spatial arrangement between the modules, which is required for achieving the optimal catalysis.

4.4. Conclusion

The conformational flexibility, molecular organization and protein stability of *Ct*XynGH30 in solution were investigated. The secondary structure analysis of *Ct*XynGH30 by CD displayed the presence of 28.25% α-helices and 40.5% β-sheets. The melting analysis showed that the Ca^{2+} ions provide extra thermal stability to the *Ct*XynGH30 structure. The ITC and mass spectrometric analyses of individual modules, *Ct*Xyn30A and *Ct*CBM6, showed that even though there is no binding interaction between them, the two modules are required to be linked together for optimal function. The mixture of *Ct*Xyn30A and *Ct*CBM6 displayed lower xylanase activity as compared to the full-length *Ct*XynGH30 protein. The lower activity of *Ct*Xyn30A alone and the mixture (*Ct*Xyn30A and *Ct*CBM6) may be due to the absence of linker region, which may be functioning as a sliding hinge between the two modules, thereby facilitating the binding to the substrate and bringing about enhanced catalytic activity of the full-length enzyme. SAXS analysis of *Ct*XynGH30 revealed a boot-shaped, fully folded structure in solution. The SREFLEX refined model of *Ct*XynGH30 superposed well with the *ab initio* modelled structure. The optimal length and composition of linker peptide is crucial for the catalytic activity. These results confirmed that the two domains are required to be attached via a linker peptide, but with enough flexibility to maintain a specific spatial organisation to achieve optimal biological function.

4.5. References

Ahmed, S., Luis, A.S., Bras, J.L.A., Ghosh, A., Gautam, S., Gupta, M.N., Fontes, C.M.G.A., Goyal, A., 2013. A Novel α-L-Arabinofuranosidase of Family 43 Glycoside Hydrolase (*Ct*43Araf) from *Clostridium thermocellum*. *PLoS One* 8, 1–10.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.*, *215*(3), 403-410.

Beckham, G.T., Ståhlberg, J., Knott, B.C., Himmel, M.E., Crowley, M.F., Sandgren, M., Sørlie, M., Payne, C.M., 2014. Towards a molecular-level theory of carbohydrate processivity in glycoside hydrolases. *Curr. Opin. Biotechnol*. 27, 96–106.

Bendtsen, J.D., Kiemer, L., Fausbøll, A., Brunak, S., 2005. Non-classical protein secretion in bacteria. *BMC Microbiol.* 5, 1–13.

Biely, P., M. Vršanská, M. Tenkanen, and D. Kluepfel., 1997. Endo-β-1, 4-xylanase families: differences in catalytic properties. *J. Biotechnol.* 57, no. 1-3, 151-166.

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

Collins, T., Gerday, C., Feller, G., 2005. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol. Rev.*, 29, 3-23.

Correia, M.A.S., Mazumder, K., Brás, J.L.A., Firbank, S.J., Zhu, Y., Lewis, R.J., York, W.S., Fontes, C.M.G.A., Gilbert, H.J., 2011. Structure and function of an arabinoxylan-specific xylanase. *J. Biol. Chem.* 286, 22510–22520.

Czjzek, M., Bolam, D.N., Mosbah, A., Allouch, J., Fontes, C.M.G.A., Ferreira, L.M.A., 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. *J. Biol. Chem.* 276, 48580–48587.

Fischer, H., De Oliveira Neto, M., Napolitano, H.B., Polikarpov, I., Craievich, A.F., 2010. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr*. 43, 101–109.

Fontes, C.M.G.A., Gilbert, H.J., 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem*. 79, 655–681.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., Svergun, D.I., 2017. ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr*. 50, 1212–1225.

Freire, F., Verma, A., Bule, P., Alves, V.D., Fontes, C.M.G.A., Goyal, A., Najmudin, S., 2016. Conservation in the mechanism of glucuronoxylan hydrolysis revealed by the structure of glucuronoxylan xylanohydrolase (*Ct*Xyn30A) from *Clostridium thermocellum*. *Acta Crystallogr. Sect. D Struct. Biol*. 72, 1162–1173.

Gallardo, Ó., Fernández-Fernández, M., Valls, C., Valenzuela, S.V., Blanca Roncero, M., Vidal, T., Díaz, P., Javier Pastor, F.I., 2010. Characterization of a family GH5 xylanase with activity on neutral oligosaccharides and evaluation as a pulp bleaching aid. *Appl. Environ. Microbiol*. 76, 6290–6294.

Ghosh, A., Luís, A.S., Brás, J.L.A., Fontes, C.M.G.A., Goyal, A., 2013. Thermostable Recombinant β-(1→4)-Mannanase from *C. thermocellum*: Biochemical characterization and manno-oligosaccharides production. *J. Agric. Food Chem*. 61, 12333–12344.

Harris, D., DeBolt, S., 2010. Synthesis, regulation and utilization of lignocellulosic biomass. *Plant Biotechnol. J*. 8, 244–262.

Jönsson, L.J., Martín, C., 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour. Technol*. 199, 103–112.

Konarev, P. V., Petoukhov, M. V., Volkov, V. V., Svergun, D.I., 2006. ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr*. 39, 277–286.

Kozin, M. B., and Svergun, D. I., 2004. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr*. 34, 33–41.

Lima, L.H.F., Serpa, V.I., Rosseto, F.R., Sartori, G.R., de Oliveira Neto, M., Martínez, L., Polikarpov, I., 2013. Small-angle X-ray scattering and structural modeling of full-length: Cellobiohydrolase I from *Trichoderma harzianum*. *Cellulose* 20, 1573–1585.

Nelson, N. (1944). A photometric adaptation of the Somogyi method for the determination of glucose. *J. Biol. Chem.*, 153(2), 375-380.

Padilha, I.Q.M., Valenzuela, S. V., Grisi, T.C.S.L., Diaz, P., De Araújo, D.A.M., Javier Pastor, F.I., 2015. A glucuronoxylan-specific xylanase from a new *Paenibacillus favisporus* strain isolated from tropical soil of Brazil. *Int. Microbiol*. 17, 175–184.

Perez-Iratxeta, C., Andrade-Navarro, M.A., 2008. K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Struct. Biol*. 8, 1–5.

Pelikan, M., Hura, G.L. and Hammel, M., 2009. Structure and flexibility within proteins as identified through small angle X-ray scattering. General Physiology and Biophysics, 28(2), p.174.

Perlman, D., Halvorson, H.O., 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol*. 167, 391–409.

Ruiz, D.M., Turowski, V.R., Murakami, M.T., 2016. Effects of the linker region on the structure and function of modular GH5 cellulases. *Sci. Rep*. 6, 1–13.

Saini, J.K., Saini, R., Tewari, L., 2015. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech* 5, 337–353.

Sainz-Polo, M.A., Valenzuela, S.V., González, B., Pastor, F.I.J., Sanz-Aparicio, J., 2014. Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity. *J. Biol. Chem*. 289, 31088–31101.

Šali, A., Potterton, L., Yuan, F., van Vlijmen, H., Karplus, M., 1995. Evaluation of comparative protein modeling by MODELLER. *Proteins Struct. Funct. Bioinforma*. 23, 318–326.

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

Somogyi, M., 1945. Determination somogyi. *J. Biol. Chem.* 160, 61–68.

St. John, F.J., Rice, J.D., Preston, J.F., 2006. Characterization of XynC from *Bacillus subtilis* subsp. subtilis strain 168 and analysis of its role in depolymerization of glucuronoxylan. *J. Bacteriol.* 188, 8617–8626.

Svergun, D.I., 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr*. 25, 495–503

Svergun, D., Barberato, C., Koch, M.H., 1995. CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr*. 28, 768–773.

Svergun, D. I., Petoukhov, M. V., and Koch, M. H. 2001. Determination of domain structure of proteins from X-Ray solution scattering. *Biophys. J*. 80, 2946–53.

Verma, A.K., Goyal, A., 2014. *In silico* structural characterization and molecular docking studies of first glucuronoxylan-xylanohydrolase (Xyn30A) from family 30 glycosyl hydrolase (GH30) from *Clostridium thermocellum*. *Mol. Biol*. 48, 278–286.

Verma, A.K., Bule, P., Ribeiro, T., Brás, J.L.A., Mukherjee, J., Gupta, M.N., Fontes, C.M.G.A., Goyal, A., 2015. The family 6 carbohydrate binding module (*Ct*CBM6) of glucuronoxylanase (*Ct*XynGH30) of *Clostridium thermocellum* binds decorated and undecorated xylans through cleft A. *Arch. Biochem. Biophys*. 575, 8–21.

Verma, A.K., Goyal, A., 2016. A novel member of family 30 glycoside hydrolase subfamily 8 glucuronoxylan endo-β-1,4-xylanase (*Ct*XynGH30) from *Clostridium thermocellum* orchestrates catalysis on arabinose decorated xylans. *J. Mol. Catal. B Enzym*. 129, 6–14.

Volkov, V. V, Svergun, D.I., 2003. Uniqueness of *ab initio* shape determination in small-angle scattering. 860–864.

Vršanská, M., Kolenová, K., Puchart, V., Biely, P., 2007. Mode of action of glycoside hydrolase family 5 glucuronoxylan xylanohydrolase from *Erwinia chrysanthemi*. *FEBS J*. 274, 1666–1677.

Webb, B., Sali, A., 2014. Protein Structure Modeling with MODELLER, in: Kihara, D. (Ed.), Protein Structure Prediction. Springer New York, New York, NY, pp. 1–15. https://doi.org/10.1007/978-1-4939-0366-5_1

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 35, 407-410.

Zdobnov, E.M., Apweiler, R., 2001. InterProScan - An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848.

Zhong, R., Teng, Q., Lee, C., Ye, Z.H., 2014. Identification of a disaccharide side chain 2-O-α-D-galactopyranosyl-α-D-glucuronic acid in arabidopsis xylan. *Plant Signal. Behav*. 9, 2–6.

Chapter 5

SAXS and comparative modeling based structure analysis of endo-β-1,4-xylanase a family 10 glycoside hydrolase from *Pseudopedobacter saltans*

5.1. Introduction

The growing interest in the utilization of lignocellulosic biomass as a renewable source for the production of biochemicals and bioenergy has become relevant for economic and environmental reasons (Jönsson and Martín, 2016.). The availability of lignocellulosic biomass at economic price and accessibility makes it suitable for the industrial applications (Harris and DeBolt, 2010.). Lignocellulosic biomass predominantly contains lignin and polysaccharides existing as a complex lignocellulosic matrix (Saini *et al*., 2010). Heteroxylan accounts for approximately, one-third of the total cell wall polysaccharides as a renewable organic carbon source (Collins *et al*., 2015). Heteroxylans are composed of a backbone of β-1,4 linked xylose residues often with decorations of arabinofuranose residues at O-2 or O-3 and 4-O-methyl-D-glucuronic acid at O-2 (Zhong et al., 2014). The substitutions and therefore the composition of xylan varies widely depending on the source. Microbial xylanolytic enzymes play an important role in bioconversion of heteroxylans into valued products such as bioethanol, xylooligosaccharides as prebiotics and therapeutics (Padilha et al., 2015). The majority of xylanases studied so far are grouped into family 10 and 11 glycoside hydrolase (GH10 and GH11) (Van Gook *et al.*, 2012). The xylanases reported in family 10 glycoside hydrolase have higher molecular weight and lower pI values than those reported in family 11 glycoside hydrolase (Meng *et al.,* 2015). Moreover, GH10 xylanases are capable in hydrolyzing the xylosidic linkage of the highly substituted xylan main chain, near the substitution, whereas GH11 xylanases cannot hydrolyze highly substituted xylan (Valenzuela *et al.*, 2012).

The xylanases belonging to family 5 and 8 glycoside hydrolase, specifically, hydrolyze arabinose substituted xylans (Correia *et al.*, 2012). So far, only a few mesophilic bacterial xylanases belonging to family 10 glycoside hydrolase have been identified and characterized (Fontes at al., 2000; Gallardo *et al.*, 2010).

Pseudopedobacter saltans comb. nov. is a mesophilic bacterium having heparinolytic and chondrotinolytic activity (Rani and Goyal, 2016). The sequenced genome of this microbe allows us to further explore various other potential enzymes such as xylanase and cellulase that can be utilized in bioprocess industries. The endo-β-1,4-xylanase (EC 3.2.1.8) of a family 10 glycoside hydrolase (*Ps*GH10A) (GeneBank: Acc No. ADY53246.1) from *P. saltans* is composed of 333 amino acid residues. The first 30 amino acid residues of *Ps*GH10A form a signal peptide and remaining amino acid residues fold to form the catalytic domain. The gene encoding *Ps*GH10A was cloned, expressed and purified. The enzyme was optimally active between 30-40°C and at the acidic pH range, 4.5–6.5. In the present study, the three-dimensional structure of *Ps*GH10A from *Pseudopedobacter saltans* was modeled by comparative modeling and the solution state was determined by Small Angle X-ray Scattering (SAXS). The secondary structure elements were determined by Circular Dichroism (CD), RaptorX property and ProMotif analysis. The residues involved in catalysis by *Ps*GH10A were discovered by docking the enzyme with various xylooligosaccharides.

5.2. Material and methods

5.2.1. Amino acid sequence retrieval and analysis

The protein sequence of family 10 glycoside hydrolase (*Ps*GH10A) from *Pseudopedobacter saltans comb. nov.* (UniProt ID: F0S6Y3) was retrieved from the UniProt database (http://www.uniprot.org/uniprot/F0S6Y3). The molecular mass analysis of *Ps*GH10A was determined by Protparam tool (https://web.expasy.org/protparam/). The PSI-BLAST tool against PDB database was used to search the putative domain (Margelevičius and Venclovas, 2005). The amino acid sequences of endo-β-1,4-xylanases; XynA2 (*Geobacillus stearothermophilus*) (Solomon et al., 2007), XynB (*Paenibacillus barcinonensis*) (Gallardo et al., 2010), *Cb*Xyn10B (*Caldicellulosiruptor bescii* DSM 6725) (Zhang et al., 2016), XynGR40 (*Aegilops speltoides*) (Zheng et al., 2016) and *Ts*XylA (*Thermoanaerobacterium saccharolyticum*) (Han et al., 2013) were obtained from the PDB database (http://www.rcsb.org/) (Rose et al., 2011). The sequences were aligned Multiple Sequence Alignment (MSA) was carried out using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al., 2011).

5.2.2. Comparative modeling, refinement and structure assessment of *Ps*GH10A

The Modeller v9.18 program was used to build the three dimensional structure of *Ps*GH10A (Eswar et al., 2008). Multiple templates based comparative modeling approach was employed to build three dimensional structure of *Ps*GH10A. The homologous xylanase sequences and structures of XynA2 (PDB ID: 2Q8X), XynB (PDB ID: 3EMC), *Cb*Xyn10B (PDB ID: 4L4O), XynGR40 (PDB ID: 5AY7) and TsXylA (PDB ID: 3W24) were selected as templates. The multiple sequence alignment of these sequences was generated using ‘salign( ) command and then the *Ps*GH10A sequence was aligned with the template structure sequences. The 50 independent models of *Ps*GH10A were built by using ‘automodel’ class command. The resulting 50 models were assessed based on their Discrete Optimized Protein Energy (DOPE) score (Shen and Sali, 2006).

5.2.3. Energy minimization and assessment of *Ps*GH10A structure

The *Ps*GH10A modeled structure with the lowest DOPE score was then chosen for the energy minimization by AMBER ff14SB force field through minimize structure command in UCSF Chimera (Pettersen et al., 2004). The stereochemical properties of the energy-minimized structure were analyzed using PROCHECK (Laskowski et al., 1993). Statistical Z-score deviation of the predicted model from high-resolution PDB deposited structures was calculated using ProSA-web integrated web server (https://prosa.services.came.sbg.ac.at/prosa.php) (Wiederstein and Sippl, 2007). The overall minimized *Ps*GH10A model with its amino acid sequence was also further validated by ERRAT plot (Colovos and Yeates, 1993) and Verify-3D (Lüthy et al., 1992). The *Ps*GH10A topology diagram was generated using PDBSum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html). The amino acid residues forming the active site for reaction mechanism was identified by superposing energy minimized *Ps*GH10A structure with its homologous xylanases structure and was analyzed by PyMol software (http://www.PyMOL.org).

5.2.4. Secondary structure analysis of *Ps*GH10A

The secondary structure of *Ps*GH10A was predicted by using RaptorX property prediction server (Wang et al., 2016). The gene encoding *Ps*GH10A from *Pseudopedobacter saltans* cloned, expressed and purified as ~42 kDa protein was used for secondary structure determination by Circular Dichroism (CD). 20 µM of purified

*Ps*GH10A in 1 ml of 50 mM sodium phosphate buffer (pH 6.0) was used for CD analysis. 20 µM of purified *Ps*GH10A in 1 ml of 50 mM sodium phosphate buffer (pH 6.0) was used for CD analysis. The Far-UV CD spectrum (190-240 nm) was recorded using quartz cuvette of path length 0.1 cm on spectro-polarimeter (J-815 Jasco Corporation, Tokyo) at 25°C. 5 scans of CD spectrum for *Ps*GH10A was recorded and averaged to determine the molar residual ellipticity (MRE deg cm^{2}dmol^{-1}) and plotted as a function of wavelength (Kelly et al., 2005). The MRE values of *Ps*GH10A were analyzed to determine the percentage of secondary structures by CONTIN programme (Sreerama and Woody, 2000) available on the DichroWeb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml). The secondary structure results obtained by CD were compared with RaptorX property prediction (Wang et al., 2016) and ProMotif prediction (Hutchinson and Thornton, 2008).

5.2.5. Molecular docking analysis of *Ps*GH10A

Molecular docking study of *Ps*GH10A was performed using AutoDock 4.2.6 linked with MGL tools 1.56 (Morris et al., 2009). The energy minimized *Ps*GH10A modeled structure was used for docking studies. 1,4-β-D-xylobiose, 1,4-β-D-xylotriose, 1,4-β-D-xylotetraose, 1,4-β-D-xylopentaose, arabinoxylobiose and arabinoxylotriose were used as the ligand for docking analysis. Ligand files were retrieved from PubChem database (http://pubchem.ncbi.nlm.nih.gov) and converted into PDB files by OpenBabel 2.3.2a software (O'Boyle et al., 2011). At first non-polar hydrogens were merged and gasteiger charges were assigned to *Ps*GH10A and ligands PDB files to save as PDBQT files. Grid box having the dimension of 68, 64 and 72 (*x*, *y*, *z* coordinates) with default grid point spacing (0.375 Å) was assigned around the active site. The docking simulation and conformational search were executed using Lamarckian

Genetic Algorithm (LGA). 100 cycles of GA were used for generating 100 different confirmation of ligands. These 100 ligand conformations were ranked by the energy of binding and best conformation showing lowest binding energy was selected to generate protein-ligand complex. The polar contacts and hydrophobic interactions were detected from protein-ligand complexes in UCSF Chimera and PyMOL (http://www.PyMOL.org). 2D interaction of protein and ligand were generated using Ligplot+ program (Laskowski and Swindells, 2011).

5.2.6. Protein melting analysis of *Ps*GH10A

The protein melting curve of *Ps*GH10A was constructed by plotting the change in the absorbance at 280 nm (A_{280}) and at different temperatures by using a double beam UV-Visible spectrophotometer with peltier (Varian, Cary 100-Bio) following the method described earlier (Ahmed et al., 2013). The purified *Ps*GH10A (50 µg/ml) in 1 ml of 50 mM sodium phosphate buffer, pH 7.5 was used for the analysis. *Ps*GH10A in quartz cuvettes (0.01 cm path-length) with matched buffer was kept in the spectrophotometer to achieve equilibration at 25°C for 20 min. A_{280} was measured by increasing the temperature from 25-100°C at 5°C/min by keeping the holding time 2 min, at an interval of 2°C, using the peltier temperature controller. Similarly, the melting curve experiment was performed in presence of 2 mM $MnCl_2$, $MgCl_2$, or $CaCl_2$, individually and in combination with EDTA. The melting profile was generated by plotting A_{280} versus temperature on Sigma plot v12.0.

5.2.7. Small Angle X-ray Scattering Analysis (SAXS) of *Ps*GH10A

The SAXS datasets were collected by using the SAXSpace instrument (Anton Paar GmbH, Graz Austria) to study the global shape of *Ps*GH10A. The SAXS pattern of *Ps*GH10A was recorded with a 1D CMOS Mythen detector (Dectris, Baden, Switzerland). The sample to detector distance was maintained at 31.7 cm. *Ps*GH10A solution at a concentration of 6, 8 and 12 mg/mL in 50 mM sodium phosphate buffer at pH 6.0 was used and the momentum transfer within the range of 0.14-5.99 nm ($q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle) was detected. Prior to SAXS data collection, the *Ps*GH10A samples and matched buffer were centrifuged at 22,000*g* for 45 min at 4°C. The incident radiation wavelength on line collimation system was set to 1.5 Å for exposing on 40 μL each *Ps*GH10A and matched buffer (50 mM sodium phosphate buffer, pH 6.0) filled in a quartz capillary of 1 mm diameter for 2 independent frames of 30 min. SAXS datasets collected on line collimation system for *Ps*GH10A different concentration and matched buffer were processed and desmeared using SAXStreat and SAXSquant software to represent scattering arising from point collimation. All processing conditions were identical for all datasets, and I(Q) profile of the respective buffer was subtracted from the I(Q) profile of the solution to obtain SAXS profile arising from the protein molecules in solution. The radiation damage or aggregation effects were checked by normalizing the scattering curves by the exposure time and protein concentration. Monodispersity was validated by Guinier analysis and indirect Fourier transform method in GNOM software was used for the radius of gyration (Rg) (Semenyuk and Svergun, 1991). The distance distribution function P(r) was analyzed by GNOM and the maximum diameter (Dmax) was assigned.

The molecular mass of *Ps*GH10A was estimated by SAXS MoW program using SAXS data (Fischer et al., 2010). 20 *ab initio* low-resolution envelopes were generated by DAMMIF (Franke et al., 2009), aligned and averaged using DAMAVER (Volkov and Svergun, 2003). SASRES program was used for the estimation of resolution of *ab initio* modeled structure (Tuukkanen et al., 2016). Theoretical scattering profiles of *ab* *initio* model and homology models were compared to experimental scattering profiles using CRYSOL (Svergun et al., 1995).

5.3. Results and Discussion

5.3.1. Sequence analysis of *Ps*GH10A

The PSI-BLAST analysis of endo-β-1,4-xylanase from *Pseudopedobacter saltans comb. nov.* confirmed the presence of 30 amino acid long signal peptide at N-terminal and family 10 glycoside hydrolase module (*Ps*GH10A) at C-terminal (Fig. 5.1). *Ps*GH10A displayed the amino acid sequence identity with previously characterized endo-β-1,4-xylanases *viz.* XynA2 (52%) from *G. stearothermophilus*, XynB (46%) from *P. barcinonensis*), *Cb*Xyn10B (48%) from *C. bescii* DSM 6725, XynGR40 (47%) from *A. speltoides* and *Ts*XylA (44%) from *T. saccharolyticum* (Table 5.1). The catalytic amino acid residues (Glu156 as Acid/Base and Glu263 as Nucleophile) were identified by multiple sequence alignment of *Ps*GH10A with above-mentioned sequences (Fig. 5.2). The conserved residues in MSA are shown in white with red color background and semi-conserved residues are highlighted in black with the yellow color background.

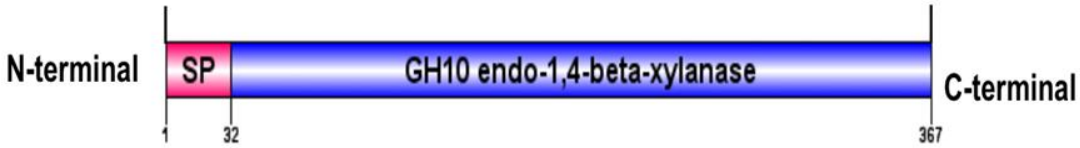
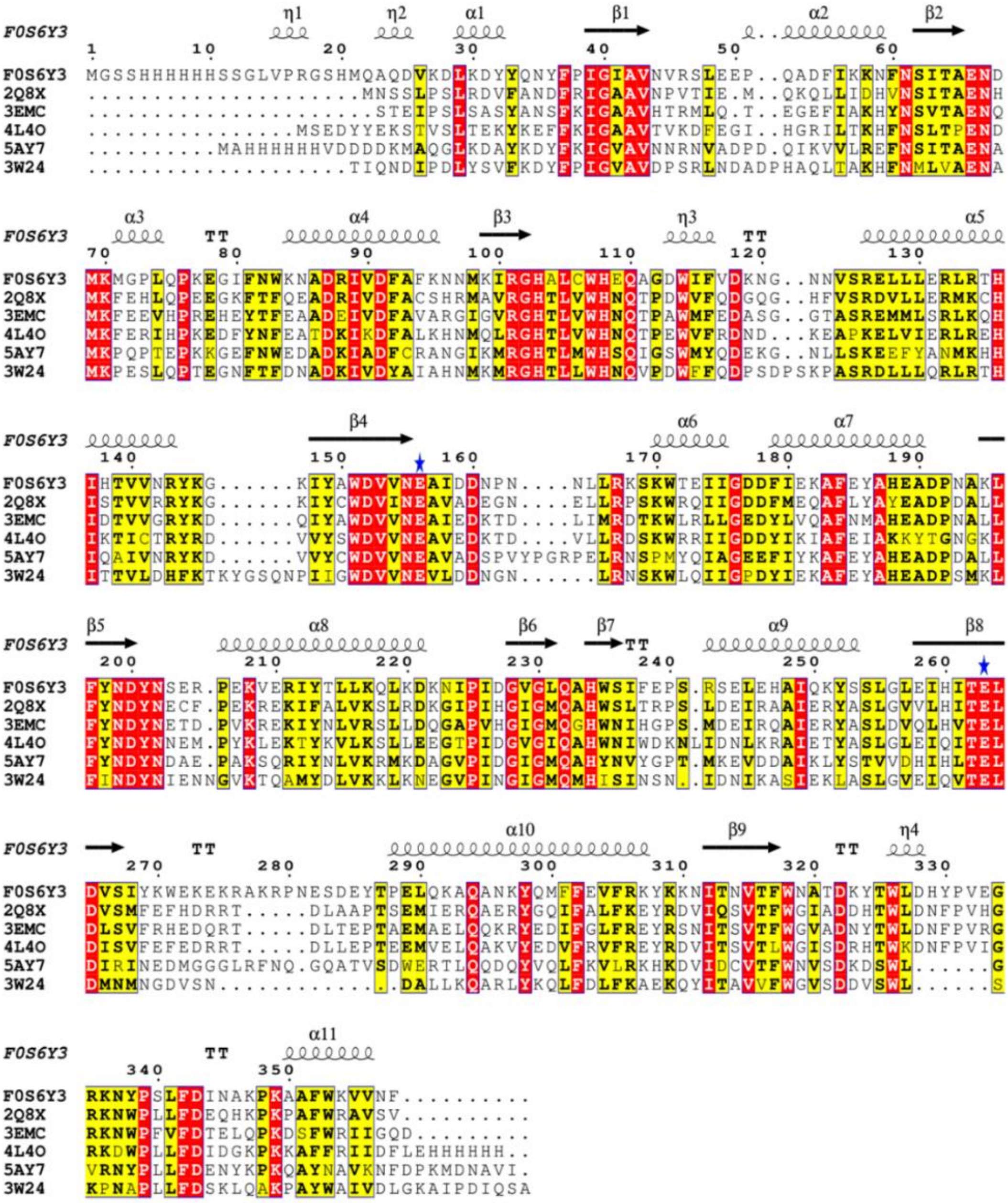


Fig. 5.1. A) Molecular architecture of *Ps*GH10A showing an N-terminal signal peptide followed by the catalytic domain. SP:Signal Peptide.

Table 5.1. PSI-BLAST analysis of *Ps*GH10A amino acid sequence.

Name	Enzyme name	Query Cover (%)	Identity (%)	E-value	Total Score	Resolution (Å)	PDB ID
G. stearothermophilus	XynA2	91	52	2e-167	471	1.45	2Q8X
P. barcinonensis	XynB	91	46	1e-155	441	2.1	3EMC
C. bescii DSM 6725	*Cb*Xyn10B	90	48	1e-150	429	2.05	4L4O
A. speltoides	XynGR40	90	47	2e-143	408	2.15	5AY7
T. saccharolyticum	*Ts*XylA	91	44	1e-143	411	1.35	3W24



```
F0S6Y3     1         10        20        30        40        50        60
F0S6Y3 MGSSHHHHHHSSGLVPRGSHMQAQDVKDLKDYYQNYFPIGIAVNVRSLEEP..QADFIKKNFNSITAEND
2Q8X   .....................MNSSLPSLRDVFANDFRIGAAVNPVTIE.M..QKQLLIDHVNSITAENH
3EMC   ......................STEIPSLSASYANSFKIGAAVHTRMLQ.T..EGEFIAKHYNSVTAENQ
4L4O   ................MSEDYYEKSTVSLTEKYKEFFKIGAAVTVKDFEGI..HGRILTKHFNSLTPEND
5AY7   ..........MAHHHHHHVDDDDKMAQGLKDAYKDYFKIGVAVNNRNVADPD.QIKVVLREFNSITAENA
3W24   ....................TIQNDIPDLYSVFKDYFPIGVAVDPSRLNDADPHAQLTAKHFNMLVAENA

F0S6Y3   70        80        90        100       110       120       130
F0S6Y3 MKMGPLQPKEGIFNWKNADRIVDFAFKNNMKIRGHALCWHEQAGDWIFVDKNG..NNVSRELLLERLRTH
2Q8X   MKFEHLQPEEGKFTFQEADRIVDFACSHRMAVRGHTLVWHNQTPDWVFQDGQG..HFVSRDVLLERMKCH
3EMC   MKFEEVHPREHEYTFEAADEIVDFAVARGIGVRGHTLVWHNQTPAWMFEDASG..GTASREMMLSRLKQH
4L4O   MKFERIHPKEDFYNFEATDKIKDFALKHNMQLRGHTLVWHNQTPEWVFRDND...KEAPKELVIERLREH
5AY7   MKPQPTEPKKGEFNWEDADKIADFCRANGIKMRGHTLMWHSQIGSWMYQDEKG..NLLSKEEFYANMKHH
3W24   MKPESLQPTEGNFTFDNADKIVDYAIAHNMKMRGHTLLWHNQVPDWFFQDPSDPSKPASRDLLLQRLRTH

F0S6Y3    140       150       160       170       180       190
F0S6Y3 IHTVVNRYKG......KIYAWDVVNEAIDDNPN....NLLRKSKWTEIIGDDFIEKAFEYAHEADPNAKL
2Q8X   ISTVVRRYKG......KIYCWDVINEAVADEGN....ELLRPSKWRQIIGDDFMEQAFLYAYEADPDALL
3EMC   IDTVVGRYKD......QIYAWDVVNEAIEDKTD....LIMRDTKWLRLLGEDYLVQAFNMAHEADPNALL
4L4O   IKTICTRYRD......VVYSWDVVNEAVEDKTD....VLLRDSKWRRIIGDDYIKIAFEIAKKYTGNGKL
5AY7   IQAIVNRYKD......VVYCWDVVNEAVADSPVYPGRPELRNSPMYQIAGEEFIYKAFEYAHEADPDALL
3W24   ITTVLDHFKTKYGSQNPIIGWDVVNEVLDDNGN......LRNSKWLQIIGPDYIEKAFEYAHEADPSMKL

F0S6Y3    200       210       220       230       240       250       260
F0S6Y3 FYNDYNSER.PEKVERIYTLLKQLKDKNIPIDGVGLQAHWSIFEPS.RSELEHAIQKYSSLGLEIHITEL
2Q8X   FYNDYNECF.PEKREKIFALVKSLRDKGIPIHGIGMQAHWSLTRPS.LDEIRAAIERYASLGVVLHITEL
3EMC   FYNDYNETD.PVKREKIYNLVRSLLDQGAPVHGIGMQGHWNIHGPS.MDEIRQAIERYASLDVQLHVTEL
4L4O   FYNDYNNEM.PYKLEKTYKVLKSLLEEGTPIDGVGIQAHWNIWDKNLIDNLKRAIETYASLGLEIQITEL
5AY7   FYNDYNDAE.PAKSQRIYNLVKRMKDAGVPIDGIGMQAHYNVYGPT.MKEVDDAIKLYSTVVDHIHLTEL
3W24   FINDYNIENNGVKTQAMYDLVKKLKNEGVPINGIGMQMHISINSN..IDNIKASIEKLASLGVEIQVTEL

F0S6Y3      270       280       290       300       310       320       330
F0S6Y3 DVSIYKWEKEKRAKRPNESDEYTPELQKAQANKYQMFFEVFRKYKKNITNVTFWNATDKYTWLDHYPVEG
2Q8X   DVSMFEFHDRRT.....DLAAPTSEMIERQAERYGQIFALFKEYRDVIQSVTFWGIADDHTWLDNFPVHG
3EMC   DLSVFRHEDQRT.....DLTEPTAEMAELQQKRYEDIFGLFREYRSNITSVTFWGVADNYTWLDNFPVRG
4L4O   DISVFEFEDRRT.....DLLEPTEEMVELQAKVYEDVFRVFREYRDVITSVTLWGISDRHTWKDNFPVIG
5AY7   DIRINEDMGGGLRFNQ.GQATVSDWERTLQQDQYVQLFKVLRKHKDVIDCVTFWNVSDKDSWL......G
3W24   DMNMNGDVSN...........DALLKQARLYKQLFDLFKAEKQYITAVVFWGVSDDVSWL......S

F0S6Y3      340       350
F0S6Y3 RKNYPSLFDINAKPKAAFWKVVNF..........
2Q8X   RKNWPLLFDEQHKPKPAFWRAVSV..........
3EMC   RKNWPFVFDTELQPKDSFWRIIGQD.........
4L4O   RKDWPLLFDIDGKPKKAFFRIIDFLEHHHHHH..
5AY7   VRNYPLLFDENYKPKQAYNAVKNFDPKMDNAVI.
3W24   KPNAPLLFDSKLQAKPAYWAIVDLGKAIPDIQSA
```

Fig. 5.2. Multiple Sequence Alignment (MSA) of *Ps*GH10A with XynA2 (PDB ID: 2Q8X), XynB (PDB ID: 3EMC), *Cb*Xyn10B (PDB ID: 4L4O), XynGR40 (PDB ID: 5AY7) and *Ts*XylA (PDB ID: 3W24) as given methods and Table 5.1. The conserved residues are displayed in white with red color background and semi-conserved residues are highlighted in black with the yellow color background.

5.3.2. Structure modeling and validation of *Ps*GH10A

The *Ps*GH10A catalytic domain structure (Fig. 5.3 A) showed the classical TIM-barrel $(\beta/\alpha)_8$-fold, similar to GH10 xylanases from *Cellvibrio mixtus,* and *Xanthomonas citri* (Pell et al., 2004; Santos et al., 2014). The catalytic domains of approximately, 50% of glycoside hydrolases comprise TIM-barrel fold and form a group of glycoside hydrolase families denoted as Clan A (Sharma et al., 2018). The internal core of *Ps*GH10A $(\beta/\alpha)_8$-barrel structure is formed by the classical arrangement of 8 β-strands followed by forming an external part by 8 α-helices. N-terminal of the internal β-barrel is linked with the external α-helical barrel by short loops, whereas, other secondary structure incorporated within the β/α motif. The eight important loops (L1–L8) connecting the C-terminal ends of the internal β-barrel with other secondary structure form the "Salad Bowl" shape with dimensions ~ 48x33x29 Å (Fig. 5.3 B). The three loops L4, L7 and L8 are organized together and form an open catalytic groove on protein surface (Fig. 5.3 C). The open catalytic groove is solvent exposed with the lining of aromatic and hydrophilic residues in the inner wall of the tunnel. The open groove is able to accommodate polysaccharide, the xylan and is involved in its catalysis. Glu156 (Acid/Base) is present at the end of β4 strand and Glu263 (Nucleophile) is located at β-bulge on the β8 strand (Fig. 5.3 D). The distance between the side chain of acid/base and the nucleophilic residue is 5.6 Å and that matches well with the distance between the residues that are involved in retaining type mechanism followed by xylanases (Rye and Withers, 2000).

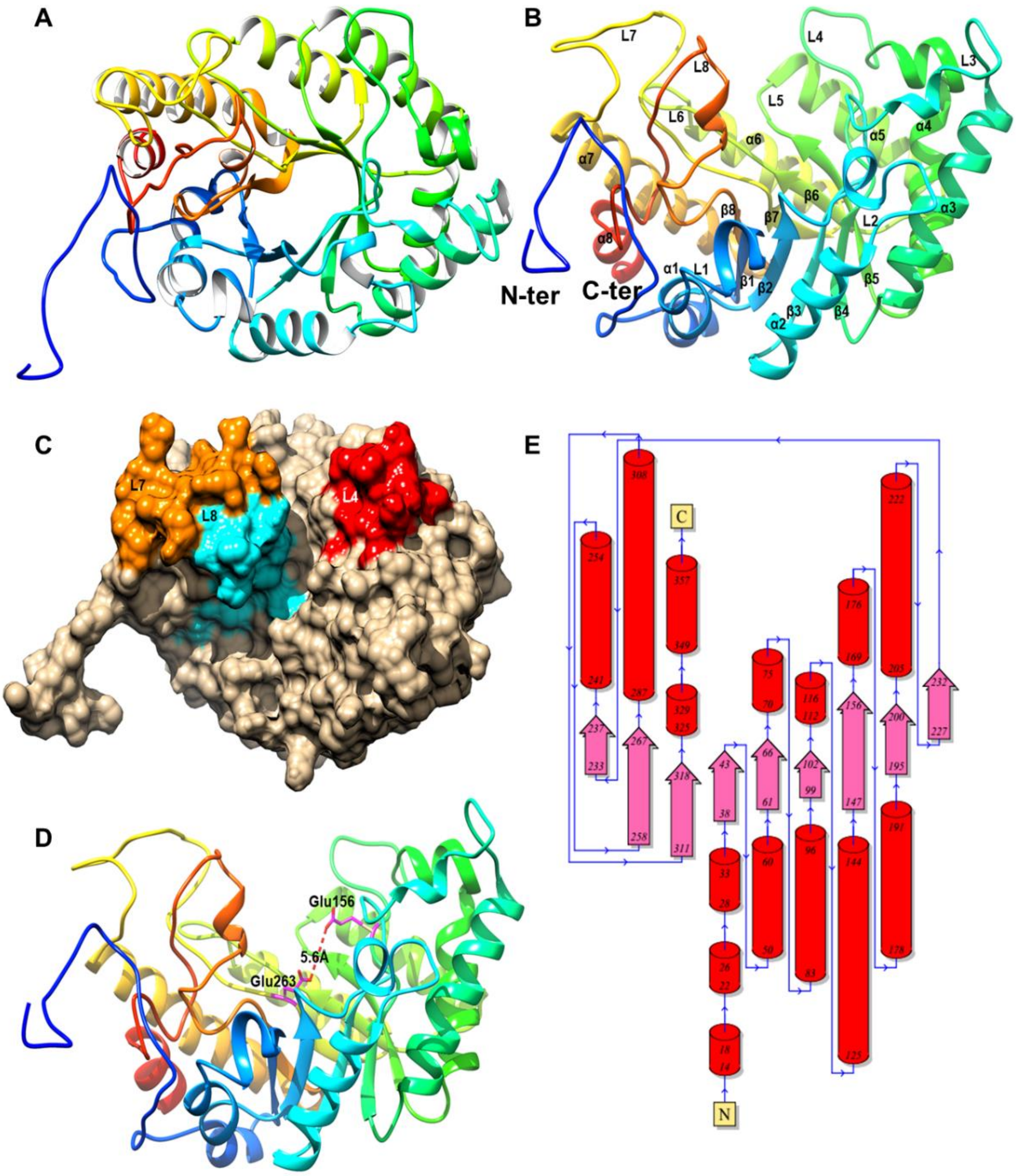


Fig. 5.3. 3-Dimensional structure analysis of *Ps*GH10A. A) 3D structure showing TIM-barrel $(\beta/\alpha)_8$ fold, B) *Ps*GH10A showing Salad Bowl shape with labeled α-helices, β-strands and connecting loops, C) Surface view of *Ps*GH10A showing Loops 4, 7 and 8 (L4, L7 and L8) forming catalytic cleft, D) The catalytic residues, Glu156 and Glu263 shown in purple color and the distance between the catalytic residues are shown as dotted line and E) Topology diagram of *Ps*GH10A showing the N and C-terminal regions with α-helix, β-strands and connecting loops.

A *Ps*GH10A topology diagram is shown in Fig. 5.3E. The numbering of the β-strands (n) and α-helices was assigned in a classical arrangement manner. α-Helices (α1–α8) at the external part of the (β/α)$_8$-fold and eight of eight β-strands (β1–β8) at the internal core of the (β/α)$_8$-fold (Fig. 5.2A). The 8 β-strands of *Ps*GH10A are constituted of the residues Ile39-Val43 (β1), Ser62-Ala65 (β2), Lys99-Gly102 (β3), Ile148-Asn155 (β4), Lys198-Asp200 (β5), Gly228-Leu231 (β6), His234-Ser236 (β7) and Glu258-Ser267 (β8). Similarly, 8 α-helices of the (β/α)$_8$-fold are constituted of the residues Glu50-Asn59 (α1), Lys84-Lys95 (α2), Arg126-Tyr144 (α3), Asp177-Ala190 (α4), Pro206-Asp221 (α5), Arg242-Leu255 (α6), Pro288-Leu307(α7) and Ala350-Val358 (α8). *Ps*GH10A contains several additional α-helices that are part of the general (β/α)$_8$-barrel fold but are present outside of the main (β/α)$_8$-barrel structure. α0-Helix is present at N-terminal region of the polypeptide (Leu29-Tyr32), just prior to β1. ɳ3 (Asp113-Phe116) and ɳ4 (Trp326-Asp328) are two short 3_{10}-helices as a part of the two loops leading to the active-site core residues. The surface topology of *Ps*GH10A matches well with the previously reported xylanase structures (Santos et al., 2014; Zhang et al., 2016; Zheng et al., 2016).

5.3.3. Energy minimization and structure validation of *Ps*GH10A

The best model due to steric clashes, unfavorable bond angle, bond lengths and contact shows high energy level, which is -77.88 MJ/mol. Therefore, the *Ps*GH10A modeled structure was subjected to geometry optimization or energy minimization by AMBER force field. The potential energy of *Ps*GH10A was improved to -81.25 MJ/mol, which conformed to the proper folding and packaging of the model. The quality of energy minimized *Ps*GH10A modeled structure was assessed by the tools available on the SAVES server. PROCHECK analysis displayed that 352 amino acid residues (99.1%) are in the favoured region, only one amino acid Tyr308 (0.3%) is in the generously allowed region, while only two residues Gln34 and Asn97 (0.6 %) are in the disallowed region (Fig. 5.4A). This confirmed that the modeled structure is following the dihedral angles, phi (ϕ) and psi (ψ) of Ramachandra plot and occupied favourable positions. ProSA-web results gave Z-score of -10.01 for *Ps*GH10A modeled structure, -10.72 for 2Q8X, -9.92 for 3EMC, -11.18 for 3W24, -9.8 for 4L4O and -9.3 for 5AY7. The Z score of *Ps*GH10A modeled structure corroborated with the homologous experimentally determined structures (Fig. 5.4 B). The analysis of *Ps*GH10A modeled structure by Verify-3D showed, that 89.69% of amino acid residues have an overall average 3D-1D score ≥ 0.2 (Fig. 5.4 C). The results of *Ps*GH10A structure refinement and validation assessment were acceptable enough to support the further studies. The active site residues involved in catalysis were identified by superposing the modeled *Ps*GH10A structure with the available homologous structure of xylanase from *Geobacillus stearothermophilus*. The superposition showed that active site residues were conserved and aligned spatially. Glu156 and Glu263 were found at the active site of *Ps*GH10A (Fig. 5.4D). The quality factor of *Ps*GH10A modeled structure by ERRAT plot was 94.45%, which further confirmed the excellent model quality of *Ps*GH10A (Fig. 5.4E).

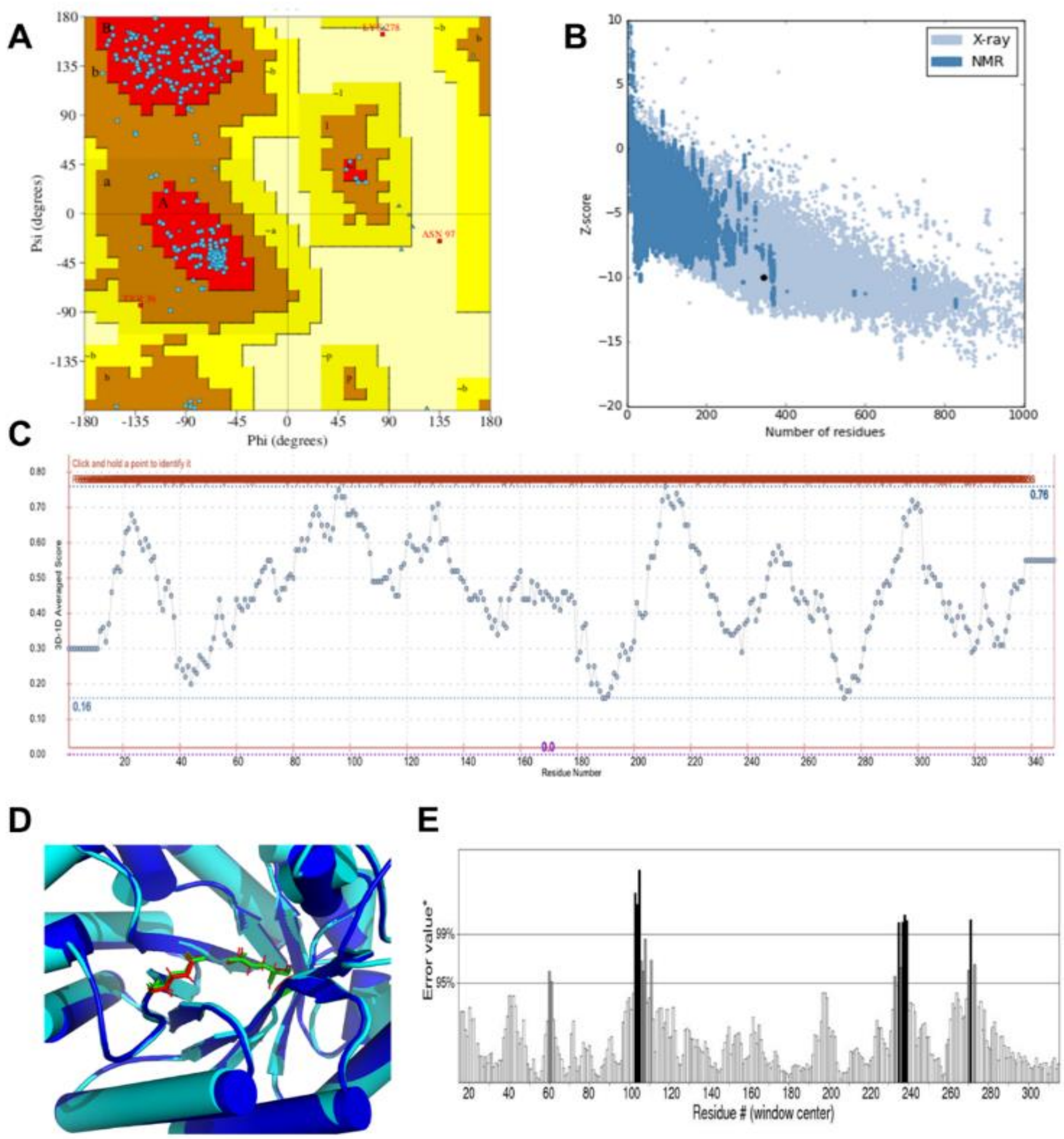


Fig. 5.4. Structure validation analysis of modeled *Ps*GH10A. A) Ramachandran Plot analysis by PROCHECK (UCLA SAVES server) showing the most favorable, additionally allowed and generously allowed regions of amino acid residues, D) ProSA-web plot showing Z-score, -10.01 for overall model quality, C) Verify-3D showing the threshold score more than 0.2, D) Structure superposition of modeled *Ps*GH10A (cyan) with xylanase from *G. stearothermophilus* (PDB ID 2Q8X) (blue) showing the binding cleft (prepared by PyMol (http://www.pymol.org) and E) ERRAT plot analysis showing the quality of the *Ps*GH10A modeled structure.

5.3.4 Secondary structure analysis of *Ps*GH10A

The RaptorX property prediction of *Ps*GH10A showed that 33.0% residues are involved in the formation of α-helices, 18% residues are arranged in β-strands and 49% of the residues occur as loops (Table 5.2). The secondary structure composition of *Ps*GH10A was confirmed by Circular Dichroism analysis (Fig. 5.5). The CD analysis showed that 31.75% of the residues form α-helix, 20.0% residues give rise to β-strands and 48.25% of the residues form loops (Table 5.2). The secondary structure prediction analysed by ProMotif also gave similar results to those of RaptorX prediction and CD (Table 5.2).

Table 5.2. Secondary structure analysis of *Ps*GH10A.

Secondary structure element	RaptorX prediction (%)	ProMotif prediction (%)	CD analysis (%)
α-helix	33.0	35.5	31.75
β-strands	18.0	14.5	20.0
Random coil	49.0	50.0	48.25

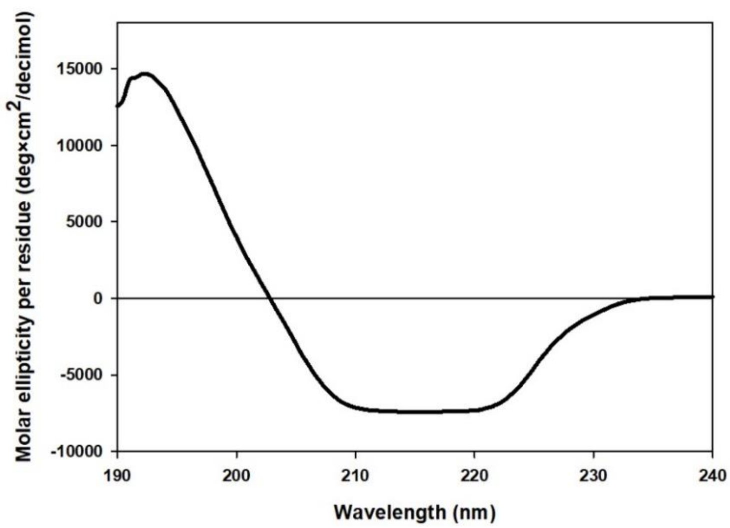


Fig. 5.5. Far-UV Circular Dichroism (CD) spectrum of *Ps*GH10A analyzed by CONTIN program available in Dichroweb.

5.3.5. Molecular docking analysis of *Ps*GH10A

Molecular docking method was used to study the interaction of *Ps*GH10A with its ligands. Interactions and free binding energies for xylobiose, xylotriose, xylotetraose, arabinoxylobiose and arabinoxylotriose are reported in Table 5.3. The xylooligosaccharides *viz*. xylobiose, xylotriose and xylotetraose showed strong binding affinity against *Ps*GH10A with -7.0, -8.3 and -8.5 kcal/mol, respectively, but as the size of ligand increased it showed comparatively, weaker binding (-4.5 and -5.3 kcal/mol for xylopentaose and xylohexaose, respectively). This indicated that the active site of *Ps*GH10A is able to accommodate up to xylotetraose. The xylotetraose and arabinoxylobiose docked on the surface of *Ps*GH10A are shown in Fig. 5.6 (A-D). In the *Ps*GH10A-xylotetraose complex, xylotetraose was accommodated within the catalytic cleft formed by the loops L4, L7 and L8 (Fig. 5.4C). Similar binding groove was found in the structure of xylanase (*Cb*Xyn10B) from *C. bescii* (Zhang et al., 2016) having 48% sequence similarity with *Ps*GH10A. The xylotetraose docked on the surface of *Ps*GH10A is shown in Fig. 5.6A. The xylotetraose is accommodated through formation of hydrogen bonds in the active cleft by Arg205, Gln232, His234, Glu239, Glu263, Asp265 and Trp330 amino acid residues and through hydrophobic interaction by Trp107, Tyr201, Ser236, Pro240, Trp271, Trp318 and Trp326 (Fig. 5.6B). These aromatic residues involved in the formation of catalytic cleft further help in the catalysis. *Ps*GH10A made six hydrogen bonds with xylose residues of xylotetraose. The arabinoxylotriose was held in catalytic cleft present on *Ps*GH10A surface by hydrogen bonds (Fig. 5.6C).

Arabinoxylotriose showed interactions with Asn67, Arg205, His234 Gln232 through hydrogen bonds and with Trp107, Gln110, Glu156, Tyr201, Trp271, Trp318, Trp326 and Tyr330 through hydrophobic interactions (Fig. 5.6D). The residues, Arg205 and His234 interacted through the hydrogen bond with arabinose unit present in arabinoxylotriose. These two residues help in the recognition and accommodating the arabinoxylotriose in the active site.

Table 5.3. Active site residues of *Ps*GH10A showing interaction with ligands.

Ligand	Binding Energy (ΔG) (kcal/mol)	Amino Acid involved in H-Bond/Polar Contacts (with residue no.)	Amino acid residues within 4 Å of active site and involved in hydrophobic/Van der Waals interaction
Xylotetraose	-8.5	Arg205, Gln232, His234, Glu239, Glu263, Asp265, Trp330	Trp107, Tyr201, Ser236, Pro240, Trp271, Trp318, Trp326
Xylotriose	-8.3	Glu66, Gln232, His234, Glu263, Trp318 and His329	Asn67, Trp107, Tyr324, Thr325, Trp326,
Xylobiose	-7.0	Glu156, Gln232, His234, Glu263 and Asp265	Trp107, Trp318, Trp326 and His329
Arabinoxylobiose	-6.5	Glu156, Arg205	Trp107, Tyr201, His234, Trp235, Ser236, Glu239, Ser267, Trp271, Trp326, Leu327 and Tyr330
Arabinoxylotriose	-7.2	Asn67, Arg205, His234 Gln232	Trp107,Gln110, Glu156, Tyr201, Trp271, Trp318, Trp326 and Tyr330

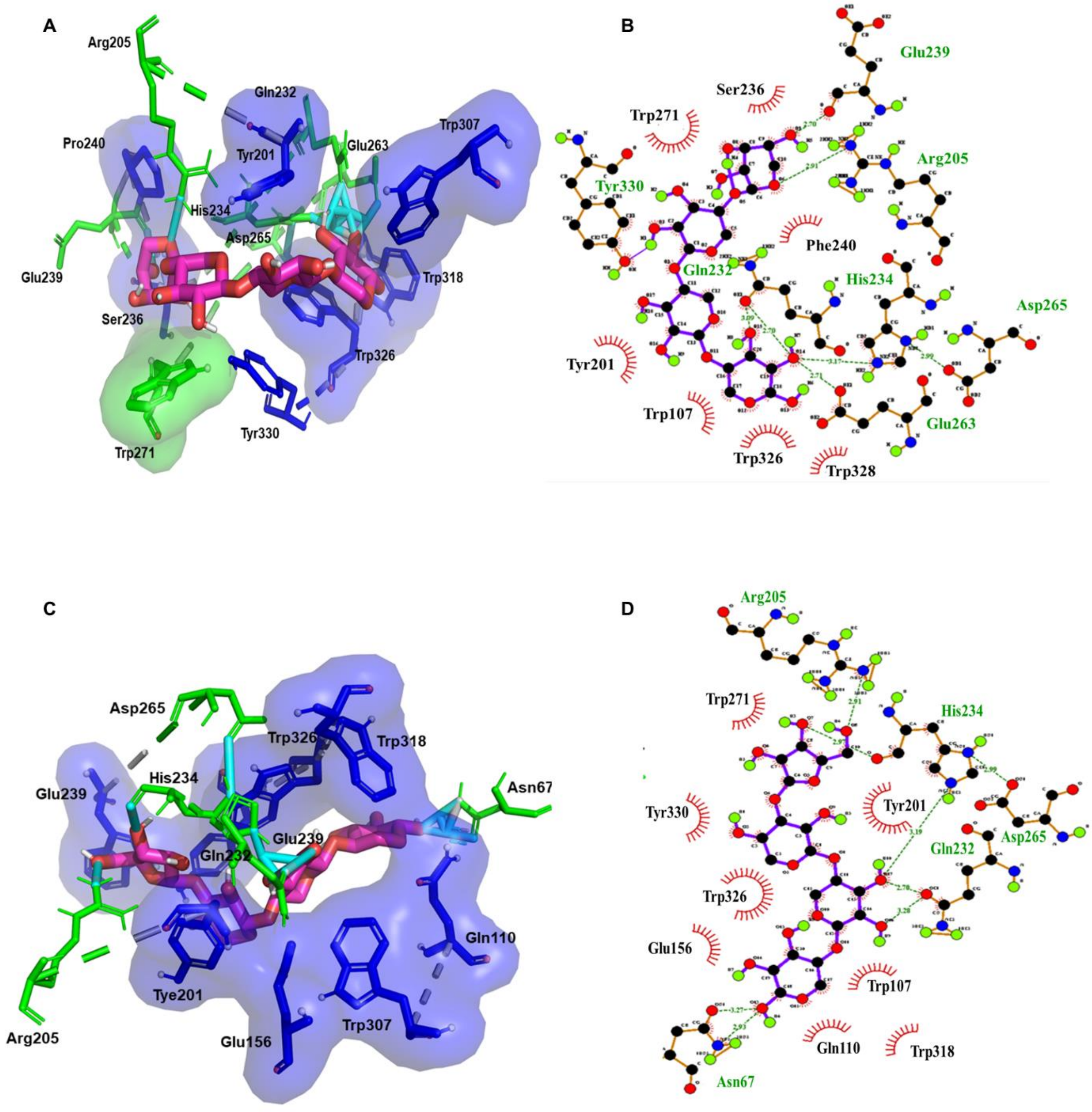


Fig. 5.6. Molecular docking analysis of *Ps*GH10A with its ligands. A) Active site showing xylotetraose ligand orientation, B) 2-Dimensional schematic presentation of xylotetraose interaction with *Ps*GH10A active site residues, C) Active site showing arabinoxylotriose ligand orientation, D) 2-Dimensional schematic presentation of arabinoxylotriose interaction with active site residues. In Fig. 5.6A and 5.6C, the hydrogen bond forming residues are shown in green and hydrophobic interaction showing residues are in blue. In Fig. 5.6B and 5.6D, the 2-Dimensional interaction image, the dashed lines showing the hydrogen bonds with bond length and the residues displayed in arc with spokes are making the hydrophobic interactions with the ligand.

5.3.6. Protein melting analysis of *Ps*GH10A

The unfolding of *Ps*GH10A started at 50°C and completely converted into primary structure at 60°C (Fig. 5.7). The biochemical characterization studies of *Ps*GH10A showed the enhancement in activity by 30% in presence of 2 mM Mn^{2+} or Mg^{2+} ions, but 2 mM EDTA lowered the activity to 64%. Therefore, the melting curve analysis of *Ps*GH10A was also performed in presence of 2 mM Mn^{2+} or Mg^{2+} ions and in combination with EDTA. However, no shift in the melting temperature was observed in presence of either of the metal ions. These results indicated that the *Ps*GH10A stability is not influenced by the metal ions though enhancing the enzyme activity.

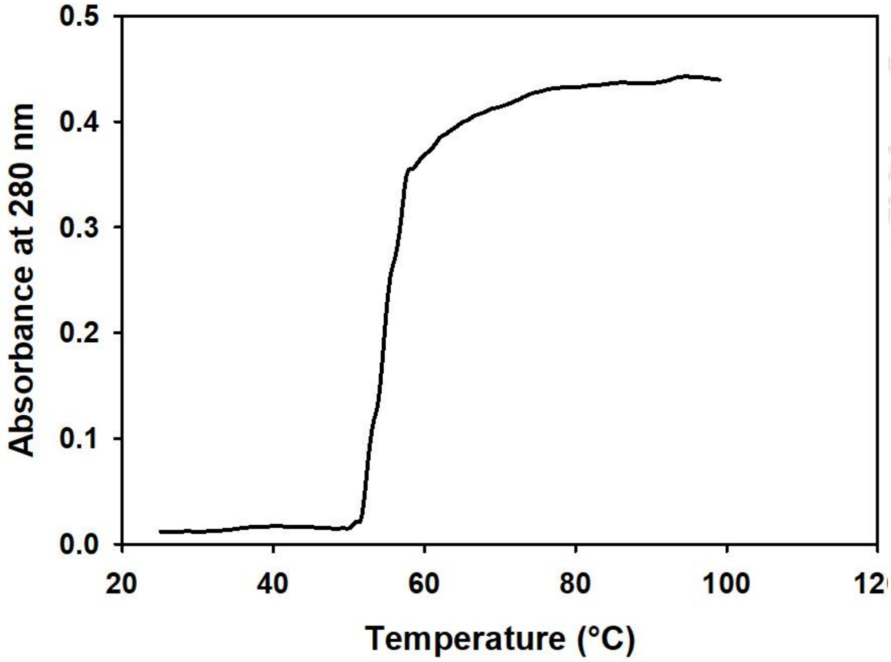


Fig. 5.7. Protein melting curve analysis of *Ps*GH10A.

5.3.7. Low resolution structure analysis of *Ps*GH10A by SAXS.

The Small Angle X-ray Scattering (SAXS) method was employed for *Ps*GH10A solution conformation analysis. The *Ps*GH10A structural parameters obtained by SAXS data processing and analysis are summarized in Table 5.4. The visual inspection and initial processing of *Ps*GH10A scattering pattern at different concentrations confirmed the absence of aggregation and interparticle correlation effect (Fig. 5.8A). Guinier plot analysis showed that the radius of gyration (Rg) for *Ps*GH10A (at all the concentrations 6, 8, 12 mg/ml) was in the range, from 2.23 (±0.15) to 2.29 (±0.16) nm and plot's linearity confirmed the protein's monodisperse state in the solution (Fig. 5.8B).

Table 5.4. SAXS data collection parameter and derived parameters of *Ps*GH10A.

Data-collection parameters	***Ps*GH10A**	***Ps*GH10A**	***Ps*GH10A**
Instrument	SAXSpace Anton-Paar		
Wavelength (Å)	1.54 Å		
Q range (nm^{-1})	0.135-5.95		
Exposure time (minutes)	30x2		
Temperature (°C)	40		
Protein Concentration mg/mL	6	8	12
Structural parameter			
Q range (nm^{-1}) for Rg analysis	0.026-0.325	0.024-0.328	0.020-0.0342
I(0) au from Guinier	93632±3496	105500±4929	125886±4614
Rg nm from Guinier	2.23±0.15	2.27±0.20	2.29±0.16
I(0) au from P(r)	90600±3125	105600±4365	126300±3748
Rg nm from P(r)	2.31±0.16	2.35±.17	2.41±0.13
D_{max} (nm)	9.25	9.40	9.70
Porod volume estimate (nm^3)	59.5	55.9	56.8
Molecular Mass (MM) determination			
I(0) au Lysozyme (14.3 kDa)	32248±898	35685±469	42773±511
MM calculated from I(0) (kDa)	41.5	42.3	42.1
Theoretical mass MM (kDa)	42.18		
Software employed			
Data processing	Primus	Primus	Primus
P(r) function calculation	GNOM	GNOM	GNOM
Ab initio modelling	DAMMIF	DAMMIF	DAMMIF
Validation and averaging	DAMAVER	DAMAVER	DAMAVER
3-D graphical representation	PyMOL	PyMOL	PyMOL
Modeling parameters			
χ^2	0.057	0.043	0.038
NSD	0.68±0.02	0.63±0.03	0.67±0.04

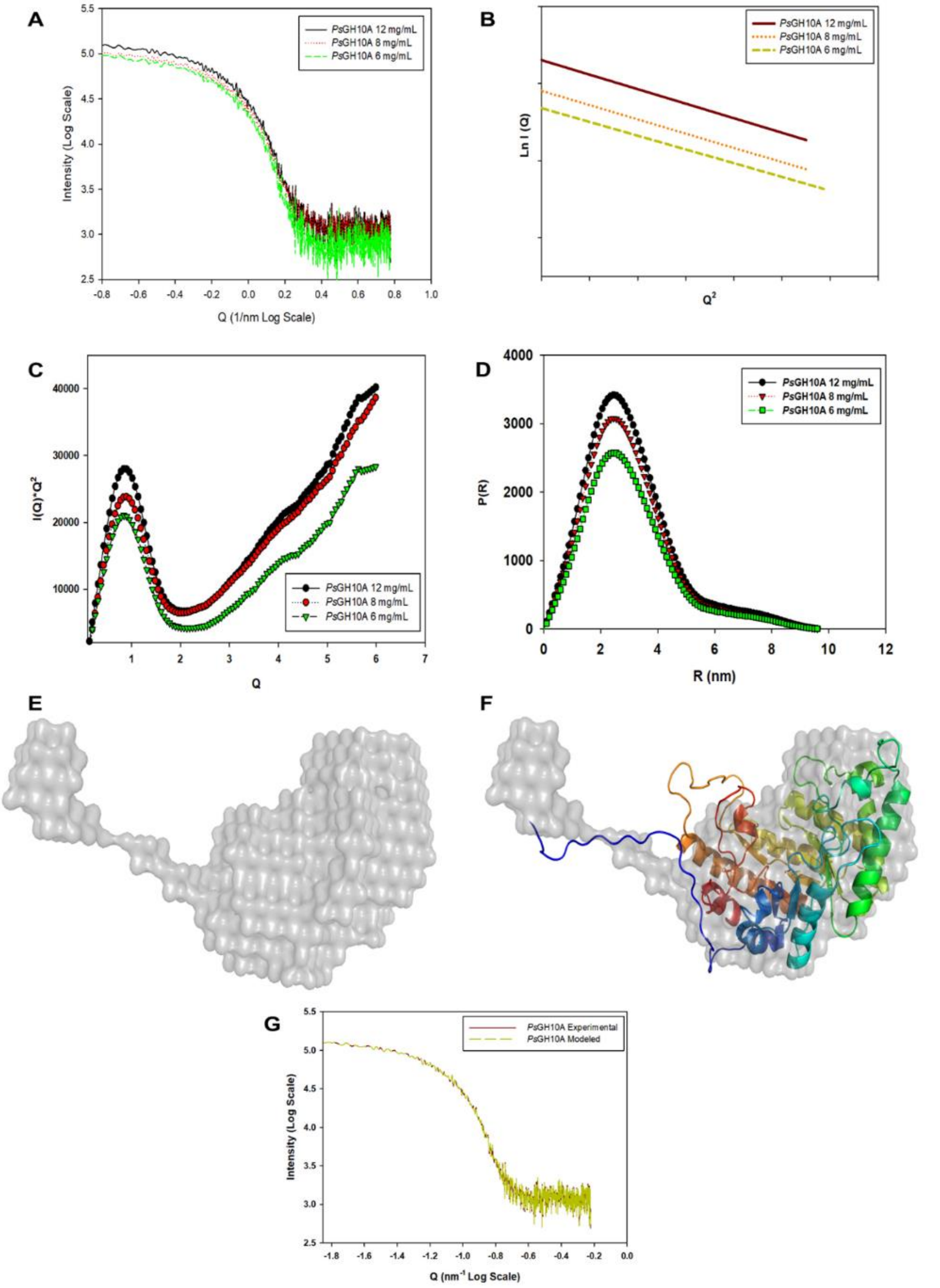


Fig. 5.8. SAXS analysis of *Ps*GH10A. A) SAXS intensity profile of *Ps*GH10A, B) Guinier plot of the SAXS intensities, C) Kratky plot analysis of *Ps*GH10A, D) P(R) curve of *Ps*GH10 as a function of R, (E) *Ab initio* shape of *Ps*GH10A, F) Superposition of *ab initio* model with homology modeled *Ps*GH10A and G) Crysol fitting of experimental data and homology modeled structure of *Ps*GH10A.

The molecular mass obtained from SAXS data using SAXSMoW tool was 42.1 kDa, similar to the theoretically determined molecular weight based on amino acid sequence calculations (42.18 kDa) (Table 5.4). The global compactness and flexibility of *Ps*GH10A were also evaluated by Kratky analysis (Fig. 5.8C), which revealed a well-defined maximum, q < 2 and that did not decay near zero at higher q values and maintained a slightly elevated baseline. Kratky analysis further displayed some flexibility at N-terminal His-tag region of *Ps*GH10A. Analysis of P(R) function showed that *Ps*GH10A is a globular molecule with the maximum dimension (Dmax) of 9.6 nm (Fig. 5.8D). The *ab initio* derived *Ps*GH10A envelope was generated by DAMMIF package without symmetry constraint (Fig. 5.8E). The *Ps*GH10A envelope was compact and clearly revealed a chicken thigh shaped single catalytic domain showing tail as the hexa-histidine tag at N-terminal with a resolution of 2.4± 0.3 nm. The *Ps*GH10A *ab initio* derived envelope showed an excellent fit with its modeled structure (Fig. 5.8F). The homology modeled structure of *Ps*GH10A superposed well with SAXS generated *ab initio* structure, but some of the loops and N-terminal His-tag are flexible and show the deviation. These results were also corroborated by the crysol fitting of modeled structure on SAXS Data (Fig. 5.8G)

5.4 Conclusion

The 3D structure of endo β-1,4 xylanase (*Ps*GH10A) of a family 10, glycoside hydrolase from *Pseudopedobacter saltans* modeled by comparative modeling was compact and stable. The secondary structure by CD analysis of *Ps*GH10A displayed presence of α-helices (31.75%), β-strands (20.00%) and random coil (48.25%). The *Ps*GH10A modeled structure showed $(\beta/\alpha)_8$ TIM barrel fold which is conserved in GH10 family. The docking studies revealed that the active site of *Ps*GH10A can accommodate linear xylooligosaccharides and arabinose substituted xylooligosaccharides. The residues, Glu153 and Glu263 are involved in catalysis. *Ps*GH10A melting analysis confirmed that the enzyme does not require any metal ion for its stability. SAXS analysis displayed the monomeric nature of *Ps*GH10A in solution form. Guinier analysis gave the radius of gyration between (Rg) 2.23±0.15 and 2.29±0.16 nm for all the protein concentrations. Kratky plot analysis of *Ps*GH10A displayed fully folded state in solution form. The *ab initio* derived DAM model of *Ps*GH10A superposed well with its comparative modeling based 3D-structure.

5.5. References

Ahmed, S., Luis, A.S., Bras, J.L.A., Ghosh, A., Gautam, S., Gupta, M.N., Fontes, C.M.G.A., Goyal, A., 2013. A novel α-L-arabinofuranosidase of family 43 Glycoside Hydrolase (*Ct*43Araf) from *Clostridium thermocellum*. *Plos One* 8(9), 1-10.

Collins, T., Gerday, C., Feller, G., 2005. Xylanases, xylanase families and extremophilic xylanases. *FEMS. Microbiol. Rev.* 29, 3-23

Colovos, C., Yeates, T.O., 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* 1511–1519.

Correia, M.A.S., Mazumder, K., Brás, J.L.A., Firbank, S.J., Zhu, Y., Lewis, R.J., York, W.S., Fontes, C.M.G.A., Gilbert, H.J., 2011. Structure and function of an arabinoxylan-specific xylanase. *J. Biol. Chem.* 286, 22510-22520.

Eswar, N., Eramian, D., Webb, B., Shen, M.Y. and Sali, A., 2008. Protein structure modeling with MODELLER. In *Structural proteomics* 145-159. Humana Press.

Fischer, H., De Oliveira Neto, M., Napolitano, H.B., Polikarpov, I., Craievich, A.F., 2010. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Cryst.* 43(1), 101–109.

Fontes, C.M.G.A., Gilbert, H.J., Hazlewood, G.P., Clarke, J.H., Prates, J.A.M., McKie, V.A., Nagy, T., Fernandes, T.H., Ferreira, L.M.A., 2000. A novel *Cellvibrio mixtus* family 10 xylanase that is both intracellular and expressed under non-inducing conditions. *Microbiology*. 146, 1959-1967.

Franke, D., Svergun, D.I., 2009. DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Cryst.* 42(2), 342–346.

Gallardo, Ó., Javier Pastor, F.I., Polaina, J., Diaz, P., Łysek, R., Vogel, P., Isorna, P., González, B., Sanz-Aparicio, J., 2010. Structural insights into the specificity of Xyn10B from *Paenibacillus barcinonensis* and its improved stability by forced protein evolution. *J. Biol. Chem.* 285, 2721–2733.

Han, X., Gao, J., Shang, N., Huang, C.H., Ko, T.P., Chen, C.C., Chan, H.C., Cheng, Y.S., Zhu, Z., Wiegel, J., Luo, W., Guo, R.T., Ma, Y., 2013. Structural and functional analyses of catalytic domain of GH10 xylanase from *Thermoanaerobacterium saccharolyticum* JW/SL-YS485. *Proteins: Struct., Funct., Bioinf.* 81(7), 1256–1265.

Harris, D., DeBolt, S., 2010. Synthesis, regulation and utilization of lignocellulosic biomass. *Plant Biotechnol J. 8*(3), 244-262.

Hutchinson, E.G., Thornton, J.M., 2008. PROMOTIF-A program to identify and analyze structural motifs in proteins. *Protein Sci*. 5(2), 212–220.

Jönsson, L.J., Martín, C., 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour. Technol.* 199, 103–112.

Kelly, S.M., Jess, T.J., Price, N.C., 2005. How to study proteins by circular dichroism. *BBA-Proteins Proteom.* 1751, 119–139.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst. 26*(2), 283-291.

Laskowski, R.A., Swindells, M.B., 2011. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.*, 51(10), 2778–2786.

Lüthy, R., Bowie, J.U., Eisenberg, D., 1992. Assessment of protein models with three-dimensional profiles. *Nature 356*(6364), 83-85.

Margelevičius, M., Venclovas, Č., 2005. PSI-BLAST-ISS: An intermediate sequence search tool for estimation of the position-specific alignment reliability. *BMC Bioinformatics*. 6, 1-10.

Meng, D.D., Ying, Y., Chen, X.H., Lu, M., Ning, K., Wang, L.S., Li, F.L., 2015. Distinct roles for carbohydrate-binding modules of glycoside hydrolase 10 (GH10) and GH11 xylanases from *Caldicellulosiruptor sp*. strain F32 in thermostability and catalytic efficiency. *Appl. Environ. Microbiol.* (81)6, 2006-2014.

Morris, G.M., Ruth, H., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 30(16), 2785-2791.

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. Open Babel: An Open chemical toolbox. *J. Cheminform.* 3(1), 1–14.

Padilha, I.Q.M., Valenzuela, S. V., Grisi, T.C.S.L., Diaz, P., De Araújo, D.A.M., Javier Pastor, F.I., 2015. A glucuronoxylan-specific xylanase from a new *Paenibacillus favisporus* strain isolated from tropical soil of Brazil. *Int. Microbiol.* 17(3), 175-184.

Pell, G., Taylor, E.J., Gloster, T.M., Turkenburg, J.P., Fontes, C.M.G.A., Ferreira, L.M.A., Nagy, T., Clark, S.J., Davies, G.J., Gilbert, H.J., 2004. The Mechanisms by Which Family 10 Glycoside Hydrolases Bind Decorated Substrates. *J. Biol. Chem*. 279, 9597–9605.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25.

Rani, A., Goyal, A., 2016. A new member of family 8 polysaccharide lyase chondroitin AC lyase (PsPL8A) from *Pedobacter saltans* displays endo- and exo-lytic catalysis. *J. Mol. Catal. B Enzym*. 134, 215–224.

Rose, P.W., Beran, B., Bi, C., Bluhm, W.F., Dimitropoulos, D., Goodsell, D.S., Prlić, A., Quesada, M., Quinn, G.B., Westbrook, J.D., Young, J., Yukich, B., Zardecki, C., Berman, H.M., Bourne, P.E., 2011. The RCSB Protein Data Bank: Redesigned web site and web services. *Nucleic Acids Res.* 39, 392–401.

Rye, C.S., Withers, S.G., 2000. Glycosidase mechanisms. *Curr. Opin. Chem. Biol*. 4, 573–580.

Saini, J.K., Saini, R., Tewari, L., 2015. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech* 5(4), 337-353.

Santos, C.R., Hoffmam, Z.B., De Matos Martins, V.P., Zanphorlin, L.M., De Paula Assis, L.H., Honorato, R.V., De Oliveira, P.S.L., Ruller, R., Murakami, M.T.,

2014. Molecular mechanisms associated with xylan degradation by *Xanthomonas* plant pathogens. *J. Biol. Chem*. 289, 32186–32200.

Semenyuk, A. V., Svergun, D.I., 1991. GNOM. A program package for small-angle scattering data processing. *J. Appl. Cryst.* 24(5), 537–540.

Sharma, K., Dhillon, A., Goyal, A., 2018. Insights into Structure and Reaction Mechanism of β-Mannanases, *Curr. Protein. Pept. Sci.*, *19*(1), 34–47.

Shen, M., Sali, A., 2006. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 2507–2524. doi:10.1110/ps.062416606.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7(1).

Solomon, V., Teplitsky, A., Shulami, S., Zolotnitsky, G., Shoham, Y., Shoham, G., 2007. Structure-specificity relationships of an intracellular xylanase from *Geobacillus stearothermophilus*. *Acta Crystallogr D, Struct Biol.* 63(8), 845–859.

Sreerama, N., Woody, R.W., 2000. Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem*. 260, 252–260.

Svergun, D., Barberato, C., Koch, M.H., 1995. CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst*. 28(6), 768-773.

Tuukkanen, A.T., Kleywegt, G.J., Svergun, D.I., 2016. Resolution of ab initio shapes determined from small-angle scattering. *IUCr J 3*(6), 440–447.

Valenzuela, S.V., Diaz, P., Javier Pastor, F.I., 2012. Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. *Appl. Environ. Microbiol.* 78, 3923-3931.

van Gool, M.P., van Muiswinkel, G.C.J., Hinz, S.W.A., Schols, H.A., Sinitsyn, A.P., Gruppen, H., 2012. Two GH10 endo-xylanases from *Myceliophthora thermophila* C1 with and without cellulose binding module act differently towards soluble and insoluble xylans. *Bioresour. Technol*. 119, 123-132.

Volkov, V. V., Svergun, D.I., 2003. Uniqueness of ab initio shape determination in small-angle scattering, in: *J. Appl. Cryst.* 36(3), 860–864.

Wang, S., Li, W., Liu, S., Xu, J., 2016. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res.* (8)44, 430–435.

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, 407-410.

Zhang, Y., An, J., Yang, G., Zhang, X., Xie, Y., Chen, L., Feng, Y., 2016. Structure features of GH10 xylanase from *Caldicellulosiruptor bescii*: implication for its thermophilic adaption and substrate binding preference. *Acta Biochim. Biophys. Sin.* 48, 948–957.

Zheng, Y., Li, Y., Liu, W., Chen, C.C., Ko, T.P., He, M., Xu, Z., Liu, M., Luo, H., Guo, R.T., Yao, B., Ma, Y., 2016. Structural insight into potential cold adaptation mechanism through a psychrophilic glycoside hydrolase family 10 endo-β-1,4-xylanase. *J. Struct. Biol.* 193, 206–211.

Zhong, R., Teng, Q., Lee, C., Ye, Z.H., 2014. Identification of a disaccharide side chain 2-O-α-D-galactopyranosyl-α-D-glucuronic acid in arabidopsis xylan. *Plant Signal Behav*. 9(2), 2-6.

Chapter 6

Extraction, characterization of xylan from neem sawdust and its application in xylanase mediated production of anticancer xylooligosaccharides

6.1. Introduction

Busy and stressful lifestyle leads to the risk of diseases such as diabetes, coronary heart disease (Deposition of fat plaques in arteries), cardiovascular diseases, increase in blood cholesterol and cancer (Sharma, Thakur & Goyal, 2019). According to the world health organization, cancer is one of the broad groups of complex and devastating diseases that affect any part of the body by uncontrolled cell growth leading cause of death worldwide. In 2018, approximately 17 million new cases and 9.5 million deaths were estimated by the International Agency for Research on Cancer (IARC). Colon cancer (CC) is the third most common cause of cancer worldwide, and exploring the new effective medicine for its treatment is still a big challenge (Tan, Chen, Liu, Yang & Li, 2018). To overcome these challenges, recently, the utilization of natural products such as polysaccharides, oligosaccharides and other antioxidative compounds with high nutritional values, which improves human health has gained great interest (Zhao et al., 2014; Tan, Chen, Liu, Yang & Li, 2018). Pharmacologically active polysaccharides have been reported from diverse sources such as dextran from *Pediococcus pentosaceus* CRAG3 (Shukla & Goyal, 2013), from *Ganoderma lucidum* (mushroom) (Zhang et al., 2010) and plants (Melo-Silveira et al., 2019), etc. The cell wall of herbs and higher plants produce a potential source of natural and pharmacologically active polysaccharides (Ebringerová, Kardošová, Hromádková, Malovíková &Hříbalová; Melo-Silveira et al., 2019). The utilization of these pharmacologically active polysaccharides and their hydrolyzed products as anti-tumor compound have been demonstrated by several studies (Hashimoto et al., 1982; Ebringerová, Kardošová, Hromádková, Malovíková &Hříbalová, 2002; Kapoor & Dharmesh, 2017; Melo-Silveira et al., 2019). Besides their anti-tumor potential, these polysaccharides are non-toxic and do not have any side effects on human health. Therefore, these pharmacologically active plant polysaccharides can be used as a potential natural candidate for the treatment of different neoplasms (Lin et al., 2013; Chang et al., 2015; Melo-Silveira et al., 2019).

Heteroxylan is the second most abundant plant cell wall hemicellulosic polysaccharide in the plant kingdom (Bendahou, Dufresne, Kaddami & Habibi, 2007). sThe structure and composition of heteroxylan depend on the plant and its tissues and method used for the extraction (Petzold-Welcke, Schwikal, Daus, & Heinze, 2014). The backbone of heteroxylan is composed of β-(1→4)-linked xylose residues, which is often substituted with methylated α- D-glucopyranosyl uronic acid and α-L-arabinofuranosyl, ferulic acid, *p*-coumaric acid in the side chain (Ordaz-Ortiz, Devaux, & Saulnier, 2005). Apart from these substitutions, heteroxylan also contains D-galactose, D-xylose, L-rhamnose, D-glucose, various acetyl groups and different dimeric and trimeric side chains (Moine et al., 2007). The heteroxylans extracted from different sources varies in their composition or percentage of side-chain substitution and affects their pharmacological properties. Methylglucuronoxylan extracted from *Castanea sativa* (Spanish chestnut) displayed the antiproliferative activity and reduced the invasion and migration potential of human epidermoid carcinoma (A431) cells (Barbat et al., 2008). Arabinoxylan (AX) extracted from the green leaves of *Litsea glutinosa* displayed the strong activation of macrophages, thymocytes and splenocytes (Das, Maiti, Maiti & Islam, 2013). In another study, arabinoxylans from the finger millet bran displayed the activation of macrophages, phagocytic and mitogenic activity. Thus, it can be used as a potent immunomodulator (Prashanth, Shruthi & Muralikrishna, 2015). The antiproliferative potential of xylan from corn cobs on human lung adenocarcinoma epithelial (A549) cells and human cervical adenocarcinoma (HeLa) cells) was reported by Melo-Silveira et al. (2019). As mentioned above, there are several studies reported on pharmacological implications of heteroxylans, but the anticancer properties of xylooligosaccharides is not stated till date.

Azadirachta indica, commonly known as Neem is one of the most widely distributed and versatile medicinal plant presents in the Indian subcontinent (Subapriya & Nagini, 2005). Different parts of the neem plant have displayed various pharmacological activities for the management of several diseases such as antimicrobial infections, diabetes, neurological disorders and cancer (Alzohairy, 2016). Owing to its utilization in various disease management, the Neem plant is known as Panacea of all Diseases (Tiwari, Verma, Chakraborty, Dhama & Singh, 2014). The potential of Neem cell wall polysaccharides is still unexplored. Therefore, the present study is focused on the isolation and characterization of xylan from Neem sawdust. The isolated polysaccharide was used for the production of enzyme-mediated xylooligosaccharides for analyzing their anticancer potential.

6.2. Material and methods

6.2.1. Materials

Neem Sawdust was collected from the wood cutting workshop of the local market. The chemicals sodium chlorite, sodium hydroxide, boric acid, dialysis bags and animal cell culture media were purchased from Himedia Pvt. Ltd., India. Glacial acetic acid, hydrochloric acid and D_2O were procured from Merck India Pvt. Ltd. Human colorectal cancer (HT-29) and mouse fibroblast (L929) cell lines were procured from NCCS Pune, India. Antibodies were purchased from Invitrogen, USA.

6.2.2. Extraction of xylan from Neem sawdust

The extraction of xylan from neem was performed by following the protocol described earlier with slight modification (Zilliox & Debeire, 1998). Neem sawdust was dried at 80ºC for 24 h to remove the moisture content. Then, the 50 g dried sawdust of neem was delignified by treating with 500 mL of sodium chlorite at a final concentration of 1.2 g/g of biomass and 2.5 mL of glacial acetic acid at 60ºC for 3 h (Rowley, Decker, Michener & Black, 2015). The deliginified neem sawdust was transferred to a muslin cloth and washed with deionized water and dried in a hot air oven at 80ºC for 24 h. The 20g of delignified neem sawdust was soaked in 200 mL of an alkaline solution containing 10% (w/v) sodium hydroxide and 1% (w/v) boric acid and the mixture filled in 1L conical flask was kept in a stirring condition at 60ºC and 2000 rpm for 3 h. The suspension was filtered using a cotton cloth to separate the biomass. The 3 volume of ice-cold 600 mL ethanol (95%, v/v) and 10% (v/v) acetic acid were added to the collected filtrate (200 mL) and incubated for overnight at 4ºC to precipitate the extracted xylan, completely. The reducing sugar and impurities from extracted xylan were completely removed by washing 2 times with ethanol and after that, the xylan was recovered by centrifugation at 10000g for 10 min. The precipitated xylan was resuspended in 100 mL Milli-Q water and dialyzed against 5 L of Milli-Q water for 2 times to remove the trace amount of reagent and salts. The dialyzed xylan was lyophilized for further analysis and characterization.

6.2.3. Carbohydrate composition analysis of extracted xylan

The carbohydrate composition was determined by hydrolyzing the extracted xylan (5 mg) with 2M trifluoroacetic acid (TFA) in a boiling water bath for 3 h (Sharma, Rajulapati & Goyal, 2019). The residual TFA from the hydrolyzed xylan was removed by evaporation in a hot air oven at 80ºC for overnight. The dried hydrolyzed sample was dissolved in 500 µL of de-gassed Mili-Q water and the un-hydrolyzed part of the sample was removed by filtration through PVDF membrane (0.22 µm pore size) syringe filter. The filtrate was transferred to HPLC vial and analyzed for the presence of different sugars by using high-performance liquid chromatography (HPLC) (UFLC, Prominence, Shimadzu, Japan) equipped with a RI detector. 10 µL of the hydrolyzed xylan sample was loaded on phenomenex Rezex ROA (H+) (specification 300 mm x7.8 mm) connected with a guard column (specification 50 mm x 7.8 mm) and the sample was eluted by using 0.005N H_2SO_4 as the mobile phase at a flow rate of 0.5 mL/min. The xylose and glucuronic acid at 1 mg/mL concentrations were used as standard sugars for the determination of sugar composition. The uronic acid content of the extracted xylan was determined by Carbazole assay method by using D-Glucuronic acid as standard (Cesaretti, Luppi, Maccari, & Volpi, 2003).

6.2.4. Average molecular mass analysis of extracted xylans

The average molecular mass of extracted xylan from neem was determined by High-performance size exclusion chromatography (HPSEC) analysis by using Phenomenex Polysep-GFC-P6000 column coupled with a guard column (Phenomenex Polysep-GFC-P) connected to HPLC system equipped with RI detector. 10 µL of extracted xylan and dextran standard molecular weight markers (10 kDa, 20 kDa, 40 kDa, 70 kDa, 100 kDa 200 kDa, 270 kDa and 500 kDa at 1 mg/mL concentration) was loaded and 100 mM $NaNO_3$ solution was used as mobile phase for the elution at a flow rate of 0.5 mL/min. The molecular mass of extracted xylan was calculated by using dextran standards.

6.2.5 Particle size determination of extracted xylan

The particle size distribution or hydrodynamic radius of extracted xylan from neem was determined by using dynamic light scattering (DLS) instrument (Litesizer™ 500, Anton Paar, Gmbh, Austria). The extracted xylan sample (1 mg/mL, w/v) was dissolved in Milli-Q degassed water by heating at 60ºC and then filtered by using 0.45 µm PVDF membrane filter. The DLS system temperature at 25ºC was maintained by a Peltier temperature controller. Total 30 numbers of scans were recorded and the average was taken by using Kalliope software for the particle size analysis.

6.2.6. Fourier Transform Infrared spectroscopic analysis of extracted xylan

The determination of functional groups present in extracted xylan from neem was performed by FT-IR spectroscopy. Approximately, 1 mg of extracted xylan was mixed with 100 mg KBr and the mixture was converted into the thin pellet by pressing the xylan-KBr mixture using a 15-ton hydraulic press. The FT-IR spectra for extracted xylan was recorded using FT-IR spectrophotometer (Perkin Elmer, Spectrum Two, USA) in transmission mode from 4000-400 cm^{-1} with a spectral resolution of 4 cm^{-1}.

6.2.7 NMR analysis of extracted xylan from neem

15 mg of extracted xylan from neem was dissolved in 600 µL of D_2O (99.96%) (Merck, Germany) for 30 min at 60ºC and further centrifuge at 13000*g* for the removal of residual xylan. Total 200 scans for ^1H, 2000 scans for ^{13}C and 16 scans for TOCSY-HSQC and HMBC High-resolution NMR spectra were recorded at 25ºC using 600 MHz Nuclear Magnetic Resonance (NMR) Spectrometer (Bruker, ASCEND 600, Karlsruhe, Germany) fitted with a 5-mm probe. Tetramethylsilane (TMS) was used as internal reference for the calibration of ^1H and ^{13}C NMR spectra. ^1H, ^{13}C and 2D-HSQC spectra were processed and analyzed using topspin 4.07 NMR software (Bruker, Karlsruhe, Germany).

6.2.8. FE-SEM analysis of extracted xylan

Surface properties of extracted xylan from neem were investigated by using a field-emission scanning electron microscope (Carl Zeiss, Model-Gemini 300, Germany). A pinch of extracted xylan powder from neem was placed and spread on a carbon tape adhered on the surface of a stub. The gold coating of the sample placed on stub was performed and kept in the vacuum chamber before imaging using FESEM (Zeiss, Sigma, Germany).

6.2.9 Thermogravimetric (TG) Analysis

The thermal degradation properties (thermogravimetric and differential thermal properties) of extracted xylan was determined by estimating the change in enthalpy and weight loss using a thermal analyzer (Netzsch, Model: STA449F3A00).

Approximately, 6 mg of extracted xylan was filled in an alumina crucible and the sample was heated in the range of 25ºC-500ºC with a heating rate of 10ºC min^{-1} and the apparatus was continuously flushed with argon at a flow rate of 60 mL/min at the atmospheric pressure.

6.2.10 Xylanase mediated production and identification of xylooligosaccharides

Endo-β-1,4-xylanase from *Pseudopedobacter saltans* (*Ps*GH10A) was expressed and purified by following earlier reported protocol (Sharma, Antunes, Rajulapati & Goyal, 2018). The hydrolytic activity of *Ps*GH10A enzyme was determined by using their optimum condition. 1% (w/v) of extracted xylan was dissolved in 10 mL of 50 mM citrate phosphate buffer, pH 6.0 and mixed with 200 µg of *Ps*GH10A. Then the reaction mixture mentioned above was incubated at 40ºC for 24 h. After that, the reaction was terminated by the addition of 3 volume of ice-cold 95% (v/v) ethanol and the residual unhydrolyzed extracted xylan was separated by centrifugation at 10000g for 5 min. The xylooligosaccharide-ethanol mixture was kept at 80ºC for 12 h to evaporate the ethanol and concentrate the xylooligosaccharides. The dried product was dissolved in 200 µl of deionized water and the xylooligosaccharides yield (total carbohydrate content) was determined by using the phenol-sulfuric acid method (Dubois, Gilles, Hamilton, Rebers & Smith, 1956). The qualitative estimation and identification of xylooligosaccharides were performed by running of 0.5 µl of concentrated sample on TLC plate using Chloroform:Acetic acid:Water in a ratio of 6:7:1 as a mobile phase. The TLC was developed by staining with visualizing solution containing 0.5% w/v α-naphthol dissolved in a solution of methanol and sulphuric acid at a ratio of 95:5. Xylooligosaccharide with degree of polymerization (DP) of 2-5 and alduronic acid mixture (1 mg/mL) was used as the standard for identifying the hydrolyzed products. The mixture of XOS was analysed by ^{1}H, ^{13}C and 2D-HSQC analysis for the structural characterization using the same parameters mentioned in section 2.7.

6.2.11 Anticancer potential of xylooligosaccharides

6.2.11.1 In-vitro cell proliferation and mode of cell death analysis

Cell cytotoxicity of xylooligosaccharides was measured by performing MTT assay on human colorectal (HT29) and mouse fibroblast (L929) cell lines. 6×10^{3} cells per well of each cell line were seeded in 96 well plates and incubated for 12h for attachment. The cells were treated with different concentrations of xylooligosaccharides ranging from 0.05 mg/mL to 0.5 mg/mL and further incubated at 37ºC under 5% CO_2 for 24h and the cell viability was determined following the standard protocol (Mosmann, 1983). The apoptotic activity of the xylooligosaccharides treated colon cancer cells was analyzed by determining the phosphatidylserine translocation and cell viability using Annexin V labeled with FITC and propidium iodide (PI) (Invitrogen, USA) dyes. Cell cycle analysis of xylooligosaccharide treated HT29 cells was performed as described previously (Riccardi & Nicoletti, 2006).

6.2.11.2 Cell imaging analysis of xylooligosaccharide treated HT-29 cells

Caspase activity and nuclear morphology and plasma membrane integrity were determined by Image-iT ™ LIVE Red Caspase Detection Kits (Molecular Probes, Invitrogen) according to the manufacturer's protocol. The change in mitochondrial membrane potential between treated and untreated HT-29 cells was monitored by using mitochondria-specific dual cationic fluorescence dye, JC-1. The immunoblotting analysis of xylooligosaccharide treated and untreated HT-29 cells were performed to analyze the expression of cleaved Poly ADP-ribose polymerase (PARP) for the caspase-mediated apoptotic killing of HT-29 cells. The genomic DNA fragmentation in the form oligonucleosomal size is the distinctive feature of apoptosis. The HT29 cells were treated with xylooligosaccharides at a different concentration ranging from 0.05 mg/mL to 0.3 mg/mL and analyzed for DNA fragmentation following reported protocol (Morla, Kumar & Kumar, 2019).

6.3. Results and Discussion

6.3.1. Extraction, composition and molecular mass analysis of extracted xylan

The extraction process gave 4.5 g of xylan from 20 g of neem sawdust, resulting 22.5% yield after lyophilization (Fig. 6.1A). The HPLC analysis of the TFA hydrolyzed extracted xylan revealed that it contains 92% (w/w) xylose units and 8% (w/w) glucuronic acid units (Fig. 6.2). The Carbazole assay displayed that the extracted xylan from neem sawdust contains ~14% of glucuronic acid (139±5 µg/mg of xylan). The presence of xylose units in a vast amount in the extracted xylan revealed that its main chain is composed of D-xylose monomers linked by the xylosidic bonds and glucuronic acid as a side chain substitution.

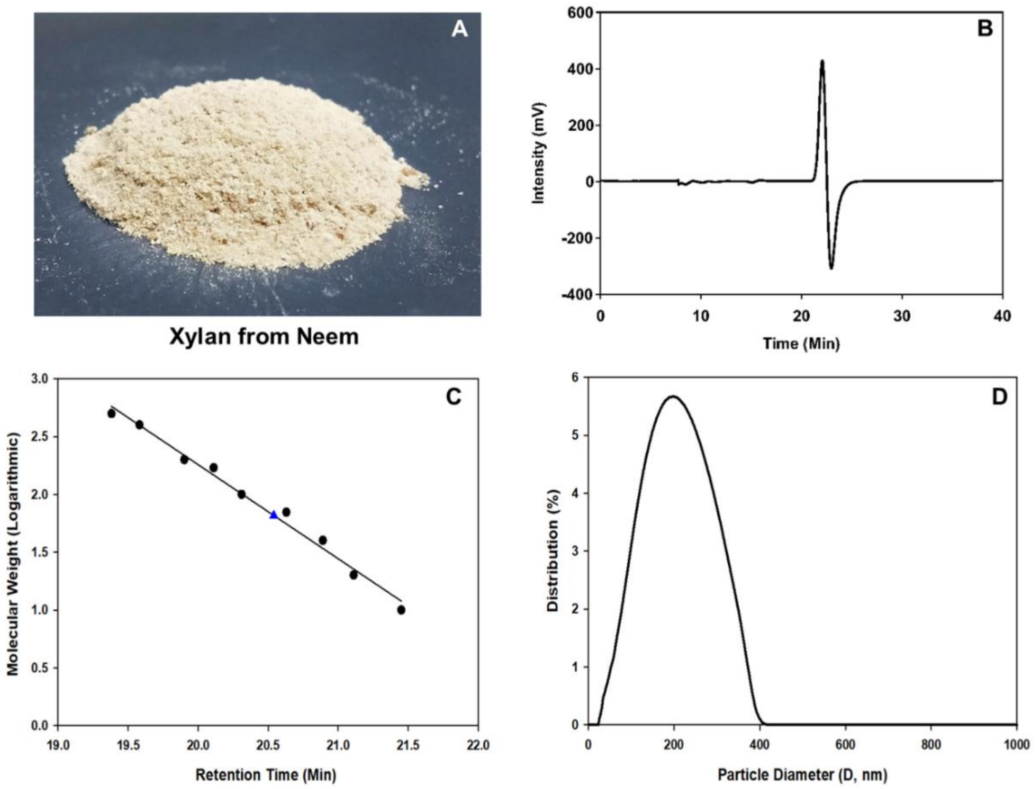


Fig. 6.1. A) Digital image of extracted xylan from neem, B) HP-SEC profile of extracted xylan, C) Hendricks plot analysis of dextran standard, D) DLS profile of extracted xylan.

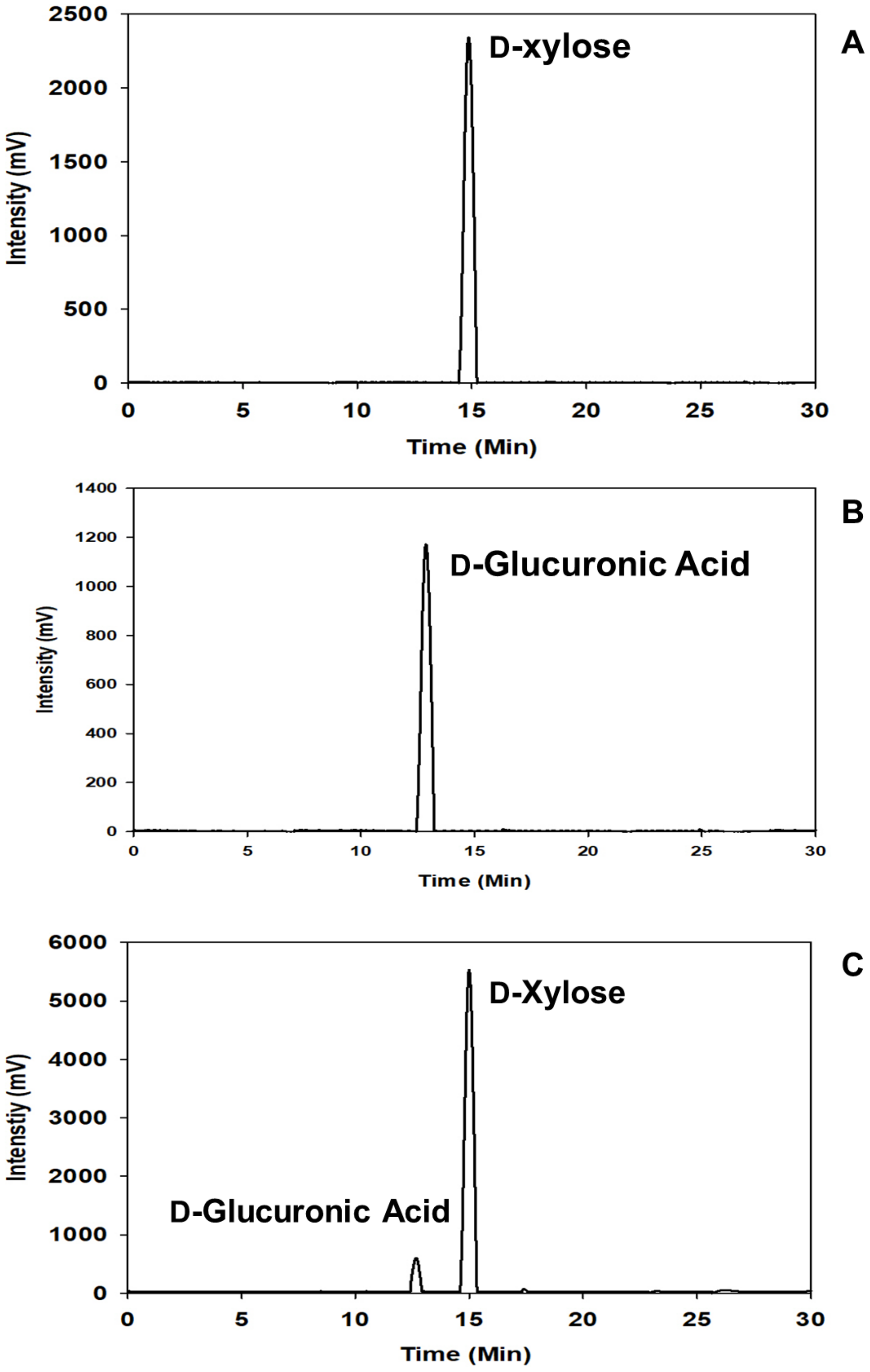


Fig. 6.2. The HPLC profiles of A) D-xylose, B) D-Glucuronic Acid and C) TFA hydrolyzed xylan extracted from neem sawdust.

Most of the xylans obtained from hardwood contain 4-O-methyl-α-D-glucuronic acid residue (MeGlcA) substitution at the O2 position of the xylan backbone are termed as 4-O-methyl-α-glucurono-D-xylan (Ebringerová, Kardošová, Hromádková, Malovíková, & Hříbalová, 2002). The ratio of Xyl/4-O-MeGlcA in 4-O-methyl-α-glucurono-D-xylan vary from 4:1 to 16:1 depending on the source and extraction conditions with an average ratio of about 10:1 (Ebringerová, Kardošová, Hromádková, Malovíková, & Hříbalová, 2002). In the present study, the molar ratio of Xyl/4-O-MeGlcA was 11:1, which is similar to results reported earlier (Sun, Wen, Xu & Sun, 2011; Zhang et al., 2016). The molecular mass analysis of extracted xylan by HPSEC displayed a single peak (Fig. 6.1B) and the molecular mass calculated by using commercial dextran standards as molecular mass marker was found to be ~66 kDa (Fig. 6.1C). The molecular mass of extracted xylan from neem sawdust is similar to glucuronoxylan isolated from the woody part of eucalyptus is of 56.60 kDa (Li, Sun, Zhou, Peng & Sun, 2015), from the apical portion of *Neolamarckia cadamba* is of 70.8 kDa (Zhao et al., 2017) and from *Cassia obtusifolia* seeds is of 49 kDa (Feng, Yin, Nie, Wan & Xie, 2018). Several other studies reported the lower molecular mass of extracted xylan 15 kDa from viscose fiber (Zhang et al., 2016) and 20.84 kDa from beechwood (Konduri and Fatehi, 2016). The DLS analysis of extracted xylan displayed monodispersity and gave a single peak with a hydrodynamic diameter of 203 nm (Fig. 6.1D). The results obtained from DLS analysis, corresponded with single peak obtained by mass by HPSEC analysis.

6.3.2. FTIR analysis of extracted xylan

The FTIR spectrum of extracted xylan displayed a distinct peak at 3429 cm^{-1} (Fig. 6.3), attributed to hydroxyl (OH) group stretching vibration, which is similar to the peak obtained by OH stretching vibration for other extracted xylans (Shi, Dong & Ding, 2014; Singh, Bhuyan, Banerjee, Muir, & Arora, 2017). The symmetric CH stretching in extracted xylan was revealed by the peak at 2922 cm$^{−1}$ as also reported earlier (Banerjee, Patti, Ranganathan & Arora, 2019). The presence of glucuronic acid component was represented by asymmetric and symmetric stretching mode of the carboxyl group of glucuronic acid at 1569 cm^{-1} and 1413 cm^{-1}, respectively as also mentioned previously (Bian, Peng, Peng, Xu & Sun, 2010; Biely, Malovíková, Hirsch, Morkeberg Krogh & Ebringerová, 2015; Corradini, et al., 2018). The peak at 1642 cm$^{−1}$ displayed the water absorption. (Bian et al., 2012; Zhang et al., 2016). The small peak at 897 cm$^{−1}$ was assigned to the absorption by a β-glycosidic linkage which is considered as anomeric region (Kacuráková, Capek, Sasinková, Wellner & Ebringerová, 2000; Zhang et al., 2016). The FTIR spectrum region between 1200-1000 cm^{-1} was dominated by the ring vibration overlapped by stretching vibration of OH side group and the C-O-C glycosidic vibrations. The peak at 1047 cm^{-1} represented the β-1,4 backbone of extracted xylan as reported earlier (Kacuráková, Capek, Sasinková, Wellner & Ebringerová, 2000).

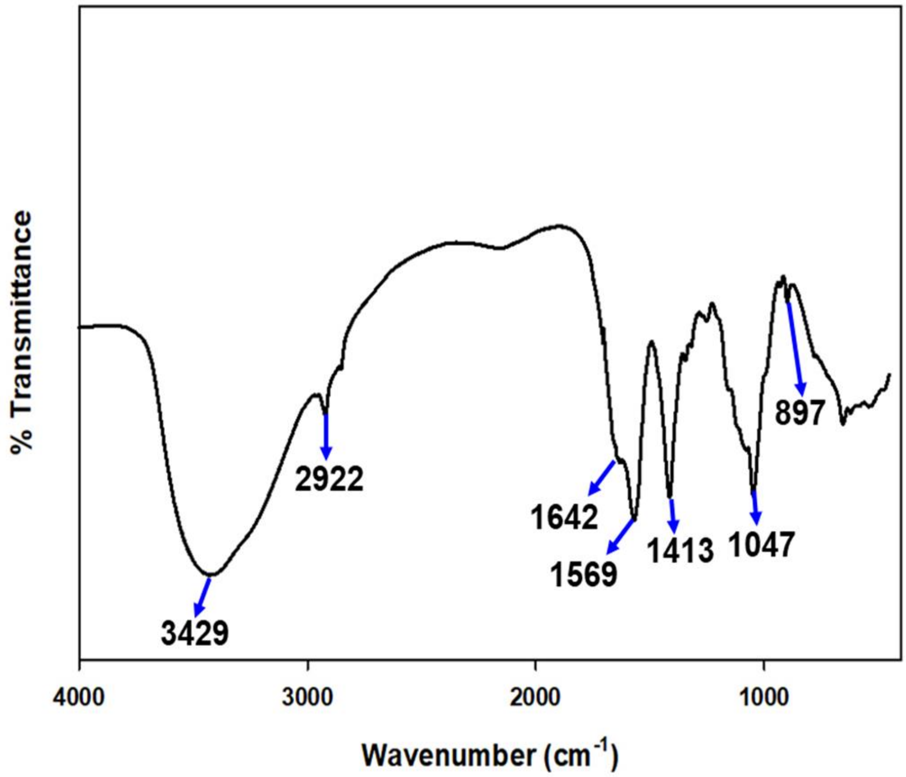


Fig. 6.3. Fourier Transform Infrared (FT-IR) spectroscopic analysis of extracted xylan (A), TOCSY-HSQC NMR (B) and HMBC NMR (C) analysis of extracted xylan.

6.3.3. NMR analysis of extracted xylan

The structure of extracted xylan from neem sawdust was determined by recording the ^{1}H (Fig. 6.4A), ^{13}C (Fig. 6.4B), 2D TOCSY-HSQC (Fig. 6.4C) and HMBC (Fig. 6.4D) NMR spectra at 600 MHz (Table 6.1).

Table 6.1. NMR analysis of extracted xylan from neem sawdust.

Linkages		Chemical shift (ppm)							
		1	2	3	4	5eq	5ax	6	-OCH$_3$
→4)-β-D-Xyl*p*-(1→	^{1}H	4.30	3.10	3.39	3.62	3.99	3.27		
	^{13}C	101.51	72.50	73.47	76.05	62.67	62.69		
→4)-β-D-Xyl*p*-2-*O*-Glc*p*A-(1→	^{1}H	4.45	3.26	3.41	3.64	3.95	3.23		
	^{13}C	101.18	75.4	71.05	76.18	62.65	62.58		
4-*O*-Me-α- D-Glc*p*A-(→	^{1}H	5.12	3.45	3.59	3.38	4.21			3.28
	^{13}C	97.32	72.02	71.70	72.17	71.54		176.8	59.61

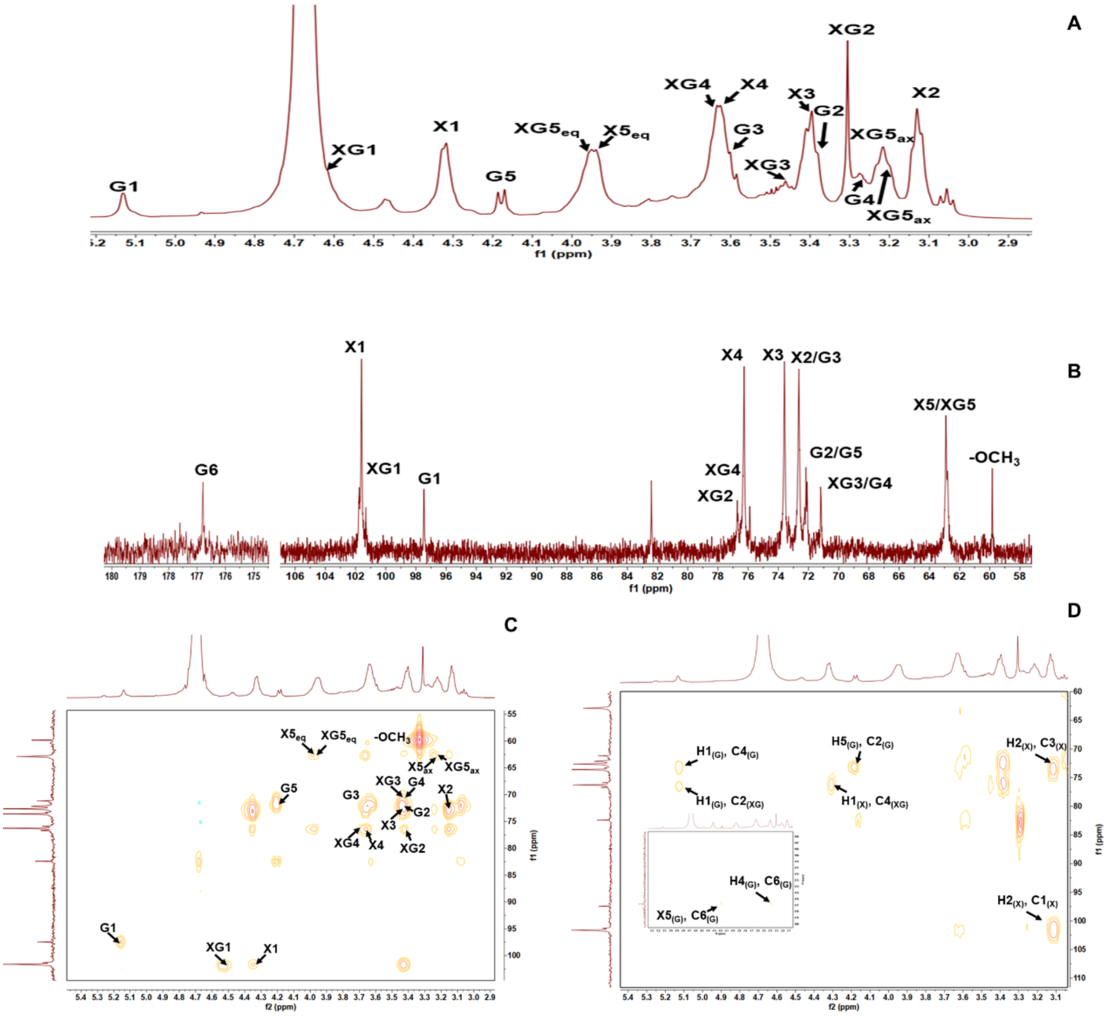


Fig. 6.4. ^{1}H NMR Spectrum (A), ^{13}C NMR Spectrum (B) TOCSY-HSQC NMR (C) and HMBC NMR (D) analysis of extracted xylan.

The characteristic signals of the chemical shift were assigned by combining the data obtained from ^{1}H, ^{13}C, 2D HSQC and HMBC NMR spectra analysis with previous reports of xylans from seed pericarp of *Argania spinosa* fruit (Habibi & Vignon, 2005)’ xylan from roots of *Cudrania tricuspidata* (Shi, Dong & Ding, 2014) and xylan from *Cassia obtusifolia* seeds (Feng, Yin, Nie, Wan & Xie, 2018). ^{1}H NMR spectrum analysis confirmed the presence of α-anomeric proton at 5.16 ppm for 4-O-methyl-glucuronic acid (G) and β-anomeric protons at 4.35 for xylose (X) and 4.55 for 4-O-methyl-glucuronic acid linked xylose residue (XG).

The integration analysis of ^{1}H NMR spectrum revealed that the ratio of xylose:4-O-methyl-glucuronic acid is 7:1. The corresponding anomeric carbon atoms identified from TOCSY-HSQC analysis were found at δ 101.51, 101.18 and 97.32 for X, XG and G, respectively. The chemical shift at 101.51 and 4.30 in TOCSY-HSQC spectrum confirms β(1→4) linkage of the non-substituted D-xylopyranosyl (X) units (Fig. 6.4C). The chemical shift at δ 3.99 ($H5_{eq}$), 3.62 (H4), 3.39 (H3), 3.27 ($H5_{ax}$) and 3.10 ppm (H2), corresponds with the cross chemical shifts of non-substituted xylose carbon atoms obtained from TOCSY-HSQC at 62.67 ($C5_{eq}$), 76.05 (C4), 73.47 (C3), 62.69 ($C5_{ax}$) and 72.50 (C2), respectively. All the chemical shifts of residue X obtained from ^{1}H and ^{13}C further confirms the β(1→4) linkage of residue X and are consistent with previous reports. The anomeric carbon chemical and proton shift of residue 4-O-methyl glucuronic acid linked xylose (XG) were at δ 101.18 and 4.45, respectively also revealed the β-linkage between X and XG (Fig. 6.4C). Chemical shifts of XG carbon atoms at δ 75.4, 71.05, 76.18, and 62.65 corresponded to C2, C3, C4 and C5 of glucuronic acid substituted xylose (→4)-β-D-Xyl*p*-2-*O*-Glc*p*A-(1→), respectively. The low intensity signal at δ 75.4for C2 position of residue XG as compared with the signal of the C2 position of X2 residue indicated the substitution of the XG residue. The corresponding proton NMR peaks of XG residues are at 3.26 (H2), 3.41 (H3), 3.64 (H4), 3.95 ($H5_{eq}$) and 3.23 ($H5_{ax}$) assigned to the β-D-xylopyranosyl units substituted with 4-O-methyl glucuronic acid at O2 position. Similarly, G residue showed the chemical shift for anomeric carbon and proton at δ97.32 (C1) and 5.12 (H1), respectively indicating the α-linkage of glucuronic acid with xylan main chain. The chemical shifts of other carbon at δ 72.02, 71.10, 72.17, 71.54, and 176.56 corresponded to C2, C3, C4, C5, and C6 of α-GlcpA-(1→, respectively. The chemical shift at δ59.77 corresponding to the methoxy group ($-OCH_3$) confirmed the methylation of glucuronopyranosyl residues at O-4 position. The cross chemical shift peak of proton present in glucuronopyranosyl residues were identified and designated as 3.45 (H2), 3.59(H3), 3.38 (H4), 4.21 (H5) and 3.28($O-CH_3$). The side chain and main chain monomeric sugar residues present in xylan extracted from neem sawdust were identified by a long-range HMBC spectrum (Fig. 6.4D). The cross-peak at δ5.15/δ76.23 demonstrated that the H1 of G residue correlated to C2 of XG residue, further confirming that the methylated glucuronopyranosyl branches are linked at the O2 position of XG residue. The correlation signal between C4 (δ76.23) of XG residue and H1 (δ4.31) of X residue indicates that the backbone of the extracted polysaccharide is composed of β-(1→4)-linked xylopyranosyl residues.

6.3.4. FE-SEM and TGA analysis of extracted xylan

The surface and morphological properties of dried extracted xylan powder were observed by the FE-SEM analysis. The surface of extracted xylan is highly porous and rough, which is composed of irregular granules (Fig. 6.5A). In the previous study xylan from corn cob (Oliveira et al., 2010), xylan from rice bran and finger millet (Palaniappan, Yuvaraj, Sonaimuthu & Antony, 2017) and also xylan from pineapple peel waste (Banerjee, Patti, Ranganathan & Arora, 2019), displayed the rough and irregular surface morphology. The thermal decomposition and thermal stability analysis of extracted xylan were studied by Thermogravimetric analysis (TGA) and Differential Thermogravimetric analysis (DTG) analysis. The TGA analysis of extracted xylan displayed the weight loss with an increase in temperature (Fig. 6.5B). The loss of mass of the extracted xylan occurred in three stages at different temperatures. The initial 7%

weight loss was observed between the range of 70ºC to 140ºC, which attributed to the loss of water content or moisture content of the extracted xylan (Fig. 6.5B).

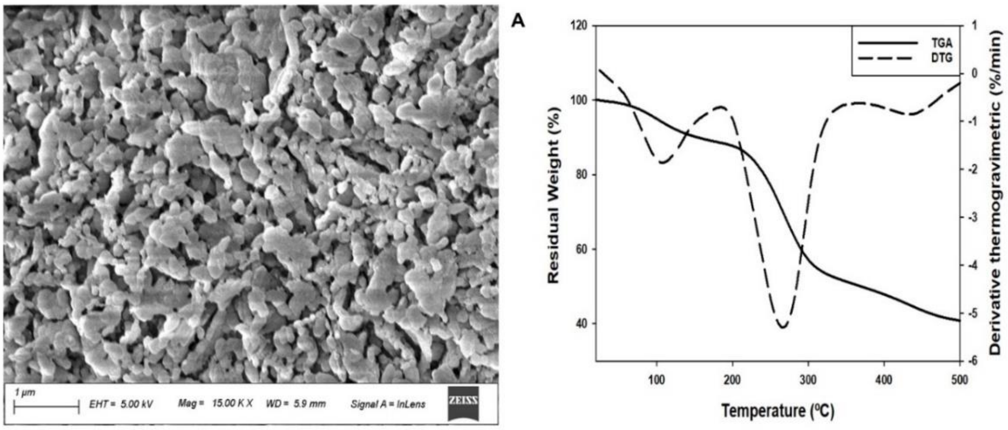


Fig. 6.5. FE-SEM analysis of extracted xylan A), Thermogravimetric analysis (TGA) and Derivative Thermogravimetric Analysis (DTG) displaying thermal degradation temperature (T_d) of 271ºC B).

The obtained results are similar to earlier reports (Zhang, Wang, Liu, Zhang & Ren, 2017; Deumaga et al., 2018). The second event of sudden weight loss started at 195ºC and 45% weight loss was observed 273ºC by fragmentation and decomposition of the extracted xylan into smaller molecules like CO_2, CO, CH_4, HCOOH and CH_3COOH as suggested by Bian et al. (2012), Zhao et al. (2017) and Deumaga et al. (2018). The third event of decomposition displayed the conversion of extracted xylan into the ash content by the pyrolysis process that retained 40% of initial weight. DTG_{max} demonstrated the maximum thermal degradation (Td) at a particular temperature, which can be used for analyzing the stability of the sample. The DTG curve of extracted xylan displayed the single sharp peak at 271°C, which may be due to the presence of less substitution in the main chain and the peak corresponded to the DTG_{max} or T_d.

6.3.5. Xylanase mediated production and identification of xylooligosaccharides

The hydrolytic activity of *Ps*GH10A against extracted xylan from neem was determined. The specific activity of *Ps*GH10A was found to be 45 U/mg. The specific activity of these enzymes on extracted xylan was comparable with the available commercial substrates *viz.* Beechwood xylan, xylan and Birchwood xylan (Sharma, Antunes, Rajulapati & Goyal, 2018). Therefore, the extracted xylan can be used as a substitute for the commercially available substrate such as beechwood xylan and birchwood xylan. *Ps*GH10A was further used for the production of xylooligosaccharides. The total yield of xylooligosaccharides from hydrolysis of 10 mL, 1% (w/v) extracted xylan by *Ps*GH10A was found to be 50±2.3 mg/mL as determined by phenol sulfuric acid analysis. The mixture of xylooligosaccharides was analyzed by TLC. The major products formed by enzymatic hydrolysis of extracted xylan were series of xylooligosaccharides with various degree of polymerization ranging from DP2 to DP7 (Fig. 6.6A).

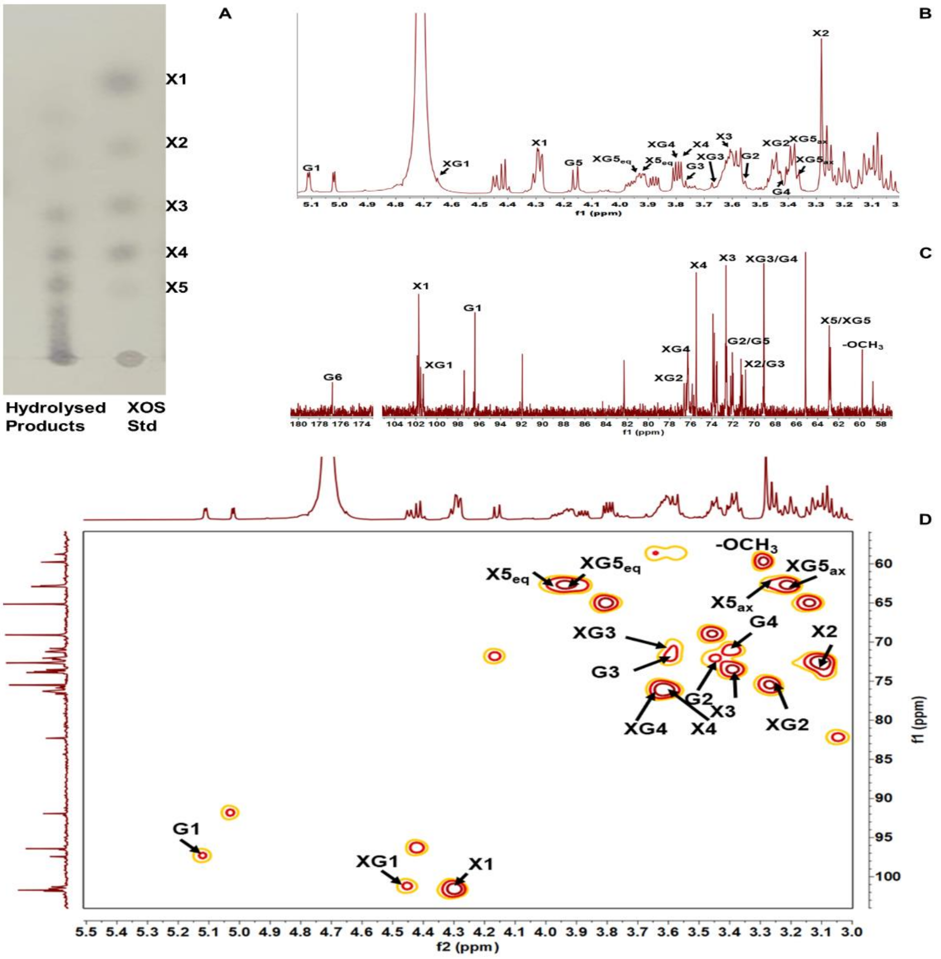


Fig. 6.6. A) TLC analysis of hydrolyzed products of extracted xylan from neem sawdust by *Ps*GH10A, Lane 1 *Ps*GH10A hydrolyzed products Lane 2 Xylooligosaccharide standards (XOS Std) where X1= xylose, X2 xylobiose, X3=xylotriose, X4=xylotetraose and X5=xylopentaose, B) ^{1}H NMR analysis of XOS, C) ^{13}C NMR analysis of XOS and D) 2D-HSQC NMR analysis of XOS.

The ^{1}H (Fig. 6.6B) and ^{13}C NMR (Fig. 6.6C) spectra analysis of XOS displayed the splitting of peaks at δ 4.45/101.18 ppm for XG, δ 4.30/101.51 ppm for X and δ 5.12/97.32 ppm for G. The splitting peaks for XG and X corresponded to β-anomeric and the peak for G corresponded to α-anomeric form. The splitting of the peaks may be due to the presence of varying lengths of XOS. The integration analysis of anomeric peaks revealed the molar ratio of xylose:4-O-Methyl glucuronic acid residue is 6.5:1.

The 2D-HSQC analysis of XOS clearly showed resolved peaks for X, XG and G (Fig. 6.6D), which when compared with the TOCSY-HSQC NMR analysis (Fig. 6.4C) results of extracted xylan were found to be similar.

6.3.6. Anticancer potential of xylooligosaccharides

6.3.6.1. In-vitro cell proliferation and mode of cell death analysis

The *in vitro* cytotoxicity analysis of xylooligosaccharide (XOS) produced by *Ps*GH10A mediated hydrolysis of extracted xylan was performed by using MTT assay. The treatment of L929 cells with 0.2 mg/mL XOS concentration displayed no significant changes and inhibition up to 15% was observed after 24h incubation (Fig. 6.7A). Therefore, the MTT analysis of XOS on normal cells displayed no cytotoxicity and that could be used for safe consumption as a food supplement. The cytotoxicity of XOS on HT-29 cells was studied with a concentration range of 0.05 mg/mL to 0.2 mg/ml. The result showed that the treatment with 0.2 mg/mL XOS causes 55% cell inhibition in HT-29 cells after 24h incubation (Fig. 6.7B). Perhaps, the mixture of XOS can be used as a prospective anticancer agent. The microscopic analysis of XOS treated and untreated L929 and HT29 cells were performed. The XOS treated at a concentration of 0.2 mg/mL and untreated L929 cells displayed no visible alteration in its shape and retained their morphology in both the conditions (Fig. 6.7C & D) as also reported earlier (Theerakittayakorn & Bunprasert, 2011). The result suggested that the XOS are non-toxic to the normal fibroblast cells. The untreated HT-29 cells form a multilayered undifferentiated and unpolarized mass (Fig. 6.7E) as reported earlier (Hekmati, Ben-Shaul & Polak-Charcon, 1990). The HT29 cells treated with 0.2 mg/mL XOS displayed irregular morphology having a disrupted cell membrane with weak adhesion to the surface of culture vessels (Fig. 6.7F). The cell imaging analysis of L929 and HT29 cells under a bright-field microscope further supported the cell viability results. The cell cycle distribution analysis of HT-29 cells treated with 0.2 mg/mL XOS resulted in a reduction of cells from 53.6% to 34.4% present in G0/G1 phase and G2/M phase from 4.4% to 3.3%. The XOS treatment of HT-29 cells induces the presence of apoptotic cell population in sub G1 phase from 10.2% to 24.7%. (Fig. 6.7G & 5H). These results revealed that the treatment of XOS retards the growth of HT-29 cells by arresting the cell cycle progression and presumably inducing apoptosis by showing changes in the expression level of some cell cycle regulatory genes.

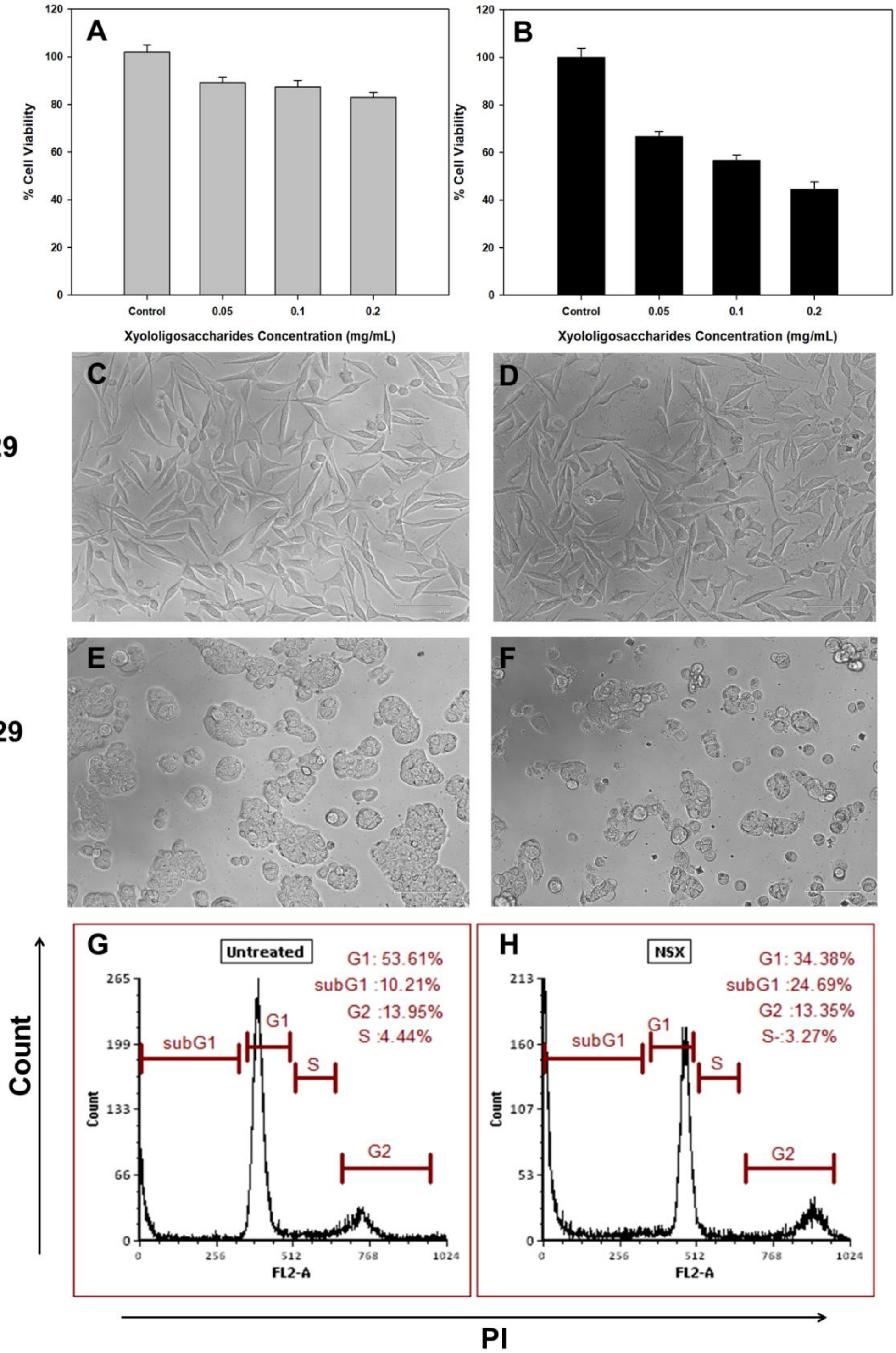


Fig. 6.7. Effect of xylooligosaccharides on L929 (Mouse fibroblast) and HT-29 (Human Colorectal cancer) cell lines, MTT analysis of L929 A), HT-29 B), cell morphology analysis of L929 cell -Untreated C), Treated D), HT-29 Cells-Untreated E) and Treated F). Cell cycle analysis of control G) and xylooligosaccharides treated HT-29 cells, H).

6.3.6.2. Cell imaging analysis of xylooligosaccharide treated HT-29 cells

The XOS treated HT-29 cells displayed the reduction in the red/green fluorescence intensity of JC-1 dye as compared with the untreated cells (Fig. 6.8A) suggesting the loss of mitochondrial membrane potential. The opening of mitochondrial permeability transition pores (PTP) resulted in the release of cytochrome c and other pro-apoptotic proteins from mitochondria to cytosol, leading to the loss of its membrane potential ($\Delta\psi m$) (Gil, Almeida, Oliveira & Rego, 2003). These results revealed the disruption of mitochondrial functions in HT-29 cells at an early stage of apoptosis by caspase-3 activation after the treatment with XOS at a concentration of 0.2 mg/mL. The similar effect was polysaccharide isolated from *Cymbopogon citratus* was previously reported on reproductive cancer cells such as Siha and LNCap (Thangam et al., 2014).

FLICA staining of XOS treated HT-29 cells displayed the caspase 3 and caspase 7 activities and giving red fluorescence as compared with the untreated cells (Fig. 6.8B). Besides, SYTOX staining of treated cells and produces green fluorescence, which indicates the presence of dead cells (Fig. 6.8B). Similarly, the Hoechst 33342 staining, demonstrated the chromatin condensation in xylooligosaccharide treated HT-29 cells (Fig. 6.8B). The immunoblotting analysis of XOS treated HT-29 cells displayed the cleavage of PARP as compared with untreated cells further supports the caspase 3 mediated apoptosis (Fig. 6.8C). DNA fragmentation analysis of XOS treated HT-29 cells revealed fragmentation of genomic DNA with the increasing concentrations of XOS. The internucleosomal DNA degradation effect is due to the activation of nuclear endonuclease in the XOS treated cells, which further confirmed the process of apoptosis (Fig. 6.8D). However, no DNA Laddering was observed in untreated cells. A similar effect of apple pectin oligosaccharide was reported earlier on MDA-MB-231 cells (Delphi, Sepehri, Khorramizadeh & Mansoori, 2015).

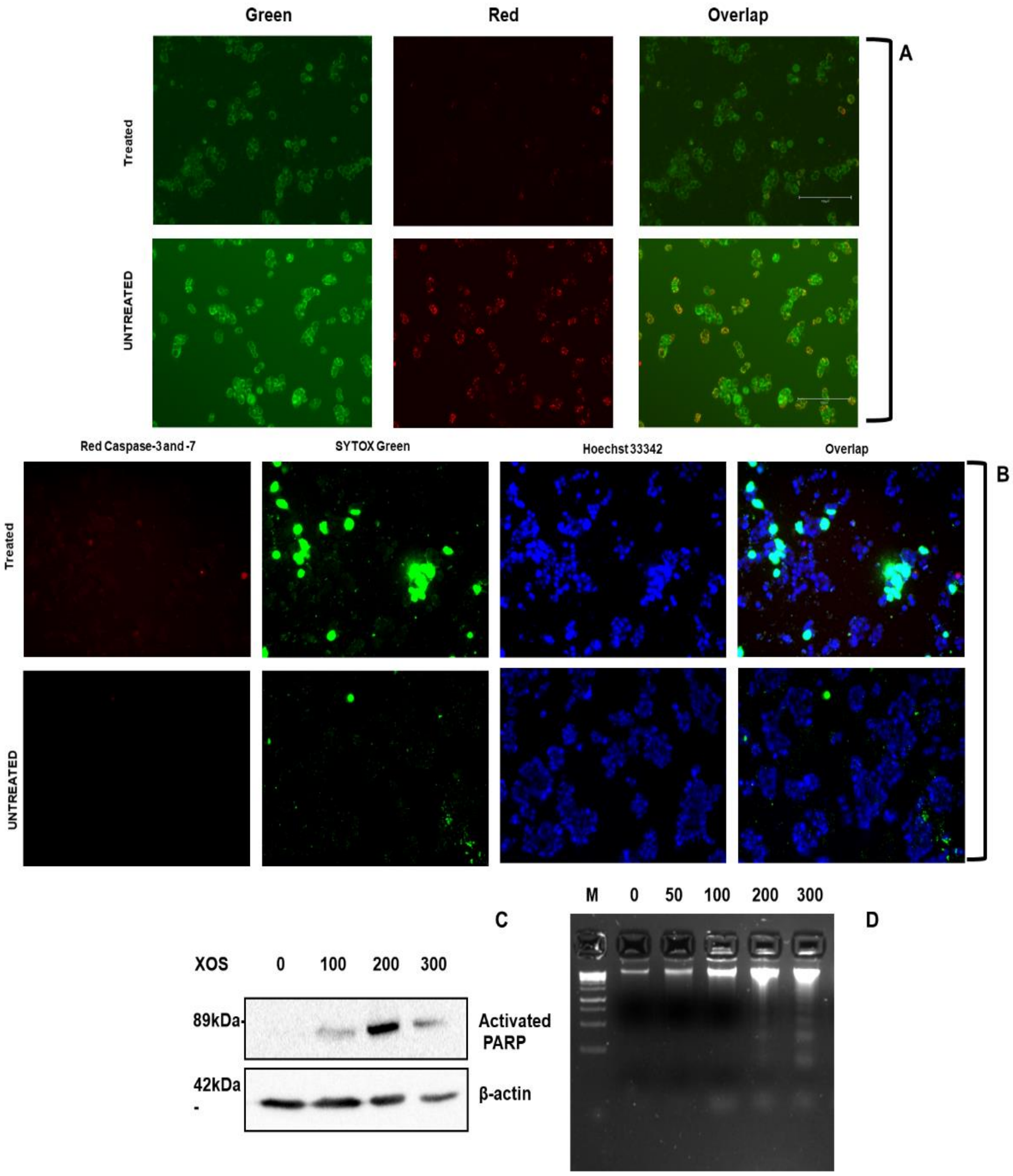


Fig. 6.8. Cell imaging analysis of untreated and treated HT-29 cells A) Image showing the decreased mitochondrial membrane potential upon xylooligosaccharides treatment B). Caspase activity, nuclear fragmentation and plasma membrane integrity analysis. Immunoblotting analysis of cleaved PARP and β-actin, where β-actin was used as a loading control C) and Agarose gel image showing DNA laddering of xylooligosaccharides treated cells D).

6.4. Conclusion

Glucuronoxylan was isolated from neem sawdust. Carbohydrate composition and structural characterization analysis of extracted xylan displayed that the main chain is composed of xylose backbone, which is substituted with a 4-O-methyl glucuronic acid side chain. The hydrolysis of extracted xylan from neem sawdust by endo-β-1,4-xylanase (*Ps*GH10A) resulted in the release of xylooligosaccharides ranging from the degree of polymerization (DP) 2-7. *In vitro* cell cytotoxicity analysis of xylooligosaccharide on mouse fibroblast (L929) cells displayed its biocompatible behavior. The HT-29 cells treated with xylooligosaccharides displayed the inhibition of cell growth. The mitochondrial membrane potential analysis of HT29 cells displayed reduction in the red/green fluorescence intensity confirming the disruption of mitochondrial membrane. FLICA, immunoblotting and nuclear fragmentation analysis revealed the caspase-3 mediated apoptotic killing mechanism of HT-29 cells. These results further contribute to the knowledge about xylooligosaccharides produced from extracted xylan gave better insights about their utilization in cancer therapy for the treatment of colon cancer.

6.5. References

Alzohairy, M.A., 2016. Therapeutics role of *Azadirachta indica* (Neem) and their active constituents in diseases prevention and treatment. Evidence-based Complement. *Altern. Med.* https://doi.org/10.1155/2016/7382506

Banerjee, S., Patti, A.F., Ranganathan, V., Arora, A., 2019. Hemicellulose based biorefinery from pineapple peel waste: Xylan extraction and its conversion into xylooligosaccharides. *Food Bioprod. Process.* 117, 38-50. https://doi.org/10.1016/j.fbp.2019.06.012

Barbat, A., Gloaguen, V., Moine, C., Sainte-Catherine, O., Kraemer, M., Rogniaux, H., Ropartz, D., Krausz, P., 2008. Structural characterization and cytotoxic properties of a 4-O-methylglucuronoxylan from *Castanea sativa*. 2. Evidence of a structure-activity relationship. *J. Nat. Prod.* 71, 1404-1409. https://doi.org/10.1021/np800207g

Bendahou, A., Dufresne, A., Kaddami, H., Habibi, Y., 2007. Isolation and structural characterization of hemicelluloses from palm of *Phoenix dactylifera* L. *Carbohydr. Polym.* 68, 601-608. https://doi.org/10.1016/j.carbpol.2006.10.016

Bian, J., Peng, F., Peng, P., Xu, F., Sun, R.C., 2010. Isolation and fractionation of hemicelluloses by graded ethanol precipitation from *Caragana korshinskii*. *Carbohydr. Res.* 345, 802-809. https://doi.org/10.1016/j.carres.2010.01.014

Bian, J., Peng, F., Peng, X.P., Xu, F., Sun, R.C., Kennedy, J.F., 2012. Isolation of hemicelluloses from sugarcane bagasse at different temperatures: Structure and properties. *Carbohydr. Polym.* 88, 638-645. https://doi.org/10.1016/j.carbpol.2012.01.010

Biely, P., Malovíková, A., Hirsch, J., Morkeberg Krogh, K.B.R., Ebringerová, A., 2015. The role of the glucuronoxylan carboxyl groups in the action of endoxylanases of three glycoside hydrolase families: A study with two substrate mutants. *Biochim. Biophys. Acta - Gen. Subj.* 1850, 2246-2255. https://doi.org/10.1016/j.bbagen.2015.07.003

Chang, W.T., Lai, T.H., Chyan, Y.J., Yin, S.Y., Chen, Y.H., Wei, W.C., Yang, N.S., 2015. Specific medicinal plant polysaccharides effectively enhance the potency of a DC-based vaccine against mouse mammary tumor metastasis. *PLoS One.* 10, e122374 https://doi.org/10.1371/journal.pone.0122374

Corradini, F.A.S., Baldez, T.O., Milessi, T.S.S., Tardioli, P.W., Ferreira, A.G., de Campos Giordano, R., de L.C. Giordano, R., 2018. Eucalyptus xylan: An in-house-produced substrate for xylanase evaluation to substitute birchwood xylan. *Carbohydr. Polym.* 197, 167-173. https://doi.org/10.1016/j.carbpol.2018.05.088

Das, D., Maiti, S., Maiti, T.K., Islam, S.S., 2013. A new arabinoxylan from green leaves of *Litsea glutinosa* (Lauraeae): Structural and biological studies. *Carbohydr. Polym.* 92, 1243-1248. https://doi.org/10.1016/j.carbpol.2012.10.052

Delphi, L., Sepehri, H., Khorramizadeh, M.R., Mansoori, F., 2015. Pectic-oligoshaccharides from apples induce apoptosis and cell cycle arrest in MDA-MB-231 cells, a model of human breast cancer. *Asian Pacific J. Cancer Prev.* 16, 5265-5271. https://doi.org/10.7314/APJCP.2015.16.13.5265

Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., Smith, F., 1956. Colorimetric Method for Determination of Sugars and Related Substances. *Anal. Chem.* 28, 350-356. https://doi.org/10.1021/ac60111a017

Ebringerová, A., Kardošová, A., Hromádková, Z., Malovíková, A., Hříbalová, V., 2002. Immunomodulatory activity of acidic xylans in relation to their structural and molecular properties. *Int. J. Biol. Macromol.* 30, 1-6. https://doi.org/10.1016/S0141-8130(01)00186-6

Feng, L., Yin, J.Y., Nie, S.P., Wan, Y.Q., Xie, M.Y., 2018. Enzymatic purification and structure characterization of glucuronoxylan from water extract of *Cassia obtusifolia* seeds. *Int. J. Biol. Macromol.* 107, 1438-1446. https://doi.org/10.1016/j.ijbiomac.2017.10.014

Gil, J., Almeida, S., Oliveira, C.R., Rego, A.C., 2003. Cytosolic and mitochondrial ROS in staurosporine-induced retinal cell apoptosis. *Free Radic. Biol. Med.* 35, 1500-1514. https://doi.org/10.1016/j.freeradbiomed.2003.08.022

Habibi, Y., Vignon, M.R., 2005. Isolation and characterization of xylans from seed pericarp of *Argania spinosa* fruit. *Carbohydr. Res.* 340, 1431-1436. https://doi.org/10.1016/j.carres.2005.01.039

Hashimoto, K., Okawa, Y., Suzuki, K., Okura, Y., Suzuki S., Suzuki, M., 1983. Antitumor activity of acidic mannan fraction from bakers yeast. *J. Pharmacobiodyn.* 6, 668-676. https://doi.org/10.1248/bpb1978.6.668

Hekmati, M., Ben-Shaul, Y., Polak-Charcon, S., 1990. A morphological study of a human adenocarcinoma cell line (HT29) differentiating in culture. Similarities to intestinal embryonic development. *Cell Differ. Dev.* 31, 207-218. https://doi.org/10.1016/0922-3371(90)90133-H

Kacuráková, M., Capek, P., Sasinková, V., Wellner, N., Ebringerová, A., 2000. FT-IR study of plant cell wall model compounds: Pectic polysaccharides and hemicelluloses. *Carbohydr. Polym.* 43, 195-203. https://doi.org/10.1016/S0144-8617(00)00151-X

Kapoor, S., Dharmesh, S.M., 2017. Pectic Oligosaccharide from tomato exhibiting anticancer potential on a gastric cancer cell line: Structure-function relationship. *Carbohydr. Polym.* 160, 52-61. https://doi.org/10.1016/j.carbpol.2016.12.046

Konduri, M.K.R., Fatehi, P., 2016. Synthesis and characterization of carboxymethylated xylan and its application as a dispersant. *Carbohydr. Polym.* 146, 26-38. https://doi.org/10.1016/j.carbpol.2016.03.036

Li, H.Y., Sun, S.N., Zhou, X., Peng, F., Sun, R.C., 2015. Structural characterization of hemicelluloses and topochemical changes in Eucalyptus cell wall during alkali ethanol treatment. *Carbohydr. Polym.* 123, 17-26. https://doi.org/10.1016/j.carbpol.2014.12.066

Lin, M., Xia, B., Yang, M., Gao, S., Huo, Y., Lou, G., 2013. Anti-ovarian cancer potential of two acidic polysaccharides from the rhizoma of *Menispermum dauricum*. *Carbohydr. Polym.* 92, 2212-2217. https://doi.org/10.1016/j.carbpol.2012.12.013

Melo-Silveira, R.F., Viana, R.L.S., Sabry, D.A., da Silva, R.A., Machado, D., Nascimento, A.K.L., Scortecci, K.C., Ferreira-Halder, C.V., Sassaki, G.L., Rocha, H.A.O., 2019. Antiproliferative xylan from corn cobs induces apoptosis in tumor cells. *Carbohydr. Polym.* 210, 245-253. https://doi.org/10.1016/j.carbpol.2019.01.073

Moine, C., Krausz, P., Chaleix, V., Sainte-Catherine, O., Kraemer, M., Gloaguen, V., 2007. Structural characterization and cytotoxic properties of a 4-O-methylglucuronoxylan from *Castanea sativa*. *J. Nat. Prod.* 70, 60-66. https://doi.org/10.1021/np060354p

Morla, S., Kumar, A., Kumar, S., 2019. Newcastle disease virus mediated apoptosis and migration inhibition of human oral cancer cells: A probable role of β-catenin and matrix metalloproteinase-7. *Sci. Rep.* 9, 10882. https://doi.org/10.1038/s41598-019-47244-y

Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods*. 65, 55-63. https://doi.org/10.1016/0022-1759(83)90303-4

Oliveira, E.E., Silva, A.E., Júnior, T.N., Gomes, M.C.S., Aguiar, L.M., Marcelino, H.R., Araújo, I.B., Bayer, M.P., Ricardo, N.M.P.S., Oliveira, A.G., Egito, E.S.T., 2010. Xylan from corn cobs, a promising polymer for drug delivery: Production and characterization. *Bioresour. Technol.* 101, 5402-5406. https://doi.org/10.1016/j.biortech.2010.01.137

Ordaz-Ortiz, J.J., Devaux, M.F., Saulnier, L., 2005. Classification of wheat varieties based on structural features of arabinoxylans as revealed by endoxylanase treatment of flour and grain. *J. Agric. Food Chem.* 53, 8349-8356. https://doi.org/10.1021/jf050755v

Palaniappan, A., Yuvaraj, S.S., Sonaimuthu, S., Antony, U., 2017. Characterization of xylan from rice bran and finger millet seed coat for functional food applications. *J. Cereal Sci.* 75, 296-305. https://doi.org/10.1016/j.jcs.2017.03.032

Petzold-Welcke, K., Schwikal, K., Daus, S., Heinze, T., 2014. Xylan derivatives and their application potential-Mini-review of own results. *Carbohydr. Polym.* 100, 80-88. https://doi.org/10.1016/j.carbpol.2012.11.052

Riccardi, C., Nicoletti, I., 2006. Analysis of apoptosis by propidium iodide staining and flow cytometry. *Nat. Protoc.* 1, p.1458. https://doi.org/10.1038/nprot.2006.238

Rowley, J., Decker, S.R., Michener, W., Black, S., 2015. Efficient extraction of xylan from delignified corn stover using dimethyl sulfoxide. *3Biotech.* 2013, 3, 433-438. https://doi.org/10.1007/s13205-013-0159-8

Prashanth, M.R.S., Shruthi, R.R., Muralikrishna, G., 2015. Immunomodulatory activity of purified arabinoxylans from finger millet (*Eleusine coracana,* v. Indaf 15) bran. *J. Food Sci. Technol.* 52, 6049-6054. https://doi.org/10.1007/s13197-014-1664-4

Sharma, K., Antunes, I.L., Rajulapati, V., Goyal, A., 2018. Molecular characterization of a first endo-acting β-1,4-xylanase of family 10 glycoside hydrolase (PsGH10A) from *Pseudopedobacter saltans* comb. nov. *Process Biochem*. 70, 79-89. https://doi.org/10.1016/j.procbio.2018.03.025

Sharma, K., Thakur, A., Goyal, A., 2019a. Green Bio-processes. Springer Singapore. https://doi.org/10.1007/978-981-13-3263-0

Sharma, K., Rajulapati V., Goyal, A., 2019b. Green synthesis of arabinoxyloglucan coated antimicrobial copper nanoparticles *Trends Carbohydr. Res.*, 2019, 11, 22-30.

Shi, L., Dong, Q., Ding, K., 2014. Structure elucidation and immunomodulatory activity in vitro of a xylan from roots of *Cudrania tricuspidata*. *Food Chem.* 152, 291-296. https://doi.org/10.1016/j.foodchem.2013.11.091

Shukla, R., Goyal, A., 2013. Novel dextran from *Pediococcus pentosaceus* CRAG3 isolated from fermented cucumber with anti-cancer properties. *Int. J. Biol. Macromol.* 62, 352-357. https://doi.org/10.1016/j.ijbiomac.2013.09.043

Singh, R.D., Bhuyan, K., Banerjee, J., Muir, J., Arora, A., 2017. Hydrothermal and microwave assisted alkali pretreatment for fractionation of arecanut husk. *Ind. Crops Prod*. 102, 65-74. https://doi.org/10.1016/j.indcrop.2017.03.017

Subapriya, R., Nagini, S., 2005. Medicinal properties of neem leaves: A review. *Curr. Med. Chem.-Anti-Cancer Agents*. 5, 149-156. https://doi.org/10.2174/1568011053174828

Sun, Y.C., Wen, J.L., Xu, F., Sun, R.C., 2011. Structural and thermal characterization of hemicelluloses isolated by organic solvents and alkaline solutions from *Tamarix austromongolica. Bioresour. Technol.* 102, 5947-5951. https://doi.org/10.1016/j.biortech.2011.03.012

Tan, H., Chen, W., Liu, Q., Yang, G., Li, K., 2018. Pectin oligosaccharides ameliorate colon cancer by regulating oxidative stress and inflammation-activated signaling pathways. *Front. Immunol.* 9, 1504. https://doi.org/10.3389/fimmu.2018.01504

Thangam, R., Sathuvan, M., Poongodi, A., Suresh, V., Pazhanichamy, K., Sivasubramanian, S., Kanipandian, N., Ganesan, N., Rengasamy, R., Thirumurugan, R., Kannan, S., 2014. Activation of intrinsic apoptotic signaling pathway in cancer cells by *Cymbopogon citratus* polysaccharide fractions. *Carbohydr. Polym.* 107, 138-150. https://doi.org/10.1016/j.carbpol.2014.02.039

Theerakittayakorn, K., Bunprasert, T., 2011. Differentiation capacity of mouse L929 fibroblastic cell line compare with human dermal fibroblast. *World Acad. Sci. Eng. Technol. 5, 51-54.*

Tiappi Deumaga, M.F., Jacquet, N., Vanderghem, C., Aguedo, M., Thomas, H.G., Gerin, P., Deleu, M., Richel, A., 2018. Fractionation and Structural Characterization of Hemicellulose from Steam-Exploded Banana Rachis. *Waste and Biomass Valorization.* https://doi.org/10.1007/s12649-018-0457-9

Tiwari, R., Verma, A.K., Chakraborty, S., Dhama, K., Singh, S.V., 2014. Neem (Azadirachta indica) and its potential for safeguarding health of animals and humans: A Review. *J. Biol. Sci.* 14, 110-123. https://doi.org/10.3923/jbs.2014.110.123

Zhang, J., Tang, Q., Zhou, C., Jia, W., Da Silva, L., Nguyen, L.D., Reutter, W., Fan, H., 2010. GLIS, a bioactive proteoglycan fraction from *Ganoderma lucidum*, displays anti-tumour activity by increasing both humoral and cellular immune response. *Life Sci.* 87, 628-637. https://doi.org/10.1016/j.lfs.2010.09.026

Zhang, Y., Yu, G., Li, B., Mu, X., Peng, H., Wang, H., 2016. Hemicellulose isolation, characterization, and the production of xylo-oligosaccharides from the wastewater of a viscose fiber mill. *Carbohydr. Polym.* 141, 238-243. https://doi.org/10.1016/j.carbpol.2016.01.022

Zhang, X., Wang, H., Liu, C., Zhang, A., Ren, J., 2017. Synthesis of Thermoplastic Xylan-Lactide Copolymer with Amidine-Mediated Organocatalyst in Ionic Liquid. *Sci. Rep.* 7, 551. https://doi.org/10.1038/s41598-017-00464-6

Zhao, T., Mao, G., Zhang, M., Zou, Y., Feng, W., Gu, X., Zhu, Y., Mao, R., Yang, L., Wu, X., 2014. Enhanced antitumor and reduced toxicity effect of *Schisanreae* polysaccharide in 5-Fu treated Heps-bearing mice. *Int. J. Biol. Macromol.* 63. 114-118. https://doi.org/10.1016/j.ijbiomac.2013.10.037

Zhao, X., Tong, T., Li, H.L., Lu, H., Ren, J., Zhang, A., Deng, X., Chen, X., Wu, A.M., 2017. Characterization of hemicelluloses from *Neolamarckia cadamba* (Rubiaceae) during xylogenesis. *Carbohydr. Polym.* 156, 333-339. https://doi.org/10.1016/j.carbpol.2016.09.041

Zilliox, C., Debeire, P., 1998. Hydrolysis of wheat straw by a thermostable endoxylanase: Adsorption and kinetic studies. *Enzyme Microb. Technol.* 22, 58-63. https://doi.org/10.1016/S0141-0229(97)00105-1

Published/accepted (*From Thesis*):

1. **Kedar Sharma**, Carlos M.G.A. Fontes, Shabir Najmudin and Arun Goyal **(2019).** Molecular organization and protein stability of the *Clostridium thermocellum* glucuronoxylan endo-β-1,4-xylanase of family 30 glycoside hydrolase in solution. *Journal of Structural Biology* 206**, 335-344.** (**JIF 3.75**)
2. **Kedar Sharma**, Inês Lobo Antunes, Vikky Rajulapati and Arun Goyal **(2018)**. Low resolution SAXS and comparative modeling-based structure analysis of endo-β-1,4-xylanase a family 10 glycoside hydrolase from *Pseudopedobacter saltans* comb. nov. *International Journal of Biological Macromolecules*. **112, 1104-1114 (JIF 4.78).**
3. Arun Goyal, Shadab Ahmed, **Kedar Sharma**, Vikas Gupta, Pedro Bule, Carlos M.G.A. Fontes and Shabir Najmudin **(2016).** Molecular determinants of substrate specificity revealed by the structure *Clostridium thermocellum* family 43_16 Arabinofuranosidase. ***Acta Crystallographica*** **D72, 1281-1289. (JIF 2.5)**

Submitted/to be submitted

4. **Kedar Sharma**, Carlos M.G.A. Fontes, Shabir Najmudin and Arun Goyal* **(2019)**. SAXS based structure, modelling and molecular dynamics analyses of family 43 glycoside hydrolase α-L-arabinofuranosidase (*Ct*Araf43) from *Clostridium thermocellum* **(Submitted after revision)**
5. **Kedar Sharma**, Sudhir Morla, Kaustubh Chandrakant Khaire, Abhijeet Thakur, Vijay Suryakant Moholkar, Sachin Kumar and Arun Goyal **(2019)**. Extraction, characterization of xylan from neem sawdust and its application in xylanase mediated production of anticancer xylooligosaccharides. **(Submitted after revision)**

Other Publications:

1. **Kedar Sharma**, Abhijeet Thakur, Rajeev Kumar and Arun Goyal (**2019**). Structure and biochemical characterization of glucose tolerant β-1,4 glucosidase (*Ht*Bgl) of family 1 glycoside hydrolase from *Hungateiclostridium thermocellum*. ***Carbohydrate Research* 483, 107750 (JIF 1.87).**
2. **Kedar Sharma**, Vikky Rajulapati and Arun Goyal (2019). Green synthesis of arabinoxyloglucan coated antimicrobial copper nanoparticles ***Trends in Carbohydrate Research*, 11(1), 22-30 (JIF 0.5)**
3. Karthika B[†]., **Kedar Sharma**[†] and Arun Goyal* (2019) Structure and dynamics analysis of a new member heparinase II/III of family 12 polysaccharide lyase from *Pseudopedobacter saltans* by computational modelling and small angle x-ray scattering. ***Journal of Biomolecular Structure and Dynamics***, https://doi.org/10.1080/07391102.2019.1622453 **(JIF 3.31) [†]Equal Contribution**
4. Dishant Goyal, Krishan Kumar, **Kedar Sharma** and Arun Goyal (2019). SAXS based structure, modeling and molecular dynamics analyses of a family 5 glycoside hydrolase first endo-mannanase (*Rf*GH5_7) from *Ruminococcus flavefaciens* FD-1 v3 ***Journal of Biomolecular Structure and Dynamics***, **https://doi.org/10.1080/07391102.2019.1680438 (JIF 3.31)**
5. Priyanka Nath, Arun Dhillon, Krishan Kumar, **Kedar Sharma**, Sumitha Banu Jamaldheen, Vijay Suryakant Moholkar and Arun Goyal* (2019). Development of bi-functional chimeric enzyme (CtGH1-L1-CtGH5-F194A) from endoglucanase (*Ct*GH5) mutant F194A and β-1,4-glucosidase (*Ct*GH1) from *Clostridium thermocellum* with enhanced activity and structural integrity. ***Bioresource Technology*. 282, 494-501. (JIF 6.66)**
6. **Kedar Sharma**, Inês Lobo Antunes, Vikky Rajulapati and *Arun Goyal (2018) Molecular characterization of a first endo-acting β-1,4-xylanase of family 10 glycoside hydrolase (*Ps*GH10A) from *Pseudopedobacter saltans* comb. nov. ***Process Biochemistry*. 70, 79-89. (JIF 2.88)**
7. **Kedar Sharma,** Arun Dhillon and Arun Goyal (2018). Insight into Structure and reaction mechanism of β-mannanases. *Current Protein and Peptide Sciences* **19, 34-47. (JIF 1.88)**.
8. Arun Dhillon, **Kedar Sharma**, Vikky Rajulapati and *Arun Goyal (2018) The multi-ligand specific family 35 Carbohydrate-binding Module (CBM35) from *Clostridium thermocellum* targets rhamnogalacturonan I and also mediates binding through a second site. *Archives of Biochemistry and Biophysics*. **654, 194-208. (JIF 3.55)**
9. Sumitha Banu Jamaldheen, **Kedar Sharma**, Vijay S. Moholkar and *Arun Goyal (2018) Comparative analysis of pretreatment methods on Sorghum (Sorghum durra) stalk agrowaste for holocellulose content. *Preparative Biochemistry and Biotechnology*. **48 (6), 457-464. (JIF 1.11).**

10. Vikky Rajulapati, **Kedar Sharma**, Arun Dhillon and Arun Goyal (2018) SAXS and homology modelling based structure characterization of pectin methylesterase a family 8 carbohydrate esterase from *Clostridium thermocellum* ATCC 27405. ***Archives of Biochemistry and Biophysics*. 641, 39-49 (JIF 3.55)**.

11. Karthika B., **Kedar Sharma**, Aruna Rani, R. Vikky and Arun Goyal (2018) Deciphering the mode of action, structural and biochemical analysis of recombinant heparinase II/III (*Ps*PL12a) a new member of family 12 polysaccharide lyase from *Pseudopedobacter saltans*. ***Annals of Microbiology*. 68(6), 409-418 (JIF 1.43).**

12. Aruna Rani, Arun Dhillon, **Kedar Sharma** and Arun Goyal* (2018) Insights into the structure and substrate binding analysis of chondroitin AC lyase (*Ps*PL8A) from *Pedobacter saltans*. ***International Journal of Biological Macromolecules*. 109, 980-991 (JIF 4.78)**.

13. S.M. Khade, S.K. Srivastava, Krishan Kumar, **Kedar Sharma**, Arun Goyal and A.D. Tripathi (2018) Optimization of clinical uricase production by *Bacillus cereus* under submerged fermentation, its purification and structure characterization. ***Process Biochemistry*, 75, 49-58. (JIF 2.88)**

14. Krishan Kumar, Márcia Correia, Virginia R. Pires, Arun Dhillon, **Kedar Sharma**, Vikky Rajulapati, Carlos M.G.A. Fontes, Ana Luísa Carvalho and Arun Goyal (2018) Novel insights into the degradation of β-1,3-glucans by the cellulosome of *Clostridium thermocellum* revealed by structure and function studies of a family 81 glycoside hydrolase. ***International Journal of Biological Macromolecules*. 117, 890-901. (JIF 4.78).**

15. Soumyadeep Chakraborty, **Kedar Sharma**, Joyeeta Mukherjee, Munishwar N. Gupta and Arun Goyal (2015) Structure, substrate binding analysis and stability studies of endo-pectate lyase (PL1B) of family 1 polysaccharide lyase from *Clostridium thermocellum*. ***Protein and Peptide Letters*. 22(6), 557-568. (JIF 1.1)**

Other Publications (Submitted)

16. Barnali Nath[†], **Kedar Sharma**[†], Komal Ahire, Arun Goyal and Sachin Kumar (2019). Structure analysis of the nucleoprotein of Newcastle disease virus: An insight towards it multimeric form. **(Submitted)** **[†]Equal Contribution**

17. Priyanka Nath, **Kedar Sharma** and Arun Goyal (2019). Combined SAXS and computational approaches for structure determination and binding characteristics of chimera (*Ct*GH-L1-*Ct*GH5-F194A) generated by assembling β-glucosidase (*Ct*GH1) and a mutant endoglucanase (*Ct*GH5-F194A) from *Clostridium thermocellum* **(Submitted after revision)**

18. Abhijeet Thakur, **Kedar Sharma**, Sumitha Banu Jamaldheen and Arun Goyal* (2019). Molecular characterization, regioselective and synergistic action of first recombinant type I α-L-arabinofuranosidase of family 43 glycoside hydrolase (*Ps*GH43A) from *Pseudopedobacter saltans* **(Submitted)**

19. Manisha Shah, **Kedar Sharma**, Arun Goyal, Sachin Kumar (2019)**.** Role of cholesterol in anatid herpesvirus type 1 infection: Potential target for an antiviral (**Submitted**)

Book Chapters

1. Abhijeet Thakur, **Kedar Sharma**, Ruchi Mutreja, Vikky Rajulapati and Arun Goyal (2019) Thermostable enzymes from *Clostridium thermocellum*. Ed. Sonali Mohapatra, Microbial Fermentation and Enzyme Technology, CRC Press. (***submitted after revision***)
2. Abhijeet Thakur, **Kedar Sharma**, Kaustubh Khaire and Arun Goyal (2018) Enzymes: Key role in conversion of waste to biofuel. Ed Sonali Mohapatra, Microbial Fermentation and Enzyme Technology, CRC Press. (***submitted after revision***)
3. **Kedar Sharma**, Abhijeet Thakur and Arun Goyal (2018), **Xylanases for food applications "Green Bio-Processes: Industrial Enzymes for Food Applications"** https://doi.org/10.1007/978-981-13-3263-0_7.
4. Abhijeet Thakur, **Kedar Sharma** and Arun Goyal (2018), **α-L-arabinofuranosidase: A potential enzyme for the food industry "Green Bio-Processes: Industrial Enzymes for Food Applications"** https://doi.org/10.1007/978-981-13-3263-0_12.
5. Arun Dhillon, **Kedar Sharma**, Vikky Rajulapati and Arun Goyal (2015) Chapter 7, Proteolytic enzymes in "Current Developments in Biotechnology & Bioengineering", Volume 7: Production, Isolation and Purification of Industrial Products, Eds. Ashok Pandey, Sangeeta Negi, Poonam Nigam, Carlos Ricardo Soccol. http://dx.doi.org/10.1016/B978-0-444-63662-1.00007-5.

1. **Kedar Sharma**, Carlos M.G.A. Fontes, Shabir Najmudin and Arun Goyal (2019). SAXS based structure, modelling and molecular dynamics analyses of family 43 glycoside hydrolase α-L-arabinofuranosidase (*Ct*Araf43) from *Clostridium thermocellum*. 16th International Conference of the Asian Crystallographic Association, 17-20 December 2019, Singapore.
2. **Kedar Sharma**, Carlos M.G.A. Fontes, Shabir Najmudin and Arun Goyal (2019). Molecular organization and protein stability of the *Clostridium thermocellum* glucuronoxylan endo-β-1,4-xylanase of family 30 glycoside hydrolase in solution. 16th International Conference of the Asian Crystallographic Association, 17-20 December 2019, Singapore.
3. **Kedar Sharma**, Kaustubh Chandrakant Khaire, Abhijeet Thakur, Vijayanand Suryakant Moholkar and Arun Goyal (2019). Isolation and characterization of glucuronoxylan from Babool as substitute of commercial xylan for xylanase activity evaluation. International Carbohydrate Conference on Emerging Frontiers in Carbohydrate Chemistry and Glycobiology, Dec. 5-7, 2019, University of Lucknow, UP, India.
4. **Kedar Sharma**, Sudhir Morla, Kaustubh Chandrakant Khaire, Abhijeet Thakur, Vijay Suryakant Moholkar, Sachin Kumar and Arun Goyal (2019) Extraction, characterization of xylan from neem sawdust and its application in xylanase mediated production of anticancer xylooligosaccharides. International Carbohydrate Conference on Emerging Frontiers in Carbohydrate Chemistry and Glycobiology, Dec. 5-7, 2019, University of Lucknow, UP, India.
5. **Kedar Sharma**, Kaustubh Chandrakant Khaire, Abhijeet Thakur, Vijayanand Suryakant Moholkar and Arun Goyal (2019) Acacia xylan as a potential commercial xylan and its application in production of xylooligosaccharides. International Conference on New Horizons in Biotechnology, November 20-24, 2019, Trivandrum, Kerala, India.
6. **Kedar Sharma**, Sudhir Morla, Kaustubh Chandrakant Khaire, Abhijeet Thakur, Vijay Suryakant Moholkar, Sachin Kumar and Arun Goyal (2019) Xylanase mediated production of xylooligosaccharides from neem sawdust xylan and its anticancer potential. International Conference on Nutraceuticals and Chronic Diseases (INCD 2019), September 23-25, 2019 Indian Institute of Technology Guwahati, Assam, India.
7. **Kedar Sharma**, Carlos M.G.A. Fontes, Shabir Najmudin and Arun Goyal (2019). Molecular organization and protein stability of the *Clostridium thermocellum* glucuronoxylan endo-β-1,4-xylanase of family 30 glycoside hydrolase in solution. Research Conclave 2019, *March 14-17, 2019***,** IIT Guwahati, India **(Best Poster Award).**
8. **Kedar Sharma**, Abhijeet Thakur, Rajeev Kumar and Arun Goyal (2018). Structure and biochemical characterization of glucose tolerant β-1,4 glucosidase (*Ht*Bgl) of family 1 glycoside hydrolase from *Hungateiclostridium thermocellum*. Bioprocessing India 2018, Dec 16-18, 2018, IIT Delhi, India.

9. **Kedar Sharma** and Arun Goyal (2018) Structure characterization of endo-β-1,4 xylanase from *Pseudopedobacter saltans* by SAXS and Molecular Dynamics simulation. 59th Annual Conference of AMI, Dec 9-12, 2018, University of Hyderabad, India.

10. **Kedar Sharma**, Abhijeet Thakur, Kaustubh Khaire and Arun Goyal (2018) Molecular characterization of halo and organic solvent stable xylanase from *Pseudopedobacter saltans* and its application in xylooligosaccharides production from Kans grass biomass. International Conference on Biotechnological Research and Innovation for Sustainable Development, 15th BRSI convention. CSIR- Indian Institute of Chemical Technology (CSIR-IICT), Nov. 22-25, 2018, Hyderabad, India.

11. **Kedar Sharma** and Arun Goyal (2018) Green synthesis of copper nanoparticles using arabinoxyloglucan isolated from water hyacinth as stabilizing agent for antimicrobial applications. International Conference on Drug Discovery: Biotechnology and Pharma at Cross Roads, 15-17th Feb 2018 Thapar University, Patiala, INDIA

12. **Kedar Sharma**, Vikky Rajulapati, Inês Lobo Antunes and Arun Goyal (2017) SAXS analysis and structure modelling of endo β-1,4 xylanase (*Ps*GH10A) from *Pedobacter saltans*. 86th Annual Meeting of Society for Biological Chemists, India, Nov. 16-19, Jawaharlal Nehru University, New Delhi, India.

13. Arun Dhillon, **Kedar Sharma**, Vikky Rajulapati and Arun Goyal (2017) Rgl-CBM35 of family 35 Carbohydrate Binding Module (CBM) from *Clostridium thermocellum* represents first CBM targeting rhamnogalacturonan I and mediating binding by two sites. 23rd INPEC (International Network of Protein Engineering Centers) Meeting Protein Structure, function and Engineering, 9-11 Nov 2017, Bose Institute, Kolkata. (**Best Poster Award**)

14. **Kedar Sharma** and Arun Goyal (2017) Biochemical characterization and deciphering the mode of action of recombinant endo β-1,4 xylanase (*Ps*GH10) from *Pedobacter saltans* DSM12145. 14th BRSI Convention and International Conference (BRSI-2017), Oct 08-10, 2017, CSIR-NEERI, Nagpur.

15. **Kedar Sharma**, Shadab Ahmed, Carlos M.G.A. Fontes, Shabir Najmudin and Arun Goyal (2017) Low-resolution structure analysis of α-L-arabinofuranosidase (*Ct*GH43) by SAXS. 24th Congress & General Assembly of the International Union of Crystallography 2017 (IUCr 2017) August 21-28, Hyderabad, India.

16. Aruna Rani, **Kedar Sharma** and Arun Goyal (2017) Insights into the structural characteristics of chondroitin AC lyase *Ps*PL8A from *Pedobacter saltans*. 12th Carbohydrate Bioengineering Meeting, April 23-26, 2017, Vienna, Austria.

17. **Kedar Sharma**, Anil Kumar Verma, Carlos M.G.A. Fontes, Shabir Najmudin and Arun Goyal (2017) Low-resolution structure of glucuronoxylan-xylanohydrolase (*Ct*XynGH30) of family 30 glycoside hydrolase from *Clostridium thermocellum* by SAXS. Annual Symposium of the Indian Biophysical Society, March 22-25, 2017, IISER Mohali, India.

18. **Kedar Sharma**, Anil Kumar Verma, Carlos M.G.A. Fontes, Shabir Najmudin and Arun Goyal (2016) Solution structure analysis of full length glucuronoxylan endo-β-1,4-xylanase from *Clostridium thermocellum* by Small Angle X-Ray Scattering. 14th International Conference of the Asian Crystallographic Association, 4-7 December 2016, Hanoi, Vietnam.

19. **Kedar Sharma** and Arun Goyal (2016) Cloning, expression and characterization of a xylanase from family 10 Glycoside hydrolase (GH10) from *Pedobacter saltans* DSM12145. 57th International Annual Conference of the Association of Microbiologists of India (AMI-2016), Nov 24-27, 2016, Gauhati University and IASST, Guwahati, Assam India.

20. Inês Lobo Antunes, **Kedar Sharma**, Vikky Rajulapati and Arun Goyal (2016) Biochemical and structure characterization of xylanase of family 10 Glycoside Hydrolase (*Ps*GH10B) from *Pedobacter saltans* DSM12145. CARBO-XXXI International Conference on "New Frontiers in Carbohydrate Chemistry & Biology", Nov 14-16, 2016, Department of chemistry, university of Delhi India.

21. **Kedar Sharma**, Bibari Boro and Arun Goyal (2015) *In silico* structure analysis of a family 12 polysaccharide lyase from *Pedobacter saltans* DSM12145. 56th International Annual Conference of Association of Microbiologists of India (AMI), December 7-10, 2015, Jawaher Lal Nehru University, New Delhi.

22. Arun Dhillon, **Kedar Sharma**, Vania O. Fernandes, Fernando M.V. Dias, José A.M. Prates, Luis M.A. Ferreira, Carlos M.G.A. Fontes, M.S.J. Centeno and Arun Goyal (2015) Biochemical characterization and deciphering the cleavage pattern of recombinant rhamnogalacturonan lyase (*Ct*RGL), a family 11 Polysaccharide Lyase (PL11) from *Clostridium thermocellum*. 56th International Annual Conference of Association of Microbiologists of India (AMI), December 7-10, 2015, Jawaher Lal Nehru University, New Delhi.

23. **Kedar Sharma** and Arun Goyal (2015) *In silico* structural characterization of a family 10 glycoside hydrolase from *Pedobacter saltans* DSM12145. 56th International Annual Conference of Association of Microbiologists of India (AMI), December 7-10, 2015, Jawaher Lal Nehru University, New Delhi.

24. **Kedar Sharma** and Arun Goyal (2015) Molecular cloning, expression and characterization of novel endo-β-1,4-mannanase of family 10 glycoside hydrolase from *Pedobacter saltans* DSM 12145. 11th Carbohydrate Bioengineering Meeting, May 10-13, 2015, Espoo, Finland.

25. **Kedar Sharma** and Arun Goyal (2014) *In silico* structure prediction of a family 10 glycoside hydrolase from *Pedobacter saltans* DSM12145. Indo-US Conference and Workshop on recent Advances in Structural Biology & Drug Discovery, October 9-11, 2014, Indian Institute of Technology Roorkee, Uttarakhand, India.

Workshops/courses attended

1. Participated in workshop on SAXS data analysis organized by “Anton Paar India Limited, on March 20-22, 2018.
2. Participated in workshop on Intellectual Property Rights and Patenting, organized by “Department of Biotechnology, Thapar Institute of Engineering and Technology Patiala and Punjab state council for Science and Technology (PSCST) Chandigarh, on February 17, 2018.
3. Participated in satellite meeting on “Phasing and Model Building” organized by IUCr Congress and Center for Chemical Biology, CSIR-Indian institute of Chemical Technology Hyderabad, on August 20-21, 2017.
4. Participated in GIAN Course “**Latest Methods in X-ray Crystallography: Lecture Series and Practical Course** organized by School of Life Sciences, Jawaharlal Nehru University, New Delhi-110067, INDIA under MHRD scheme on Global Initiative for Academic Network on **November 14th to 25th, 2016.**
5. Participated in “4th edition of **Computational Biotechnology at the Nanoscale: CCP4 Workshop 2016**” Jointly organized by CCP4, AIIMS, ICGEB, JNU, NII and UNESCO-Regional Centre for Biotechnology, at Regional Biotech Science Cluster NCR Biotech Science Cluster 3rd Milestone, Faridabad-Gurgaon Faridabad 121001 on 15/2/2016 - 20/2/2016.
6. Participated in “**Symposium cum Workshop on "Advances in Computational Biology and Computer Aided Drug Design**” organized by Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati on **June 24th to 26th, 2015.**

VITAE

The author was born on September 14, 1990 in the city of Tijara, (Rajasthan). He passed the Secondary Examination (10th Class) conducted by Board of Secondary Education Rajasthan, Ajmer in 2005 and Higher Secondary Examination (12th Class) conducted by Board of Secondary Education Rajasthan, Ajmer in 2007. He completed B.Sc. (Biotechnology) from Govt. Rajrishi College Alwar, affiliated to University of Rajasthan, Jaipur (Rajasthan) in July, 2010. He completed M.Sc. (Biotechnology) from Guru Ghasidas Vishwavidyalaya Bilaspur, Chhattisgarh in July, 2012.

Mr. Kedar Sharma joined the M.Tech program in July, 2013 at Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781 039, Assam, India. After the successful completion of coursework with 8.03/10 CPI, He converted his M.Tech. into Ph.D. program in August 2014. He received Institute Fellowship (IIT Guwahati), for M.Tech (Ist year) and whole Ph.D. duration under the scheme run by the Ministry of Human Resource and Development (MHRD), New Delhi. He delivered the open (PhD Synopsis) Seminar on September 4, 2019 and presented his thesis work before the Doctoral Committee and his performance was satisfactory. He submitted the PhD thesis in September 2019.