Structural, biochemical and functional analyses of modular recombinant glucuronoxylan-xylanohydrolase (*Ct*XynGH30) of family 30 Glycoside Hydrolase and its truncated derivative *Ct*Xyn30A and associate family 6 carbohydrate binding module *Ct*CBM6 from *Clostridium thermocellum*

PhD Thesis

by

Anil Kumar Verma

March 2015

Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Guwahati, Assam, India

Dedicated to
My Parents,
Brother
And Teachers

Structural, biochemical and functional analyses of modular recombinant glucuronoxylan-xylanohydrolase (*Ct*XynGH30) of family 30 Glycoside Hydrolase and its truncated derivative *Ct*Xyn30A and associate family 6 carbohydrate binding module *Ct*CBM6 from *Clostridium thermocellum*

A Thesis

Submitted for the Award of the Degree

of

DOCTOR OF PHILOSOPHY

by

Anil Kumar Verma

Under the supervision of

Prof. Arun Goyal

Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Guwahati 781 039, Assam, India

March 2015

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

STATEMENT

I do hereby declare that the content embodied in this thesis entitle "**Structural, biochemical and functional analyses of modular recombinant glucuronoxylan-xylanohydrolase (*Ct*XynGH30) of family 30 Glycoside Hydrolase and its truncated derivative *Ct*Xyn30A and associated family 6 carbohydrate binding module *Ct*CBM6 from *Clostridium thermocellum*".** 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.

March, 2015

Anil Kumar Verma
(9610621)

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

CERTIFICATE

It is certified that the work described in this thesis entitled **"Structural, biochemical and functional analyses of modular recombinant glucuronoxylan-xylanohydrolase (*Ct*XynGH30) of family 30 Glycoside Hydrolase and its truncated derivative *Ct*Xyn30A and associated family 6 carbohydrate binding module *Ct*CBM6 from *Clostridium thermocellum*."** by Mr. **Anil Kumar Verma (09610621)** 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 and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India and this work has not been submitted elsewhere for a degree.

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

ACKNOWLEDGEMENTS

The journey towards producing this thesis was full of new experiences for me. I take this opportunity to express my sincere gratitude to all those who were there with me and contributed in many different ways to the success of this study and made this journey a wonderful and unforgettable experience for me. I take the pleasant opportunity to express my sincere gratitude and heartfelt obligation towards all the personages who have helped me in this endeavor.

At this moment of accomplishment, I am extremely indebted 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 am indebted to him for introducing me into the realm of this interesting work, giving me the freedom in designing and conducting the experiments.

Under his guidance I successfully overcame many difficulties and learned a lot. I can't forget how patiently he listened to my problems and provided the necessary instructions. Without his meticulous planning, foresight, valuable suggestions, encouragement and constructive criticism this work couldn't have assumed the form it is today. His true scientific spirit has inspired me a lot during the research work. There were many frustrating moments when the "graphs" and "figures" didn't seem to lead anywhere and without his vision I would have never come out of the maze. He always used to review my thesis progress, give me valuable suggestions and made corrections numerous times. I shall forever remain grateful to him for all that he has done.

I would also like to express my sincere gratitude to all my doctoral committee members Dr. Utpal Bora, Dr. Ranjan Tamuli, Dr. A. B. Kunnumakkara and Prof. T. Punniyamurthy for their continuous support, encouragement and valuable suggestions during my seminars and progress reviews that has led to the successful completion of my thesis.

I owe my profound gratitude to the present and previous heads of the Department of Biosciences and Bioengineering, IIT Guwahati, Prof. V. V. Dasu and Prof. Arun Goyal for providing me with the necessary facilities.

I am thankful to Department of Biosciences & Bioengineering and Central Instrumentation Facility (CIF), IITG for providing instruments facility for research work. I would also like to acknowledge MHRD for fellowship.

I would like to thank all the staff of Department of Biosciences & Bioengineering for their timely help and assistance whenever sought for.

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 crystallization work. I feel myself lucky to being 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 suggestion during crystallization and data analysis. I would also like to thanks Filipe Freire and Pedro Bule for their contribution, valuable suggestions and untiring help especially during the crystallization of CtXyn30A mutant.

Everybody wants a home away from home and all these years, the lab was like a home to me and it wouldn't be the same without such wonderful seniors and lab mates so here I seek this opportunity to thank Dr. Shadab Ahmed, Dr. Rishikesh Shukla, Dr. Shraddha Shukla, Dr. T.J.M. Rao, Dr. Saprativ P. Das, Dr. Deeplina Das and Dr. Arabinda Ghosh, Suchi Singh, Damini Kothari and Soumyadeep Chakraborty for their help and suggestions. In the same breath I would like to thank my research group members Aruna, Arun, Rwivoo, Vikky, Kedar and Ashutosh for their help and keeping pleasant working environment in the lab.

I would like to thank all my friends and batchmates Himangshu, Anil Kumar, Prakash, Kimjolly, Saravanan, Rohitas, Manjeet, Supriyo, Mitun, Gauri, Sushant, Aadi Moolam, Digar and Soma Sekhar for a wonderful company. *My greatest appreciation and friendship goes to my closest friend Anil Kumar and Himangshu who were always a great support in all my struggles and frustrations during my stay in IIT Guwahati.*

I would also like to thank my seniors in the department Dr. Chokalingam, Dr. Tushar, Dr. Abhay Narayan Singh, Dr. Pojul Loying, Dr. Krishnamoorthy Hegde, Dr. Naresh Kasoju and Dr. Vijay Kumar Mishra for their help and suggestions.

I would also like to thank my friends Narendra Kumar, Shadab Khan and Deepak Kumar for their constant appreciation and support.

Thank you Mousami, Mohan, Nayan Moni, Suman, Balaji, Ruchira, Radhika, Ankita, Deepika, Arghya, Swagata, Anuj, Garima, Muthu and Santosh for all the cheer and light moments during my stay in IIT Guwahati.

Lastly I am thankful to my family members whose continuous encouragement, moral support and blessings helped me in successful completion of my research work. I owe my achievements to my parents, elder brother and sister-in-law for their motivation.

I pray almighty GOD to give me the strength to keep going.

Anil Kumar Verma
IIT Guwahati

SYNOPSIS

Introduction

In recent times the advent of novel xylanase from different microbial sources at the expense of genetic manipulation has led to great advances and development of carbohydrate active enzymes. 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 us with sufficient amounts of animal feed, food, fuel, and chemicals. 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 this regard the present study deals with the cloning, structure and function characterization of modular glucuronoxylan-xylanohydrolase from *Clostridium thermocellum.* In the present study 3-dimentional structural features of catalytic module *Ct*Xyn30A of family GH30 and associated module *Ct*CBM6 of family 6 from C*lostridium thermocellum* was characterized by employing both computational approach and X-ray crystallography technique to understand the architect of substrate binding cleft in *Ct*CBM6 while the overall structure feature of *Ct*Xyn30A. Docking study with xylo-oligosaccharide ligands and ligand bound crystal structures will provide the necessary information regarding the key amino acids involved in substrate targeting in *Ct*CBM6, whereas, in case of the catalytic module it infers regarding the

key residues responsible for substrate catalysis respectively. This study helps to generate desired mutants of family 6 carbohydrate binding modules that participate in substrate binding. The affinity gel electrophoresis and ITC results of different mutants of *Ct*CBM6 as compared to wild type have been investigated in order to establish structure function relationship. Moreover, in catalytic *Ct*Xyn30A by employing site directed mutagenesis approach, we can create the desired mutant with particular amino acid residues which predicted *in silico* study helps us to establish and understand the role of particular residues in the substrate recognition and substrate catalysis of enzyme glucuronoxylan-xylanohydrolase family GH30. Therefore, structural and functional investigation of *Ct*Xyn30A is important in order to understand the differences as well as similarities among the proteins characterized so far from the same sub-family as well as across the families. Moreover GH family 30 (sub-family 8), (glucuronoxylan-xylanohydrolase) is known to effectively depolymerize acidic xylans (major component of secondary cell wall of hardwood plant) into simple sugars that can be utilized for bioethanol production.

Present work

The present investigations are carried out on the **"Structural, biochemical and functional analyses of modular recombinant glucuronoxylan-xylanohydrolase (*Ct*XynGH30) of family 30 Glycoside Hydrolase and its truncated derivative *Ct*Xyn30A and associated family 6 carbohydrate binding module *Ct*CBM6 from *Clostridium thermocellum*"**. The thesis work comprises 5 Chapters.

Chapter 1 is the General Introduction starts with the structures of different structural polysaccharides. An effort was made to understand the basic structural monomeric units of the natural polysaccharides such as cellulose, hemi-cellulose and lignin. This chapter also includes different types of Carbohydrate-active enzymes with special emphasis on glycoside hydrolase, its modular nature and reaction mechanism. The current status of the members belonging to family GH30 as well as subfamily 8 has been also reviewed in detail with respect to the enzyme glucuronoxylan-xylanohyrolase (*Ct*XynGH30) from *Clostridium thermocellum*. The chapter describes the various application of the glucuronoxylan-xylanohyrolase. In this chapter the state of art of carbohydrates binding modules (CBMs) belongs to the family 6 have been described. The different approaches in the classification of CBMs, their function and various applications have also been discussed. This chapter also mentions why enzyme from *clostridium thermocellum* was selected for present study. In brief the significance of the cellulosomes and the crucial roles of cellulosomal enzymes in the degradation of the recalcitrant and chemically complex polysaccharides efficiently into the biofuel production have been also discussed.

Chapter 2 The modular protein *Ct*XynGH30 having gene accession number ABN54208 (http://www.ncbi.nlm.nih.gov/protein) and uniprot ID A3DJS9 is a carbohydrate active enzyme. *Ct*XynGH30 consisted of an N-terminal, catalytic module from family 30 glycoside hydrolase (GH30) named *Ct*Xyn30A and a C-terminal, family 6 carbohydrate binding module (*Ct*CBM6) connected via a flexible linker of 6 amino acid residues (Fig.1). The presence of dockerin type I module at C-terminal implies that *Ct*XynGH30 is a part of cellulosome of *C. thermocellum*.

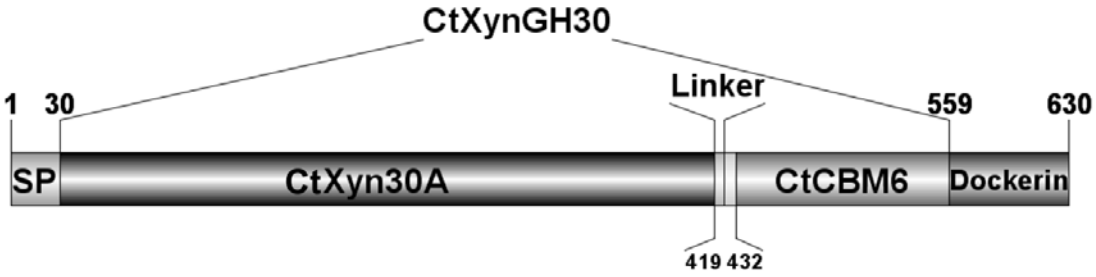


Fig. 1 Molecular architecture of *Ct*XynGH30 showing boundaries and designation of different domains.

The genes encoding *Ct*XynGH30 and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6 were amplified from the genomic DNA of *Clostridium thermocellum* ATCC 27405 (GenBank Accession No: ABN54208). The amplified PCR products carrying the desired restriction sites successfully cloned initially into the pGEM-T Easy vector and subsequently, cloned into expression vector pET-28a(+). *Ct*XynGH30 and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6 proteins showed the band of over-expression in response to IPTG when recombinant clone plasmid transformed into the BL21 cells. The catalytic enzymes purified by IMAC displayed molecular sizes of 60 kDa (*Ct*XynGH30) and 45 kDa (*Ct*Xyn30A) whereas non-catalytic carbohydrate binding modules (*Ct*CBM6) showed molecular sizes of 14 kDa, on SDS-PAGE. The amount of protein obtained from 100 ml of cultures of *Ct*XynGH30 and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6 after purification were 4.2 mg, 4.8 mg and 3.9 mg. These purified proteins were used for further studied for biochemical, functional and structural characterization.

Chapter 3 In this chapter the enzymes identified and characterized are the catalytic domains *Ct*XynGH30 and *Ct*Xyn30A of *C. thermocellum.* The substrate specificity analysis of *Ct*XynGH30 and *Ct*Xyn30A clearly suggested that it acts on xylan based

polysaccharides with xylose as a monomeric unit and displayed the maximum activity of 34 and 31 U/mg on beechwood xylan respectively. Zymogram analysis of *Ct*XynGH30 and *Ct*Xyn30A with beechwood xylan also confirmed its xylanase activity. Biochemical experiments demonstrated that *Ct*XynGH30 and *Ct*Xyn30A to be active over broad range of pH from neutral to alkaline with an optimum pH at 6.0. The thermo-stability experiments showed that *Ct*XynGH30 and *Ct*Xyn30A is a thermophilic xylanase displaying stability at a higher range of temperature (from 60 to 80ºC) with an optimum temperature of 70 ºC. The effect of metal ions and chemical agents on the activity of *Ct*XynGH30 and *Ct*Xyn30A was studied at low (1mM) as well as high concentration (10mM) which displayed the presence of Li^{2+} ions only showed bit elevated activity in both at low and high concentrations. The other mental ions studied do not show positive effect on the enzyme activity. The Cu^{2+}, Ni^{2+}, Fe^{2+} ions and SDS slightly inhibited the enzyme activity at 1 mM and significantly at 10 mM. The activity of *Ct*XynGH30 and *Ct*Xyn30A was not significantly affected by Mg^{2+}, Ca^{2+} and β-mercaptoethanol. The effect of chemical agent and detergents was also studied and found that the activity of both the module was decreased by NaCl, SDS and urea at low as well as high concentration. Kinetic studies of *Ct*XynGH30 and *Ct*Xyn30A with beechwood xylan gave a K_m of 2.01 mg/ml and 2.184 mg/ml and turnover number (K_{cat}) of 1.107×10^5 min^{-1} and 1.205×10^5 min^{-1} respectively. The TLC analysis of hydrolysed products of beechwood xylan by treatment with *Ct*XynGH30 produced a series of higher xylo-oligosaccharides showing that it is an endo-xylanase. Whereas, on prolonged treatment of xylohaxaose by *Ct*Xyn30A did not lead to formation of any lower xylo-oligosaccharides or monomer xylose. This suggested that *Ct*Xyn30A does not act on linear xylo-oligosaccharides and has an

absolute requirement of substitution of glucuronic acid or 4-O-methyl glucuronic acid on xylans for enzymatic cleavage.

Chapter 4 deals with the identification of the active site residues and understanding of their role in hydrolysis of substrate. *Ct*Xyn30A showed sequence and structural similarity with XynC from *B.subtilis*, and both were reassigned in the family GH30 glucuronoxylan-xylanohydrolase; subfamily H. The 3-D structure of *Ct*Xyn30A was modeled and validated. Molecular dynamic simulation ensured the structural integrity of energy minimized *Ct*Xyn30A. Structure analysis of *Ct*Xyn30A displayed that it has a $(\beta/\alpha)_8$ barrel fold structure along with an associated side β-structure resembling to CBM, attached to the main catalytic core. The side β-structure helps in substrate selection and crucial for the activity of catalytic core. Our findings inferred that carboxylic group of amino acid residue Glu136 act as a catalytic acid/base whereas Glu225 act as a nucleophile, moreover these residues are 5.2 Å apart from each other which confirmed the presence of retention type mechanism in *Ct*Xyn30A. These key active site residues were found to be conserved within GH30 family. The catalytic core of *Ct*Xyn30A consisted of β-barrel and connecting loops toward the C-terminal side, formed by aromatic amino acid residues Trp53, Trp173, Tyr139, Tyr230, Trp111, Trp293 and Tyr294 which found to be conserved. The three dimensional structure of *Ct*Xyn30A contained a central $(\beta/\alpha)_8$ TIM barrel to which a side β-sheet domain is tightly connected. Glu136 and Glu225 are the key residues for catalytic activity of *Ct*Xyn30A. Trp81, Tyr139, Trp143, Gln173, Tyr200, Tyr227, Ser231, Trp264, Tyr265, Arg268 and Tyr270 are important residues for ligand binding. A comparison of *Ct*Xyn30A with XynA and XynC structures with various bound ligands confirmed that equivalent residues are involved in substrate recognition

providing a common consensus in the catalytic site. Thus, GH30-8 contain specificity pockets near the active site, enabling GH30-8 members to bind to a number of ligands that may mimic the different decorations of xylose in the GX_n chains of their substrates. So this chapter described the complete structural features and the conserved active site residues along with conserved aromatic residues which overall architect the catalytic core of glucuronoxylan-xylanohydrolase *Ct*Xyn30A from *Clostridium thermocellum*.

Chapter 5 deals with the structural and functional aspect of *Ct*CBM6 which is cloned, expressed and purified as a soluble ~14 kDa protein. *Ct*CBM6 demonstrated affinity towards α-L-arabinosyl and glucuronic acid substituted xylans. The equilibrium association constant (K_a) determined by affinity electrophoresis and isothermal titration calorimetry suggests that *Ct*CBM6 displays similar affinities against highly substituted xylans or decorated xylans (wheat- and rye- arabinoxylans) and poorly substituted xylans or undecorated xylans (birch-, beech- and oat spelt-xylan). *Ct*CBM6 is unable to interact with other plant cell wall carbohydrates. The protein melting of *Ct*CBM6 at 83°C conveyed the thermostable nature of the protein. The presence of Ca^{2+} ions increased only marginally the structure stability of *Ct*CBM6 by shifting the melting peak by 2°C to 85°C.

The structure of *Ct*CBM6 was modelled and refined by molecular dimension. CD spectrum of *Ct*CBM6 displays 44% β-strands and rest random coils and supported the predicted structure. The *in silico* analysis revealed that *Ct*CBM6 display β-jelly roll β-sandwich fold with two potential binding sites; cleft A and B. Cleft A, located on the loops connecting the two β-sheets is decorated by the side chains of Phe-84, Tyr-28 and Asn-112 whereas cleft B is located in the concave surface of the protein where

residues Tyr-34 and Tyr-104 might contribute to ligand recognition. The functional importance of the two putative clefts for carbohydrate recognition was investigated by mutagenesis analysis. The data revealed that two cleft A protein mutant derivatives, the single mutant F84A and the double mutant Y28A/F84A, completely lost the ability to recognize xylans, whereas, the single mutant Y28A was able to recognize glucuronic acid and arabinose decorated xylan polysaccharides but with significantly reduced affinity. These results confirm that cleft A, through the action of Tyr28 and Phe84, constitutes the functional carbohydrate binding platform of *Ct*CBM6. The overall data confirms that within the CBM6 family a subset of proteins have evolved the capacity to recognize xylans through cleft A, while in this case cleft B remain non-functional.

CONTENTS

Chapter 1

General Introduction

1.0 Carbohydrates

Carbohydrates are the most abundant biomolecules present on the earth's surface and represent the largest renewable source of organic carbon (Gilbert, 2010). They are the main source of energy and constitute a major organic component in various fruits, vegetables, legumes and cereal grains. Carbohydrates chemically comprise carbon, hydrogen and oxygen and are defined as polyhydroxy aldehydes or ketones (Eisenstein, 2006a). The hydroxyl groups of carbohydrates are susceptible to substitution by different groups which lead to formation of different combinations. They can be modified into esters and ethers or be replaced by de-oxy and amino groups. The carbohydrates also form covalent bonds with protein and lipid molecules and form glyco-conjugates such as proteoglycans, glycoproteins and glycolipids (Eisenstein, 2006b).

1.1 Structural polysaccharides in plants

The most important function of plant structural polysaccharides is to provide mechanical strength and maintain the structural integrity. The structural polysaccharides found in different species of plants include cellulose, hemicellulose and lignin. The main function of these structural polysaccharides is to maintain the structural organization and provide mechanical support to the plants. They also carry out internal transport of water, nutrients and products of photosynthesis through the plant. The amount of these structural components; cellulose, hemicellulose and lignin vary significantly and depend upon the plant species and the cell type within the tissue. The figure 1.1 demonstrates the cell wall model to understand how different types of plant polysaccharides are interconnected and form a highly ordered network.

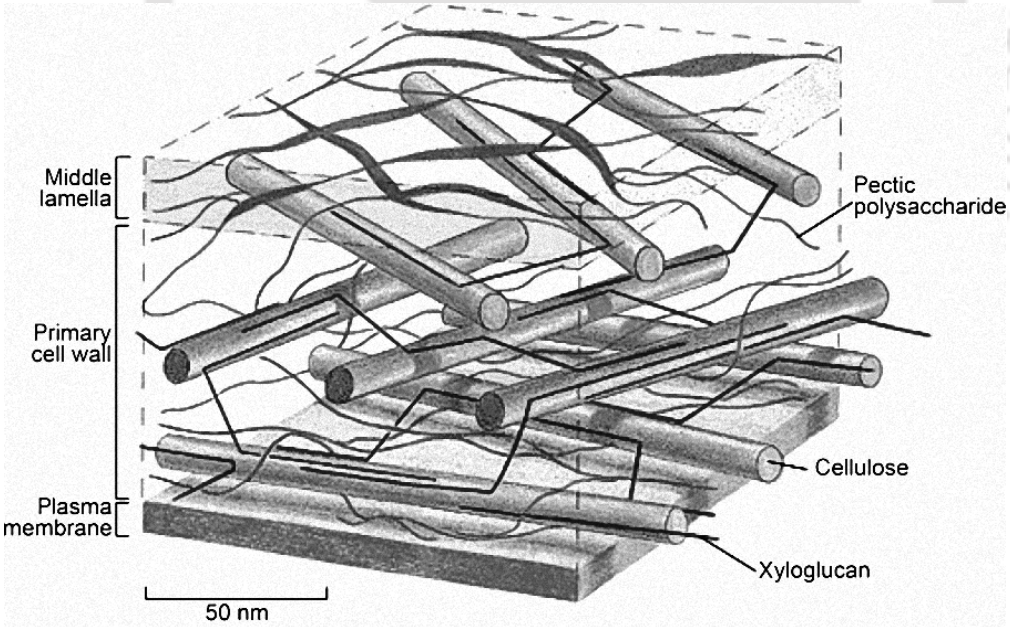


Fig. 1.1 Model of the primary cell wall in higher plants (adapted from McCann & Roberts, 1991).

The role of different plant polysaccharides is important to understand and therefore it is essential to know the basic structural units.

1.1.1 Cellulose

Cellulose is the most abundant organic polymer on Earth (Klemm *et al*., 2005). It is also the most important and integral part of primary cell walls of all gymnosperms and angiosperms. The seed hairs of cotton represent almost pure form and contain 91% of cellulose, in contrast, to wood cellulose. The cellulose is a linear (unbranched) and long homo-polymer of a repeated monomeric unit of β-D-glucopyranose residues joined by β-(1→4)-glycosidic linkages as shown in figure 1.2 (Klemm *et al*., 2005). The average degree of polymerization is approximately, 6000 glucose units in primary walls and up to 14000 in secondary walls. The cellulose microfibrils are hydrophobic and can be highly crystalline, contributing greatly to the recalcitrance of biomass (Harris *et al*., 2009).

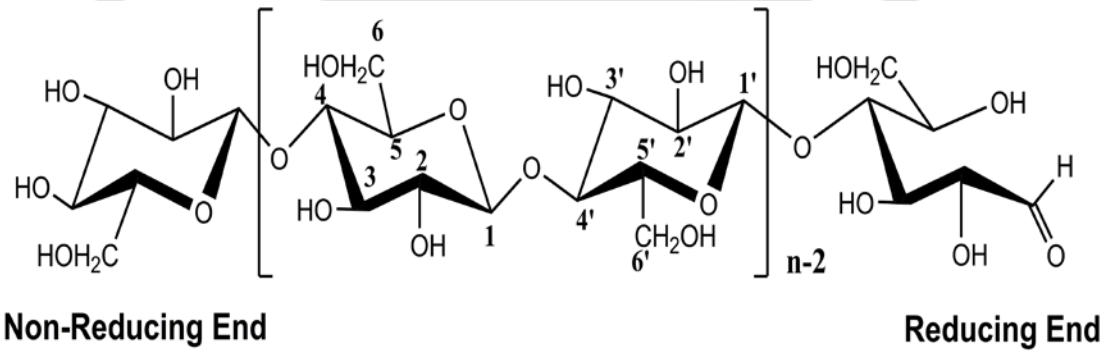


Fig. 1.2 Chemical structure of cellulose showing β-(1→4) glycosidic linkages, reducing end and non-reducing end (Klemm *et al*., 2005).

1.1.2 Hemicelluloses

Hemicelluloses are branched, hetero-polysaccharides formed by β-(1→4)-linked backbones of glucose, mannose or xylose. This stereo-irregular, non-cellulosic polysaccharide include xyloglucans, mannans, glucomannans, xylans and β-(1→3,1→4)-glucans (Scheller and Ulvskov, 2010). Unlike cellulose, hemicelluloses are amorphous, branched, single-chain polysaccharides, with 500-

3000 average degree of polymerization and are less recalcitrant as compared with cellulose (Davison, 2013). The most important biological role of hemicelluloses is to contribute to the strength of the cell wall by interaction with cellulose and lignin (Scheller and Ulvskov, 2010).

1.1.2.1 Xylans

After cellulose, xylans are the major 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 various degree of substitution depending on the type of plant tissue and species. The common side chain substitutions on the backbone are arabinose, glucuronic acid, acetic acid, ferulic acid, p-coumaric acid or 4-O-methyl glucuronosyl. Depending on the relative abundance of their substitutions within the xylan, it is further categorized into the following types.

1.1.2.1.1 Glucuronoxylan

Glucuronoxylan (GX_n) is composed of a linear backbone of β-(1→4)-linked D-xylopyranose units 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. GX_n is the major component of secondary cell wall of hardwood plant and crop residues hence constitutes a major part of plant biomass (Hayashi *et al*., 1996). The amount of GlcA (or 4-OMe-GlcA) varies depending upon the source and species of plant biomass. The molar ratio of GlcA (or 4-OMe-GlcA) to xylose may vary from 5 to 20. In glucuronoxylan the ratio of the glucuronic acid and β-(1→4)-linked D-xylopyranose units is reported to be 1:10, which means on an average every tenth xylose residue, is substituted at C-2 or C-3 positon by a 4-*O*-methyl α-D-glucuronic acid (MeGlcA) residue (Hurlbert and Preston, 2001).

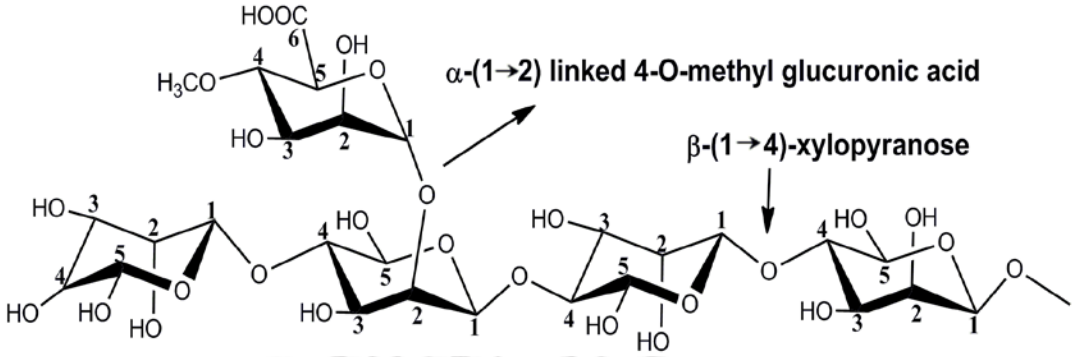


Fig. 1.3 Chemical structure of glucuronoxylan displaying linear chain of β-(1→4)-linked D-xylopyranose substituted with α-D-glucuronic acid (GlcA) or 4-*O*-methyl α-D-glucuronic acid (MeGlcA) residue at O2 position.

1.1.2.1.2 Arabinoxylans

Arabinoxylans have the backbone of β-(1→4)-linked xylose residues with the substitution of one or more L-arabinofuranosyl residues, at position 2 or 3 (Fig. 1.4). Arabinoxylans are present in primary cell walls of grasses and monocot plants cereals like wheat, rye, barley, oat, rice, corn and sorghum. They are the major hemicellulose component of flour and bran (Brett and Waldron, 1990). Arabinoxylans are important in the baking industry. The arabinose units bind water and produce viscous compounds that affect the consistency of dough, the retention of gas bubbles from fermentation in gluten-starch films and the final texture of baked products (Saha, 2003).

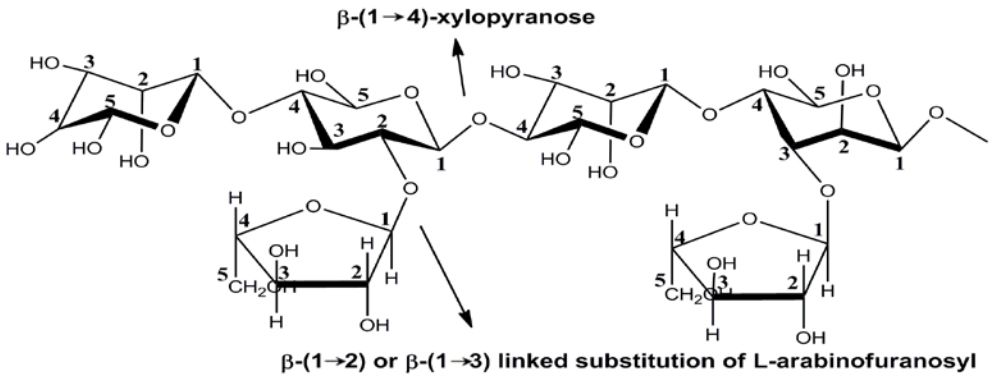


Fig. 1.4 Chemical structure of arabinoxylan displaying the β-(1→4)-linked xylose main chain substituted at O-2 and O-3 with L-arabinofuranosyl residues.

1.1.2.1.3 Xyloglucans

Xyloglucan has a backbone of β-(1→4)-glucose units, substituted with β-(1→6)-D-xylose side chains. Xyloglucan mainly occurs in the primary cell wall of dicotyledonous plants. Xyloglucan has a property to bind with cellulose microfibrils and therefore, xyloglucan-cellulose interactions are important determinants of the mechanical strength as well as the growth of cell wall (Cosgrove, 2005).

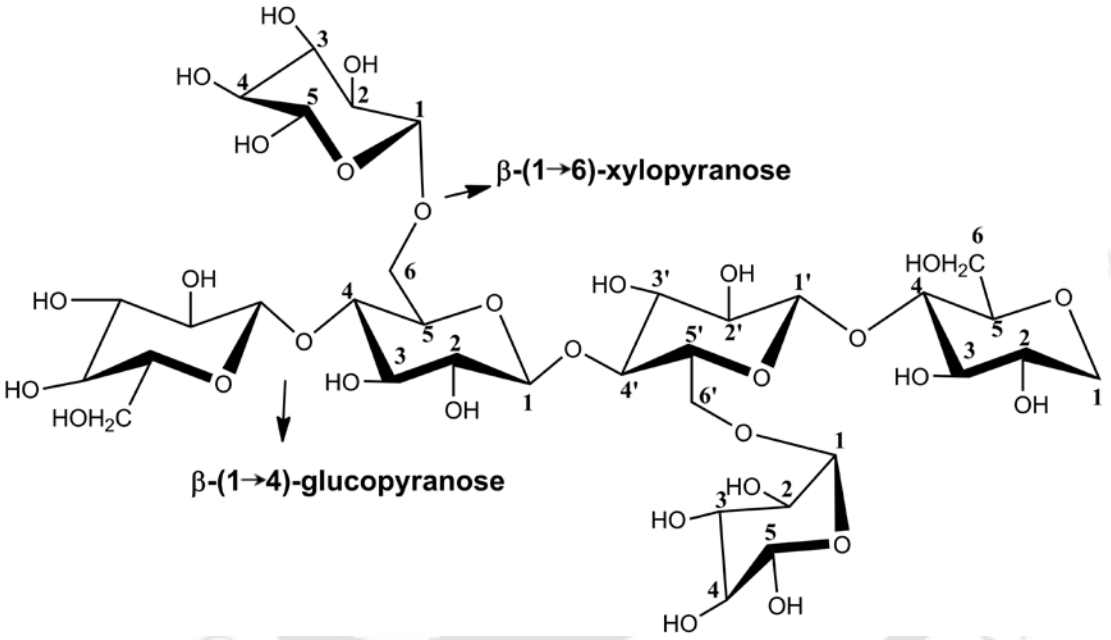


Fig. 1.5 Chemical structure of xyloglucan showing main chain β-(1→4)-glucose units substituted at β-(1→6) position with D-xylopyranose sugars side chains (Brennan and Harris, 2011).

1.1.2.1.4 Arabinans

Arabinans are important structural components of the cell wall and widely present in various plant tissues. They are composed of α-(1→5)-linked-L-arabinofuranosyl main chain residues to which other L-arabinofuranosyl residues are 1,3-α and 1,2-α linked in either a comb-like or a ramified arrangement (Fig 1.6) (Pitson *et al.*, 1997). Arabinan degrading enzymes are very much helpful in the food industries by processing the fruits, vegetable and agriculture by-products.

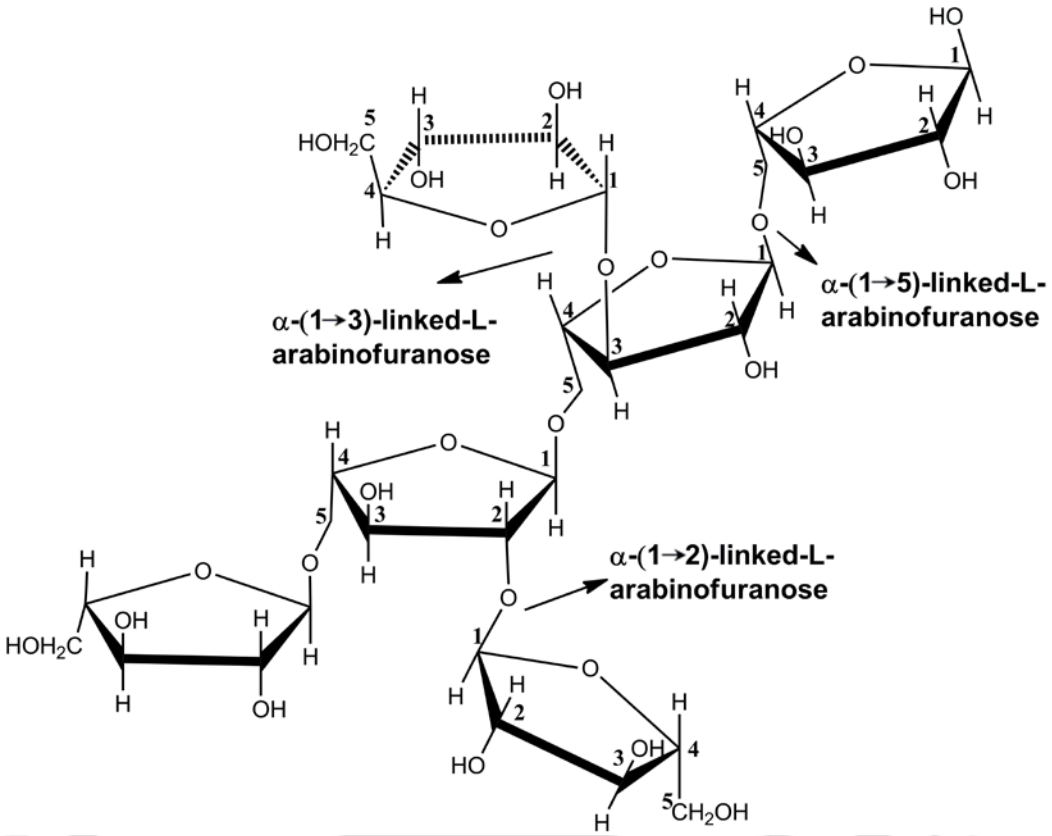


Fig. 1.6 Chemical structure of arabinan showing linear main chain of α-(1→5)-linked L-arabinofuranosyl substituted with L-arabinofuranosyl residues at 1,2 or 1,3 position.

1.1.2.1.5 Galactans

Galactans are polymers of β-D-galactopyranose (or galactose). Polysaccharides like galactans, arabinans and arabinogalactans are included in the heterogeneous group of hemicellulose as they show certain basic differences from other hemicellulases (Scheller and Ulvskov, 2010). The two basic differences from other hemicellulases are **i)** these polysaccharides show similarity with the pectin polysaccharides in their initial stage of synthesis thus appear as a part of pectin molecule. **ii)** They do not share the equatorial β-(1→4)-linked backbone structure like other hemicellulosic polysaccharides (Fig. 1.7).

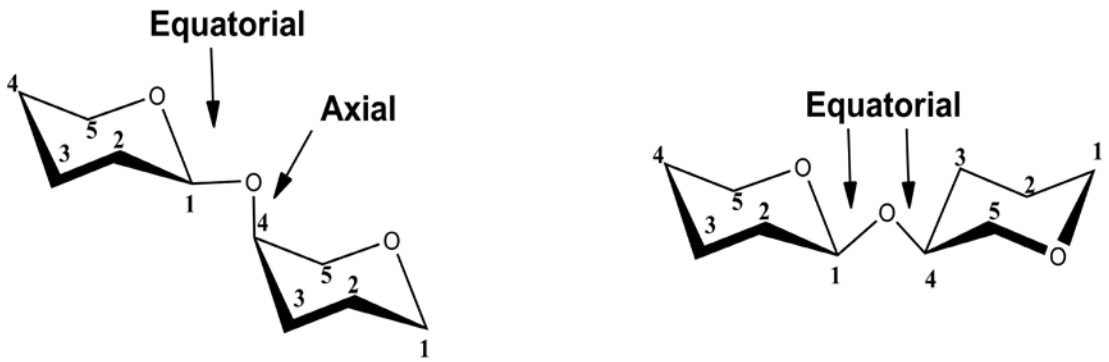


Fig. 1.7 Schematic presentation of axial and equatorial configuration at C1 and C4 position. Hemicelluloses having β-(1→4)-linked backbone showed either axial or equatorial configuration. Galactan has an axial configuration at C4 position.

1.1.2.1.6 Mannans

Mannans are the predominant hemi-cellulosic polysaccharides present in Charophytes (obligate water plants) (Popper, 2008; Popper and Fry, 2003). Mannans are heteropolymers containing β-(1→4)-D-mannopyranose backbone. The β-(1→4)-D-mannopyranose backbone is usually substituted with galactose or glucose. The important role of different mannans is that they serve as storage compounds in seeds (Meier, 1982). In gymnosperms, galacto-glucomannans are major components of the secondary cell walls (Ebringerov *et al*., 2005).

1.1.2.1.7 Galactomannans

Galactomannans are composed of β-(1→4)-D-mannopyranose backbone chain substituted with α-1-6-linked D-galactopyranose (Fig. 1.8). In leguminous plants where galctomannan serve as a main constitute of endospermic cell walls of seed, served as a storage food constitute more than 30% of the seed dry weight (Buckeridge, 2010).

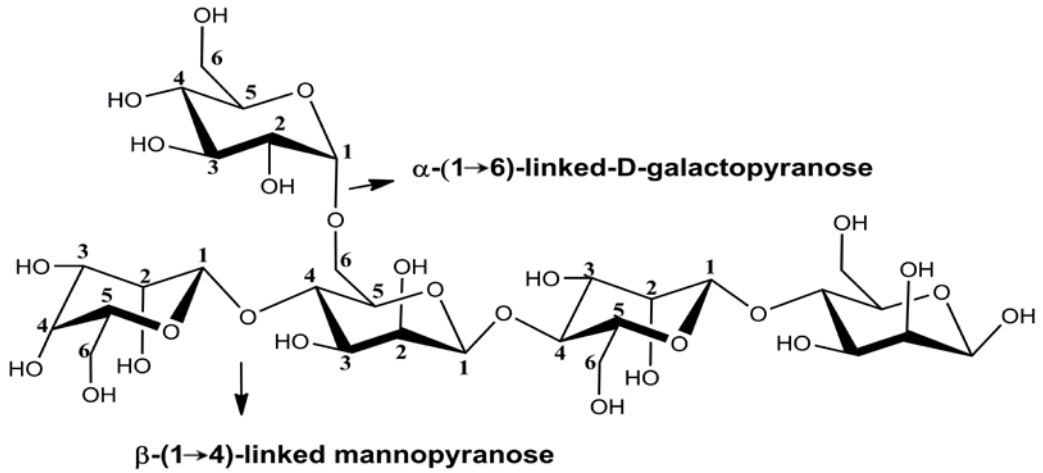


Fig. 1.8 Chemical structure of galactomannan exhibiting the repeating units of β-(1→4)-D-mannopyranose making the back bone substituted with α-(1→6)-linked-D-galactopyranose side chain.

1.1.2.1.8 Glucomannan

Glucomannans are generally a straight chain polymer of β-(1→4)-D-mannopyranose backbone with a limited branches of β-1,3 D-glucopyranosyl units (Fig. 1.9). The main source of glucomannan is a corm (root, not seed) of konjak plant (*Amorphophallus konjac*). It is also present in the cell wall of gymnosperms. It is a soluble dietary fibre used as a food additive and nutritional supplements to use as therapeutic agents in case of constipation, obesity, high cholesterol, acne valgaris and type 2 diabetes (Passaretti *et al*., 1991; Vuksan *et al*., 1999; Walsh *et al*., 1984).

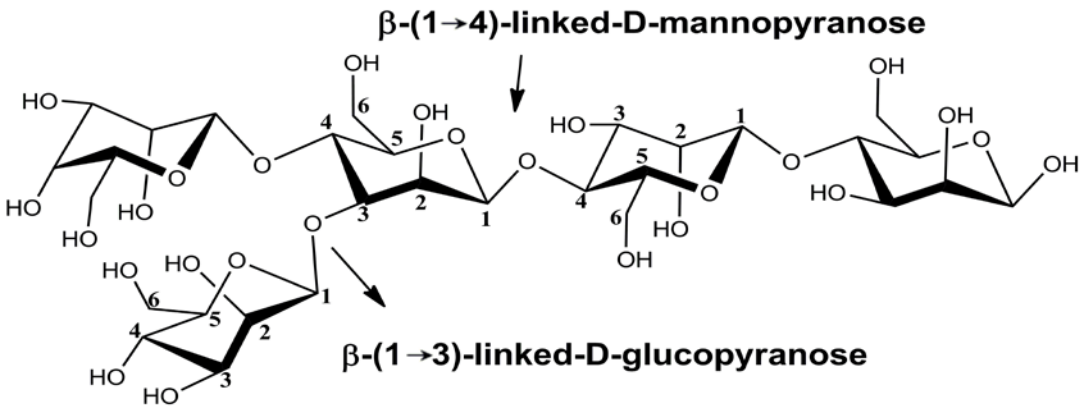


Fig. 1.9 Chemical structure of glucomannan displaying β-(1→4)-D-mannopyranose backbone substituted with β-(1→3) linked D-glucopyranose.

1.1.3 Lignins

Lignin contains a complex dendritic network of aromatic heteropolymers derived mainly from three hydroxycinnamyl alcohol monomers differing in their degree of methoxylation as shown in the figure 1.10 known as *p*-coumaryl (H), coniferyl (G), and sinapyl (S) alcohols (Fig. 1.10) (Freudenberg, 1959). Lignin associated with secondary cell walls represents the one of the major component. It provides mechanical support for the transportation of water and nutrients to the xylem in vascular plants. Lignin also act as a chemical barrier to microbial pathogens by its recalcitrance nature (Boerjan *et al*., 2003).

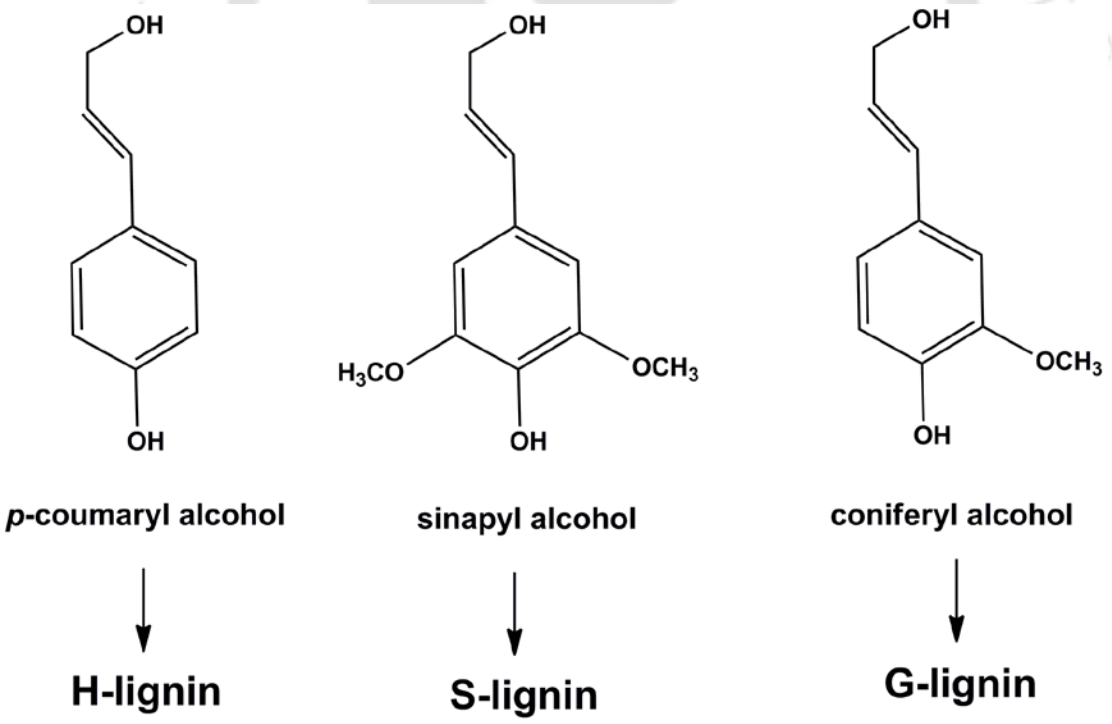


Fig. 1.10 Three hydroxycinnamyl alcohol monomers differing in their degree of methoxylation, *p*-coumaryl (H), sinapyl (S) and coniferyl (G) alcohols (Freudenberg, 1959)

1.1.4 Pectin

Pectin is a complex set of polysaccharides mainly present in the primary cell wall and middle lamella of all terrestrial plant cell wall. It contains a high proportion of D-galacturonic acid residues joined by α-(1→4) glycosidic linkages. Pectins are broadly grouped into three categories based on different substitutions *viz*, homogalacturonan, rhamnogalacturonan-I and substituted galacturonans.

1.1.4.1 Homogalacturonan

Homogalacturonan (HG) are composed of linear chain of α-D-galactosyluronic residues joined by α-(1→4) glycosidic linkages with some degree of esterification at the carboxylic acids of the galacturonates by methanol (methyl esterified) or L-rhamnose (α-6-deoxyhexose) residues (Fig. 1.11) (Ridley *et al*., 2001).

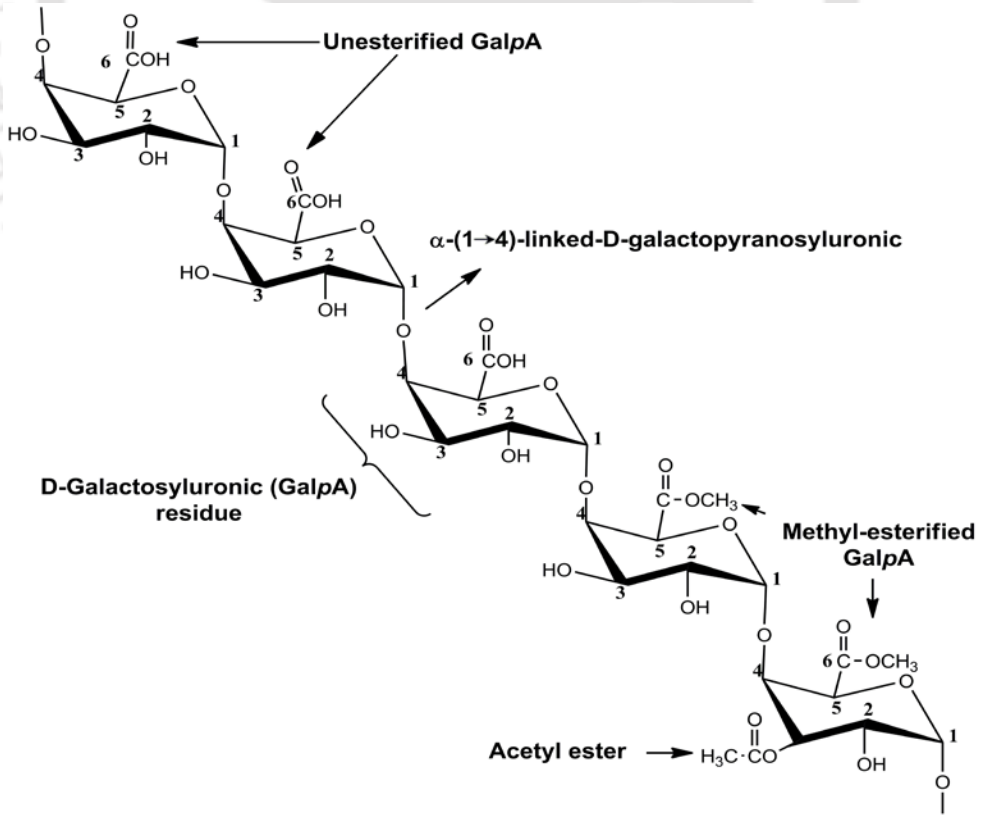


Fig. 1.11 Structure of homogalacturonan showing α-(1→4) linked D-galactopyranosyl-uronic acid residues (GalpA). Carboxyl groups of GalpA are often methyl-esterified and some hydroxyl groups may be O-acetylated (Ridley *et al*., 2001).

1.1.4.2 Heterogalacturonan

1.1.4.2.1 Rhamnogalacturonan-I

Rhamnogalacturonan-I (RG-I) represents 20-30% of the total pectin. It contains backbone of the repeating disaccharide unit composed of galacturonic acid and rhamnose with linkage of $[\rightarrow 4)$-α-D-GalpA-$(1\rightarrow 2)$-α-L-Rhap-$(1\rightarrow]_n$ as shown in Fig. 1.12. RG-I showed the substitution at O-3 and O-4 positions of α-L rhamnose residues with α-$(1\rightarrow 5)$-L arabinans and β-$(1\rightarrow 4)$-D galactans, arabinogalactans-I (AG-I), arabinogalactans-II (AG-II) and possibly galacto-arabinans (Jens *et al*., 2004).

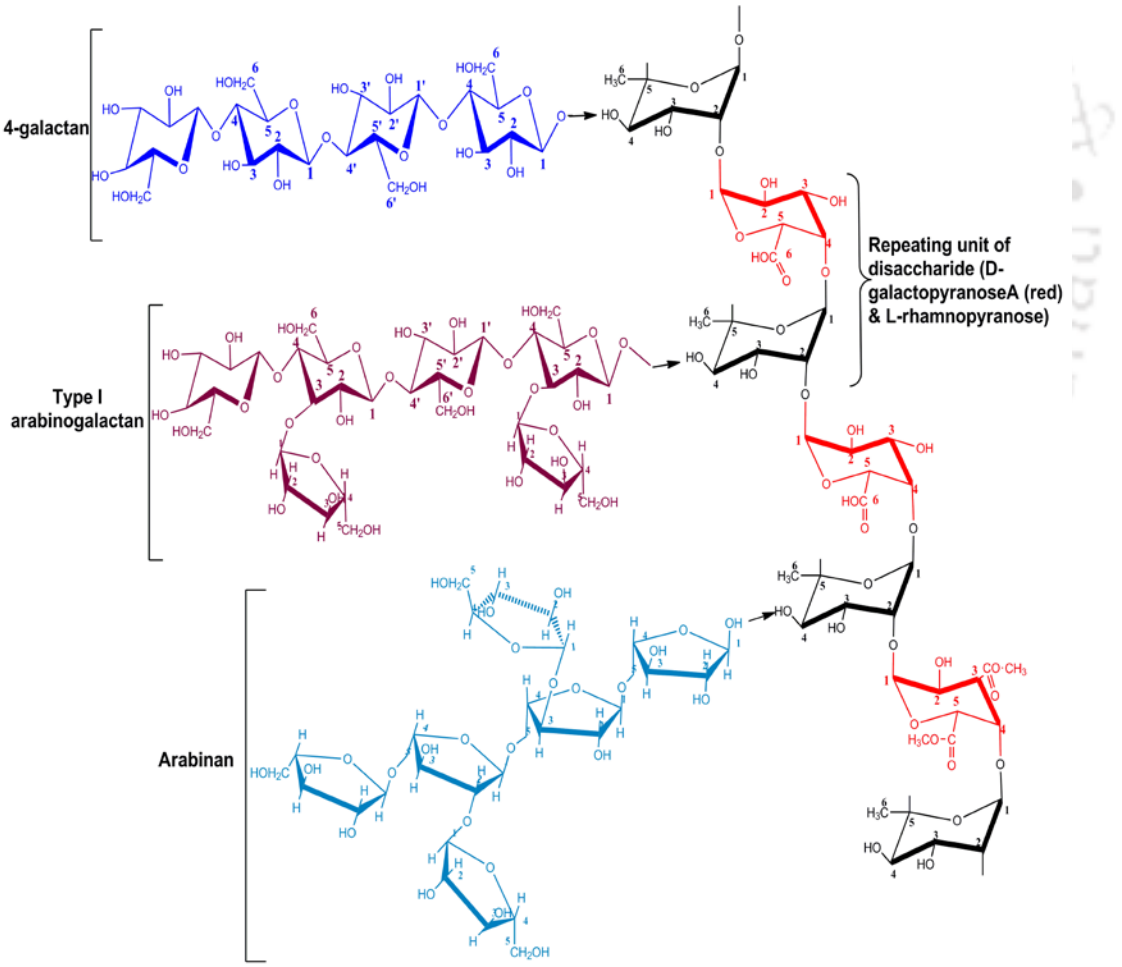


Fig. 1.12 Structure of rhamnogalacturonan I (RG-I) displaying repeating disaccharides units of α-D-galactopyranose (in Red) and α-L-rhamno-pyranose substituted with α-L-arabinofuranose and β-D-galactopyranose residues are linked to C-4 of some of rhamnopyranose residues.

The suggested function of RG-I to provide the support by linking other pectic polysaccharides such as HG and Rhamnogalacturonans-II (RG-II), that are covalently attached as side chains (Caffall and Mohnen, 2009).

1.1.4.2.2 Rhamnogalacturonan-II

Rhamnogalacturonan-II (RG-II) is a structurally complex pectic polysaccharide having a main chain backbone of 1,4-linked α-D-galacturonic acid (GalA) residues (O'Neill *et al*., 2004). Four structurally different oligosaccharide side chains are linked to the RG-II. They are termed as A, B, C and D. Side chain A is octasaccharide whereas, structurally diverse side chain B showed variation in the presence or absence of substituents linked to O-2 and/or O-3 of the Arap residue (Fig. 1.13). Both the side chains (A and B) attached to the C-2 carbon of the backbone GalA residue. The side chains C and D comprise two structurally different disaccharides attached to C-3 of the backbone GalA residues as illustrated in the figure 1.13. RG-II is predominantly present in the primary cell walls of both gymnosperms and angiosperms (O'Neill *et al*., 1990; Ridley *et al*., 2001). These polysaccharides HG, RG-I and RG-II are covalently linked to one another and form a pectin network, which gives strength, flexibility and functionality to the cells.

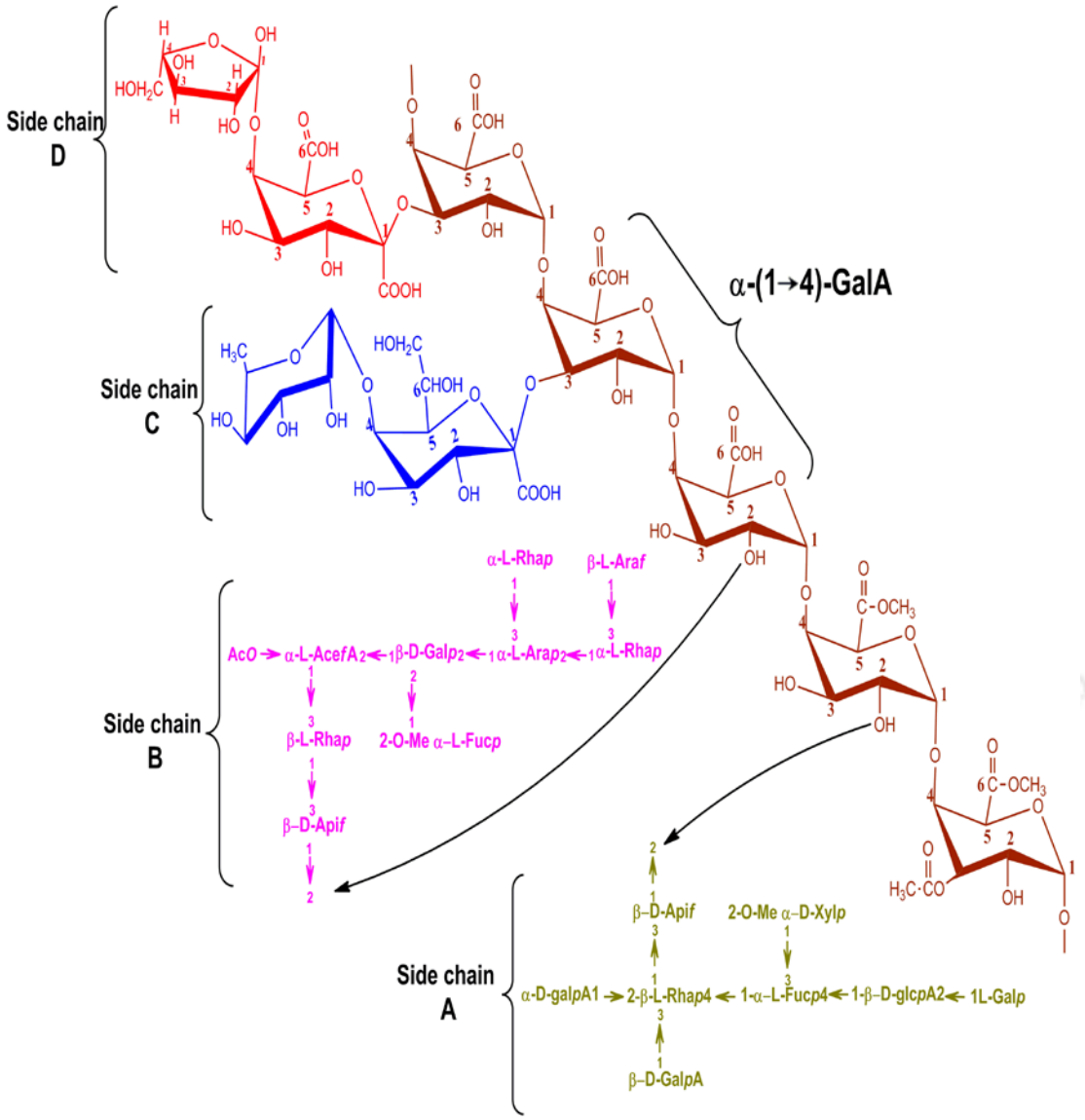


Fig. 1.13 The primary structure of rhamnogalacturonan II composed of α (1→4)-linked D-galacturonic acid (GalA) residues. Four structurally different oligosaccharide side chains (A, B, C and D) are linked to the rhamnogalacturonan II backbone. The abbreviations used for different units are 3-C-carboxy-5-deoxy-l-xylofuranosyl (*α*Ace*f*), arabinofuranosyl (Ara*f*), arabinopyranosyl (Ara*p*), fucopyranosyl (Fuc*p*), galactopyranosyl(Gal*p*), galactopyranosyluronic acid (Gal*p*A), glucopyranosyluronic acid (Glc*p*A), 2-*O*-methylfucopyranosyl (2-*O*-Me-Fuc*p*), 2-*O*-methylxylopyranosyl (2-*O*MeXyl*p*) OAc = O-acetyl ester; O-Me = O-methyl ether (Ishii *et al*., 1999).

1.2 Carbohydrate-active enzymes

Carbohydrates representing the most abundant complex biopolymer on the planet earth require variety of enzymes for their synthesis as well as breakdown. The enzymes which can either synthesize or breakdown the complex carbohydrates and glycol-conjugates are known as carbohydrate-active enzymes. They are mainly of four types: i) Glycoside hydrolases (GHs) hydrolyze the glycosidic linkage of glycosides, resulting in the formation of sugar hemiacetal or hemiketal and the corresponding free aglycon, ii) Glycosyltransferases (GTs) are involved in formation of glycosidic bond or biosynthesis of carbohydrates, iii) Carbohydrate esterases (CEs) remove the ester based modifications present in mono-, oligo- and polysaccharides and iv) Polysaccharide Lyases (PLs) cleave glycosidic bonds of uronic acid-containing polysaccharide chains (http://www.cazypedia.org/).

Carbohydrates are found in a large number of variations because of their structure flexibility to adapt to different stereoisomer by a variety of monosaccharides (epimers). Therefore, different carbohydrate active enzymes are required for breakdown of different substrates. These carbohydrate active enzymes evolved from a limited number of ancestors. To cope up with this variety of specificities required they acquired their specificities through subtle structural differences among them (Henrissat *et al*., 2008). The variety of enzymes acting on the glyco-conjugates, oligo- and polysaccharides probably required proper classification and indexing. The Carbohydrate-Active Enzymes database (CAZy; http://www.cazy.org) fulfils this requirement. This database was launched in 1999 and is constantly updated with the new CAZy sequences, their substrate specificity and structure information. Some of the salient features of CAZymes database are as follows,

1. The family classification system is based on protein sequence and structure similarities.
2. The database updated on the regular basis incorporating new family members and their biochemical and 3-D structure details extracted from the literature or protein data base (PDB). CAZy database covers approximately, 300 protein families (regularly updated) which are divided into different classes of enzyme activities.

1.2.1 Glycosyltransferases

Glycosyltransferases (GTs) (EC: 2.4.x.y) catalyse the transfer of sugar moieties of activated donor molecules to specific acceptor molecules, forming glycosidic bonds. GTs are involved in the biosynthesis of glycosidic bonds from phospho-activated sugar donors to synthesize different disaccharides, oligo-and polysaccharides (Henrissat, 2009; Sinnott, 1990; Yip *et al*., 2006). Glycosyltransferases follows retaining or inverting mechanism depending upon the stereo-chemistry of the substrates and reaction products. GTs contributed approx. 41% of total CAZymes at present and is virtually present in every single organism (Cantarel *et al*., 2009).

1.2.2 Polysaccharide lyases

Polysaccharide lyases (PLs) (EC: 4.2.2.x) or eliminases are a class of enzymes (EC 4.2.2.-) that cleave glycosidic bonds of uronic acid-containing polysaccharide chains. These enzymes act via a β-elimination mechanism to generate an unsaturated hexenuronic acid residue and a new reducing end. PLs contributed approx. 1.5% of total CAZymes but despite its less in number PLs are important in various biotechnological and biomedical applications (Lombard *et al*., 2010).

1.2.3 Carbohydrate esterase

The carbohydrate esterases (CEs) catalyze the de-O or de-N-acylation of substituted polysaccharides. They remove the ester based modifications present in mono-, oligo- and polysaccharides. It includes two types of substitution one in which sugar plays the role of the "acid", such as pectin methyl esters for 4-O-methyl-glucuronoyl methylesterase from *Schizophyllum commune* (Li *et al*., 2007) and secondly in which the sugar behaves as the alcohol, such as acetylated xylan for acetyl xylan esterase (family 1 and 2 CEs) from *Clostridium thermocellum* ATCC 27405 (Montanier *et al*., 2009).

1.2.4 Glycoside hydrolase

Glycoside hydrolase (GHs), (EC: 3.2.1.x) represent the highest number of enzyme classified and contributed 47% of total CAZy. It includes glycosidases as well as transglycosidases which can hydrolyse and/or transglycosylate the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (Henrissat, 1991; Henrissat and Bairoch, 1993, 1996). There are 133 glycoside hydrolase families as on date (March 2015) with more than 30000 entries in the CAZy database (www.cazy.org). On the basis of protein structure folds certain families of GHs were further categorized into 14 ‘clans’, knowing the fact that the protein folds are better conserved than their sequences, clans or ‘superfamilies’ classification was introduced. Different clans named alphabetically from A to N, suggested about the structure topology of the catalytic core region, like family 30 glycoside hydrolase lies in the clan ‘A’ (GH-A) which represents $(\beta/\alpha)_8$ topology. Similarly, GH5 and GH26 also belong to the same clan, as they have the same topology. GHs classification into different clans also reflects the fact they passage a

common evolutionary origin of their genes as well as it also gave information about the anomeric configuration of cleaved glycosidic bonds and the molecular mechanism of catalysis (either inverting, or retaining) (Naumoff, 2011). The Table1.1 provides the classification of GHs clans with structure folds.

Table 1.1 GH Clans with structure folds of related families (Naumoff, 2011).

GH Clans	Structure fold	Anomeric configuration	Associated Families
GH-A	$(\beta/\alpha)_8$	retained (eq)	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113, 128
GH-B	β-jelly roll	retained (eq)	7, 16
GH-C	β-jelly roll	retained (eq)	11, 12
GH-D	$(\beta/\alpha)_8$	retained (ax)	27, 31, 36
GH-E	6-fold β-propeller	retained (eq)	33, 34, 83, 93
GH-F	5-fold β-propeller	inverted (eq)	43, 62
GH-G	$(\alpha/\alpha)_6$	inverted (ax)	37, 63
GH-H	$(\beta/\alpha)_8$	retained (ax)	13, 70, 77
GH-I	α+β	inverted (eq)	24, 46, 80
GH-J	5-fold β-propeller	retained (β-furanoside)	32, 68
GH-K	$(\beta/\alpha)_8$	retained (eq)	18, 20, 85
GH-L	$(\alpha/\alpha)_6$	inverted (ax)	15, 65, 125
GH-M	$(\alpha/\alpha)_6$	inverted (eq)	8, 48
GH-N	β-helix	inverted (ax)	28, 49

eq = equatorial, ax= axial

1.2.4.1 Modular nature of glycoside hydrolases

Over the years different types of GHs have been identified and 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 (Henrissat and Davies, 2000). Consideration of different modules of particular GHs is the importance for correct open reading frame (ORF) annotation and functional prediction (Bourne and Henrissat, 2001).

1.2.4.2 Mechanism of action of Glycoside hydrolase

Hydrolysis of a glycosidic bond occurs through one of two possible stereochemical outcomes either the retention or by inversion of anomeric configuration. The similarity between these two reaction mechanisms is that both form an oxocarbenium-ion-like transition state and require a pair of carboxylic acids at the catalytic site. Whilst there is a subtle difference in both the mechanisms like in inverting glycoside the distance between the two carboxylic acid residues one acting as an electrophile/acid and the other as a nucleophile/base is ~10 Å and the reaction occurs via a single-displacement mechanism (Fig. 1.14).

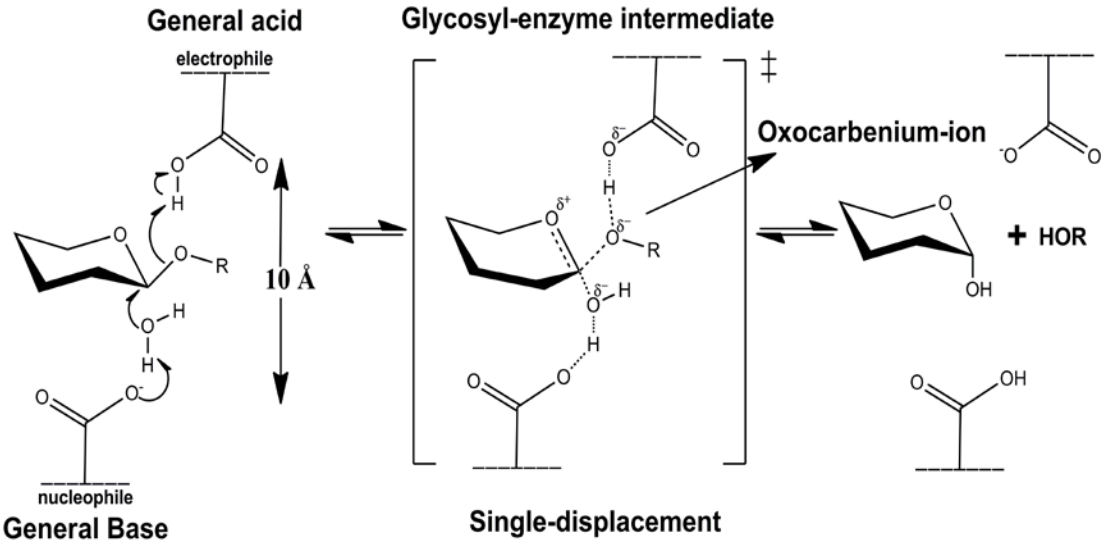


Fig. 1.14 Inverting mechanism *via* an oxocarbenium ion-like transition state leads to inversion of stereochemistry at the anomeric carbon.

In the retaining mechanism the distance between the two carboxylic acids is ~ 5.5 Å and the reaction occurs via a double displacement mechanism (Fig. 1.15).

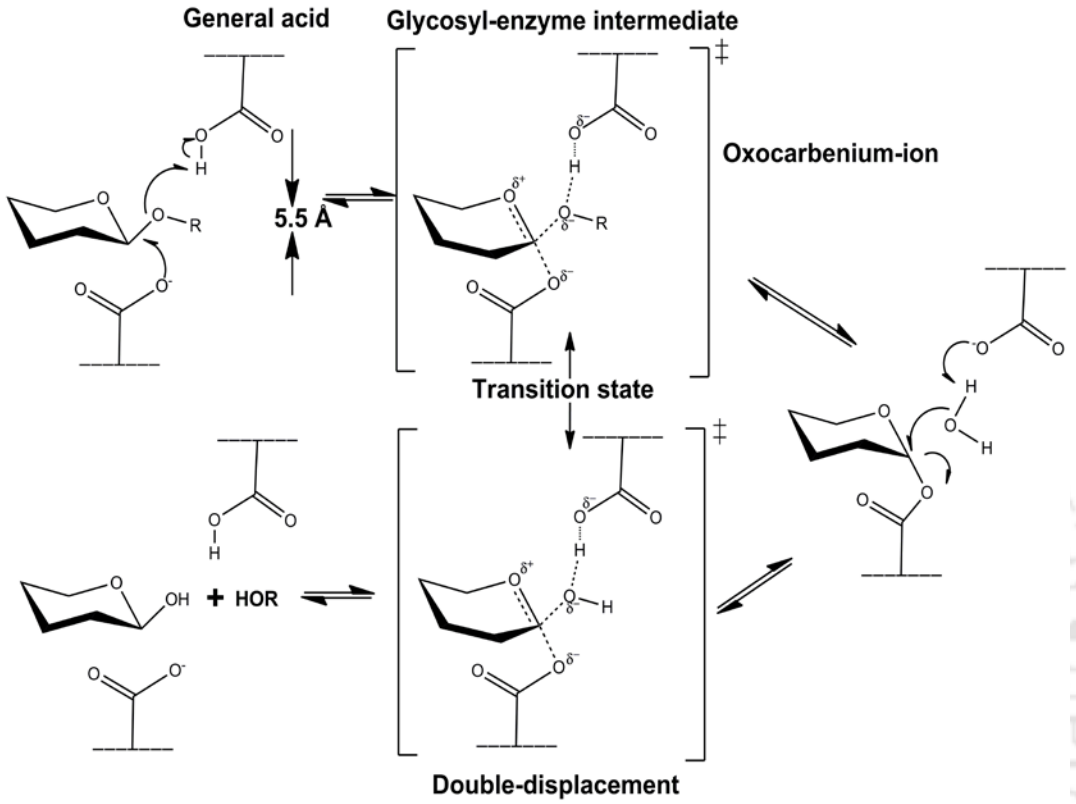


Fig. 1.15 Retaining type mechanism occurs via double displacement leading to retaining of the stereochemistry of the anomeric carbon.

1.2.4.3 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-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.-). A snapshot of the CAZy database is shown in the figure 1.16 displays the current status of the members and their respective numbers belonging to family 30 GHs.

Fig. 1.16 A snapshot of CAZy database displaying basic information and members of different classes (and their respective numbers) belonging to family GH30.

Currently, family GH30 entries in CAZy database include 561 bacterial and 68 protein entries from eukaryotes. They were further classified into 555 subfamilies (http://www.cazy.org/GH30.html). The family 30 glycoside hydrolases (GH30) belong to the GH-A clan having $(\beta/\alpha)_8$ topology of the catalytic core region. The mechanism of action of GH30 is of retention type where two acidic amino acids (generally glutamate) are involved in which one acts as acid/base and the other acts as a nucleophile.

1.2.4.4 GH30 sub-family 8 and *Ct*XynGH30

Initially, *Ct*XynGH30 was placed in Glycoside Hydrolase (GH) family 5 in the CAZy database. After amino acid sequence alignment and phylogenetic analysis it has been reclassified in GH family 30 glucuronoxylan-xylanohydrolase, subgroup H (in CAZy database it is in sub-family 8) (St John *et al*., 2010). In this sub-family, so far five enzymes have been structurally characterized. The two are XynC from *B. subtilis* (PDB id: 3GTN) and XynA (PDB id: 1NOF) from Gram-negative phytopathogenic *Erwinia chrysanthemi*. Both of these proteins were studied in their native as well as ligand bound crystal structures (Larson *et al*., 2003; St John *et al*., 2009). The other two protein structures were recently solved for Xyn30D (PDB id: 4QAW) from *Paenibacillus barcinonensis* which efficiently hydrolyzed glucuronoxylans and methyl-glucuronic acid branched xylo-oligosaccharides (Sainz-Polo *et al*., 2014), and *Cp*C71 (PDB id: 4FMV) from *Clostridium papyrosolvens* which displayed less activity towards glucuronic acid substituted xylans (St John *et al*., 2014). The fifth structure from sub-family 8 was of *Ct*Xyn30A (PDB id: 4QUE) the first glucuronoxylan-xylanohydrolase from *Clostridium thermocellum* (Verma *et al*., 2013). The other proteins belonging to the family GH30 subfamily 8 which were bio-

chemically characterized *viz.* Xyn5B from *Bacillus sp*. strain BP-7 showing activity on neutral, non-substituted xylo-oligosaccharides (Gallardo *et al*., 2010). The Xyn30D from *Paenibacillus barcinonensis* as mentioned earlier showed activity on methyl-glucuronic acid substituted xylans as well as on acidic xylo-oligosaccharides (Valenzuela *et al*., 2012). Xyn30A from *Bacillus licheniformis* SVD1 was characterized as glucurono-arabinoxylan endo-1,4-β-xylanase (Keppler and Showalter, 2010) and XynD as xylanase from *Aeromonas caviae* ME-1 (Sakka *et al*., 2012). The glycoside hydrolase enzyme glucuronoxylan-xylanohyrolase can be described as a xylanase enzyme which cleaves β-(1→4) linkages of 4-O-methylglucuronoxylan ($MeGX_n$) as directed by the position along the xylan chain of an α-(1→2)-linked 4-O-methylglucuronate (MeGA) moiety (St John *et al*., 2010). Chemically, glucuronoxylan (GXn) is composed of a linear backbone of β-(1→4)-linked D-xylopyranose units having a substitute of α-D-glucuronic acid (GlcA) or 4-O-methyl α-D-glucuronic acid (MeGlcA) residue at O2 position by α-(1→2) glycosidic bond. GXn is the major component of secondary cell wall of hardwood plant and crop residues hence constitutes a large part of plant biomass (Suzuki *et al*., 1997).

The xylanases of GH30 subfamily H are specific for hydrolysis of particular types of xylan such as glucuronoxylan, hence they differ from other xylanases belonging to families GH5, GH10 and GH11 which cleave internal xylosidic bonds of mainly linear xylo-oligosaccharides displaying endo-1,4-β-xylanase activity (Biely *et al*., 1997). Glucuronoxylan-xylanohydrolase differs from xylanases of family GH39 and GH43, which showed β-xylosidase activity hence producing mainly xylose as the main product (Collins *et al*., 2005; Santos *et al*., 2012). Glucuronoxylan-

xylanohydrolase also differs from family GH 67 α-glucuronidase that cleaves α-(1→2) glycosidic bond of GlcA or MeGlcA residues, substituted on the main chain of xylan (Zaide *et al*., 2001). Xylanases of family GH30 require a xylan decorated with GlcA or MeGlcA moiety. They cleave the β-(1→4) xylosidic bonds along with α-(1→2) substituted GlcA or MeGlcA moiety (Preston, 2003). Therefore; complete degradation of glucuronoxylan requires a cumulative action of endo-1,4-β-xylanase, β-xylosidase and α-glucuronidase enzymes.

1.2.5 Applications of glucuronoxylan-xylanohydrolase

Glucuronoxylan-specific xylanase cleaves β-(1→4) linkages of xylan hemicellulose same as endo-exo xylanases. Therefore, glucuronoxylan-xylanohydrolase has almost the same applications which are associated with the endo and exo acting xylanases. The specific function associated with GH30 glucuronoxylan-xylanohydrolase is that, they play an important role in complementing the action of family GH10 and GH11 enzymes in the depolymerization of glucuronoxylans present in plant biomass into simple fermentable sugars. The other well documented functions of xylanases are as follows: it is of great importance to the pulp and paper industries as the hydrolysis of the xylan facilitates release of lignin from paper pulp and reduces the level of usage of chlorine as the bleaching agent (Shoham, 1992). Xylanases play important role in protoplastation (cell wall maceration for the production of plant protoplasts) of plant cells as well as in clarification of juices and wine (Biely, 1985). They are used in liquefaction of coffee mucilage for making liquid coffee, recovery of oil from sub-terranian mines, extraction of flavors, pigments and starch (McCleary, 1986). Potential applications of xylanases also include bioconversion of lingocellulosic

material and agro-wastes to fermentative products, clarification of juices, improvement in consistency of beer and the digestibility of animal feed stock (Wong *et al*., 1988). In bakeries the xylanases act on the gluten fraction of the dough and help in the even redistribution of the water content of the bread thereby significantly improving the desirable texture, loaf volume and shelf life of the bread (Kulkarni *et al*., 1999). Xylanases in combination with other enzymes works more effectively in certain applications such as xylanases in conjuction with pectinases for degumming of bast fibers such as flax, hemp, jute and ramie (Borisjuk *et al*., 1999).

1.3 Carbohydrate binding modules

Initially the non-catalytic polysaccharide-recognizing modules of glycoside hydrolases were described as cellulose-binding domains (CBDs) because the first described module showed affinity towards crystalline cellulose. Later on more diverse carbohydrate binding domains were identified and then the term, carbohydrate binding module (CBM) was adopted as they reflected the diverse ligand specificities. (Tomme *et al*., 1988; Van Tilbeurgh *et al*., 1986). CBMs have an ability to fold and express independently of their associated catalytic modules. The well-defined function of this non-catalytic module is to increase the effective concentration of substrate in the vicinity of associated catalytic module by bringing the catalytic modules in close proximity to the substrate and thus to improve its performance (Shoseyov *et al*., 2006a). CBMs can be situated at either C- or N- terminal and occasionally occupy central position sandwiched by catalytic modules. It usually contains up to 200 amino acids and is found as single, double or multiple modules in one protein.

1.3.1 Classification of carbohydrate binding modules

Carbohydrate binding modules (CBMs) are currently grouped into 69 families based on the sequence similarity (http://www.cazy.org/Carbohydrate-Binding-Modules.html). The structure fold of proteins is better conserved than their sequences; so CBMs are further subdivided into three types based on mechanism of ligand binding and binding site topology (Boraston *et al*., 2004; Hashimoto, 2006). The Type A CBMs are described as 'surface binding' CBMs. They bind highly crystalline polysaccharides such as chitin or cellulose. The binding site architecture of type 'A' CBMs, is flat or platform-like formed by aromatic amino acid to accommodate flat surfaces of insoluble polysaccharides (chitin or cellulose) (Nagy *et al*., 1998). The type 'B' (glycan chain-binding), CBMs show affinity towards soluble polysaccharides. Binding site is situated in the grooves formed by the aromatic amino acid like Type A, in a planer, twisted or sandwich form to accommodate the ligand molecule. Type 'C' (small sugar binding), binds the smaller oligosaccharides by recognizing the terminal sugars of these molecules. The binding site is formed by the grooves as in case of type B but is smaller.

Another classification on the basis of 'fold similarity' was suggested (Czjzek *et al*., 2001; Sunna *et al*., 2001). In this approach 22 different families were classified into seven 'fold families'. The seven different folds include the β-sandwich fold, the β-trefoil fold, oligonucleotide/oligosaccharides binding fold (OB), the hevein fold, the knottin fold (cysteine knot), the hevein-like fold and a unique fold (Boraston *et al*., 2004). The most abundant fold in terms of total number is the β-sandwich fold. This fold comprises two β-sheets, each consisting of three to six antiparallel β-strands. The second most frequent fold is β-trefoil fold. It is present in ricin toxin B-chain. This

fold contains 12 strands of β-sheet, forming six hairpin turns. (Rutenber and Robertus, 1991, Murzin *et al*., 1992).

1.3.2 Functions of carbohydrate binding modules

Apart from the increasing effective concentration by bringing the biocatalyst into prolonged and close vicinity with its substrate, the other well described function of CBMs are as follows:

a) ***Substrate Binding and Selectivity***: CBMs show affinity towards insoluble as well as soluble polysaccharides. CBMs also functions as selectively targeting the hydrolytic enzymes towards a specific substrate with in the dense plant cell wall (McLean *et al*., 2002).

b) ***Nonhydrolytic Substrate Disruption***: CBMs are also known to disrupt the substrate without possesing any hydrolytic activity (Southall *et al*., 1999). The interesting application associated with this property is removal of dental plaque polysaccharides (mainly fructan and glucan), by applying CBM (Fuglsang and Tsuchiya, 2011).

c) ***Surface/Interfacial Modifications***: CBMs also showed the phenomenon by which it is capable of changing the surface of the polysaccharides. It was shown that on treatment of cotton fibres with CBM alters their affinity for dyeing (Paulo *et al*, 1999).

1.3.3 Applications of carbohydrate binding modules

Three basic features of CBMs have made them perfect candidates for several applications: **i)** CBMs are usually independently folding units and therefore can function autonomously in chimeric proteins; **ii)** the attachment matrices are abundant and inexpensive and have excellent chemical and physical properties; and **iii)** the

binding specificities of CBMs can be controlled by modifying the structure by site directed mutagenesis and therefore the right solution can be adopted for a particular task (Abbott *et al*., 2009).

The important application of CBMs includes, construction of an expression vector containing a CBM as a fusion tag (Novy *et al*., 1997). CBMs are used in immobilized affinity ligand technology for purification of biomolecules (Greenwood *et al*., 1992). The high thermo-stability of CBMs play significant role in separating different carbohydrate analyte (cello- and xylo-oligomers) by retaining them at higher temperatures for better separation of oligosaccharides (Johansson *et al*., 2006). CBMs are also used in making biosensors such as CBM of *Cellulomonas fimi* was used for glucose sensing in bioreactor (Levy and Shoseyov, 2002; Phelps *et al*., 1995). Whole-cell immobilization by cellulosic material was first demonstrated when an *E. coli* surface anchored CBM, derived from *Cellulomonas fimi*, was attached to cellulose (Francisco *et al*., 1993; Wang *et al*., 2001). The cell immobilization technique using CBM, was also explored for bioremediation purpose such as detoxification of nerve gas by immobilizing the organo-phosphorus hydrolase as well as removal of heavy metal contamination from polluted air (Shoseyov *et al*., 2006; Wang *et al*., 2002).

1.3.4 Carbohydrate Binding Modules of family 6

The overall status of the family 6 CBM is shown in the Fig. 1.17. The snapshot from the CAZy database is displaying the current information related to CBM6 *viz.* the affinity shown by different characterized members, the dominant 3-D structure topology and proper indexing of present members across the different strata of hierarchy (www.cazy.org/CBM6.html). The family 6 CBM shows diverse binding affinities with different polysaccharides therefore it is interesting to investigate the

protein-carbohydrate interaction and structural aspect of the protein which makes family 6 flexible enough to interact with soluble as well as insoluble polysaccharide having different monomeric units.

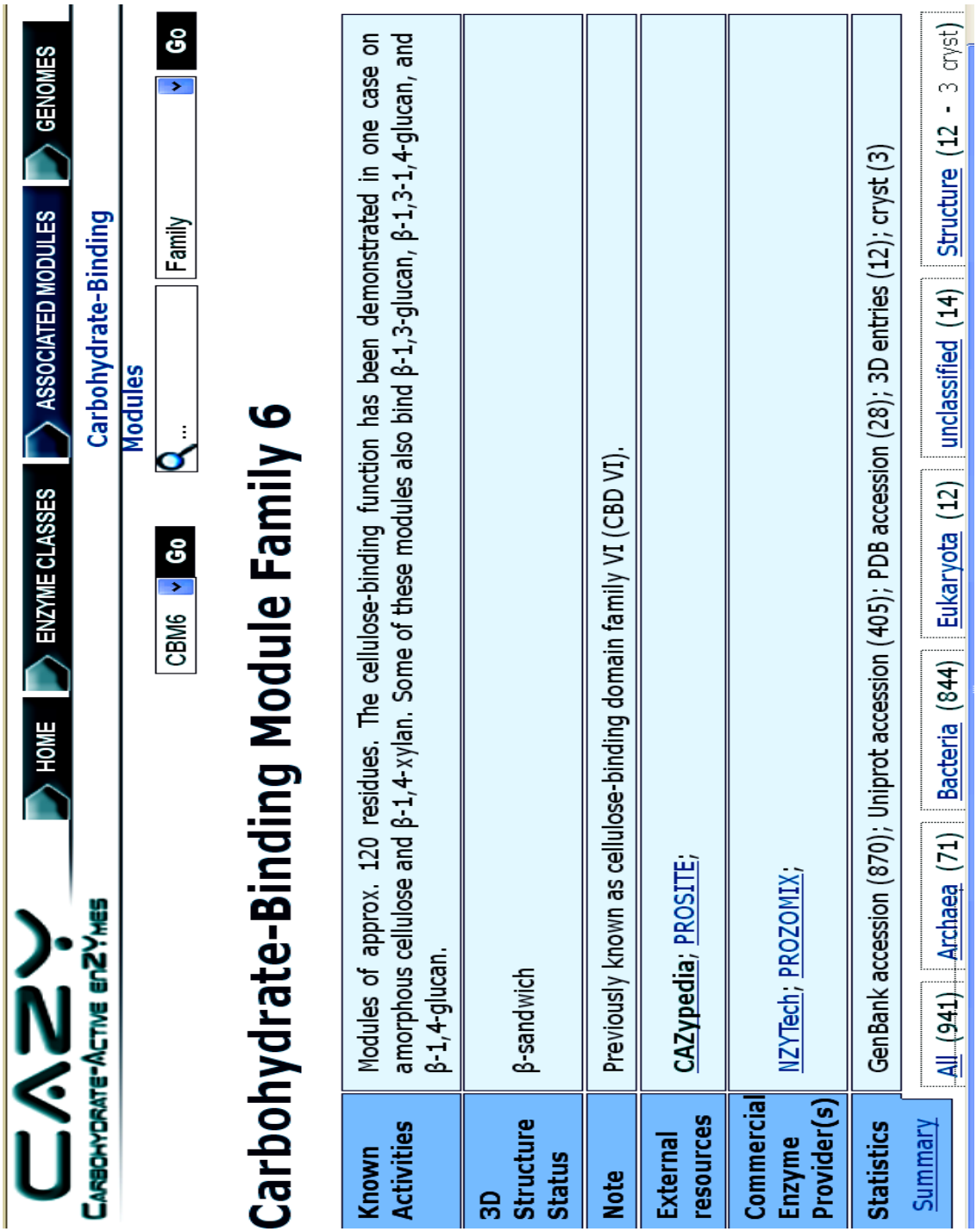


Fig. 1.17 A snapshot from CAZy database displaying basic information and all members of different classes (and their respective numbers) from CBM6.

1.4 The microorganism

Clostridium thermocellum is an anaerobic, thermophilic, cellulolytic and ethanologenic, Gram-positive bacterium capable of directly converting cellulose 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. 1.18)

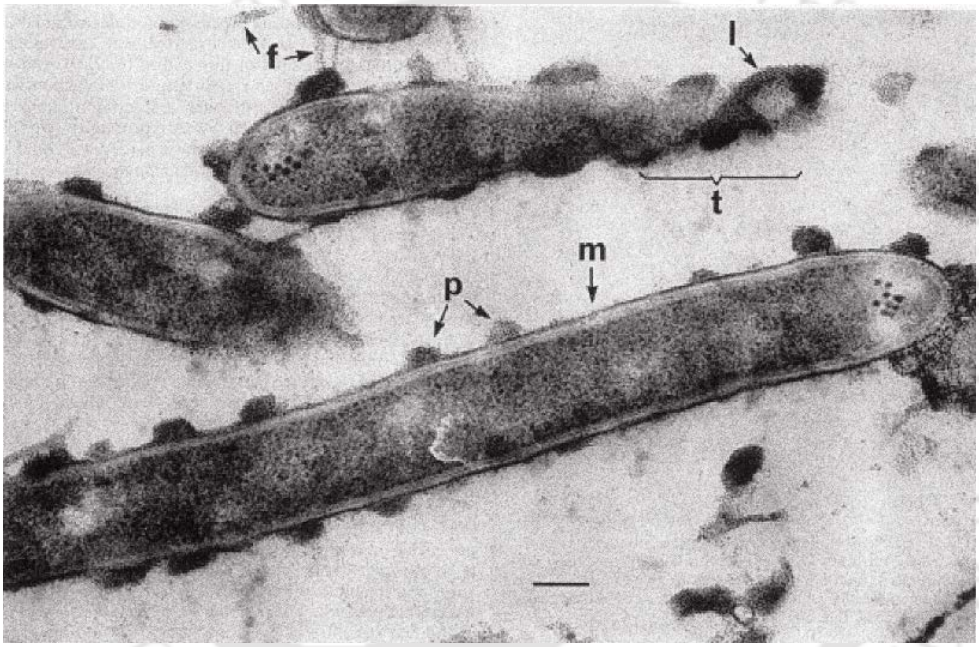


Fig. 1.18 Transmission electron microscope (TEM) image of cationized ferritin (CF) stained *Clostridium thermocellum* YS grown on cellobiose. The three major types of labelling: the monolayer (m) of CF particles enveloping the entire cell surface, the fibrous structures (f) which connecting two adjacent 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 labelling of the protuberances, and the ferritin particles appeared to be packed very tightly over the entire structure (adapted from Bayer and Lamed, 1986).

1.5 Cellulosomes

The cellulosomes are defined as an extracellular supramolecular machine that can efficiently degrade crystalline cellulosic substrates and associated plant cell wall polysaccharides. The cellulosome arrangement can also promote adhesion to the insoluble substrate, thus providing individual microbial cells with a direct competitive advantage in the utilization of the soluble hydrolysis products (Shoham *et al*., 1999). The cellulosome of C*lostridium thermocellum* contains different types of glycosyl hydrolases which include cellulases, hemicellulases and carbohydrate esterases bound to a polypeptide called scaffolding (also known as the cellulosome integrating protein, CipA). C*lostridium thermocellum* is considered to show one of the highest rates of cellulosome utilization by any bacterium, therefore detailed studies were carried out on the organization cellulosome and its component. (Bayer *et al*., 2004; Demain *et al*., 2005; Johnson *et al*., 1982). The conversion of lignocellulosic biomass into the fermentable sugars is still a rate limiting step in bioconversion to ethanol, so the effective conversion of lignocellulose into fuels require more efficient systems (Lynd *et al*., 2002). Cellulosome is also described as one of nature's most elegant and elaborate nano-machines. The ability of this multicomponent system to degrade the recalcitrant and chemically complex polysaccharides efficiently showed its potential for use in biofuel production (Fontes and Gilbert, 2010).

1.6 Significance of investigation

In recent times the advent of novel xylanase from different microbial sources at the expense of genetic manipulation has led to great advances and development of carbohydrate active enzymes. 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 us with sufficient amounts of animal feed, food, fuel, and chemicals. 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. The genome sequence of several industrially important organisms and available advance genetic engineering tools have revolutionised the overall approach for screening the efficient enzyme candidates. In several carbohydrate active enzymes, including xylanases predominantly belonging to GH10 and GH11 families have been cloned and characterized. A potential carbohydrate active enzyme will be screened from thermophilic *Clostridium thermocellum* knowing the fact that its genome contains a complex cellulosome containg a large number of enzymes, capable of degrading cellulosic and hemicellulosic polysaccharides effectively. Furthermore, the cellulosomal enzyme complex is 50 times more efficient in degrading polysaccharides, as reported by Fontes and Gilbert (2010). The proposed study will involve a xylanase known as glucuronoxylan-xylanohydrolase belonging to family GH30 (sub-family 8) from *Clostridium thermocellum*. In this regard the present study

deals with the cloning, structure and function characterization of modular glucuronoxylan-xylanohydrolase from *Clostridium thermocellum.*

1.7 Objectives of the present study

The protein having gene accession number ABN54208 and uniprot ID A3DJS9, belongs to family GH30 (sub-family 8) from *Clostridium thermocellum* was chosen for the study along with a carbohydrate binding module of family 6. Initially it is belonged to family GH5, later it was classified in GH30 and sub-family 8 after sequence and phylogenetic analysis. The other member of this subfamily was described in the Section 1.2.4.4 in this chapter. The aim of the current study is to clone all the three modules full-length (*Ct*XynGH30) truncated catalytic module (*Ct*Xyn30A) and carbohydrate binding module (*Ct*CBM6) and later on biochemically structurally and functionally characterize them. The gene sequences will be retrieved and there module boundaries will be assigned. The specific primers will be designed to amplify the respective genes and they will be subsequently cloned into an expression vector followed by expression in *E. coli*. All three modules *Ct*XynGH30 (full-length), *Ct*Xyn30A (truncated catalytic module) and *Ct*CBM6 will be overexpressed and purified via immobilized metal ion affinity chromatographic technique (IMAC). The substrate specificity of full-length and truncated catalytic module enzymes will be determined. The reaction conditions for enzyme assay such as optimum pH, temperature and ionic strength of the buffer will be optimized. The pH and temperature stability will be studied. The effects of different metal ions, chemicals and additives on enzyme activity will be investigated. The enzyme activity with different substrate including the synthetic *p*NP (para nitro phenyl) glycosides will be addressed. Zymogram analysis will be carried out to confirm the activity. The

kinetic parameters of the enzymes will be determined. Purification and characterization of substrates degradation products will be carried out by a chromatographic technique including TLC followed by analysis with the help of mass- and NMR-spectrophotometer.

In the present study 3-dimentional structure of catalytic module *Ct*Xyn30A of family GH30 and associated module *Ct*CBM6 of family 6 from *Clostridium thermocellum* will be studied by employing both computational approach and X-ray crystallography technique to understand the architect of substrate binding cleft in *Ct*CBM6 while the overall structure feature of *Ct*Xyn30A. Docking study with xylo-oligosaccharide ligands and ligand bound crystal structures will provide the necessary information regarding the key amino acids involved in substrate targeting in *Ct*CBM6, whereas, in case of the catalytic module it infers regarding the key residues responsible for substrate catalysis, respectively. This study will help in generating the desired mutants of CBM6 that participate in substrate binding. In catalytic *Ct*Xyn30A by employing site directed mutagenesis approach, desired mutants with key amino acid residues modified will be created in order to establish the role of these residues in catalysis. Therefore, structural and functional investigation of *Ct*Xyn30A is important in order to understand the differences as well as similarities among the proteins characterized so far from the same sub-family as well as across the families. Moreover GH family 30 (sub-family 8), (glucuronoxylan-xylanohydrolase) is known to effectively depolymerise acidic xylans (major component of secondary cell wall of hardwood plant) into simple sugars that can be utilized for bioethanol production.

1.7.1 Specific objectives

1. Cloning, expression and purification of family 30 glucuronoxylan-xylanohydrolase *Ct*XynGH30 (full-length) and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6 from *Clostridium thermocellum.*
2. Biochemical and functional characterization of *Ct*XynGH30 and *Ct*Xyn30A from *Clostridium thermocellum.*
3. Homology modelling and molecular docking study of *Ct*Xyn30A from *Clostridium thermocellum*.
4. Crystallization and three dimensional structure determination of glucuronoxylan-xylanohydrolase (*Ct*Xyn30A) from *Clostridium thermocellum.*
5. Functional and biochemical characterization and ligand binding specificity of family 6 carbohydrate binding module (*Ct*CBM6) of *Ct*XynGH30.
6. *In silico* structure determination and molecular docking of family 6 carbohydrate binding module (*Ct*CBM6) of *Ct*XynGH30.

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, 10395-10404.

Bayer, E.A., Belaich, J.P., Shoham, Y., Lamed, R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annual Review of Microbiology, 58, 521-554.

Bayer, E.A., Kenig, R., Lamed, R. (1983) Adherence of *Clostridium thermocellum* to cellulose. Journal of Bacteriology, 156, 818-827.

Biely P., V.M., Tenkanen M., Kluepfel D. (1997) Endo-β-1,4 xylanase families: Differences in catalytic properties. Journal of Biotechnology, 57, 151-166.

Biely, P. (1985) Microbial xylanolytic systems. Trends in Biotechnology, 3, 286-290.

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

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

Borisjuk, N.V., Borisjuk, L.G., Logendra, S., Petersen, F., Gleba, Y., Raskin, I. (1999) Production of recombinant proteins in plant root exudates. Nature Biotechnology, 17, 466-469.

Bourne, Y., Henrissat, B. (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. Current Opinion in Structural Biology, 11, 593-600.

Brennan, M. and Harris, P.J. (2011) Distribution of fucosylated xyloglucans among the walls of different cell types in monocotyledons determined by immunofluorescence microscopy. Mol. Plant. 4, 144-56.

Brett, C., Waldron, K. (1990) Physiology and Biochemistry of Plant Cell Walls. Springer Science.

Buckeridge, M.S. (2010) Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. Plant Physiology, 154, 1017-1023.

Caffall, K.H., Mohnen, D. (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydrate Research, 344, 1879-1900.

Cantarel, B.L. 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 Research, 37, D233-238.

Collins, T., Gerday, C., Feller, G. (2005) Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiology Reviews, 29, 3-23.

Cosgrove, D.J. (2005) Growth of the plant cell wall. Nature Reviews Molecular Cell Biology, 6, 850-861.

Coutinho, P.M., Rancurel, C., Stam, M., Bernard, T., Couto, FM., Danchin, E.G. J., Henrissat, B. (2009) Carbohydrate-Active Enzymes Database: principles and classification of glycosyltransferases. Bioinformatics for Glycobiology and Glycomics. John Wiley & Sons, Ltd, pp. 89-118.

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

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. John Wiley & Sons, Ltd.

Demain, A.L., Newcomb, M., Wu, J.H. (2005) Cellulase, clostridia, and ethanol. Microbiology and Molecular Biology reviews, 69, 124-154.

Ebringerov, A., Hrom Z., Heinze T. (2005) Hemicellulose. Advances in Polymer Science. Springer Science Business Media, pp. 1-67.

Eisenstein, R.S. (2006a) structure nomenclature and properties of carbohydrats Biochemical, Physiological, and Molecular Aspects of Human Nutrition, 2nd Edition ed. Ovid Technologies (Wolters Kluwer Health), Medicine & Science in Sports & Exercise, 38, 50-64 (2182).

Eisenstein, R.S. (2006b) Biochemical, Physiological, and Molecular Aspects of Human Nutrition, 2nd Edition ed. Ovid Technologies (Wolters Kluwer Health).

Fontes, C.M., Gilbert, H.J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annual Review of Biochemistry 79, 655-681.

Francisco, J.A., Stathopoulos, C., Warren, R.A., Kilburn, D.G., Georgiou, G. (1993) Specific adhesion and hydrolysis of cellulose by intact *Escherichia coli* expressing surface anchored cellulase or cellulose binding domains. Biotechnology (N Y), 11, 491-495.

Freudenberg, K. (1959) Biosynthesis and constitution of lignin. Nature, 183, 1152-1155.

Fuglsang, C.C., and Tsuchiya R. (2011) Cellulose binding domains (CBDs) for oral care products. In: patent, U.S. (Ed.)

Gallardo, O., Fernandez-Fernandez, M., Valls, C., Valenzuela, S.V., Roncero, M.B., Vidal, T., Diaz, P., Pastor, F.I. (2010) Characterization of a family GH5 xylanase with activity on neutral oligosaccharides and evaluation as a pulp bleaching aid. Applied and Environmental Microbiology, 76, 6290-6294.

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

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

Harris, P.J., Stone, B.A. (2009) Chemistry and Molecular Organization of Plant Cell Walls. Biomass Recalcitrance. Blackwell Publishing Ltd., pp. 61-93.

Hashimoto, H. (2006) Recent structural studies of carbohydrate-binding modules. Cellular and Molecular Life Sciences, 63, 2954-2967.

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. Letters in Applied Microbiology, 22, 293-298.

Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. The Biochemical Journal, 280, 309-316.

Henrissat, B., Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. The Biochemical Journal, 293, 781-788.

Henrissat, B., Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. The Biochemical Journal, 316, 695-696.

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

Henrissat, B., Sulzenbacher, G., Bourne, Y. (2008) Glycosyltransferases, glycoside hydrolases: surprise, surprise! Current Opinion in Structural Biology, 18, 527-533.

Hurlbert, J.C., Preston, J.F., 3rd, (2001) Functional characterization of a novel xylanase from a corn strain of *Erwinia chrysanthemi*. Journal of Bacteriology 183, 2093-2100.

Ishii, T., Matsunaga, T., Pellerin, P., O'Neill, M.A., Darvill, A., Albersheim, P., 1999. The plant cell wall polysaccharide rhamnogalacturonan II self-assembles into a covalent cross-linked dimer. Journal of Biological Chemistry. 274, 13098-13109.

Jens, O.B., Harholt, J., Scheller, H.V., Orfila, C. (2004) Rhamnogalacturonan I in Solanum tuberosum tubers contains complex arabinogalactan structures. Phytochemistry 65, 1429-1438.

Johansson, R., Gunnarsson, L.C., Ohlin, M., Ohlson, S. (2006) Thermostable carbohydrate-binding modules in affinity chromatography. Journal of Molecular Recognition, 19, 275-281.

Johnson, E.A., Sakajoh, M., Halliwell, G., Madia, A., Demain, A.L. (1982) Saccharification of complex cellulosic substrates by the cellulase system from *Clostridium thermocellum*. Applied and Environmental Microbiology, 43, 1125-1132.

Keppler, B.D., Showalter, A.M. (2010) IRX14 and IRX14-LIKE, two glycosyl transferases involved in glucuronoxylan biosynthesis and drought tolerance in Arabidopsis. Molecular Plant, 3, 834-841.

Klemm, D., Heublein, B., Fink, H.P., Bohn, A. (2005) Cellulose: fascinating biopolymer and sustainable raw material. Angewandte Chemie, 44, 3358-93.

Kulkarni, N., Shendye, A., Rao, M. (1999) Molecular and biotechnological aspects of xylanases. FEMS Microbiology Reviews, 23, 411-456.

Larson, S.B., Day, J., Barbade la Rosa, A.P., Keen, N.T., McPherson, A. (2003) First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis. Biochemistry, 42, 8411-8422.

Levy, I., Shoseyov, O. (2002) Cellulose-binding domains: biotechnological applications. Biotechnology Advances, 20, 191-213.

Li, X.L., Spanikova, S., de Vries, R.P., Biely, P. (2007) Identification of genes encoding microbial glucuronoyl esterases. FEBS Letters, 581, 4029-4035.

Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M., Henrissat, B. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. The Biochemical Journal, 432, 437-444.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiology and Molecular Biology Reviews, 66, 506-577.

McCann, M.C. and Roberts, K. (1991) Architecture of the primary cell wall. In The Cytoskeletal Basis of Plant Growth and Form (ed. C.W. Lloyd), Academic Press, London, pp. 109–129.

McCleary, B.V. (1986) Enzymatic modification of plant polysaccharides. International Journal of Biological Macromolecules, 8, 349-354.

McLean, B.W., Boraston, A.B., Brouwer, D., Sanaie, N., Fyfe, C.A., Warren, R.A., Kilburn, D.G., Haynes, C.A. (2002) Carbohydrate-binding modules recognize fine substructures of cellulose. The Journal of Biological Chemistry, 277, 50245-50254.

Meier, H., Reid, J. S. G. (1982) Reserve polysaccharides other than starch in higher plants (Plant Carbohydrates I), Springer Berlin Heidelberg pp. 418-471.

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

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

Nagy, T., Simpson, P., Williamson, M.P., Hazlewood, G.P., Gilbert, H.J., Orosz, L. (1998) All three surface tryptophans in Type IIa cellulose binding domains play a pivotal role in binding both soluble and insoluble ligands. FEBS Letters, 429, 312-316.

Naumoff, D.G. (2011) Hierarchical classification of glycoside hydrolases. Biochemistry (Moscow), 76, 622-635.

Novy, R., Yaeger, K., Monsma, S., McCormick, M., Berg, J., Shoseyov, O., Shpigel, E., Seigel, D., Goldlust, A., Efroni, G., Singer, Y., Kilburn, D., Tomme, P.,

Gilkes, N. (1997) Cellulose binding domain expression vectors for the rapid, low cost purification of CBD-fusion proteins. FASEB Journal, 11, 1715.

O'Neill, M.A., Albersheim, P., Darvill, A. "The pectic polysaccharides of primary cell walls," in Methods in Plant Biochemistry, D. M. Dey, Ed., pp. 415–441, Academic Press, London, UK, 1990.

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

Passaretti, S., Franzoni, M., Comin, U., Donzelli, R., Rocca, F., Colombo, E., Ferrara, A., Dinelli, M., Prada, A., Curzio, M., et al. (1991) Action of glucomannans on complaints in patients affected with chronic constipation: a multicentric clinical evaluation. The Italian Journal of Gastroenterology, 23, 421-425.

Paulo, A.C., Morgado, J., Andreaus, J., Kilburn, D. (1999) Interactions of cotton with CBD peptides. Enzyme and Microbial Technology, 25, 639-643.

Phelps, M.R., Hobbs, J.B., Kilburn, D.G., Turner, R.F. (1995) An autoclavable glucose biosensor for microbial fermentation monitoring and control. Biotechnology and Bioengineering, 46, 514-524.

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

Popper, Z.A. (2008) Evolution and diversity of green plant cell walls. Current Opinion in Plant Biology, 11, 286-292.

Popper, Z.A., Fry, S.C. (2003) Primary cell wall composition of bryophytes and charophytes. Annals of Botany, 91, 1-12.

Preston, J.F., Hurlbert, J.C., Rice, J.D., Ragunathan, A., St. John, F.J. (2003) Microbial Strategies for the depolymerization of glucuronoxylan: Leads to biotechnological applications of endoxylanases. ACS Symposium Series. American Chemical Society, pp. 191-210.

Ridley, B.L., O'Neill, M.A., Mohnen, D. (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry, 57, 929-967.

Rutenber, E., Robertus, J.D. (1991) Structure of ricin B-chain at 2.5 A resolution. Proteins, 10, 260-269.

Saha, B.C. (2003) Hemicellulose bioconversion. Journal of Industrial Microbiology and Biotechnology, 30, 279-291.

Sainz-Polo, M.A., Valenzuela, S.V., Gonzalez, B., Pastor, F.I., Sanz-Aparicio, J., (2014) Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity. Journal of Biological Chemistry, 289, 31088-31101.

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 and Microbial Technology, 51, 193-199.

Santos, C.R., Polo, C.C., Correa, J.M., Simao Rde, C., Seixas, F.A., Murakami, M.T. (2012) The accessory domain changes the accessibility and molecular topography of catalytic interface in monomeric GH39 beta-xylosidases. Acta Crystallographica. Section D, Biological Crystallography, 68, 1339-1345.

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

Shoham, Y., Lamed, R., Bayer, E.A. (1999) The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends in Microbiology, 7, 275-281.

Shoham, Y., Schwartz, Z., Khasin, A., Gat, Orit., Zosim, Z., Rosenberg, E. (1992) Delignification of wood pulp by a thermostable xylanase from *Bacillus stearothermophilus* strain T-6. (Microorganisms to Combat Pollution), Springer Netherlands, pp. 83-94.

Shoseyov, O., Shani, Z., Levy, I. (2006) Carbohydrate binding modules: biochemical properties and novel applications. Microbiology and Molecular Biology Reviews, 70, 283-295.

Sinnott, M.L. (1990) Catalytic mechanism of enzymic glycosyl transfer. Chemical Reviews, 90, 1171-1202.

Southall, S.M., Simpson, P.J., Gilbert, H.J., Williamson, G., Williamson, M.P. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. FEBS Letters, 447, 58-60.

St John, F.J., Dietrich, D., Crooks, C., Pozharski, E., Gonzalez, J.M., Bales, E., Smith, K., Hurlbert, J.C. (2014) A novel member of glycoside hydrolase family 30 subfamily 8 with altered substrate specificity. Acta Crystallographica, Section D, Biological Crystallography, 70, 2950-2958.

St John, F.J., Godwin, D.K., Preston, J.F., Pozharski, E., Hurlbert, J.C. (2009) Crystallization and crystallographic analysis of *Bacillus subtilis* xylanase C. Acta Crystallographica, Section F, Structural Biology and Crystallization Communications, 65, 499-503.

St John, F.J., Gonzalez, J.M., Pozharski, E. (2010) Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. FEBS Letters, 584, 4435-4441.

Sunna, A., Gibbs, M.D., Bergquist, P.L. (2001) Identification of novel beta-mannan- and beta-glucan-binding modules: evidence for a superfamily of carbohydrate-binding modules. The Biochemical Journal, 356, 791-798.

Suzuki, T., Ibata, K., Hatsu, M., Takamizawa, K., Kawai, K. (1997) Cloning and expression of a 58-kDa xylanase VI gene (xynD) of *Aeromonas caviae* ME-1 in *Escherichia coli* which is not categorized as a family F or family G xylanase. Journal of Fermentation and Bioengineering, 84, 86-89.

Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., Claeyssens, M. (1988) Studies of the cellulolytic system of Trichoderma reesei QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. European Journal of Biochemistry FEBS, 170, 575-581.

Valenzuela, S.V., Diaz, P., Pastor, F.I. (2012) Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. Applied and Environmental Microbiology, 78, 3923-3931.

Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R., Pettersson, G. (1986) Limited proteolysis of the cellobiohydrolase I from Trichoderma reesei: Separation of functional domains. FEBS letters, 204, 223-227.

Verma, A.K., Goyal, A., Freire, F., Bule, P., Venditto, I., Bras, J.L., Santos, H., Cardoso, V., Bonifacio, C., Thompson, A., Romao, M.J., Prates, J.A., Ferreira, L.M., Fontes, C.M., Najmudin, S. (2013) Overexpression, crystallization and preliminary X-ray crystallographic analysis of glucuronoxylan xylanohydrolase (Xyn30A) from *Clostridium thermocellum*. Acta Crystallographica, Section F, Structure Biololgy and Crystallization Communication, 69, 1440-1442.

Vuksan, V., Jenkins, D.J., 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. International Journal of Obesity, 8, 289-293.

Wang, A.A., Mulchandani, A., Chen, W. (2001) Whole-cell immobilization using cell surface-exposed cellulose-binding domain. Biotechnology Progress, 17, 407-411.

Wang, A.A., Mulchandani, A., Chen, W. (2002) Specific adhesion to cellulose and hydrolysis of organophosphate nerve agents by a genetically engineered *Escherichia coli* strain with a surface-expressed cellulose-binding domain and organophosphorus hydrolase. Applied and Environmental Microbiology, 68, 1684-1689.

Wong, K.K., Tan, L.U., Saddler, J.N. (1988) Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. Microbiological Reviews, 52, 305-317.

Yip, V.L., Withers, S.G. (2006) Breakdown of oligosaccharides by the process of elimination. Current Opinion in Chemical Biology, 10, 147-155.

Zaide, G., Shallom, D., Shulami, S., Zolotnitsky, G., Golan, G., Baasov, T., Shoham, G., Shoham, Y. (2001) Biochemical characterization and identification of catalytic residues in alpha-glucuronidase from *Bacillus stearothermophilus* T-6. Europian Journal of Biochemistry, 268, 3006-3016.

Chapter 2

Cloning, expression and purification of glucuronoxylan-xylanohydrolase (*Ct*XynGH30) and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6 from *Clostridium thermocellum*

2.1 Introduction

Cell wall polysaccharides are most abundant component of plant biomass on earth and represent the largest renewable source of organic carbon (Gilbert, 2010). Thus, efficient bioconversion of this biomass into biofuel is a subject of interest for many researchers and also has considerable industrial importance (Dodd and Cann, 2009). Primary plant cell walls are composed of cellulose, hemicellulose and pectin, whereas secondary cell walls contain complex heterogeneous polymers of aromatic alcohols known as lignin (Beg *et al*., 2001). Heteroxylans are a major component of hemicellulose comprising a backbone of xylopyranose units linked by β-1,4 xylosidic bonds. These xylose units are decorated to various degrees by side chains like arabinose, glucuronic, acetic, galactose and their derivatives, like ferulic and *p*-coumaric acids, depending on the source of plant species and tissues (Kulkarni *et. al*., 1999, Shallom *et. al*, 2003). Complete degradation of plant polysaccharides requires the cumulative action of a wide range of microbial hydrolase enzymes. Therefore,

some anaerobic microbes have evolved a highly efficient multi-enzyme complex system called cellulosome, which brings together these enzymes promoting enzyme synergism and stability (Fontes and Gilbert, 2010; Shoham *et al*., 1999). The complex nature of hetero-xylans and crystalline nature of cellulosic polysaccharides required a synergetic action of various plant polysaccharides degrading enzyme to achieve maximum conversion of plant biomass into fermentable sugars. In the figure 2.1.1 showed the site of action of different xylanase enzymes acting on a heteroxylan substrate.

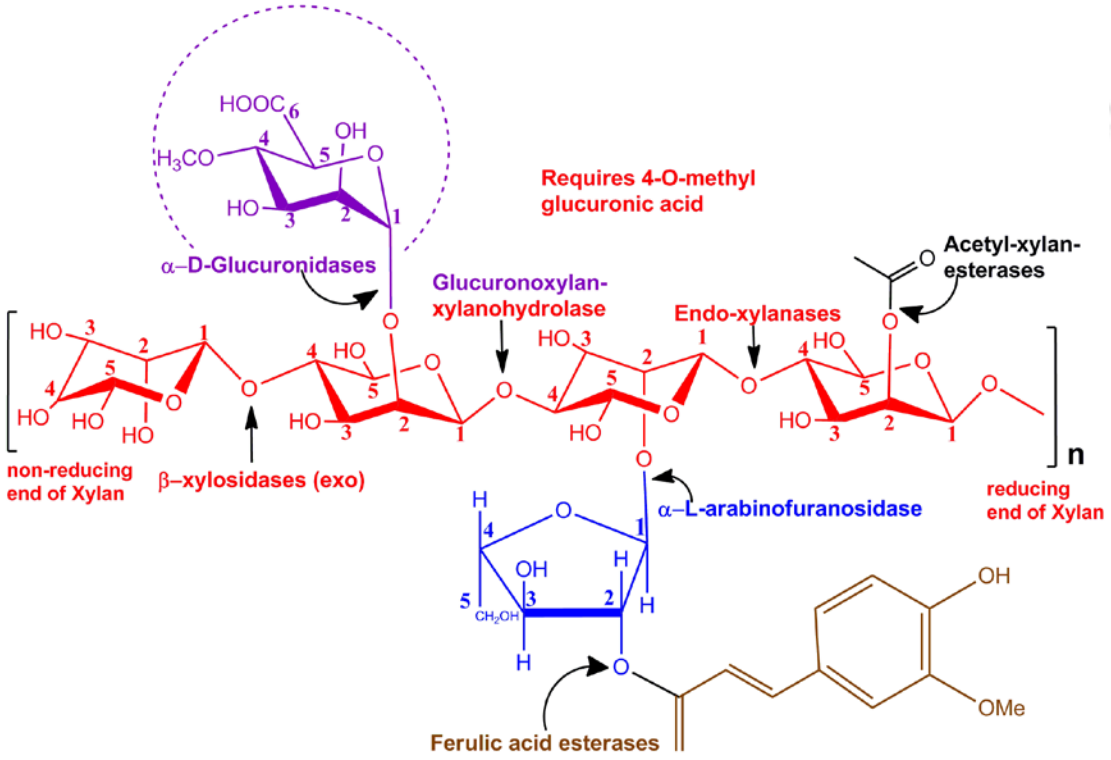


Fig. 2.1.1 Site of action of different hemicellulases responsible for degradation of xylans.

The endo-1-4-β-xylanase (EC 3.2.1.8) acts on a xylan backbone to produce various length xylooligomers whereas, β-xylosidase (EC 3.2.1.37) hydrolyses xylooligomers to xylose monomers. The different substitution of side chains in the

xylan main chain are removed by the action of α-D-glucuronidases (EC 3.2.1.139), α-L-arabinofuranosidases (EC 3.2.1.55) and acetylesterases (EC 3.1.1.6) etc. The main chain hydrolysing xylanase encounters the problem in cleaving the xylan backbone close to these modifying groups such as substitutions of 4-O-methyl glucuronic acid, arabinose, acetic and galactose therefore required a cumulative synchronizing action of different xylanases to efficiently hydrolysed hemicellulosic polysaccharides into fermentable sugars (Biely *et. al*, 1977; Zaide *et. al*, 2001).

The modular carbohydrate-active enzyme having gene accession number ABN54208 and uniprot ID A3DJS9 belonging to glycoside hydrolase family 30 (GH30) from *Clostridium thermocellum* (*Ct*XynGH30) is a cellulosomal protein which plays an important role in plant cell-wall degradation. The gene and protein sequence was retrieved from the CAZy database and analyzed by different computational approach which displayed (as shown in Fig. 2.1.2) protein *Ct*XynGH30 contains an N-terminal signal peptide followed by a catalytic module (*Ct*Xyn30A), a C-terminal carbohydrate-binding module belonging to family 6 (*Ct*CBM6) and a dockerin module.

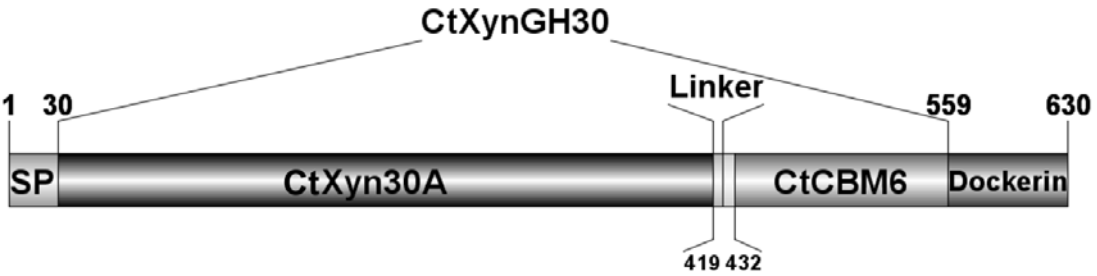


Fig. 2.1.2 Molecular architecture of *Ct*XynGH30 showing boundaries and designation of different domains.

The molecular architecture of modular carbohydrate active enzyme *Ct*XynGH30 showed, 1–30 aa signal peptide, catalytic module *Ct*Xyn30A of 391 aa,

carbohydrate binding module of 127 aa, dockerin type I of 71 aa and a linker of 6 aa connecting the CBM and the catalytic module. In the present study the gene encoding *Ct*XynGH30 and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6 were cloned. All the proteins expressed in *E. coli* and purified by immobilized metal ion affinity chromatography (IMAC) for further biochemical, functional and structural characterization.

2.2 Materials and Methods

2.2.1 Chemicals, reagents and kits

The oligonucleotide primers for PCR amplification of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were procured from Eurofins Genomics Pvt. Ltd. India. *Pfu* DNA polymerase was supplied by Stratagene, USA. The DNA marker, dNTPs and $MgCl_2$ were procured from Bioline, USA. PCR tubes (0.2 ml) used were from Axygen, Germany. Restriction enzymes *Nhe*I and *Xho*I were purchased from Fermentas, Germany. The cloning vector pGEM-T Easy kit was purchased from Promega, USA. The expression vector pET-28a(+), was purchased from Novagen, EMD4 BioScience, Germany. T_4 DNA ligase, 10x ligase buffer were purchased from Promega, USA. RNAse solution (20 mg/ml), X-gal (5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside), glacial acetic acid (99.9 % pure), Trizma base (Tris free base), ethidium bromide, Bradford reagent, DNAse-RNAse free water (pH 8.0) and components of polyacrylamide gel electrophoresis were purchesed from Sigma-Aldrich Pvt. Ltd. USA. The protein staining dye, Coomassie Brilliant Blue R250 was from Himedia, India, whereas, methanol was supplied by Merck, India. The miniprep plasmid isolation kit was purchased from Sigma-Aldrich Pvt. Ltd. India. The gel extraction kit was purchased from Qiagen, Germany. The DNA was detected using SeaKem LE agarose (Cambrex Bio Science, USA). The DNA marker *viz.*Hyperladder I was purchased from Bioline, UK. The Page Ruler protein marker was procured from Fermentas, Germany. Disodium ethylenediamine tetra acetate salt (EDTA), glucose, sodium hydroxide, sodium dodecyl sulphate (SDS), LB medium and SOC medium components were purchased from Himedia, India. The antibiotics, ampicillin was supplied by Himedia, India, whereas, kanamycin was from Sigma-Aldrich Pvt. Ltd.

USA. SDS-PAGE was performed using Mini-PROTEAN Tetra Cell purchased from Bio-Rad Laboratories (India) Private Limited. Coomassie Brilliant Blue G-250 to stain the SDS-PAGE gel was purchased from Amresco LLC, USA. The genomic DNA of *Clostridium thermocellum* ATCC 27405 was procured from DSMZ, Germany.

2.2.2 Microorganisms

E. coli (DH5α) cells were procured from Invitrogen (USA). *E. coli* BL-21 (DE3) cells for expression of recombinant proteins were obtained from Novagen, EMD4 BioScience, Germany.

2.2.3 PCR amplification of gene encoding for *Ct*XynGH30 and truncated derivatives *Ct*Xyn30A and *Ct*CBM6

The oligonucleotide primers were designed and the DNA encoding the full length gene *Ct*XynGH30 (GenBank Accession No: ABN54208 and uniprot ID A3DJS9) and the truncated derivative *Ct*Xyn30A and *Ct*CBM6 were amplified with primers containing *Nhe*I and *Xho*I restriction sites underlined in Table 2.2.1 using *Pfu* DNA polymerase (Stratagene). The 50 µl PCR reaction mixture and PCR cycles for amplification are mentioned in Tables 2.2.2 and 2.2.3, respectively. The PCR amplification was performed using a thermal cycler (Applied Biosystems, GeneAmp PCR System 9700). The amplified PCR fragments were detected on 0.8 (%, w/v) agarose gel as mentioned in Section 2.2.4.

Table 2.2.1 Oligonucleotide primers used for cloning of full-length *Ct*XynGH30 and truncated derivatives *Ct*Xyn30A and *Ct*CBM6 from *Clostridium thermocellum*. The bold and underlined sequence represents the restriction enzyme site.

Construct	Primer sequence
***Ct*XynGH30**	Forward: 5'-CTCT**GCTAGC**GCAACAATCAACTTGTCG-3' Reverse: 5'-CACA**CTCGAG**TTACTCCTGTTTTGACTC-3'
***Ct*Xyn30A**	Forward: 5'-CTCT**GCTAGC**GCAACAATCAACTTGTCG-3' Reverse: 5'-CACA**CTCGAG**TTATACAAAAGTTGTAACGC-3'
***Ct*CBM6**	Forward: 5'-CTCT**GCTAGC**AAGATAGAATGCGAAGAA -3' Reverse: 5'-CACA**CTCGAG**TTACTCCTGTTTTGACTC-3'

Table 2.2.2 PCR mixture for amplification of full-length *Ct*XynGH30 and truncated derivatives *Ct*Xyn30A and *Ct*CBM6 from *Clostridium thermocellum.*

PCR components	Volume (µl)	Final Concentration
10x reaction buffer	5.0	1x
dNTP mix (100 mM)	0.8	1.6 mM
Forward primer (15 µM)	1.5	0.45 µM
Reverse primer (15 µM)	1.5	0.45 µM
Sigma water, (nuclease free) pH 8.0	39.2	--
Genomic DNA (15.4 µg/ml)	0.5	7.7.ng
Pfu DNA polymerase (2.5 U/µl)	0.5	1.25 U
Total	50.0	--

Table 2.2.3 PCR cycles for amplification of full-length *Ct*XynGH30 and truncated derivative *Ct*Xyn30A and *Ct*CBM6 from *Clostridium thermocellum*.

Steps		Time (min)
I. Denaturation at 94°C		4
II. 30 cycles of	i) Denaturation at 94°C	1
	ii) Annealing at 55°C	1
	iii) Extension at 72°C	2 (0.5 in *Ct*CBM6)
III. Final extension at 72°C		10

2.2.4 Agarose gel electrophoresis of PCR amplified and other DNA

The PCR amplified products were run on 0.8% agarose gel prepared in 1x TAE buffer. A stock solution of TAE buffer was prepared by keeping the concentrations of components to 10x (400 mM Tris-acetate, 10 mM EDTA, pH 8.0) according to Sambrook and Russell (2001). The gel was prepared by dissolving agarose (400 mg for 0.8% and 500 mg for 1.0% gel) in 50 ml of 1x TAE buffer by heating in a microwave oven for few min to get a clear solution. Then 5.0 µl of ethidium bromide (5.0 mg/ml) was added when the solution temperature was around 50°C. The solution was mixed and poured on the casting apparatus, combs were placed and the gel was allowed to set for 30 min. For the separation of PCR amplified DNA, 1x TAE (Tris-acetate-EDTA) buffer was used for preparation of agarose gels and as electrophoresis running buffer (Sambrook and Russel, 2001). The DNA sample and DNA loading dye were mixed in 4:1 ratio and the gel was run at 60 Volts for 1h. The bands were then visualized under UV illumination in a gel documentation system (Kodak, Gel Logic 1500).

2.2.4.1 DNA loading dye

The DNA or sample loading dye was prepared by mixing the components mentioned below in Table 2.2.4. A 5x stock solution of DNA loading dye was prepared and was mixed with 4 volumes of DNA to make it to 1x before loading on to agarose gel. The final pH of the DNA loading dye adjusted to pH 8.0.

Table 2.2.4 Composition of 5x DNA loading dye.

Components	Final concentration (5x)
Tris-HCl	50 mM
Glycerol	25% (w/v)
EDTA	5.0 mM
Bromophenol blue	0.2% (w/v)
Xylene cyanol	0.2% (w/v)

2.2.5 Extraction of DNA from agarose gel

The PCR amplified DNA or other plasmid DNA was purified from agarose gel using a kit (Qiagen, QIAquick Gel Extraction Kit), following the protocol provided by the manufacturer as discussed in Section 2.2.6.1. The extracted DNAs were eluted in 50 µl DNAse free water (Sigma-Aldrich Pvt. Ltd. USA).

2.2.5.1 DNA gel extraction protocol

1. 1.5 ml sterile, empty microcentrifuge tubes were weighed and the empty weight was noted.
2. The PCR or plasmid DNAs were excised from gel using sharp sterile scalpel and transferred to empty microcentrifuge tubes. The tubes were weighed again and the weight of excised gel was determined by subtracting the empty tube weight (noted above).
3. Now, 3 volumes of buffer QG were added to every 1 volume of gel (100 mg ~ 100 µl).
4. The microcentrifuge tubes containing excised gel were incubated at 50°C for 10 min (or until the gel slice has completely dissolved). When the gel slice dissolved completely, the colour of the solution became yellow (similar to Buffer QG without dissolved agarose).

5. 1 gel volume of isopropanol was added to above yellow colour solution containing of PCR amplified DNAs or recombinant plasmids (For higher yield of DNA fragments <500 bp and >4 kb).
6. QIAquick spin column (DNA binding column) was placed on 2 ml collection tube provided with the kit. The above solution containing PCR-amplified or plasmid DNAs (750 µl) were added to DNA binding columns and centrifuged at 17,900***g*** for 1 min at 25°C, and the flow through was discarded. If the volume was more than 750 µl, the remaining solution was centrifuged similarly and again the flow through was discarded.
7. 500 µl of buffer QG was added to each QIAquick spin column and the mixtures were centrifuged again at 17900***g*** for 1 min at 25°C, and the flow through was discarded.
8. Now, 750 µl of buffer PE was added to each column containing PCR DNA or recombinant plasmid DNAs and the mixture was centrifuged at 17,900***g*** for 1 min at 25°C. The flow through was discarded and the column was given an additional spin of 1 min at 17900***g***, to completely remove the residual ethanol.
9. Now the column containing bound DNA was placed on a fresh 1.5 ml sterile microcentrifuge tube. 30 µl of DNAse free water (Sigma-Aldrich Pvt. Ltd.) or elution buffer (10 mM Tris-Cl, pH 8.5) was added at the centre of the column. The column was incubated for 2 min at 25°C and centrifuged at 17900***g*** for 1 min.
10. The PCR-amplified or plasmid DNAs were eluted from QIAquick spin columns in 1.5 ml sterile microcentrifuge tube. The PCR-amplified or plasmid DNAs were stored at -20°C for further use.

2.2.6 Cloning of PCR Products into a pGEM-T Easy Vector (TA cloning)

The gel eluted amplified PCR products were initially cloned into pGEM-T Easy vector. The vector contains 3´-thymidine at both ends which provides a compatible overhang for PCR amplicons generated by certain thermo-stable DNA polymerases. These PCR amplified products were directly used for TA cloning. The size of pGEM-T Easy vector is 3015 bp (Fig. 2.2.1). It also allows easy selection Blue/White colonies of recombinant clones. Furthermore, it is a high-copy-number vector containing T7 and SP6 RNA polymerase promoters. It also provides multiple options to incorporate insert as it contains different restriction sites, present in the multiple cloning site (MCS).

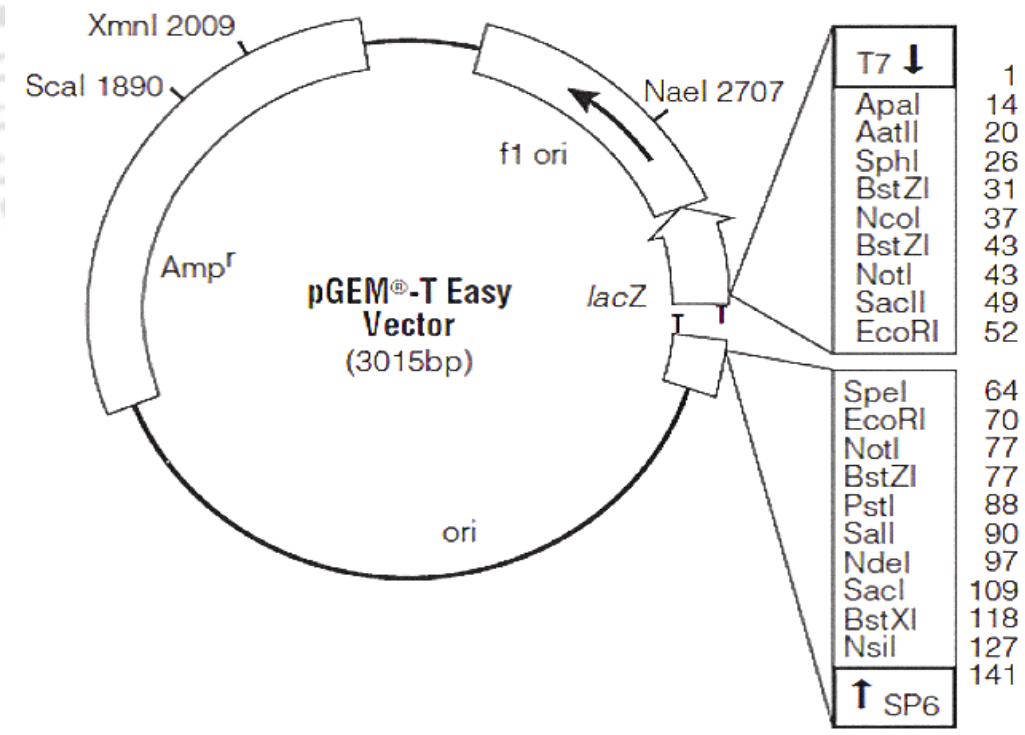


Fig. 2.2.1 pGEM-T Easy vector map showing T7 RNA polymerase transcription initiation site (1), multiple cloning region (10-113), pUC/M13 reverse sequencing primer binding site (161-177), *lac*Z start codon (165), *lac* operator (185-201), β-lactamase coding region (1322-2182), phage f1 region (2365-2820), *lac* operon sequences (2821-2981, 151-380), pUC/M13 forward sequencing primer binding site (2941-2957), T7 RNA polymerase promoter (2984-3). Amp^r represent ampicillin resistance gene.

2.2.7 Restriction enzyme digestion of the pGEM-T Easy clone

The cloned pGEM-T Easy plasmid DNA was isolated by following the protocol mentioned in the Section 2.2.14. The recombinant DNAs were digested with restriction enzymes (RE) *Nhe*I and *Xho*I as per the following set up given in Table 2.2.5. The reaction mixtures were incubated at 37°C in a water bath (Grant, SUB36) for 90 min. The digested DNAs were run on 1.0% agarose gel and the desired fragments were purified using gel extraction kit as mentioned in Section 2.2.6 and eluted in 30 µl of sterile DNAse free water pH 8.0.

Table 2.2.5 Restriction enzyme digestion set up of TA cloned DNAs of *Ct*XynGH30 and truncated derivatives *Ct*Xyn30A and *Ct*CBM6.

RE digestion set up	1x (µl)
10x buffer	6.0
TA cloned DNAs	20 (1-3 µg)
DNAse free water	23
*Nhe*I (10 U/µl)	0.5
*Xho*I (10 U/µl)	0.5
Total	50

2.2.8 Generation of pET-28a(+) expression vector for cloning of genes encoding *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6

The pET vectors, in recent times, have become the most commonly used expression vector systems for the cloning and expression of recombinant proteins in *E. coli*. It is based on the T7 promoter-driven system originally developed by Studier and colleagues (Studier and Moffatt, 1986; Studier *et al.*, 1990). It is also possible to control expression levels simply by manipulating the concentration of the inducer. The desired genes to be cloned in pET vectors are under the influence of bacteriophage T7 transcription and the expression is induced by T7 RNA polymerase in the host cell. An additional advantage with pET system is its ability to maintain target genes transcriptionally silent in the uninduced state. The His-Tag is useful for

easy purification in a single step by making use of affinity chromatography. The pET-28a(+) vector, employed for cloning of all the three modules, had an N-terminal T7-Tag sequence configuration in addition to a C-terminal His-Tag. The cloning and expression region of the coding strand transcribed by T7 RNA polymerase is shown in vector map (Fig. 2.2.2). The thrombin sequence tag can be used to remove the His-Tag, in case if activity of protein without the tag is needed to be checked. The location of His-Tag, T7 coding sequence, T7 terminator, kanamycin coding region and f1 origin are depicted in the figure 2.2.2.

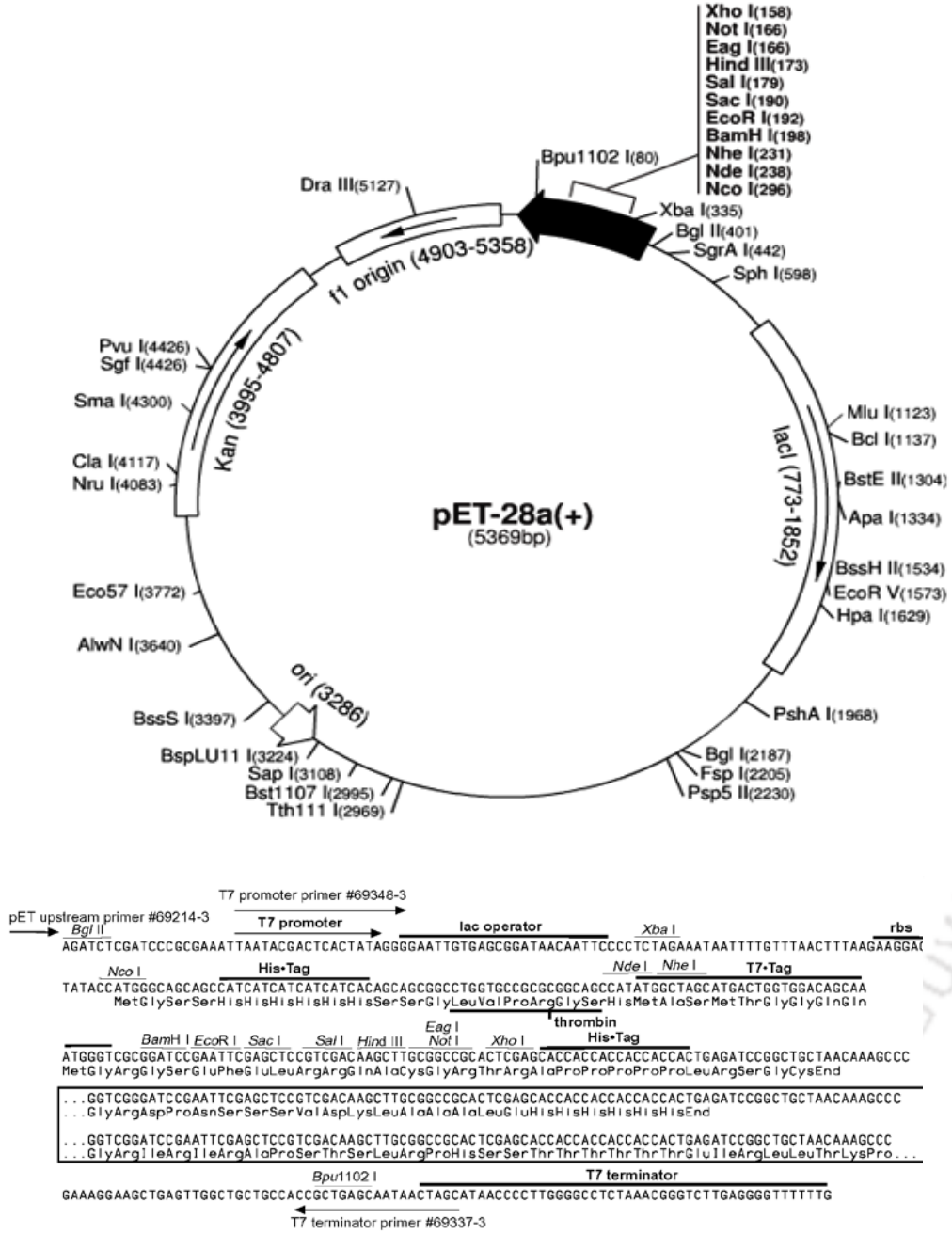


Fig. 2.2.2 Restriction map of the pET-28a(+) expression vector showing multiple cloning site (158-203 bp), restriction enzyme sites, N-terminal His•Tag coding sequence (270-287 bp), C-terminal His•Tag coding sequence (140-157 bp), T7 promoter (370-386), T7 terminator (26-72 bp), pBR322 origin (3286 bp), kanamycin marker (3995-4807 bp) and a f1 origin (4903-5358). *Nhe*I cuts at 229, *Bgl*II at 401 and *Xho*I at 158.

The pET-28a(+) vector was digested with *Nhe*I-*Xho*I to prepare it for cloning of inserts *CtXynGH30*, *CtXyn30A* and *CtCBM6* obtained from TA clone after

digestion with *Nhe*I-*Xho*I. A stock solution (100 ng/µl) was prepared from the supplied stock of 10 µg of pET-28a(+) vector using DNAse free water (pH 8.0). The restriction enzyme digestion of pET-28a(+) vector was then carried out as mentioned in Table 2.2.6. The digestion mixtures were incubated at 37°C in a water bath for 90 min. The *Nhe*I-*Xho*I digested pET vector was purified from agarose gel as described in Section 2.2.6 and resuspended in the 30 µl nuclease free water.

Table 2.2.6 Restriction enzyme digestion set up of pET-28a(+).

RE digestion set up	1x (µl)
10x buffer	2.0
vector (pET-28a(+) 100 ng/µl	10.0
DNAse free water (Sigma-Aldrich)	7.0
*Nhe*I (10 U/µl)	0.5
*Xho*I (10 U/µl)	0.5
Total	20.0

2.2.9 Ligation of inserts released from recombinant pGEM-T Easy clones to pET-28a(+) vector

The *Nhe*I-*Xho*I digested TA cloned DNA as mentioned in Section 2.2.7 were cloned to pET28a(+) vector, which was also digested with same restriction enzymes as described in Section 2.2.8. The concentration of *Nhe*I-*Xho*I digested pET-28a(+) vector, after gel extraction was adjusted to 100 ng/µl using DNAse free Sigma water. The insert: vector molar ratio was kept at 3:1 for all the fragments. Now, the amount of *Nhe*I-*Xho*I digested insert DNA of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 required for cloning were calculated as using the following formula (Engler and Richardson, 1982),

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert :vector molar ratio} = \text{ng of insert}$$

$$\frac{100 \text{ ng} \times 1.578 \text{ kb}}{5.36 \text{ kb}} \times \frac{3}{1} = 88.32 \text{ ng of insert (}\textit{CtXynGH30}\text{)}$$

$$\frac{100 \text{ ng} \times 1.158 \text{ kb}}{5.36 \text{ kb}} \times \frac{3}{1} = 64.81 \text{ ng of insert (}\textit{CtXyn30A}\text{)}$$

$$\frac{100 \text{ ng} \times 0.369 \text{ kb}}{5.36 \text{ kb}} \times \frac{3}{1} = 20.65 \text{ ng of insert (}\textit{CtCBM6}\text{)}$$

The following set up was used for ligation of digested fragments into pET-28a(+) vector as shown below in Table 2.2.7

Table 2.2.7 Ligation set up for restriction enzyme digestion of pET-28a (+) vector.

Ligation set up	1x (µl)
10x buffer	3.0
pET-28a(+) vector (100 ng)	1.0
RE digested DNA insert (calculated accordingly in ng)	--
T_4 DNA ligase (3 U/µl)	1.5
Total	30.0

The ligation mixtures were then incubated overnight at 16°C in a cooling water bath (GE Healthcare, MultiTemp III).

2.2.10 Preparation of *E. coli* (DH5α) competent cells

Day 1

1. 50 µl of culture of *E. coli* (DH5α) from glycerol stock was inoculated into 5.0 ml LB medium (Sambrook *et al.*, 1989), as shown in Section 2.2.11, in test tube grown at 37ºC and 180 rpm for overnight.
2. 0.1 M $CaCl_2$ solution was prepared and filter-sterilized by passing through 0.22 µm filter in laminar air flow and kept in refrigerator.

Day 2

3. 1.0 ml of overnight grown culture from above was taken and inoculated into 100 ml LB medium contained in a 250 ml conical flask. The flask was incubated at 37ºC and 180 rpm till cell absorbance at 550 nm reached 0.5-0.6.
4. Micro-centrifuge tubes, 50 ml centrifuge tubes (round bottom) and micro tips were autoclaved and kept on ice bath which was placed under laminar air flow hood.
5. 40 ml from 100 ml grown *E. coli* culture was transferred aseptically to round bottom centrifuge tubes (the centrifuge tubes were weighed and balanced before placing on cooling centrifuge).
6. The tubes were centrifuged at 4ºC with 4000***g*** for 10 min.
7. Previous step was repeated for making pellet from the remaining culture.
8. The supernatant was discarded and the cell pellet was resuspended gently in 3-4 ml of sterile, ice-chilled 0.1 M $CaCl_2$ solution, and then the remaining solution was added to 20 ml final volume. The re-suspended cells in centrifuge tubes were kept on ice for 10 min.
9. The tubes were centrifuged again at 4000***g*** at 4ºC for 10 min.
10. The supernatant was carefully removed and the pellet was resuspended gently in 3.0 ml of sterile ice chilled 0.1 M $CaCl_2$ solution.
11. 200 µl of competent cells were aliquoted into each 1.5 ml microcentrifuge tube containing 10% (v/v) glycerol (final concentration) and kept at -80ºC for further use.
12. The cells prepared could be used only once after taking out from ultra-freezer (-80ºC).

2.2.11 Preparation of Luria-Bertani (LB) medium

The LB medium was prepared by using the components mentioned in Table 2.2.8. The contents were dissolved in 800 ml of deionized water, the pH was adjusted to 7.0 using NaOH (2N, Normality) and the final volume was made up to 1 litre. 100 ml of LB medium was then transferred to each of 250 ml conical flask and the flasks were autoclaved at 121°C at 15 psi for 20 min. It is the most commonly used medium for the growth of recombinant *E. coli* cells. The filter sterilized antibiotics ampicillin (100 µg/ml) or kanamycin (50 µg/ml) were added to autoclaved and cooled LB medium prior to inoculation.

Table 2.2.8 Composition of Luria-Bertani medium (Sambrook *et al*., 1989).

Components	Final concentration (%, w/v)
Tryptone	1.0
Yeast extract powder	0.5
Sodium chloride	1.0

2.2.11.1 Preparation of LB-agar medium

LB agar medium was prepared by dissolving the components mentioned in Table 2.2.8 and in addition agar-agar type 1 was also added to a final concentration of 2% (w/v). The medium was autoclaved as described in Section 2.2.11. The medium was cooled to around 50-55°C and then antibiotic was added in appropriate amount under laminar air flow and mixed well. Immediately around 25 ml of medium supplemented with antibiotics was poured in sterile petri-plates and the medium was allowed to solidify for 15- 20 min.

2.2.12 Preparation of SOC medium

The SOC (super optimal medium with catabolic repression) was prepared using components as described in Table 2.2.9. It is a modified SOB medium with addition of glucose (Hanahan, 1983; Deutscher, 2008). Bactotryptone, yeast extract powder and NaCl were autoclaved separately. 1 M stock solutions of KCl, $MgCl_2$, $MgSO_4$ and glucose were filter-sterilized and the required volume of each was added to above solution in the laminar hood to finally make SOC medium as mentioned in Table 2.2.9.

Table 2.2.9 Composition of SOC medium (Sambrook*et al*., 1989).

Components	Final concentration
Bactotryptone	2.0 (%, w/v)
Yeast extract powder	0.5 (%, w/v)
NaCl	10 mM
KCl	2.5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
Glucose	20 mM

2.2.13 Transformation of ligated DNA using *E. coli* (DH5α) cells

The *E. coli* (DH5α) competent cells were transformed with recombinant plasmid DNAs of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6. The step-wise transformation protocol is described below:

1. The microcentrifuge tube containing competent cell (200 µl) was taken out from -80°C and kept on ice for 5 min, then 10 µl of ligation mixture was added to it (the mixture was given a very short spin prior to use).
2. The tube was gently tapped 4-5 times and kept on ice for 30 min.
3. The cells were given a heat shock at 42°C for 40s.

4. The cells were immediately transferred back to ice for 2-3 min.
5. 1.0 ml of super optimal medium with catabolite repression (SOC) (previously incubated at 37°C) (Hanahan, 1983; Sambrook *et al*., 1989; described in Section 2.2.12) was added to the transformed cells.
6. The transformed cells were incubated at 37°C in a shaking incubator at 200 rpm for 1h.
7. The cells were harvested by centrifugation at 2000***g*** at 25°C for 3 min.
8. 1.0 ml supernatant was discarded and the cell pellet was then resuspended in remaining 200 µl supernatant.
9. The cells were spread on LB agar plates (as described in Section 2.2.11.1) supplemented with 50 µg/ml kanamycin.
10. The LB agar plates were incubated at 37°C and the cells were grown overnight.
11. The transformation efficiency was calculated by using the following formula,

$$\text{Transformation efficiency} = \frac{\text{No. of colonies on LB agar plate} \times 1000}{\text{ng of insert DNA}} = \text{cfu/µg}$$

The colonies were picked at random from the LB agar plates containing transformed *E. coli* (DH5α) cells and inoculated in 5 ml LB medium supplemented with kanamycin, and grown at 37ºC and at 180 rpm for 12h. The plasmid DNA was isolated to check for positive clones.

2.2.14 Isolation of recombinant plasmid DNA

Recombinant plasmid DNAs of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were isolated from the transformed *E. coli* (DH5α) cells using alkaline lysis method (Sambrook *et al.*, 1989) and eluted in 30 µl of sterile DNAse free water pH 8.0

(Sigma-Aldrich Pvt. Ltd. USA) with RNAseA (20 µg/ml). The plasmid DNA was isolated following a protocol as described in Section 2.2.14.1.

2.2.14.1 Plasmid miniprep (alkaline lysis) protocol

1. The colonies were randomly picked from the LB agar plates and each colony was inoculated into 5.0 ml of LB medium supplemented with appropriate antibiotics and incubated at 37°C with shaking at 180 rpm for 12h.
2. 1.5 ml culture was taken in micro-centrifuge tubes and centrifuged at 16,000***g*** for 1 min. The supernatant was removed and the above step was repeated with fresh 1.5 ml culture so as to process a total of 3.0 ml for plasmid isolation and rest was kept for glycerol stock preparation.
3. The cell pellet was resuspended in 200 µl of GTE solution (50 mM Glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 8.0) containing RNAse at a final concentration of 20 µg/ml.
4. Then 200 µl of lysis solution (0.2 M NaOH, 1% SDS) was added to it and the tubes were gently inverted 5-6 times.
5. 300 µl of ice chilled 5 M potassium acetate solution (pH 4.8) was added immediately and the tubes were gently inverted again 5-6 times. This solution neutralizes NaOH added in the previous lysis step, while precipitating the genomic DNA and SDS in to an insoluble white, rubbery precipitate. The tubes were kept on ice for 3-5 min.
6. The mixture was then centrifuged at 16,000***g*** and at 4°C for 5 min.
7. The supernatant was transferred to new tube (2.0 ml micro-centrifuge tube), while taking care not to pick up any white precipitate.

8. The plasmid DNA was extracted with 1 volume of PCI (phenol/chloroform/isoamylalcohol; 25:24:1) removing the protein contaminations. The mixture was mixed by a vortex for 30s and then centrifuged at 16000***g*** and at 4°C for 10 min. The organic PCI layer stayed at the bottom of tube after the spin.
9. The upper aqueous layer containing the plasmid DNA was carefully transferred to fresh microcentrifuge tube (2.0 ml), avoiding the white precipitated protein layer above the organic PCI layer.
10. The plasmid DNA was precipitated by adding 2 volumes of absolute ethanol at 25°C. The solution was mixed well and kept again at 25°C for 2 min. Thereafter, the plasmid DNA was centrifuged at 16,000***g*** and 4°C for 5 min.
11. Ethanol was removed by micropipette. The plasmid DNA was washed with 700 µl of 70% (v/v) ethanol and DNA pellet was made by spinning at 16,000***g*** and 4°C for 5 min. Ethanol was removed again by micropipette and the microcentrifuge tube containing plasmid DNA was air dried for 5 min.
12. The plasmid DNA was resuspended in 30 µl 10 mM Tris buffer, pH 8.0 or DNAse free Sigma water (pH 8.0). 10 µl of RNAseA solution was added (20 µg/ml final concentration) and the plasmid DNA was mixed well and stored at -20°C.

2.2.15 Screening of recombinant plasmid DNA for identification of positive clone

10 µl of recombinant plasmid DNA of all the derivatives *viz.* *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6, isolated by alkaline lysis method as described in Section 2.2.14, was taken in a fresh sterile 1.5 ml microcentrifuge tube for restriction enzyme digestion analysis. The recombinant DNA of each of the above mentioned derivatives was digested with restriction enzymes, *NheI* and *Xho*I, to check for positive clones

following the reaction mixture set up given in Table 2.2.10. The reaction mixture for each recombinant derivative was incubated at 37ºC in a water bath (Grant, SUB36) for 90 min. The digested products were run on 1% agarose gel as described in Section 2.2.5. The digested fragments of pET-28a(+) vector and the respective DNAs insert were visualized by placing the gel under UV trans-illuminator. The digested fragments (insert and vector) were analyzed to check whether the size of insert DNA and linearized vector, pET-28a(+), were same as determined earlier in Sections 2.2.7 and 2.2.8. Based on this observation, the positive clones for the respective recombinant derivatives were identified. Positive clones were also confirmed by DNA sequencing of the recombinant plasmids of each of the derivative. The plasmid DNAs were sent for sequencing (Xcelris Labs, Ahmedabad, India) to confirm that the recombinant plasmid harboured the desired gene. Glycerol stocks of *E. coli* (DH5α) cells harbouring the recombinant plasmids of each of the derivative were prepared by keeping final concentration of glycerol to 20-25% (v/v). The glycerol stocks were stored at -80°C. The positive clones for each of the recombinant derivatives *viz.Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were identified and were preserved.

Table 2.2.10 Restriction enzyme digestion set up of recombinant plasmid DNA.

Digestion set up	**1x (µl)**
10x buffer	2.0
DNAse free water (Sigma-Aldrich)	7.0
Recombinant plasmid DNA (approx. 200 ng)	10.0
*Nhe*I (10 U/µl)	0.5
*Xho*I (10 U/µl)	0.5
Total	20.0

2.2.16 Preparation of *E. coli* (BL-21) competent cells

The *E. coli* BL-21 (DE3) competent cell was prepared by Calcium chloride method following the protocol as discussed in Section 2.2.10. Finally, 10% (v/v) glycerol (final concentration) was added to competent cells and 200 µl aliquots were made in sterile 1.5 ml microcentrifuge tubes and stored at -80°C for further use.

2.2.17 Isolation of recombinant plasmid DNA

E. coli (DH5α) cells containing the recombinant derivatives *viz. Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were taken from the glycerol stocks and inoculated in 5 ml LB medium containing appropriate antibiotics. The culture was grown at 37°C with shaking at 180 rpm overnight. The plasmid DNA of each recombinant derivative were isolated on next day using plasmid miniprep kit (Sigma-Aldrich Pvt. Ltd. USA) following manufacturer's protocol as described in Section 2.2.17.1.

2.2.17.1 Plasmid isolation protocol (Sigma-Aldrich)

1. 1.5 ml from each of the grown *E. coli* (DH5α) culture was taken and was transferred to 1.5 ml microcentrifuge tube aseptically.
2. The cells were then centrifuged at 13000***g*** for 1 min and the process was repeated with another 1.5 ml of grown culture.
3. The resulting cell pellet of each recombinant derivative was resuspended in 200 µl resuspension solution by using a vortex. RNaseA in final concentration of 20 µg/ml was added to the re-suspension solution prior to use.
4. 200 µl of lysis solution was added to each tube and the tubes were gently inverted 5-6 times. The tube was allowed to stand for 2-5 min.
5. 350 µl of neutralization solution was added to the mixture and the tubes were inverted again for 4–6 times to mix properly.

6. The mixture was then centrifuged at 16,000***g*** for 10 min.
7. The DNA binding columns were prepared or activated by adding 500 µl of column preparation solution to binding column and centrifuging at 13,000***g*** for 1 min. The flow through accumulated in collection tube was discarded.
8. The clear lysate was then transferred to activate DNA binding column and centrifuged at 13,000***g*** for 1 min and the flow through in the collection tube was discarded.
9. The plasmid DNAs bound to column was washed with wash solution and spun at 13,000***g*** for 1 min. The flow through was discarded and the column was given another 1 min spin at 13,000***g*** for removing the wash solution completely.
10. The DNA binding column was transferred to a fresh sterile 1.5 ml microcentrifuge tube and 30 µl of DNAse free water was added at the centre of binding column. The microcentrifuge tube was allowed to stand for 2 min at room temperature and then plasmid DNA was eluted by centrifugation at 13,000***g*** for 1 min. The plasmid DNA was collected in the sterile microcentrifuge tube.
11. The eluted plasmid DNA in sterile microcentrifuge tube was stored at -20°C.

2.2.18 Transformation of plasmid DNA using *E. coli* BL-21 cells

10 µl of each of the recombinant plasmid DNA isolated by above mentioned method was used for transformation of *E. coli* BL-21 (DE3) competent cells for expression of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 proteins. The transformation protocol was similar to as described in Section 2.2.13. The transformed cells were spread on LB agar plates having 50 µg/ml kanamycin and incubated at 37ºC under static condition for 12h. The colonies were randomly picked from LB agar plates containing cells and inoculated further in 5.0 ml LB medium supplemented with 50

µg/ml kanamycin as mentioned in Section 2.2.13 and incubated at 37ºC in a shaking incubator at 180 rpm for analysis of protein expression.

2.2.19 Overexpression of recombinant proteins

E. coli BL-21(DE3) cells were transformed with recombinant plasmids (10 µl) of *CtXynGH30, CtXyn30A* and *CtCBM6* as described in Section 2.2.18. The colonies grown on LB agar plates supplemented with kanamycin 50 µg/ml were picked and inoculated in 5 ml LB medium as stated in Section 2.2.13. The cells were incubated at 37ºC with 180 rpm to grow up to mid-exponential phase till cell absorbance at 550 nm (A_{550}) reached ≈0.6 (Taylor *et al*., 2005). 1.5 ml of this grown culture containing uninduced cells were taken out for making glycerol stocks (after expression analysis) and for the sample preparation for analysis by SDS-PAGE. The remaining 3.5 ml culture was then induced with isopropyl-1-thio-β-D-galactopyranoside (IPTG) at 1.0 mM final concentration for overexpression of recombinant proteins and further incubated at 24ºC at 180 rpm for 24h (Ichinose *et al.*, 2008). The expression of recombinant proteins were checked by loading respective uninduced and induced cell samples on polyacrylamide gel (SDS-PAGE) as described in Section 2.2.20. The *E. coli* BL-21 (DE3) harbouring the genes encoding *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were preserved as glycerol stocks by keeping the final glycerol concentration to 20-25% (v/v) as described in Section 2.2.15 and finally stored at -80°C in ultra-deep freezer.

2.2.20 Sodium dodecyl sulphate Polyacrylamide gel electrophoresis (SDS-PAGE) analysis of recombinant proteins

The recombinant proteins were run on SDS-PAGE gel for analysis of their respective molecular size. SDS-PAGE is an analytical method used to separate components of a protein mixture based on their size (Laemmli, 1970). The PAGE makes use of the fact that a charged molecule migrates in an electric field in the direction of an electrode with opposite charge. But this method cannot be used to determine the molecular weight of proteins because the mobility of a substance in the gel depends on both charge and size. Therefore, the proteins are treated with SDS so that they acquire uniform charge, then the electrophoretic mobility depends primarily on size. The proteins being covered by SDS are negatively charged and when loaded onto a gel and placed in an electric field, they migrate towards the anode and are separated based on size. The migrated proteins were then visualized using a staining solution containing Coomassie brilliant blue R-250 dye which binds with proteins.

2.2.20.1 Preparation of SDS-PAGE gel

The polyacrylamide gels are prepared by copolymerization of acrylamide and bis-acrylamide (“bis,” N,N'-methylene-bisacrylamide). The copolymerization reaction is a vinyl addition reaction initiated by a free radical-generator *viz.* Ammonium per sulphate (APS) in presence of N,N,N',N'-tetramethylethane-1,2-diamine (TEMED) which acts as a catalyst (Chrambach, 1985). The main components of a SDS-PAGE gel are acrylamide 30% (w/v), resolving gel (Tris-HCl, pH 8.8), a stacking gel (Tris-HCl, pH 6.8), SDS 10% (w/v), APS 10% (w/v), TEMED, sample loading buffer (pH 6.8) and electrophoresis running or tank buffer (pH 8.3-8.5). The composition of each

component of SDS-PAGE gels and buffers are described below in Sections 2.2.20.2 to 2.2.20.5.

2.2.20.2 Preparation of acrylamide 30% (w/v) solution

0.8 g of bis-acrylamide was weighed and transferred into an amber colour bottle and dissolved in 50 ml of ultra-pure deionized water collected at 18 MΩcm (Millipore, Milli-Q water purification system) on a magnetic stirrer (IKA, C-MAG HS7). After completely dissolving bis-acrylamide, 29.2 g of acrylamide was added to it and stirred on a magnetic stirrer till the solution was clear. The final volume was adjusted to 100 ml with ultra-pure water as mentioned above by keeping the measuring cylinder (100 ml) wrapped with aluminium foil as acrylamide is light sensitive. The acrylamide solution was then filtered (Whatman No. 1) under dark condition and stored at 4°C.

2.2.20.3 Polymerization of SDS-PAGE gel

The resolving gel and stacking gels were prepared following protocols from Sambrook *et al*. (1989) using the composition as described in Tables 2.2.11 and 2.2.12 given below. The resolving gels were prepared by adding all the components in the order as mentioned in Table 2.2.11, in a 25 ml beaker, by keeping acrylamide concentration to 13% (w/v) for *Ct*XynGH30 and *Ct*Xyn30A and 14% (w/v) for *Ct*CBM6. Similarly, the stacking gel prepared by dissolving all the components mentioned in Table 2.2.12. The acrylamide concentration in the stacking gel was kept at 4% (w/v).

Table 2.2.11 Composition of SDS-PAGE components for preparation of resolving gel.

Components	13% gel volume (ml)	14% gel volume (ml)
Acrylamide solution (30%,w/v)	4.33	4.67
Deionized water	0.36	0.03
SDS (10%,w/v)	1.0	1.0
Glycerol (50%,v/v)	1.0	1.0
1.5 M Tris-HCl (pH 8.8)	3.3	3.3
APS (10%,w/v)	0.1	0.1
TEMED	0.01	0.01

Table 2.2.12 Composition of SDS-PAGE components for preparation of stacking gel.

Components	4% gel volume (ml)
Acrylamide solution (30%,w/v)	0.7
Deionized water	2.8
SDS (10%,w/v)	0.5
0.5 M Tris-HCl (pH 6.8)	1.0
APS (10%,w/v)	0.05
TEMED	0.005

2.2.20.4 Preparation of SDS-PAGE running buffer

The SDS-PAGE gels were run using a 1x running or tank buffer prepared from the 5x stock solution as described in Table 2.2.13. 15.14 g of Tris free base and 94 g of glycine were dissolved in 800 ml of deionized water. To this 50 ml of 10% (w/v) SDS was added and the final volume was adjusted to 1 litre. The final pH of the buffer was adjusted to 8.3. The 5x buffer was filtered (Whatman, Filter No. 1) and stored at 4°C.

Table 2.2.13 Composition of 5x Tris-Glycine, running or tank buffer.

Components	Final concentration (5x buffer)
Tris base	0.125 M
Glycine	1.25 M
SDS	0.5 % (w/v)

2.2.20.5 Preparation of sample buffer

5x sample loading buffer was prepared by dissolving the components while keeping the concentration of components as described in Table 2.2.14 and the pH of the buffer was adjusted to 6.8. The components were dissolved in the order as mentioned in Table 2.2.14 to make 5x sample buffer. However, the final concentration while loading to a SDS-PAGE gel was kept to 1x by mixing 4 volumes of sample (protein) with 1 volume of 5x sample buffer.

Table 2.2.14 Composition of 5x sample loading buffer (Laemmli, 1970).

Components	Final concentration (5x buffer)
Tris-HCl (pH 6.8)	62.5 mM
Glycerol	20.0 (%,v/v)
SDS	2.0 (%,w/v)
Bromophenol blue	0.025 (%, w/v)
β-mercaptoethanol	5.0 (%, w/v)

2.2.20.6 Preparation of staining and destaining solutions

The proteins on the SDS-PAGE gel were visualized using a staining solution that contained Coomassie Brilliant Blue (CBB) R-250 dye, which is disulfonated triphenyl methane. The CBB R-250 dye (detection range of 100-1000 ng of protein) forms a non-covalent complex with proteins, involving van der Waals forces and electrostatic interactions (Neuhoff *et al.*, 1985). The negatively charged anionic form of the dye is stabilized by formation of a blue colour protein-dye complex which may then be seen on gel (Meyer and Lambert, 1965). The staining solution (100 ml) was prepared by dissolving 250 mg or 0.25% (w/v), of CBB R-250 dye in 50 ml of deionized water in an amber colour bottle by keeping on a magnetic stirrer for overnight. The solution was filtered (Whatman, Filter No. 1) and then 40 ml of

methanol and 10 ml of glacial acetic acid were added to finally make the ratio 5:4:1 (deionized water: methanol: glacial acetic acid). The destaining solution was prepared dissolving deionized water: methanol: glacial acetic in 5:4:1 ratio. The gels were destained by immersing the gel in destaining solution under gentle shaking condition with change of destaining solution every 15 min, until the background stain is removed or protein bands were clear.

2.2.21 Purification of recombinant proteins

The *E. coli* BL-21(DE3) harbouring recombinant plasmids were grown in 100 ml LB medium in 250 ml flask supplemented with 50 µg/ml kanamycin as mentioned in Section 2.2.10.2. The recombinant proteins contained a His_6 tag at N-terminal in pET-28a(+) vector. A single step purification method based on immobilized metal-ion affinity chromatography (IMAC) for His_6-tag containing proteins was employed using 1.0 ml sepharose columns (GE Healthcare, HiTrap chelating HP). The compositions of binding as well as elution buffers used for affinity column purification of proteins are mentioned Table 2.2.15. All the buffers (binding, elution and cleaning buffer), 0.1 M $NiSO_4$ solution and water was filtered through 0.45 µm membranes and degassed to avoid back pressure and clogging of column.

Table 2.2.15 Composition of buffers required for purification recombinant proteins by affinity purification (IMAC).

Buffers	Composition
Equilibration buffer	20 mM sodium phosphate, pH 7.4 500 mM NaCl, 60 mM Imidazole
Elution buffer	20 mM sodium phosphate, pH 7.4 500 mM NaCl, 300 mM Imidazole
Cleaning buffer	20 mM sodium phosphate, pH 7.4 500 mM NaCl, 50 mM EDTA

The bacterial cells (100 ml culture) were harvested by centrifugation at 10,000***g*** using a centrifuge (Sigma, 4K15) and the resulting cell pellets were re-suspended in 5 ml of 20 mM sodium phosphate buffer pH 7.4. The cells were sonicated (Sonics, Vibra cell) on ice for 8 min (9s on/off pulse; 33% amplitude) and again centrifuged at 19,000***g*** at 4ºC for 30 min to get the crude cell free extract. The cell free extract was filtered through a 0.45 µm membrane before loading onto HiTrap chelating HP column (GE Healthcare, USA). Initially the column was washed with 5 volumes of filtered and degassed water to remove the alcohol. The column was activated using 2-3 ml 0.1 M $NiSO_4$ solution and the unbound Ni^{2+} ions were washed away with 5 column volumes of water. The column was then equilibrated with 10 volumes of equilibration buffer (Table 2.2.15). The filtered cell free extract (5) ml of recombinant protein was loaded on to the column at a flow rate of 1 ml/min. The column was then washed with 50-60 column volumes of equilibration buffer to remove the unbound proteins. The retained recombinant protein was then eluted with elution buffer and 1 ml fractions were collected (Carvalho *et al.,* 2004; Taylor *et al.,* 2005). The column was cleaned using cleaning buffer as mentioned in Table 2.2.15, washed with 5 column volumes of water and incubated in 1N NaOH at 4°C for 2h. The column was then washed with 20 volumes of water to remove NaOH and finally stored in 20% (v/v) ethanol at 4°C. The collected fractions were analyzed for protein content by taking 20 µl of purified protein and 80 µl of Bradford reagent in a 1.5 ml microcentrifuge tube. Development of blue colour complex indicated the presence of protein. The purified recombinant *Ct*XynGH30 and *Ct*Xyn30A proteins were dialyzed against 20 mM sodium phosphate buffer, pH 6.0 and *Ct*CBM6 was dialyzed with 20 mM sodium phosphate buffer, pH 7.0. The purity and molecular size of

proteins were verified by SDS-PAGE as described in Section 2.2.14. The purified proteins were quantified as described in Section 2.2.22.

2.2.22 Protein estimation by Bradford method

The concentration of the recombinant proteins was determined by the Bradford's method measuring the absorbance at wavelength of 595 nm (A_{595}) (Bradford, 1976). Bovine serum albumin (BSA) purchased from Sigma-Aldrich Pvt. Ltd. was used as standard protein. A standard plot of A_{595} *versus* different concentration of BSA (1.25–10 μg/ml) was prepared.

2.2.22.1 Preparation of Bradford reagent

The Bradford assay uses the spectral properties of Coomassie Brilliant Blue G-250 to estimate the amount of protein in a solution (Bradford, 1976). 100 mg 0.01% (w/v) Coomassie Brilliant Blue G-250 was weighed and dissolved in 50 ml 95% ethanol (in an amber colour bottle). 100 ml 85% (w/v) phosphoric acid was added to it. A magnetic bead was placed inside the bottle and the contents were mixed properly by keeping on magnetic stirrer until the dye completely dissolved. The dye solution was finally diluted to 1 liter with deionized water, filtered through (Whatman, No. 1 paper) under dark conditions and stored at 4°C. Commercial Bradford reagent (Sigma-Aldrich Pvt. Ltd., USA) was also used for protein content determination. The amount of protein was estimated using the following equation,

$$[\text{Protein}] = \frac{\Delta A_{595} \times V \times C}{v}$$

Where,

ΔA_{595} = change in absorbance of the sample
V = volume of the reaction mixture (ml)
C = 1 OD equivalent of BSA from standard plot (mg/ml)
v = volume of the enzyme used for assay (ml)

2.3 Results and Discussion

2.3.1 Defining boundaries to modular full length *Ct*XynGH30

The complete domain-wise distribution with proper designations of the different modules along with their amino acid range was described earlier in the Section 2.1**.**

2.3.2 PCR amplification of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 encoding genes

CtXynGH30, *CtXyn30A* and *CtCBM6* (genes) were PCR amplified from the genomic DNA of *Clostridium thermocellum* using the conditions as mentioned in Section 2.2.3 and were detected by 0.8% agarose gel. The intense band of PCR DNA were observed at position approx. 1578 bp, 1158 bp and 369 bp for *Ct*XynGH30 (Fig. 2.3.1 A), *Ct*Xyn30A (Fig. 2.3.1 B) and *Ct*CBM6 (Fig. 2.3.1 C), respectively. The PCR products were purified from gel using gel extraction kit (Qiagen) as mentioned in Section 2.2.6 and were stored at -20°C for subsequent cloning experiments.

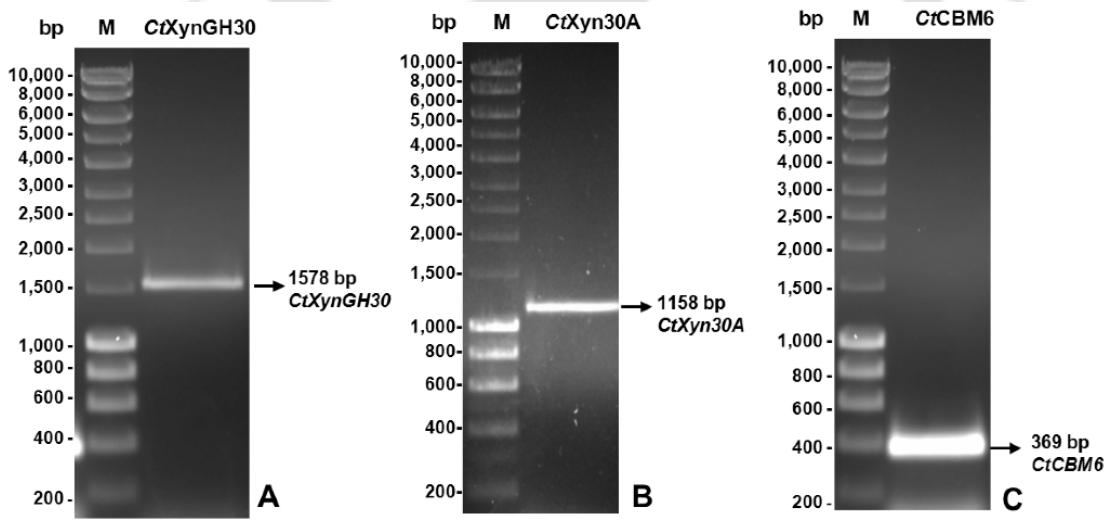


Fig. 2.3.1 Agarose (0.8%) gel showing PCR amplified products, **A)** 1578 bp *CtXynGH30* **B)** 1158 bp *CtXyn30A* and **C)** 369 bp *CtCBM6*. Lane M in the above figures represents DNA ladder

2.3.3 Cloning of PCR products into a pGEM-T Easy vector

The PCR products were ligated into pGEM-T Easy vector system by following the protocol mentioned in the Section 2.2.2. The primary screening of positive clones was based on the blue/white selection (colour of transformed bacterial colony). The possible positive colonies appear to be white whereas blue colour colonies indicate the intact *lacZα* gene (active β-galactosidase) that has no insert which ligated with the vector. The restriction enzyme digestion results of plasmid DNA isolated from these colonies further confirmed the present of positive clones. The restriction digestion results are shown in the Fig. 2.3.2, displaying approx. 3 kb band of pGEM-T Easy vector whereas, insert release of 1578 bp of *Ct*XynGH3*0* (Fig. 2.3.2 A), 1158 bp of *Ct*Xyn30A (Fig. 2.3.2 B) and 369 bp of *Ct*CB*M6* (Fig. 2.3.2 C) were observed.

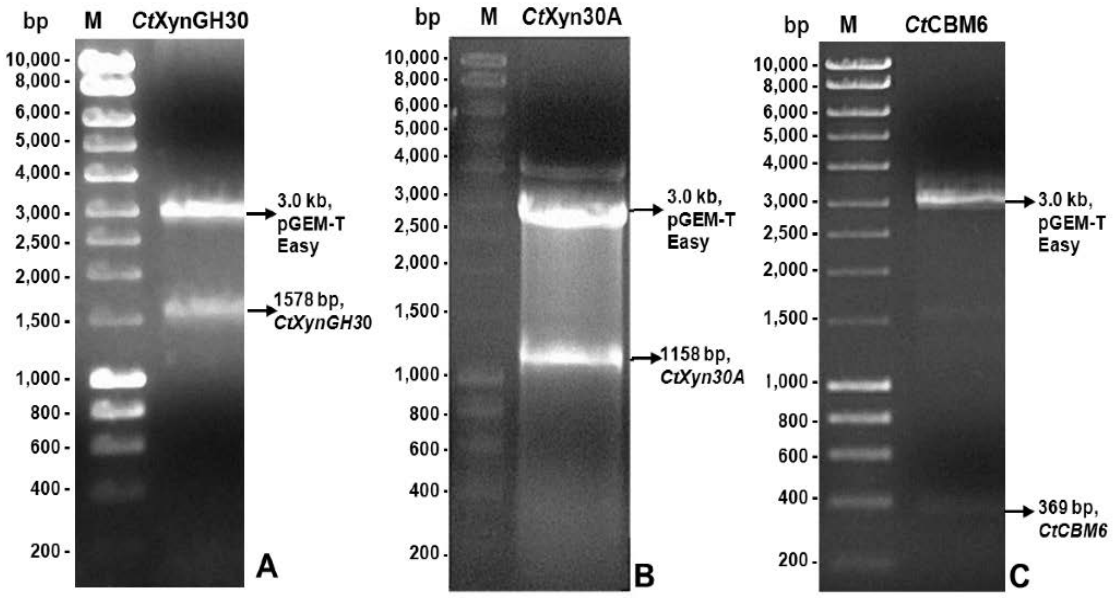


Fig. 2.3.2 Agarose gel (1.0%) showing restriction digestion of **A)** *Ct*XynGH30 **B)** *Ct*Xyn30A and **C)** *Ct*CBM6 containing pGEM-T Easy vector. Lane M- represent DNA ladder (Hyperladder I, Bioline)

2.3.4 Cloning and expression of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6

2.3.4.1 Cloning of NheI-XhoI digested fragments into pET-28a(+) expression vector

Both the *NheI-XhoI* digested inserts (from positive TA clones) as well as pET-28a(+) vector both were ligated by following the procedure described in the section 2.2.9. The ligated products *viz*. *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 containing plasmid DNAs were transformed using *E. coli* (DH5α) cell as described in section 2.2.13. The cells were spreaded on kanamycin supplemented agar plates and incubated at 37°C for 12-16 h which resulted in growth of number of colonies.

2.3.4.2 Isolation of plasmids DNA from colonies

The plasmid DNAs were isolated for each recombinant derivative from the cells grown in 5.0 ml LB medium containing kanamycin described in Section 2.2.17.

2.3.4.3 Screening of plasmid DNAs for positive clones

The plasmid DNA was isolated from the colonies of each recombinant derivative. The restriction digestion profile of isolated recombinant plasmid DNAs by *NheI* and *XhoI* are shown in figure 2.3.3.

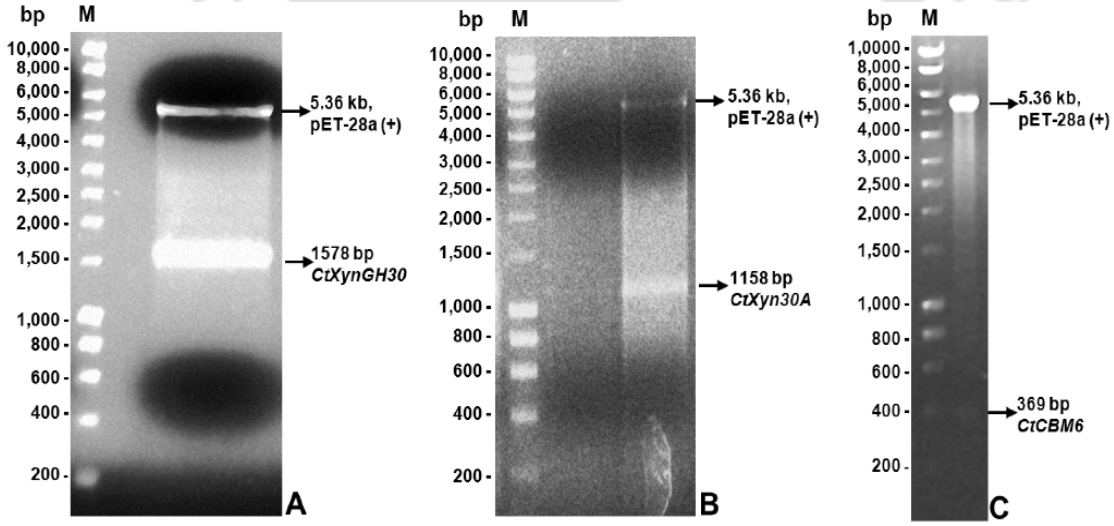


Fig. 2.3.3 Agarose gel (1%) showing confirmation of cloned recombinant plasmids by restriction digestion (*NheI/XhoI*) of **A)** *Ct*XynGH30 **B)** *Ct*Xyn30A and **C)** *Ct*CBM6, Lane M- DNA ladder.

The *NheI-XhoI* digested plasmid DNA of each positive clone showed two distinct bands, one of approx. 5.36 kb for pET-28a(+) vector and other band of 1578 bp for *Ct*XynGH30 (Fig. 2.3.3 A), 1158 bp for *Ct*Xyn30A (Fig. 2.3.3 B) and 369 bp band for *Ct*CBM6 (Fig. 2.3.3 C).

2.3.4.4 Sequencing of positive recombinant plasmid DNA

The final confirmation of clones was carried out by DNA sequencing of catalytic module *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6. The sequencing results of *Ct*CBM6 confirmed that the recombinant plasmid harbours the desired gene is well within the frame without any mutation. However, the sequencing results of *Ct*Xyn30A indicated there was change in the nucleotide bases at two places which leads to change in the amino acid. The amino acid Phe77 was mutated to Ser77 and Tyr315 was converted into Cys315.

2.3.5 Hyper-expression analysis of recombinant proteins

After confirmation by sequencing and restriction digestion analysis the recombinant plasmid DNAs were used for transformation of *E. coli* BL-21 cells (DE3). The *E. coli* BL-21 cells harbouring desired recombinant plasmids were initially grown in 5 ml LB medium supplemented with kanamycin, as described in Section 2.2.13. The cells were induced by IPTG for protein expression at mid exponential stage as described in Section 2.2.19. Hyper-expression of protein was analyzed on SDS-PAGE, by loading uninduced and the induced cells in adjacent wells, as shown in the figure 2.3.4. The results showed the expression of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 proteins (Fig. 2.3.4 A, B&C). Glycerol stocks of *E. coli* BL-21 cells harbouring recombinant plasmids were made and kept at -80ºC. The cells in 100 ml LB medium were grown for purification of enzyme.

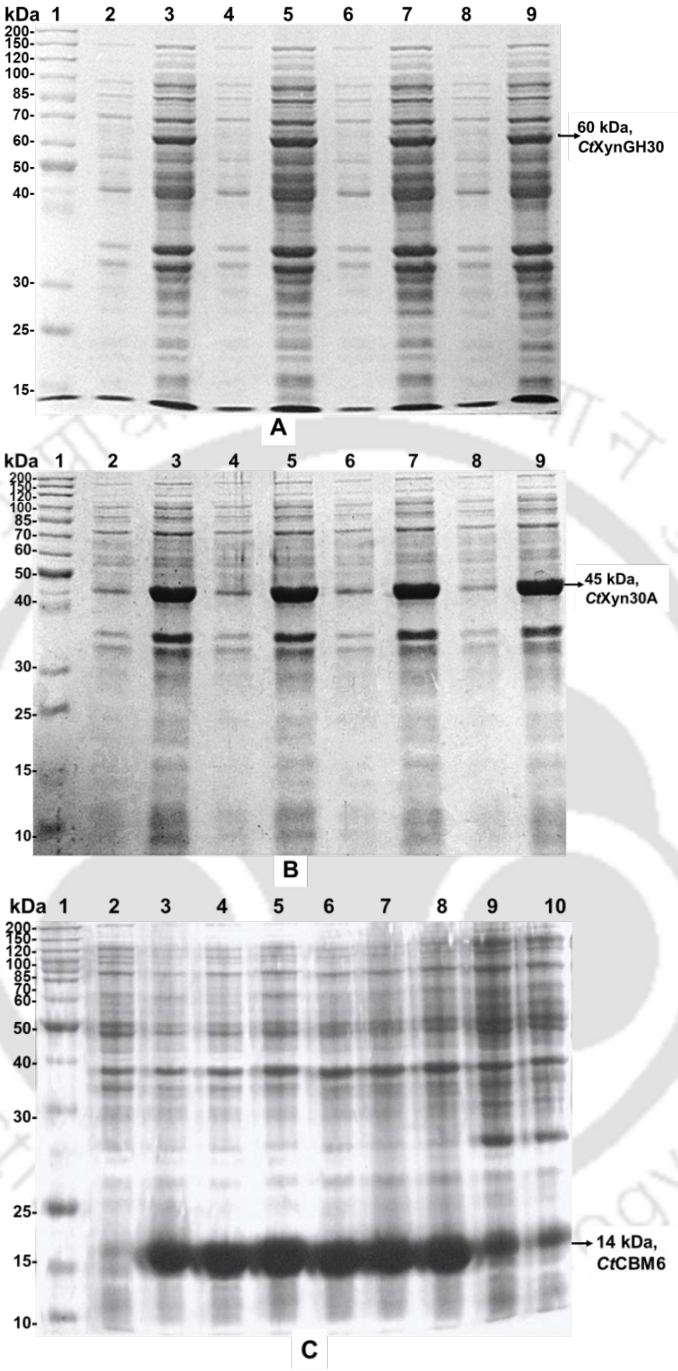


Fig. 2.3.4 SDS-PAGE gels showing hyper-expression of recombinant proteins, **A)** 13.0% SDS-PAGE gel showing hyper-expression of *Ct*XynGH30- Lane 1: protein marker, Lane 2,4,6,8 Uninduced: *Ct*XynGH30 cell, lane 3,5,7,9 Induced *Ct*XynGH30 cells. **B)** 12.0% SDS-PAGE gel showing hyper-expression of *Ct*Xyn30A where Lane 1: protein ladder), Lane 2,4,6,8 Uninduced: *Ct*Xyn30A cell, lane 3,5,7,9 Induced *Ct*Xyn30A cells. **C)** 14.0% SDS-PAGE gel showing hyper-expression of *Ct*CBM6 where Lane 1: protein marker, Lane 2: Uninduced *Ct*CBM6 (*E. coli* BL-21), Lane 3-10: induced *Ct*CBM6 (*E. coli* BL-21) cells.

2.3.6 Purification of *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6

*2.3.6.1 Purification profile of Ct*XynGH30

The recombinant proteins, full length *Ct*XynGH30, *Ct*Xyn30A and *Ct*CBM6 were purified by affinity chromatography as described in the Section 2.2.21. The purification profile of *Ct*XynGH30 in the figure 2.3.5. showed the purified homogenous band of approx. 60 kDa molecular size.

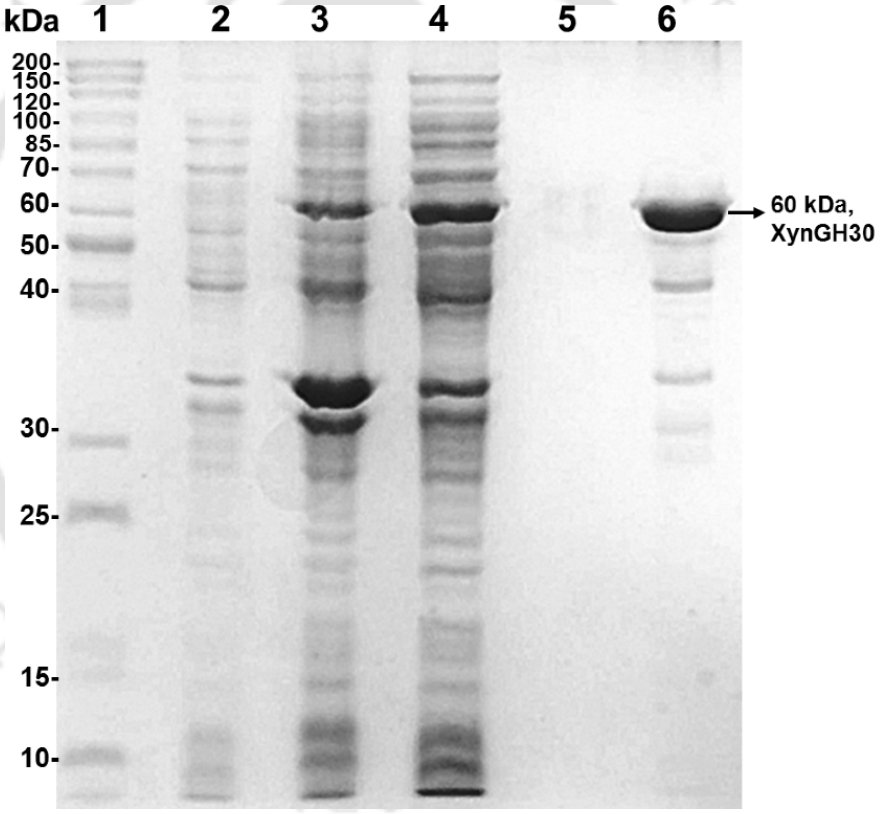


Fig. 2.3.5 SDS-PAGE (13.0%) showing expression and purification of *Ct*XynGH30, Lane 1: protein ladder, Lane 2: Uninduced *Ct*XynGH30 (*E. coli* BL-21), Lane 3: cell pellet of *E. coli* BL-21 cells, Lane 4: induced *E. coli* BL-21 cells, Lane 5: last wash from column, lane 6: purified protein *Ct*XynGH30 (~60 kDa).

*2.3.6.2 Purification of Ct*Xyn30A

The recombinant *Ct*Xyn30A contains an N-terminal His_6 tag (MGSSHHHHHHSSGLVPRGSHMAS), giving a total of 409 amino acid residues with a molecular mass of 45.18 kDa as shown in the figure 2.3.6. The purified *Ct*Xyn30A displayed a homogenous band of approx. 45 kDa. *Ct*Xyn30A expression as soluble protein as a most of the protein could be seen in the cell free extract (lane 3, Fig. 2.3.6)

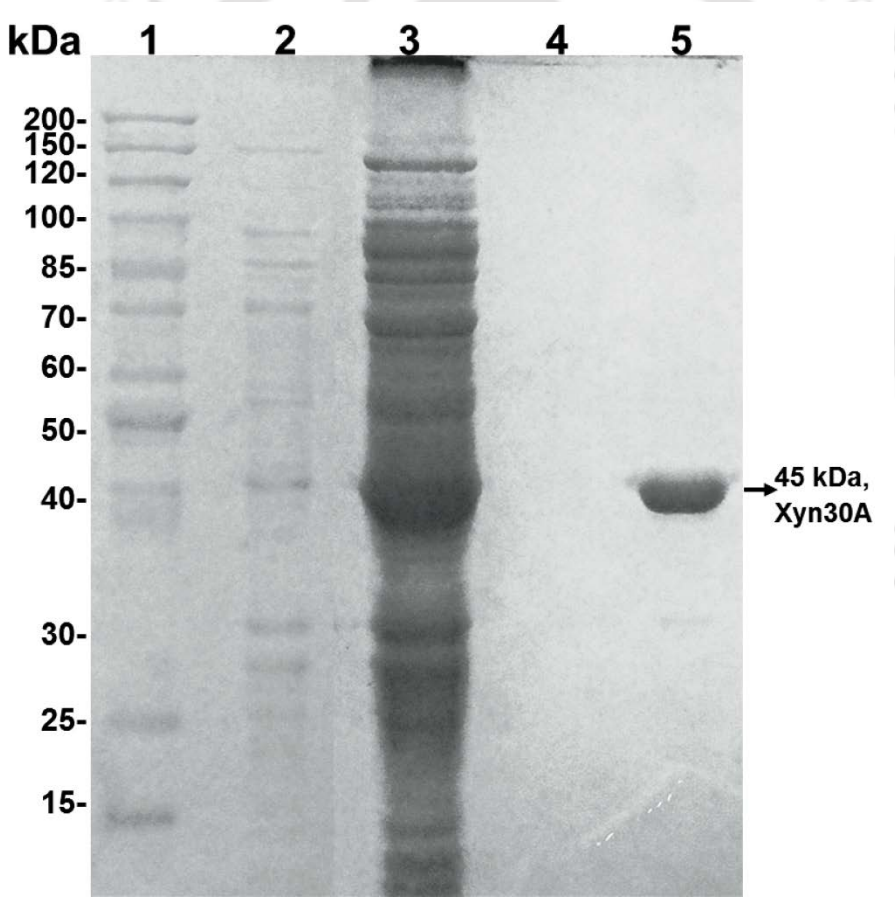


Fig. 2.3.6 SDS–PAGE 13% (w/v) showing over-expression and purification of *Ct*Xyn30A. Lane 1, Page Ruler protein marker (labelled in kDa); lane 2, uninduced *Ct*Xyn30A cells; lane 3, cell-free extract; lane 4, last wash from column; lane 5, purified *Ct*Xyn30A (~45 kDa).

2.3.6.3 Purification of CtCBM6

The purification profile of protein *Ct*CBM6 is shown in the figure 2.3.7. The figure clearly displayed the purified band of molecular size approximately 14 kDa. The molecular size of *Ct*CBM6 predicted by sequence analysis using pepstats (http://emboss.bioinformatics.nl/) server was 13.8 kDa including the Histidine tag (His_6 tag). However, this did not include a few amino acids from N-terminal which are present in the protein from to the pET-28a(+) vector. The theoretical molecular size of *Ct*CBM6 was in the close agreement with the size displayed by SDS-PAGE

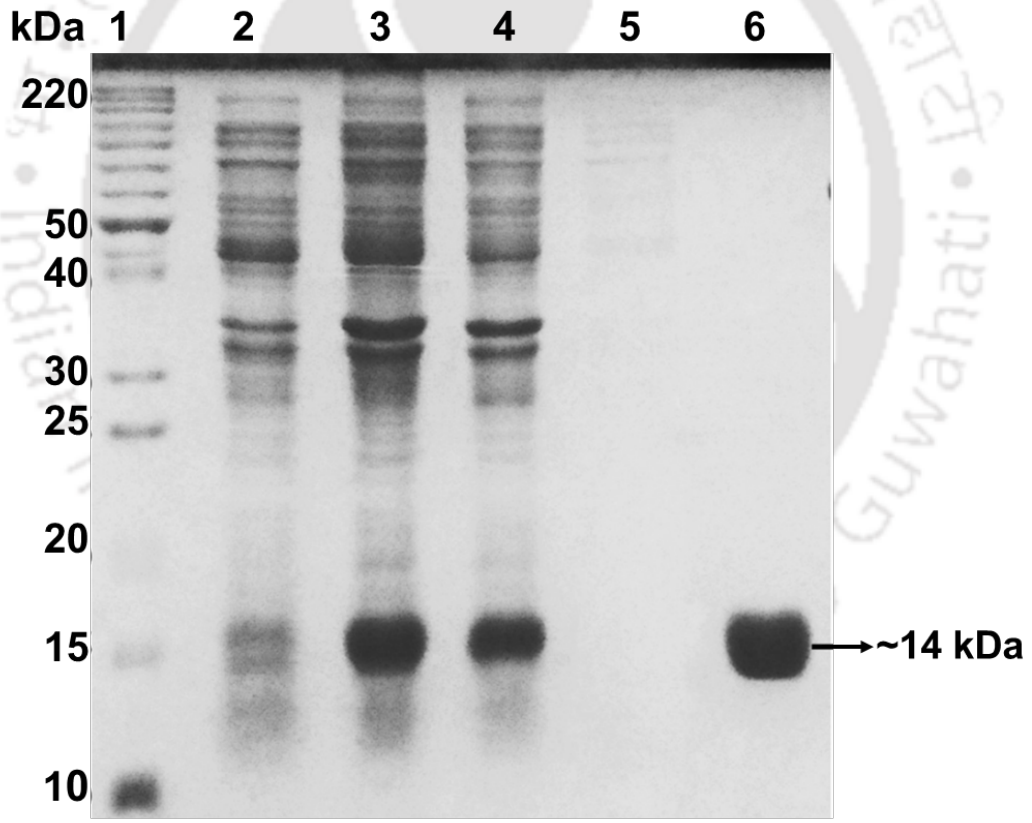


Fig. 2.3.7 SDS-PAGE (14%) showing over-expression and purification profile of recombinant *Ct*CBM6, Lane 1: protein marker, Lane 2: uninduced *E. coli* BL-21 cells, Lane 3: induced *E. coli* BL-21 cells, Lane 4: cell pellet, Lane 5: last wash from column and lane 6: purified *Ct*CBM6 protein.

2.3.7 *Protein estimation of expressed and purified recombinant derivatives*

The amount of purified recombinant proteins obtained from 100 ml of grown cultures was calculated from the standard plot of BSA (bovine serum albumin) and are listed in Table 2.3.1. The concentration of recombinant proteins purified after HiTrap column and after dialysis against respective buffers were in the range of 2-2.4 mg/ml (Table 2.3.1). The total amount of recombinant enzymes present in 2 ml buffer was 4.2 mg (*Ct*XynGH30), 4.8 (*Ct*Xyn30A) and 3.9 (*Ct*CBM6) as displayed in Table 2.3.1.

Table 2.3.1 Amount of purified recombinant proteins obtained from 100 ml cultures.

Recombinant proteins	Protein concentration (mg/ml)	Volume of purified protein (ml)	Total amount of purified protein (mg)
*Ct*XynGH30	2.1±0.10	2.0	4.2±0.1
*Ct*Xyn30A	2.4±0.12	2.0	4.8± 0.12
*Ct*CBM6	1.95±0.05	2.0	3.9± 0.05

Conclusions

The modular protein *Ct*XynGH30 having gene accession number ABN54208 (http://www.ncbi.nlm.nih.gov/protein) and uniprot ID A3DJS9 is a carbohydrate active enzyme. *Ct*XynGH30 consisted of an N-terminal, catalytic module from family 30 glycoside hydrolase (GH30) named *Ct*Xyn30A and a C-terminal, family 6 carbohydrate binding module (*Ct*CBM6) connected via a flexible linker of 6 amino acid residues. The presence of dockerin type I module at C-terminal implies that *Ct*XynGH30 is a part of cellulosome of *C. thermocellum*. The genes encoding *Ct*XynGH30 and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6 were amplified from the genomic DNA of *Clostridium thermocellum* ATCC 27405 (GenBank Accession No: ABN54208). The amplified PCR products carrying the desired restriction sites successfully cloned initially into the pGEM-T Easy vector and subsequently, cloned into expression vector pET-28a(+). The recombinant plasmids corroborating the desired genes (*Ct*XynGH30 and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6) were transformed into the *E.coli* BL21 cells. Over-expression of all the three proteins was achieved after induction with IPTG. The catalytic enzymes *Ct*XynGH30 and *Ct*Xyn30A purified by IMAC displayed molecular sizes of 60 kDa and 45 kDa respectively. The non-catalytic carbohydrate binding modules (*Ct*CBM6) showed molecular size of 14 kDa. The amount of protein obtained from 100 ml of cultures of *Ct*XynGH30 and its truncated derivatives *Ct*Xyn30A and *Ct*CBM6 after purification were 4.2 mg, 4.8 mg and 3.9 mg respectively. These purified proteins were used for further studies for biochemical, functional and structural characterization.

References

Beg, Q.K., Kapoor, M., Mahajan, L., Hoondal, G.S. (2001) Microbial xylanases and their industrial applications: a review. Applied Microbiology and Biotechnology, 56, 326-338.

Biely, P, Vrsanska, M, Tenkanen, M, Kluepfel, D. (1997) Endo-beta-1,4-xylanase families: differences in catalytic properties. Journal of Biotechnology, 57, 151-166.

Bradford, M. (1976) A Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, 72, 248-254.

Carvalho, A.L., Goyal, A., Prates, J.A.M., Bolam, D.N., Gilbert, H.J., Pires, V.M.R., Ferreira, L.M.A., Planas, A., Romão, M.J., Fontes, C.M.G.A. (2004) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4 and beta-1,3-1,4-mixed linked glucans at a single binding site. Journal of Biological Chemistry, 279, 34785-34793.

Chrambach, A. (1985) The practice of quantitative gel electrophoresis (1st ed). Deerfield Beach, FL, VCH, John Wiley & Sons. 9-18, 85-99, 101-104.

Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. Current Opinion in Microbiology, 11, 87–93.

Dodd, D., Cann, I.K., (2009) Enzymatic deconstruction of xylan for biofuel production. Global Change Biology Bioenergy, 1, 2-17.

Engler, M.J., Richardson, D.C. (1982) DNA ligases. In P.D. Boyer (ed.), The Enzymes. Academic Press, San Diego. 15, 3-30.

Fontes, C.M., Gilbert, H.J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annual Review of Biochemistry, 79, 655-681.

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

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology, 166, 557-580.

Ichinose, H., Yoshida, M., Fujimoto, Z., Kaneko, S. (2008) Characterization of a modular enzyme of exo-1,5-alpha-L-arabinofuranosidase and arabinan binding module from *Streptomyces avermitilis* NBRC14893. Applied Microbiology and Biotechnology, 80, 399-408.

Kulkarni, N., Shendye, A., Rao, M. (1999) Molecular and biotechnological aspects of xylanases. FEMS Microbiology Reviews, 23, 411-456.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227, 680-685.

Meyer, T.S., Lambert, B, L. (1965) Use of coomassie brilliant blue R250 for the electrophoresis of microgram quantities of parotid saliva proteins on acrylamide-gel strips. Biochimica et Biophysica Acta, 107, 144-145.

Neuhoff, V., Stamm, R., Hansjorg, E. (1985) Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels: a systematic analysis. Electrophoresis, 6, 427-448.

Park, J.M., Vinuselvi, P., Lee, S.K. (2012) The mechanism of sugar-mediated catabolite repression of the propionate catabolic genes in *Escherichia coli*. Gene, 504, 116-121.

Sambrook, J., Russel, D.W (2001) In (3rd ed.) Molecular Cloning: A Laboratory Manual, Vol. 1. Cold Spring Harbor Laboratory Press, Woodbury, New York.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) In (2nd ed.) *Molecular Cloning: A Laboratory Manual*, Vol. 1. Plainview, Cold Spring Harbor Laboratory Press, Woodbury, New York.

Shallom, D., Shoham, Y. (2003). Microbial hemicellulases. Current Opinion in Microbiology, 6, 219-228.

Shoham, Y., Lamed, R., Bayer, E.A. (1999) The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends in Microbiology, 7, 275-281.

Studier, F.W., Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology, 189, 113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., Dubendorff, JW. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods in Enzymology, 185, 60-89.

Taylor, E., Goyal, A., Guerreiro, C.I.P.D., Prates, J.A.M., Money, A.V., Ferry, N., Morland, C., Planas, A., Macdonald, J.A., Stick, R.V., Gilbert, H.J., Fontes, C.M.G.A., Davies G.J. (2005) How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides? Structure and activity of a *Clostridium thermocellum* lichenase, *Ct*Lic26A. Journal of Biological Chemistry, 280, 32761-32767.

Zaide, G., Shallom, D., Shulami, S., Zolotnitsky, G., Golan, G., Baasov, T., Shoham, G., Shoham,Y. (2001) Biochemical characterization and identification of catalytic residues in alpha-glucuronidase from *Bacillus stearothermophilus* T-6. European Journal of Biochemistry, 268, 3006-3016.

Chapter 3

Biochemical and functional characterization of catalytic module the full-length *Ct*XynGH30 and its truncated derivative *Ct*Xyn30A from *Clostridium thermocellum*

3.1 Introduction

Bioconversion of plant biomass into biofuel and many other value added products is a subject of high interest these days (Dodd and Cann, 2009). The cheap and widespread availability of plant polysaccharides, which represents the largest renewable organic carbon source on earth, makes it a promising candidate for biofules (Gilbert, 2010). Hemicellulose constituted the second highest percentage after cellulose in which heteroxylans constitute a major part in land plants (Scheller and Ulvskov, 2010). Heteroxylans have a xylose (a pentose sugar) as the main chain monomer linked by β-1,4 xylosidic bonds and chain containing branching of arabinose, glucuronic acid and galactos. Heteroxylans are complex in nature therefore it requires a cumulative action of various specific xylanases for its efficient and complete degradation (Shallom and Shoham, 2003). Xylanases characterized so far predominantly belong to family 10 and 11 glycoside hydrolase (GH10 and GH11) furthermore, they are also present in GH5, 30 and 43, wherein families such as GH5,

10 and 30 are the members of the same superfamily or clan i.e. GH-A. The family GH30 follows the retention type mechanism and their structure comprises of $(\beta/\alpha)_8$ TIM-barrel shape catalytic core like other members of GH-A clan (CAZy; http://www.cazy.org). A distinctive structure feature associated with GH30-8 subfamily is that along with the $(\beta/\alpha)_8$ catalytic core motif, an essential side β-strand rich structure known as "side β-structure" resembling CBMs is also present (St John *et al*., 2010). The other distinguishing property of GH30 subfamily H xylanase is the requirement of free 4-O-methyl-D-glucuronosyl (MeGlcA) residues as the side chain to act on a xylan main chain hence also called 'appendage dependent xylanase'. So far all the glucuronoxylan-specific enzymes under the same subfamily such as XynC from *B. subtilis* 168 (St John *et al*., 2006), *Cp*Xyn30A from *C. papyrosolvens* (*Cp*C71) (St John *et al*., 2014), XynA from *E. chrysanthemi* (Larson *et al*., 2003) and Xyn30D from *P. barcinonensis* (Sainz-Polo *et al*., 2014) are from mesophilic bacteria. However, *C. thermocellum* is known to be thermophilic, anaerobic, cellulolytic, ethanologenic and Gram-positive bacterium capable of directly converting plant biomass into ethanol (Bayer *et al*., 1983). A variety of carbohydrate active enzymes from *C. thermocellum* have been isolated and characterized but no glucururonoxylan specific enzyme has been reported. Thus, the biochemical and functional investigation of a glucuronoxylan-xylanohydrolase from a thermophile will provide important information regarding the differences as well as the similarities between the enzymes from the same sub-family as well as across the families.

In the present study the purified recombinant proteins, *Ct*XynGH30 and *Ct*Xyn30A modules were biochemically and functionally characterized. Substrate specificity of both the enzymes was determined. The analysis of hydrolyzed products

of substrates by these enzymes was carried out to determine the mode of action i.e. endo- or exo- acting xylanase. The biochemical properties such as optimum temperature and pH for activity of both enzymes were studied. The effects of various metal ions and chemical reagents on the enzyme activity have been also studied. Specific mutants of *Ct*Xyn30A were generated by altering the key amino acid residues by site-directed mutagenesis in order to study its structure and functional relationship.

3.2 Materials and Methods

3.2.1 Substrates and chemicals

Natural polysaccharides such as rye arabinoxylan, wheat arabinoxylan (soluble and insoluble), arabinogalactan, sugar beet arabinan, rhamnogalactouronan, carob galactomannan and different oligosaccharides such as xylotetraose, xylopentaose, xylohexaose and aldouronic acid mixture were purchased from Megazyme International, Ireland. Oat spelt xylan, birchwood xylan, beechwood xylan, 4-O-methyl glucoronoxylan, barley β-D-glucan, carboxy methylcellulose, curdlan, pullulan, pustulan, pectic galactan (apple), pectic galactan (citrus), pectic galactan (lupin), xyloglucan, glucomannan (konjac and locust bean) and galactan (lupin) were purchased from Sigma-Aldrich Chemicals Pvt. Ltd., USA. Synthetic substrates, 4-nitrophenyl β-D-xylopyranoside, 4-nitrophenyl α-D-xylopyranoside, and 4-nitrophenyl α-D-glucopyranoside were purchased from Sigma-Aldrich Chemicals Pvt. Ltd., India. D-xylose, Bradford reagent was procured from Sigma-Aldrich Pvt. Ltd., USA. Coomassie brilliant blue G-250 was purchased from Amresco LLC, USA. Disodium 2-[2-carboxylatomethyl (carboxymethyl)amino] ethyl (carboxymethyl) amino] acetate (disodium EDTA), sodium carbonate, sodium potassium tartarate, sodium bicarbonate, sodium sulphate, sodium phosphate (monobasic), sodium phosphate (dibasic) and salts of metal ions *viz.* Na^{+}, Ca^{2+}, Mg^{2+}, Ni^{2+}, Zn^{2+}, Mn^{2+}, Cu^{2+}, Co^{2+}, Hg^{2+}, Fe^{3+} and Al^{3+} were procured from Himedia Laboratories Pvt. Ltd., India. Sodium arsenate, ammonium molybdate, sulphuric acid, hydrochloric acid, acetone, acetonitrile, acetic acid were purchased from Merck Limited, India. The TLC plate was purchased from Merck KGaA, Darmstadt, Germany.

3.2.2 Cloning, over-expression and purification of *Ct*XynGH30 and *Ct*Xyn30A

The detailed procedure and protocol of cloning, overexpression and purification has been described in Chapter 2.

3.2.3 Enzyme activity assay

The enzyme assay were initially performed in 100 µl of reaction mixture containing 20 mM sodium phosphate buffer pH 7.0 and 1% (w/v) substrate using 10 µl of purified enzyme (*Ct*XynGH30 and *Ct*Xyn30A, 1.7 mg/ml each). The released reducing sugar was determined by the Nelson and Somogyi method (Nelson, 1944; Somogyi, 1945). The reaction was carried out at 50°C for 15 min. After the analysis of the un-optimized assays, the results showed that the enzymes preferably hydrolyzed xylan based substrates. Afterwards, the assays were carried out using beechwood xylan under optimized conditions of pH 6.0 and temperature, 70°C.

3.2.3.1 Preparation of reagents for reducing sugar estimation

Reagent A

Sodium carbonate anhydrous	6.25 g
Sodium potassium tatarate	6.25 g
Sodium bicarbonate	5.00 g
Sodium sulphate anhydrous	50.0 g

The above mentioned components were dissolved in 100 ml of deionized water and the final volume was adjusted to 250 ml. The solution was filtered (Whatman No. 1) and stored at a temperature between 30-37°C.

Reagent B

Reagent B was prepared by dissolving 15 g of copper sulfate ($CuSO_4$) in 50 ml deionized water and one or two drops of concentrated sulphuric acid was added to it. The final volume was made up to 100 ml with deionized water and the solution was filtered (Whatman No. 1) and stored at room temperature.

Reagent C

Reagent C was prepared in two steps by under dark condition as it is light sensitive. First, 2.5 g of ammonium molybdate was dissolved in 45 ml of deionized water in 100 ml beaker and 2.1 ml of concentrated sulphuric acid was added to it. In another beaker 0.3 g of sodium arsenate was dissolved in 2.5 ml of deionized water. Now, this solution was added to ammonium molybdate solution and the contents were mixed (total volume was around 50 ml). The solution was filtered (Whatman No. 1) under dark conditions and stored at 37°C. The solution was used after 24h incubation.

Reagent D

Reagent D was prepared by mixing reagent A and reagent B in the ratio 25:1. Reagent D was always prepared freshly for use in the assay.

3.2.3.2 Generation of standard plot of D-xylose

The standard plot was prepared by varying the concentration of D-xylose from 10-250 µg/ml. The reaction volume was kept 100 µl with 20 mM sodium phosphate buffer pH 7.0. The reaction mixture containing buffer and D-xylose was incubated (in a 1.5 ml micro-centrifuge tube) at 70°C for 15 min and then 100 µl of solution D (Section 3.2.2.2) was added to it. The reaction mixture was then boiled for 20 min and cooled. 100 µl of solution C (Section 3.2.2.2) was added and the contents were mixed. The colour development was rapid and completed after the evolution of carbon dioxide stopped. Then 700 µl of deionized water was added to make the final volume 1 ml. The absorbance at 500 nm wavelength (A_{500}) was measured using UV-Visible spectrophotometer (Varian, Carry 100 Bio) against a buffer as blank. A standard plot of A_{500} *versus* D-xylose concentration (µg/ml) was generated and one absorbance unit

equivalent of D-xylose (µg/ml) was calculated. The one absorbance unit equivalent of D-xylose (µg/ml) was converted in mg/ml for calculation of enzyme activity.

3.2.3.3 Calculation of enzyme activity

The activity of the enzyme was expressed as U/ml and the specific activity as U/mg of protein. One unit (U) of enzyme activity is defined as the amount of enzyme that liberates 1 µmole of reducing sugar per min. The enzyme activities of *Ct*XynGH30 and *Ct*Xyn30A were calculated as described below,

$$\text{Enzyme activity (U/ml)} = \frac{\Delta A_{500} \times C \times V}{150 \times t \times v} = (\mu \text{ mole/min/ml})$$

where,

ΔA_{500} = change in absorbance of the sample at 500 nm

C = 1 OD equivalent D-xylose concentration from standard plot

V = volume of the reaction mixture (ml)

t = time of reaction (min)

150 = molecular weight of D-xylose

v = volume of the enzyme taken in assay (ml) for reducing sugar estimation.

3.2.3.4 Assay of CtXynGH30 and CtXyn30A with natural and synthetic p-nitrophenylglycoside substrates

*Ct*XynGH30 and *Ct*Xyn30A were assayed against several natural polysaccharides, in order to determine the substrate specificity. The assays were performed in 100 µl reaction mixture using (1%, w/v) beechwood xylan and 10 µl of purified enzyme (*Ct*XynGH30 or *Ct*Xyn30A, 1.7 mg/ml each) in 20 mM sodium phosphate buffer (pH 6.0) at 70°C for 10 minute following the procedure described in 3.2.1.3. The enzyme activity was determined by measuring the released reducing

sugar by the Nelson and Somogyi method (Nelson, 1944; Somogyi, 1945). The concentration of reducing sugar was estimated using a standard curve of xylose as *Ct*XynGH30 and *Ct*Xyn30A showed maximum activity against the substrates with xylose residues in the main chain. One unit of activity was defined as the amount of enzyme which produced 1 µmole of xylose per min under the optimized condition of temperature and pH. The assay with synthetic *p*-nitrophenylglycoside substrates was performed in 1.0 ml 20 mM sodium phosphate buffer (pH 6.0) containing 50 µM of *p*-nitrophenylglycoside (*p*NP glycoside) and 10 µl of purified enzymes (*Ct*XynGH30 or *Ct*Xyn30A, 1.7 mg/ml each) and incubated at 70°C for 10 min. The enzyme activity was determined by measuring the release of 4-nitrophenol (*p*NP) by monitoring the absorbance at 405 nm (A_{405}) using a UV-Visible spectrophotometer (Varian, Cary 100 Bio) as described earlier (Ahmed *et al*., 2013). One unit of enzyme activity was defined as the amount of enzyme liberating 1 µmole of *p*-nitrophenol per min.

3.2.3.5 Substrate specificity of CtXynGH30 and CtXyn30A against a range of substrates

The activity of recombinant enzymes *Ct*XynGH30 and *Ct*Xyn30A was analyzed against several natural polysaccharides. The polysaccharides were chosen on the basis of substrates reported so far for the family 30 and GH family 5 such as different xylans, celluloses, pectins and other hemicelluloses substrates. The assay was performed as mentioned in the section 3.2.3.4.

3.2.4 Activity staining of *Ct*XynGH30 and *Ct*Xyn30A

The zymogram analysis of *Ct*XynGH30 and *Ct*Xyn30A was carried out in the presence of beechwood xylan to confirm the xylanase activity by following the method as previously described by Ghosh *et al*., (2013). 10 µg of each *Ct*XynGH30

and *Ct*Xyn30A was run on a SDS-PAGE (12.5%) gel containing 0.2% beechwood xylan. After the run SDS was removed by constantly shaking the gel in 2.5% (v/v) Triton X-100 for 30 min at 25°C followed by incubation in 50 mM sodium phosphate buffer (pH 6.0) at 60ºC for 1 h. The gel was stained with 0.1% (w/v) Congo red solution for 15 min and then washed with 1 M NaCl until a clear zone of xylanase activity was visible against red background. Counter staining was performed with 1 M hydrochloric acid as described previously (Songsiriritthigul *et al*., 2010).

3.2.5 Determination of optimum pH and pH stability of *Ct*XynGH30 and *Ct*Xyn30A

A wide range of pH (from 2.0 to 11.0) was chosen to find the optimum pH of *Ct*XynGH30 and *Ct*Xyn30A. Since no single buffer system can work from pH 2.0 to pH 11.0, three different buffer systems were chosen for this experiment: McIlvaine buffer (0.1 M citric acid and 0.2 M disodium phosphate) pH range from 2 to 8, glycine-NaOH buffer pH 8.0 to pH 10.0 and Na_2HPO_4.HCl buffer for pH 11.0. The assays were performed in 100 µL reaction mixtures using 17 µg of *Ct*XynGH30 or *Ct*Xyn30A with beechwood xylan (1%, w/v) at different pH at 70°C for 10 min. In order to determine the pH which gives maximum stability, 50 µl (85 µg) of *Ct*XynGH30 or *Ct*Xyn30A were incubated at different pH varying from 2.0 to 12.0 for 30 min. Subsequently, 10 µl aliquots of enzyme were taken for determining the residual enzyme activity and plotted by taking the highest activity as 100%.

3.2.6 Temperature optima and thermal stability of *Ct*XynGH30 and *Ct*Xyn30A

The enzyme assays were performed in 100 µl reaction mixtures (1%, w/v) beechwood xylan at pH 6.0 for 10 min using 17 µg of *Ct*XynGH30 or *Ct*Xyn30A at different temperatures ranging from 20°C to 90°C in order to determine the optimum

temperature. The thermal stability of *Ct*XynGH30 and *Ct*Xyn30A were determined by incubating 50 µl (85 µg) of each enzyme at different temperatures for 30 min. An aliquot (10 µl) of enzyme was taken from 100 µl reaction mixtures and residual activity was determined as described in section 3.2.3.4.

3.2.7 Kinetic parameters determination of *Ct*XynGH30 and *Ct*Xyn30A.

The kinetic parameters such as K_m, turn over number (k_{cat}, min^{-1}) and catalytic efficiency (k_{cat}/K_m, min^{-1} mg^{-1} mL^{-1}) were determined by performing assays with varying concentration of beechwood xylan by following the methodology described by Ahmed *et al*., (2012). The concentration of beechwood xylan was varied from 0.05 to 3% (w/v) and beyond that concentration no further increase in activity was observed, keeping the same enzyme concentration. The assays were performed in 100 µl reaction mixtures containing 17 µg *Ct*XynGH30 and *Ct*Xyn30A in 20 mM sodium phosphate buffer (pH 6.0) at 70°C for 10 min. A blank was used with equivalent substrate concentrations, but without the enzyme. All determinations of enzyme activity were made in triplicate. The Michaelis-Menten constant (K_m) was calculated by Lineweaver-Burk plot.

3.2.8 Effects of metal ions and chemical agents on the activity of *Ct*XynGH30 and *Ct*Xyn30A

The effect of different metal ions and chemical agents on the activity of *Ct*XynGH30 and *Ct*Xyn30A were analysed. The enzyme activity was determined by adding different metal salts (1 mM and 10 mM; NaCl, KCl, $CaCl_2$, LiCl, $CoCl_2$, $NiSO_4$, $CuSO_4$, $MgSO_4$, $FeCl_3$, $MnSO_4$, $ZnSO_4$, or $HgCl_2$) and other reagents (SDS, EDTA or β-mercaptoethanol) to the assay mixtures of 100 µl containing beechwood xylan (1%, w/v) along with 17 µg *Ct*XynGH30 or *Ct*Xyn30A. The blank with

substrates and the respective metal ions and chemical agents were also assayed in parallel. The enzyme activity determined in the absence of metal ion was taken as 100%.

3.2.9 TLC analysis of *Ct*XynGH30 hydrolysed products from substrate

Thin layer chromatography (TLC) was employed for qualitative analysis of hydrolyzed products following the method as described by Chakraborty *et al*., (2015). The enzymatic reaction was performed taking beechwood xylan (1%, w/v) as a substrate in 100 µl reaction mixtures containing 17 µg of enzyme (*Ct*XynGH30) at optimum pH 6, however temperature was kept 10°C lower than the optimized 70°C because of higher stability of the enzyme during longer incubation period. The protein and uncut polysaccharides were precipitated by adding equal volume of acetone (100 µl) with gentle mixing and subsequently centrifuged at 13000g for 5 min. The supernatant was collected and concentrated by heat drying. The 0.2 µl of sample as well as of standard solutions (xylo-oligosachharides mixture and aldouronic acids mixture oligosaccharide, 1.0 mg/ml) were loaded on the TLC plate (TLC Silica gel 60 F254 20 × 20 cm, Merck). The developing chamber was saturated prior with the developing solution (mobile phase) constitute of acetic acid-butanol-water in a ratio of 3:2:2. The plate was air dried for few minutes after loading the samples or standard and then kept carefully in the developing chamber. The mobile phase was allowed to migrate as much as possible up to the edge of the TLC plate and after that it was air dried for 5 min. The visualizing solution was prepared from diphenylamine, aniline and phosphoric acid by mixing 1 ml of 37.5% HCl, 2 ml aniline, 10 ml of 85% H_3PO_4, 100 ml of ethyl acetate and 2 g of diphenylamine reagent. The migrated analyte was visualized by overlaying the visualizing solution over the TLC plate. The

TLC plates were than dried in a hot-air oven at 90°C for 20 min. The migrated spots of samples were analyzed corresponding to the standards.

3.2.10 ^{1}H-NMR spectroscopic analysis of hydrolyzed products

^{1}H Nuclear magnetic resonance (NMR) of hydrolysed products obtained by enzyme *Ct*XynGH30 treatment of substrate beechwood xylan was carried out using a spectrometer (Bruker, 600 MHz) as described by Ghosh *et al.*, (2014). The enzymatic reaction was performed by taking beechwood xylan (1%, w/v) as a substrate in 200 µl reaction mixtures containing 34 µg of enzyme (*Ct*XynGH30) at pH 6 and 50°C. The protein and uncut polysaccharides were precipitated by adding equal volume of acetone (200 µl) with gentle mixing and subsequently centrifuged at 13000g for 5 min. The supernatant solution was lyophilized. The 10-15 mg of this lyophilized sample was dissolved in the 0.75 ml of D_2O. The sample was placed in the NMR tube (0.5 mm, Sigma-Aldrich) and the spectra were recorded at 600 MHz.

3.2.11 Site-directed mutagenesis of *Ct*Xyn30A

Site-directed mutagenesis was carried out using the PCR-based NZY mutagenesis kit (NZY Tech Ltd., Portugal) according to the manufacturer's instructions, using recombinant plasmid containing gene encoding *Ct*Xyn30A as a template. The sequences of the primers used for generation of following mutants are displayed in Table 3.2.1.

Table 3.2.1 Oligonucleotide primers for site-directed mutagenesis of *Ct*Xyn30A.

Construct	Primer sequence
WT	Fwd:5'-GCCATTTCAGTTCAGAACGAACCAGACTATGCCCATGAG-3'
Glu-136-Ala	Fwd:5'-GCCATTTCAGTTCAGCGCGAACCAGACTATGCCCATGAG-3'
	Rev: 5'- CTCATGGGCATAGTCTGGTTCGCGCTGAACTGAAATGGC-3'
WT	Fwd:5'-GAGCTGTGGATGACAGAGGTTTATGTTCCAAACAGC-3' (WT)
Glu-225-Ala	Fwd:5'-GCCATTTCAGTTCAGCGCGAACCAGACTATGCCCATGAG-3'
	Rev: 5'-CTCATGGGCATAGTCTGGTTCGCGCTGAACTGAAATGGC-3'

3.2.11.1 RCR reaction set-up for mutants

The PCR conditions were chosen according to the manufacturer's instructions. The reaction was carried out in sterile, nuclease free micro-centrifuge tube as mentioned in Table 3.2.2.

Table 3.2.2 PCR mixture reagent for site-directed mutagenesis of *Ct*Xyn30A

PCR components	Volume (µl)	Final concentration
10x reaction buffer	5.0	1x
25 mM dNTP mix	0.5	0.25 mM
Forward primer (100 mM)	1.0	2.5 ng
Reverse primer (100 mM)	1.0	2.5 ng
Sigma water, pH 8.0	40.5	--
*Ct*Xyn30A plasmid DNA (50 ng/µl)	1.0	1.0 ng
NZYDNA change (2.5 U/µl)	1.0	2.5 U
Total	50.0	--

The PCR reaction setup was mixed gently and subjected to short spin for 10 second before amplification.

3.2.11.2 PCR amplification condition

The PCR conditions were chosen according to the manufacturer's instructions. The PCR amplification was performed using a thermal cycler (Applied Biosystems, GeneAmp PCR System 9700). The PCR cycles for amplification are mentioned in Table 3.2.3

Table 3.2.3 PCR cycles for generating mutants of *Ct*Xyn30A.

Steps		Time (min)
I. Denaturation at 95°C		1.5
II. 18 cycles of	i) Denaturation at 95°C	1
	ii) Annealing at 60°C	1
	iii) Extension at 68°C	13
III. Final extension at 68°C		10

3.2.11.3 DpnI endonuclease digestion of PCR amplified product

The amplified PCR product reaction tube was first kept on ice for 2 min. The restriction digestion of non-mutated DNA (parental DNA) was achieved by adding 1 µl of *Dpn*I directly to the reaction tube. The reaction tube was gently mixed, spins down and incubated at 37°C for 1 h to digest the wild type template DNA. The *Dpn*I specifically acted on the methylated DNA therefore it chop-up the template DNA while does not digest newly amplified RCR product which is devoid of methylation.

3.2.11.4 Transformation of DpnI digested product

The 10 µl *Dpn*I treated PCR amplified DNA was transformed to 100 µl ultra-competent *E. coli* BL-21 cells following the protocol as described in detail earlier in Chapter 2 Section 2.2.13, in Brief,

1. *Dpn*I treated PCR amplified DNA in tube was incubated for 30 min on ice.
2. Heat shock at 42°C for exactly 40 s to cells in tube.
3. Tube was kept on the ice for 2 min.
4. 900 µl pre-warmed (37°C) SOC medium was added to the tube.
5. Incubated the tube at 37°C at 200 rpm for 1 h.
6. Centrifuged at 2000 rpm for 1 min, gently removed 900 µl of supernatant.
7. Re-suspended cells by gently pipetting and plated 100 µl of cells on LB agar plates containing 50 µg/µl kanamycin.
8. Incubated inverted plates overnight at 37°C

3.2.11.5 Screening of positive mutants

The mutant plasmid DNAs were sequenced from GATC Biotech AG**,** European Custom Sequencing Centre, Cologne, Germany to confirm if the appropriate mutations were incorporated.

3.3 Results and Discussion

3.3.1 Substrate specificity of *Ct*XynGH30 and *Ct*Xyn30A with natural polysaccharides

In order to determine the substrate specificity of *Ct*Xyn30A and *Ct*Xyn30A, 26 different natural polysaccharides having different main chain monomeric units or different side chain substitutions and from different sources were screened. The specific activity of *Ct*XynGH30 and *Ct*Xyn30A with various polysaccharides is reported in Table 3.3.1. The maximum specific activity of *Ct*XynGH30 and *Ct*Xyn30A was found to be 31 and 34 U/mg, respectively, with beechwood xylan (Table 3.3.1). *Ct*XynGH30 and *Ct*Xyn30A also displayed noticeable activity against birchwood xylan (31 and 28 U/mg) and 4-O-methyl glucuronoxylan (30 and 27 U/mg) (Table 3.3.1). It is evident from the activity data that both *Ct*XynGH30 and *Ct*Xyn30A predominantly act on polysaccharides having xylan as a main chain substituted with varying amounts of glucuronic acids such as beechwood-, birchwood- and 4-O-methylglucurono-xylan. Both the enzymes also displayed activity against polysaccharides containing xylan backbone substituted with side chain of arabinose residues. However, the activity of *Ct*Xyn30A and *Ct*XynGH30 was comparatively lesser on oat spelt xylan (10 and 10 U/mg), wheat arabinoxylan (insoluble) (9.0 and 9.5 U/mg), wheat arabinoxylan (soluble) (7.0 and 7.0 U/mg) and rye arabinoxylan, (6.0 and 6.7 U/mg), respectively, as compared with glucuronic acid substituted xylans (Table 3.3.1). *Ct*XynGH30 and *Ct*Xyn30A did not show any activity against polysaccharides having a main chain of D-glucose unit such as carboxy methyl- and ethyl-cellulose, curdlan, pullulan or pustulan. No activity was

observed with galactan (Lupin), pectic galactans and any of synthetic *p*NP-glycoside substrates.

Table 3.3.1 Substrate specificity of *Ct*XynGH30 and *Ct*Xyn30A from *Clostridium thermocellum*.

Substrates	Specific Activity (U/mg) *Ct*XynGH30	Specific Activity (U/mg) *Ct*Xyn30A
Xylan (Beechwood)	34±0.8	31±0.8
Xylan (Birchwood)	31±0.7	28±0.7
4-O-methyl glucuronoxylan	30±0.7	27±0.7
Xylan (Oat spelt)	10±0.6	10±0.6
Arabinoxylan (Wheat, insoluble)	9.5±0.5	9.0±0.5
Arabinoxylan (Wheat, soluble)	7.0±0.4	7.0±0.4
Arabinoxylan (Rye)	6.7±0.3	6.0±0.3
Arabinogalactan	5.0±0.3	4.5±0.3
Arabinan (Sugar beet)	4.2±0.3	4.0±0.3
Galactomannan (Locust bean)	2.8±0.2	2.7±0.2
Xyloglucan	2.2±0.2	2.2±0.2
Glucomannan (Konjac)	1.5±0.2	1.4±0.2
Other Substrates*	ND	ND

ND = *No activity detected.*

* = *Carob galactomannan, Galactan (Lupin), β-D-Glucan (Barley), Carboxy methyl cellulose, Pecticgalactan (Apple), Pecticgalactan (Citrus), Pecticgalactan (Lupin), Rhamnogalactouronan, Curdlan, Pullulan and Pustulan.*

All assays were performed at 70°C using 20 mM sodium phosphate buffer (pH 6.0) in triplicates.

3.3.2 Zymography of *Ct*XynGH30 and *Ct*Xyn30A

Zymogram analysis of *Ct*XynGH30 performed on beechwood xylan as substrate showed the clear zone of substrate hydrolysis at position approximately, 60 kDa corresponding to the molecular size of *Ct*XynGH30 (Fig. 3.3.1).

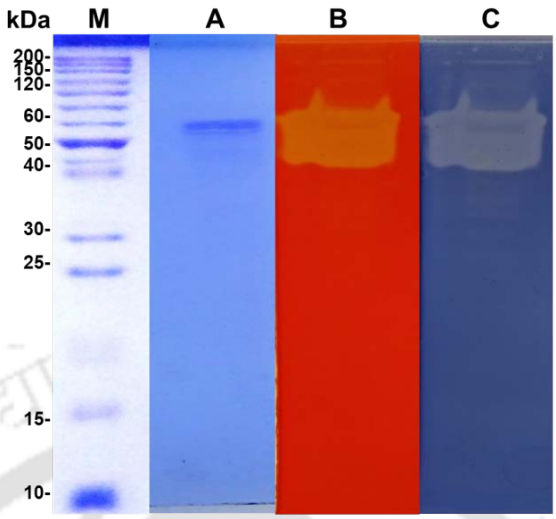


Fig. 3.3.1 Zymogram analysis of *Ct*XynGH30 for xylanase activity against beechwood xylan. Lanes A) purified *Ct*Xyn30A, B) congo red staining and C) counter stain by acid, Lane M, Molecular mass marker.

The clear zone of *Ct*Xyn30A activity band of approximately, 45 kDa was observed as an orange color against a red background and clear zone against the blue background after the counter staining (Fig. 3.3.2). The zymogram analysis of *Ct*Xyn30A confirmed its xylanase activity (Fig. 3.3.2).

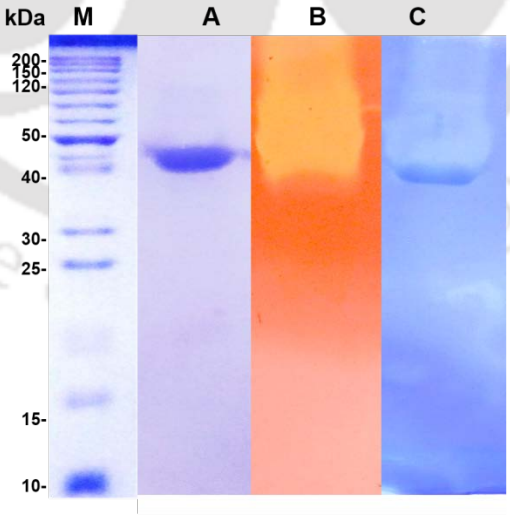


Fig. 3.3.2 Zymogram analysis of *Ct*Xyn30A for xylanase activity against beechwood xylan. Lanes A) purified *Ct*Xyn30A, B) congo red staining and C) counter stain by acid, Lane M, Molecular mass marker.

3.3.3 Effect of pH and temperature on the enzyme activity of *Ct*XynGH30 and *Ct*Xyn30A

The pH and temperature optima and the stability are important biochemical parameters of an enzyme.

3.3.3.1 pH optima and pH stability of CtXynGH30 and CtXyn30A

The optimum pH for the *Ct*XynGH30 was found to be 6.0 and it retained more than 80% of its activity till pH 8 (Fig. 3.3.3). The pH profile showed that *Ct*XynGH30 loses the activity gradually with increase in the pH from beyond 8.0. The pH stability profile of the enzyme displayed it is maximally stable at near neutral pH (near 6 and 7).

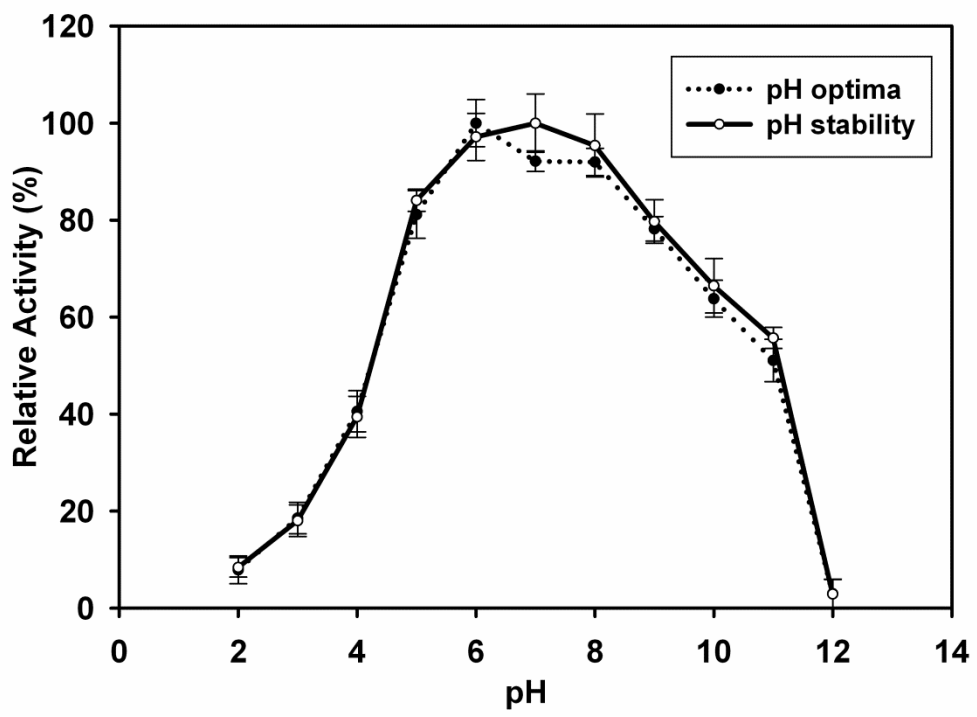


Fig. 3.3.3 Effect of pH on the activity of *Ct*XynGH30, Filled black circle (•) representing the pH optima profile and empty circle represent (–○–) the pH stability profile of *Ct*XynGH30.

The pH optima and stability profiles of *Ct*Xyn30A are shown in the figure 3.3.4. *Ct*Xyn30A showed maximum activity at pH 6.0 and it retained more than 75% activity up to pH 9.0. *Ct*Xyn30A displayed maximum stability at pH 6.0 and retained 90% residual activity up to pH 8.0 (Fig. 3.3.4).

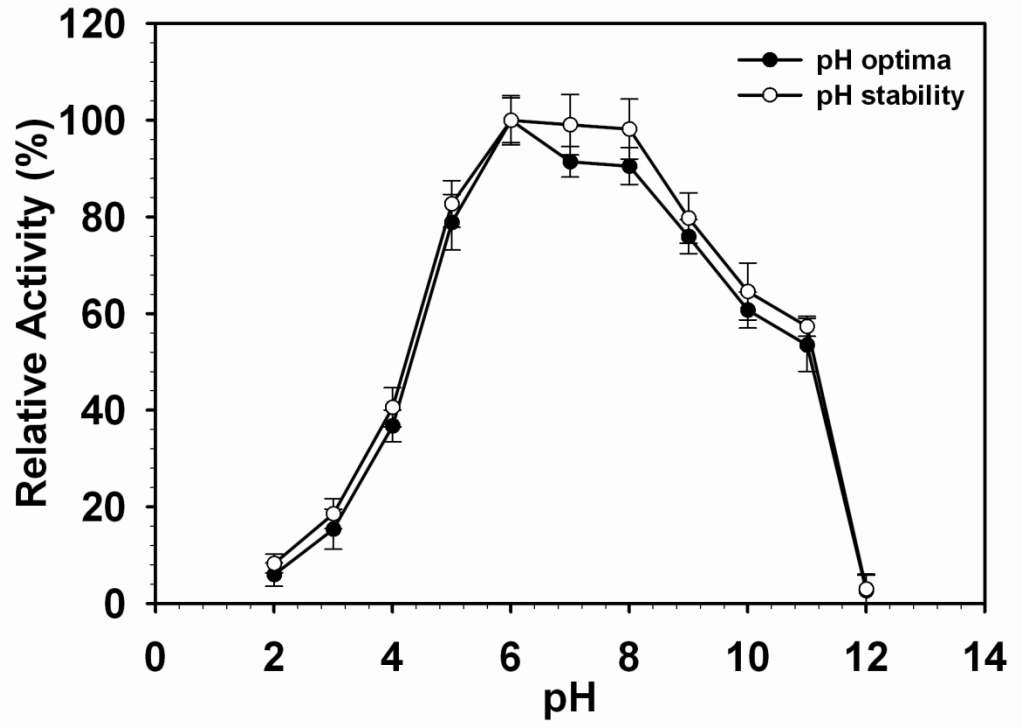


Fig. 3.3.4 Effect of pH on the activity of *Ct*Xyn30A, Filled black circle (•) representing the pH optima profile and empty circle represent (–○–) the pH stability profile of *Ct*Xyn30A.

3.3.3.2 Temperature optima and thermo-stability of CtXynGH30 and CtXyn30A

The temperature optima study of *Ct*XynGH30 showed it has highest activity at 70ºC and retained more than 75% residual activity till 80ºC (Fig. 3.3.5). The temperature stability study showed that it retained around 65% residual activity when incubated at 70ºC for 30 min (Fig. 3.3.5).

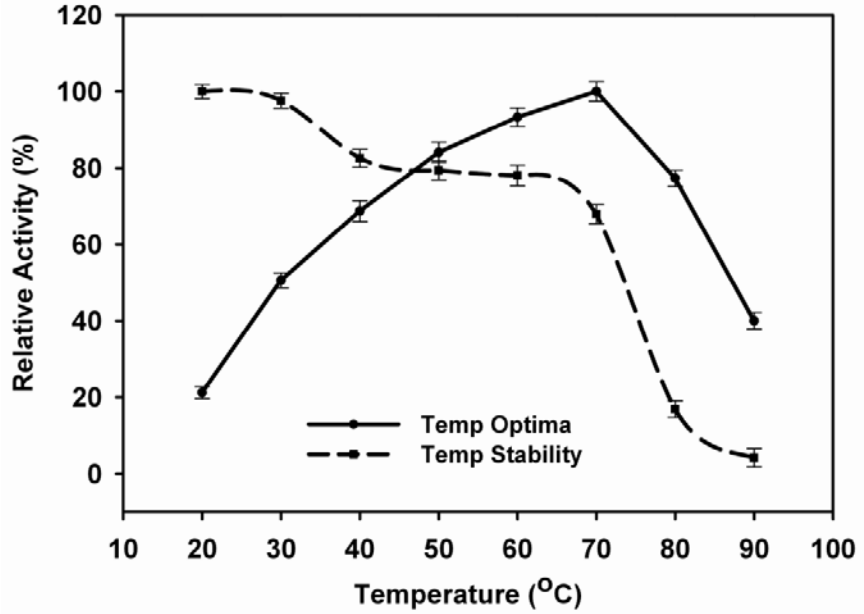


Fig. 3.3.5 Temperature optima and thermo-stability profile of *Ct*XynGH30.

*Ct*Xyn30A displayed maximum activity at 70ºC and retained more than 80% residual enzyme activity with in temperature range, 60-80ºC (Fig. 3.3.6). *Ct*Xyn30A was stable up to 60ºC and retained 75% enzyme activity for 30 min.

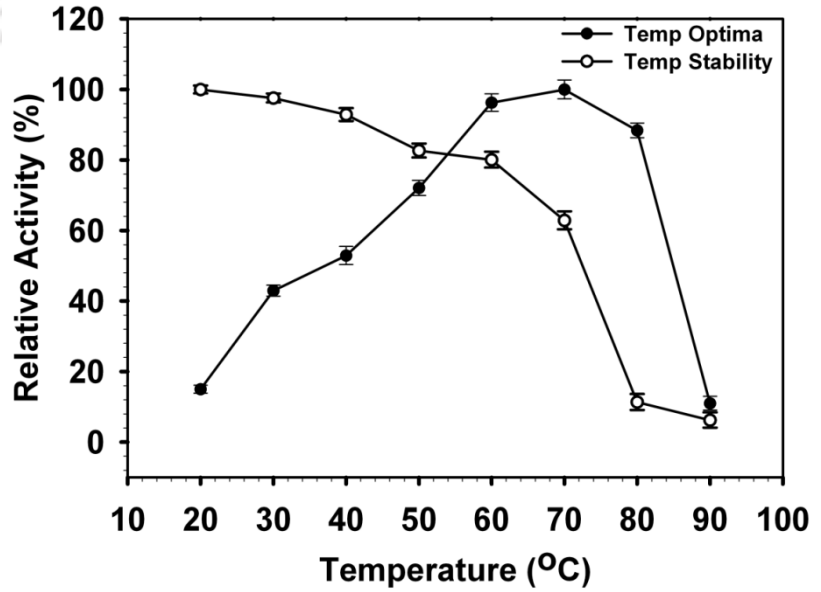


Fig. 3.3.6 Temperature optima and thermo-stability profile of *Ct*Xyn30A.

3.3.4 Effects of metal ions and chemical agents on *Ct*XynGH30 and *Ct*Xyn30A activity

The effect of metal ions and chemical agents on the activity of *Ct*XynGH30 and *Ct*Xyn30A was studied at low (1 mM) as well as high concentration (10 mM) as mentioned in the Table 3.3.2. The enzyme activity of *Ct*XynGH30 and *Ct*Xyn30A was enhanced by Li^{2+} ions by ~19-26% at 1 mM as well as 10 mM. The other mental ions studied do not showed positive effect on the enzyme activity. Surprisingly, the activities of both enzymes were elevated in the presence of EDTA at low concentration and on the contrary it gets remarkably decrease at high concentration (Table 3.3.2). Cu^{2+}, Ni^{2+}, Fe^{2+} ions and SDS slightly inhibited the enzyme activity at 1 mM and significantly at 10 mM. The activity of *Ct*XynGH30 and *Ct*Xyn30A was not significantly affected by Mg^{2+}, Ca^{2+} and β-mercaptoethanol. The effects of chemical agents and detergents were also studied and it was found that the activity of both the modules was decreased by NaCl, SDS and urea at low as well as high concentrations as mentioned in the Table 3.3.2.

Table 3.3.2 Effect of metal ions and chemical reagents on *Ct*XynGH30 and *Ct*Xyn30A activity.

Metal ion /Reagent additives	Relative activity (%)[a] (*Ct*XynGH30)		Relative activity (%)[a] (*Ct*Xyn30A)	
	1 mM	**10 mM**	**1 mM**	**10 mM**
Control	100	100	100	100
Li^{2+}	119±3.1	126±2.2	123±3.1	121±2.2
Co^{2+}	82±2.8	36±1.3	98±2.8	76±1.3
Ca^{2+}	95±3.3	86±1.4	97±3.3	84±1.4
Mg^{2+}	96±2.2	85±1.9	94±2.2	90±1.9
Ni^{2+}	90±1.5	26±2.1	92±1.5	33±2.1
Fe^{2+}	91±1.7	60±1.7	98±1.7	42±1.7
Cu^{2+}	87±1.1	16±1.8	80±1.1	7±1.8
Hg^{2+}	0	0	0	0
EDTA	117±3.6	69±3.2	126±3.6	68±3.2
β-merceptoethanol	96±4.0	47±3.9	98±4.0	72± 3.9
SDS	86±1.2	44±1.1	86±1.2	44±1.1
NaCl	84± 3.1 (50 mM)	72± 2.4 (250 mM)	83± 3.1 (50 mM)	72± 2.4 (250 mM)
Urea	89± 1.4 (0.5 M)	67± 1.7 (4M)	92± 1.4 (0.5 M)	67± 1.7 (4M)

[a] *Values are the mean ± SD (n=3) relative to the control without any metal ions or reagent added.*

3.3.5 Kinetic parameters of *Ct*XynGH30 and *Ct*Xyn30A with natural substrates

The kinetic parameters of *Ct*XynGH30 and *Ct*Xyn30A were determined by taking beechwood xylan as a substrate and the values of V_{max} were 40.52 U/mg and 36.15 U/mg, respectively. The other parameters are summarized in the Table 3.3.3. The enzyme activity of enzymes belonging to the same sub-family and structurally close to *Ct*Xyn30A was 59.9 U/mg of XynC from *Bacillus subtilis* which displayed highest sequence similarity with *Ct*XynGH30 (St John *et al.* 2006). The enzyme activity of Xyn30D from *Paenibacillus barcinonensis* was 30.3 U/mg (Valenzuela *et al.*, 2012).

Table 3.3.3 Kinetic properties and catalytic efficiencies of *Ct*XynGH30 and *Ct*Xyn30A with natural substrates.

Recombinant derivatives	Natural polysaccharides	K_m (mg/ml)	K_{cat} (min^{-1})	K_{cat}/K_m ($min^{-1}M^{-1}$)
*Ct*XynGH30	Beechwood xylan	2.0± 0.002	1.411×10^5	7.01×10^4
*Ct*Xyn30A	Beechwood xylan	2.2± 0.001	1.205×10^5	5.5×10^4

3.3.6 Activity of *Ct*XynGH30 and *Ct*Xyn30A towards synthetic substrates

The activity assed performed with synthetic *p*NP substrates to determine the type of cleavage by these enzymes. The results are shown in the Table 3.3.4. Both the enzymes lack exo-clevage activity and are unable to release the *p*-nitrophenyl unit (*p*NP) from their respective substrates. Therefore, no activity was detected with any of the synthetic *p*NP substrates.

Table 3.3.4 Assay of *Ct*XynGH30 and *Ct*Xyn30A with synthetic *p*NP substrates.

Synthetic *p*NP derivative	Type of enzyme activity	*Ct*XynGH30 activity	*Ct*Xyn30A activity
p-nitrophenyl-α-D-xylopyranoside	α-xylosidase	ND	ND
p-nitrophenyl- β-D-xylopyranoside	β-xylosidase	ND	ND
p-nitrophenyl-α-L-arabinofuranoside	α-L-arabinosidase	ND	ND
p-nitrophenyl-α-L-arabinopyranoside	α-L-arabinosidase	ND	ND
p-nitrophenyl-β-D-glucuronide	β-glucuronidase	ND	ND

ND= Not Detected

3.3.7 TLC analysis of *Ct*XynGH30 hydrolyzed products from beechwood xylan

The time dependent TLC profile of hydrolyzed products of beechwood xylan by treatment with *Ct*XynGH30 produced a series of higher xylo-oligosaccharides, thus implying the endo-acting cleavage by the enzyme (Fig. 3.3.7). The analysis showed that there is no change in the pattern of hydrolyzed products with incubation time. In another experiment when the xylohaxaose was treated with *Ct*XynGH30 even for a prolonged period did not lead to formation of any lower xylo-oligosaccharides or

monomer xylose. This suggested that *Ct*XynGH30 does not act on linear xylo-oligosaccharides and has an absolute requirement of substitution of glucuronic acid or 4-O-methyl glucuronic acid on xylan main chain for enzymatic cleavage. Similar results were observed in XynA from *Erwinia chrysanthemi* (Vrsanska *et al*., 2007; Urbanikova *et al*., 2011) and Xyn30D from *Paenibacillus barcinonensis* (Valenzuela *et al*., 2012). On the contrary *Cp*Xyn30A, a family GH30-8 from *Clostridium papyrosolvens* showed activity against linear xylo-oligosacharides, but less or no activity against the substituted xylans (St John *et al*., 2014).

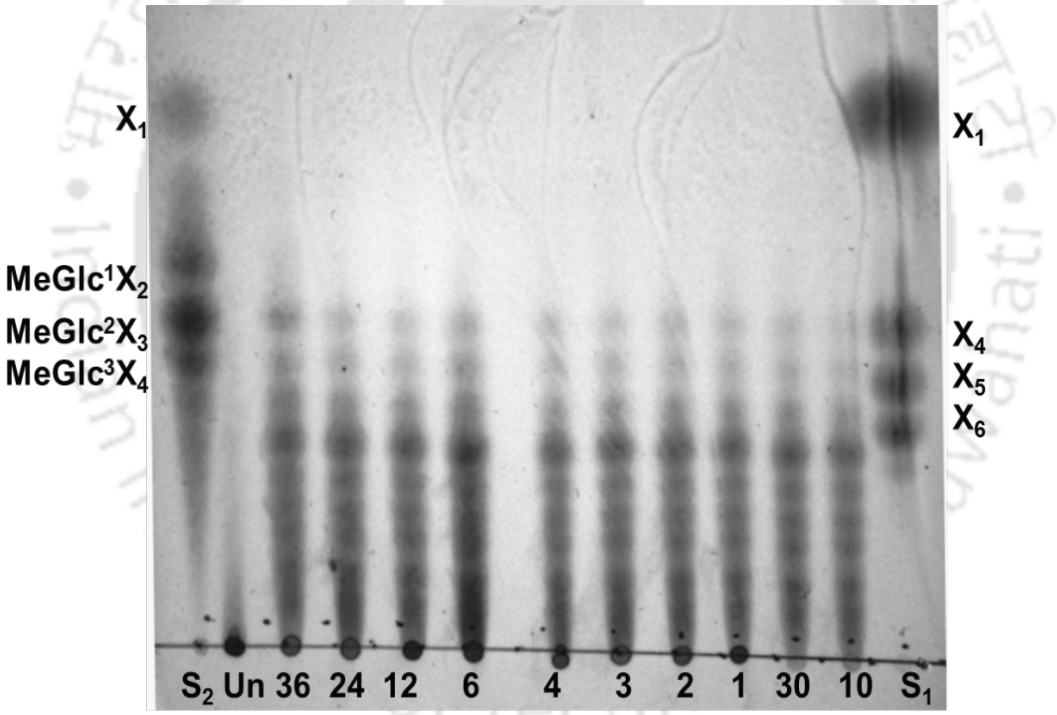


Fig. 3.3.7 Thin layer chromatography (TLC) analysis of time dependent hydrolysis of beechwood xylan (1%, w/v) incubated with *Ct*XynGH30 for 10 min to 36 h. Standards used **S_1**: **X_1**-xylose, **X_4**-xylotetraose, **X_5**-xylopentaoase **X_6**-xylohexaoase, standard **S_2:** $MeGlc^1X_2$-aldotriouronic, $MeGlc^2X_3$-aldotetraouronic, $MeGlc^3X_4$-aldotpentauronic, **Un** represents the control (uncut polysaccharide).

3.3.7.1 TLC analysis of *Ct*XynGH30 hydrolyzed products from different polysaccharides

It was evident from the activity data that both the catalytic modules (*Ct*XynGH30 and *Ct*Xyn30A) showed highest activity towards the polysaccharides having xylan main chain substituted with glucuronic acid (or 4-O-methyl glucuronic acid). Moverover, both the catalytic modules also displayed significant activity towards the polysaccharides having xylan main chain decorated with arabinose side chain. Therefore a range of substrates showing significant activity were treated with *Ct*XynGH30 and analyzed by TLC. The result displayed in the Fig. 3.3.8 showed that the enzyme is active against different polysaccharides and produce a series of oligosaccharides, except with arabinan (sugar beet) and xyloglucan.

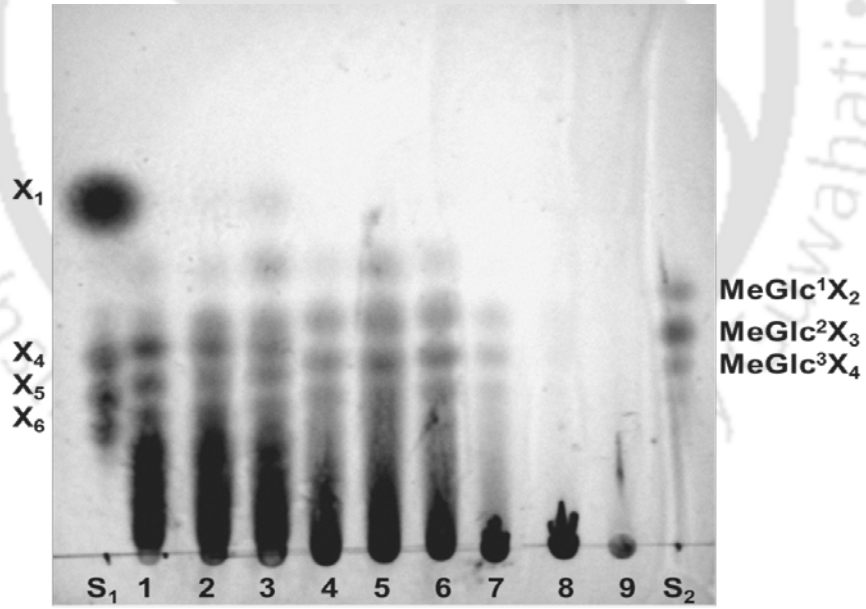


Fig. 3.3.8 Thin layer chromatography (TLC) analysis of different substrate (1%, w/v) incubated with *Ct*XynGH30 for 1 h. Substrate used from 1 to 9; are 1) beechwood xylan, 2) birchwood xylan, 3) 4-O-methyl glucuronoxylan, 4) oat splet xylan, 5) wheat arabinoxylan, 6)rye arabinoxylan, 7) arabino galactan, 8) arabinan (sugar beet) and 9) xyloglucan. Standards used **S_1**: **X_1**-xylose, **X_4**-xylotetraose, **X_5**-xylopentaoase, **X_6**-xylohexaoase, standard **S_2:** $MeGlc^1X_2$-aldotriouronic, $MeGlc^2X_3$-aldotetraouronic, $MeGlc^3X_4$-aldotpentauronic.

3.3.8 Analysis of ^{1}H-NMR of *Ct*XynGH30 hydrolyzed products of beechwood xylan

The ^{1}H-NMR spectra of products released after the hydrolysis of beechwood xylan by the treatment of *Ct*XynGH30 for 1 h analyzed by the TLC (Fig. 3.3.7) is shown figure 3.3.8. The characteristic signals and their respective peak assignment were carried out by method as described previously by Kardosova (Kardosova *et al.*, 1998) and Rhee (Rhee *et al.*, 2013) and the results are summarized in Table 3.3.5.

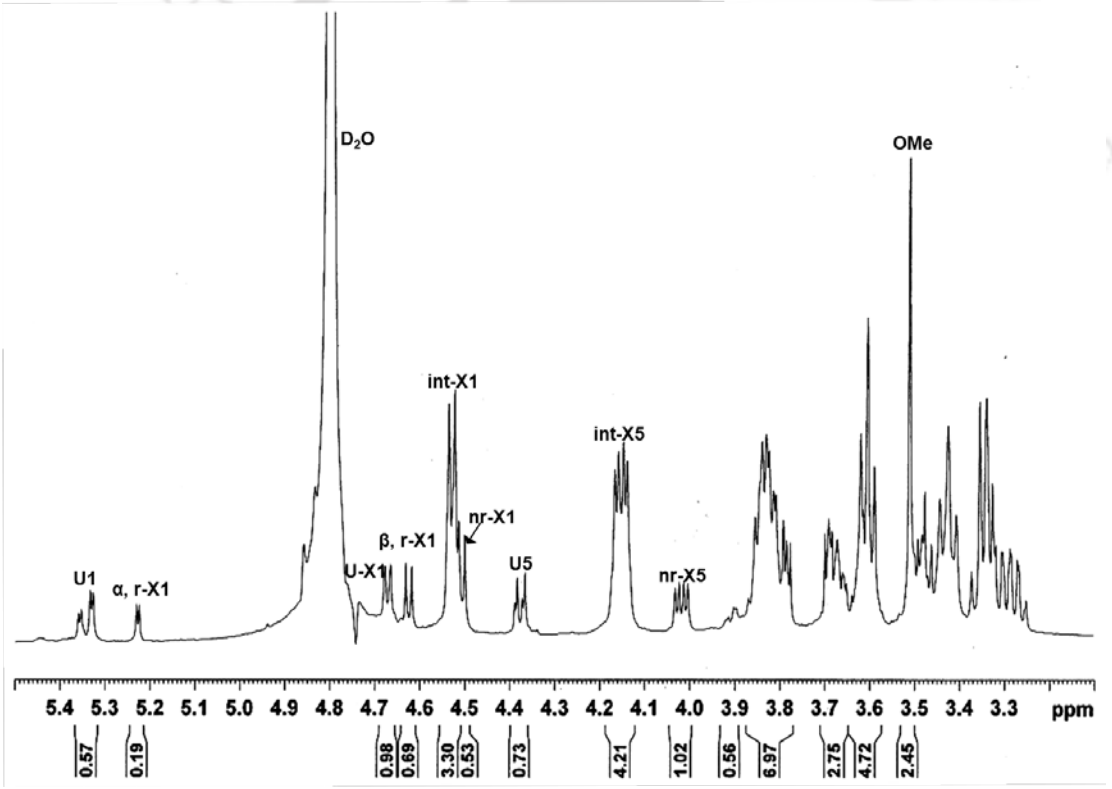


Fig. 3.3.8 1D ^{1}H-NMR spectrum (600 MHz) of hydrolyzed products (obtained from 1%, w/v beechwood xylan by treatment with *Ct*XynGH30 for 1h) in D_2O at 25ºC, showing different peaks (in ppm) and their respective integral values.

Table 3.3.5 Assignments of peaks present in xylo-oligosaccharides.

Atoms	acidic xylo-oligosaccharides (Rhee *et al*. 2013)	Present study (In ppm)
U1	5.27-5.33	5.32-5.37
α, r-X1	5.0-5.1	5.21-5.24
U-X1	4.5	4.66-4.69
β, r-X1	4.4-4.47	4.61-4.64
int-X1	4.31-4.35	4.51-4.57
nr-X1	4.3	4.5
U-5	4.32-4.34	4.35-4.39
int-X5	4.08-4.14	4.12-4.18
nr-X5	3.95-3.98	4.0-4.05
OMe	3.46	3.5
D_2O (as reference)	4.65	4.8

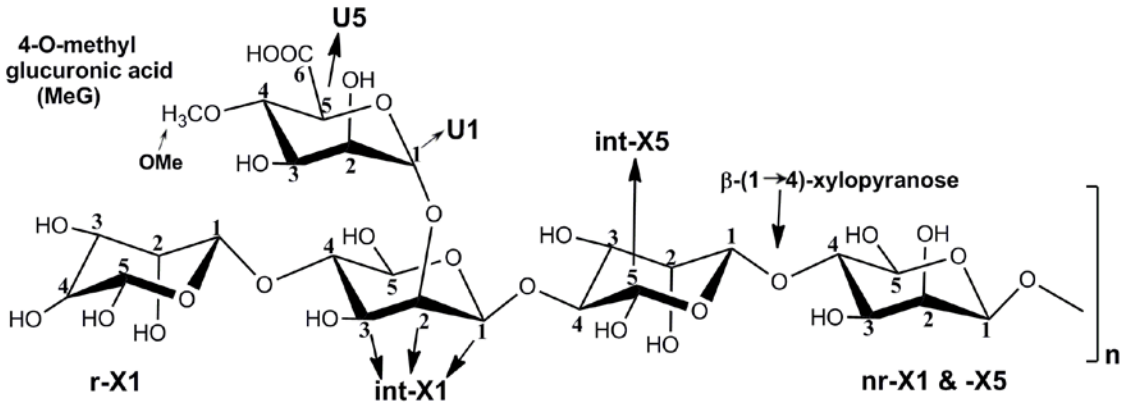


Fig. 3.3.9 Schematic representation of product generated by the *Ct*XynGH30 showing the position of different protons.

U1 =	(uronate or uronic acid) ^{1}H atoms linked to the uronate C-1 of MeG (methyl glucuronic acid) residue showed characteristic a single-doublet
α, r-X1 =	xylose in the alpha configuration at the reducing terminus
U-X1 =	internal proton of the α-(1→2) linked uronate
β, r-X1 =	xylose in the beta configuration at the reducing terminus
int-X1 =	internal proton of the β-(1→4), linked xylose
nr-X1 =	internal proton of the non reduring terminus β-(1→4), linked xylose
U-5 =	^{1}H atoms linked to the C-5 of MeG (U-5)
int-X5 =	^{1}H atoms linked to the C-5 (X-5) of internal β-(1→4), linked xylose (including reducing terminal)
nr-X5 =	^{1}H atoms linked to the C-5 (X-5) of the nonreducing (nr) terminal xylose
OMe =	^{1}H atoms C-4 linked -OCH_3 of MeG

The average degree of substitution of MeG over the xylose residues was determined by finding the ^{1}H integrals using the formula

(int-X1+ nr-X1+U-X1+ αr-X1+ βr-X1)/U1

i.e. (3.3+0.53+0.98+0.19+0.69)/0.57=9.982

This showed the average degree of substitution of MeG over xylose residues in a polymeric $MeGX_n$ (methyl glucuronoxylan).

The ratio of ^{1}H integrals (int-X1+nr-X1+U-X1+ αr-X1+ βr-X1)/ αr-X1+ βr-X1 was found to be (3.3+0.53+0.98+0.19+0.69)/0.88= 6.46

The ^{1}H integrals ratio obtained by above mentioned formula suggested that there are approximately, three residues of MeG on every two U-XOS (uronic acid substituted xylo-oligosachharide). A prominent peak at 3.46 ppm was observed indicating that ^{1}H atoms of the C-4 linked $O\text{-}CH_3$ of MeG residue (Fig. 3.3.8). This represents that *Ct*XynGH30 produced the acidic xylo-oligosaccharides having methylated glucuronic acid as substitution on the xylose moiety. The ratio of the integration of $^{1}H\text{-}U\text{-}O\text{-}CH_3$ to $^{1}H\text{-}U\text{-}5$ is 2.45/0.73 = 3.351 and if the integration of $^{1}H\text{-}U\text{-}5$ considered as 1, the ratio become 3.351:1 and this become 1.118:1 (3.351/3) for single hydrogen of $^{1}H\text{-}U\text{-}O\text{-}CH_3$ that represent all the C-4 carbons on the glucoronate residues contained $–OCH_3$ groups. At 5.32-5.37 ppm a split doublet was observed which signify that this sample contained an oligosaccharide moiety substituted with methyl glucuronic acid residue penultimate to the reducing end terminal at xylose (Fig. 3.3.8). Similar results were observed in the hydrolysis products obtained by the enzyme XynC (displayed highest sequence similarity towards the *Ct*XynGH30) from *Bacillus subtilis* where the main products formed a 4-O-methylglucuronate moiety linked to a xylose penultimate of the reducing end

terminal xylose and some number ahead of xylose residues towards the non-reducing terminal (St John *et al*., 2011). This result also matches with the previously described ^{1}H NMR spectra of acidic xylo-oligosaccharides where MeGA is substituted penultimate to the reducing terminal xylose as shown in the Fig. 3.3.9 (Cavagna *et al*., 1984; St John *et al*., 2006).

3.3.9 Assessment of the role of catalytic residues in *Ct*Xyn30A by site-directed mutagenesis

The members of family GH30 follow the retention type of mechanism (http://www.cazy.org/GH30.html). The key amino acid residues of *Ct*Xyn30A were predicted by *in silico* study by means of multiple sequence alignment, structural superposition and docking study as described in Chapter 4 and also published (Verma and Goyal, 2014). The amino acid Glutamate 136 (E136) was found to be catalytic acid/base while Glutamate 225 (E225) acted as a catalytic nucleophile in *Ct*Xyn30A. To establish the validity of the results obtained by the *in silico* study the mentioned predicted key catalytic residues were mutated with alanine. The sequencing result confirmed that the E136 and E225 were replaced with alanine. The activity assay performed taking beechwood xylan as substrate with the mutant (E136A and E225A) as described in the section 3.2.3.4 showed that both the mutants did not display any activity and were dead mutants. As predicted the Glu136 as a catalytic general acid (electrophile) during the reaction mechanism, the side chain of carboxylic group of Glu136 provide the necessary transfer of electron (Fig. 3.3.10A) protonating the glycosidic oxygen and successively forming glycosyl enzyme complex that will lead to the product formation. Whereas, in case of mutant E136A and E226A the side chain of alanine has the methyl group (-CH_3) which is incapable to act as a general

acid (Fig. 3.3.10B) or as a nucleophile in case of E225A (Fig. 3.3.10C), therefore, no transfer of the electron taken place and no successive intermediate is formed. This study confirmed the predicted residues in *in silico* study i.e. Glu136 and E225 are the key catalytic amino acid residues of *Ct*Xyn30A.

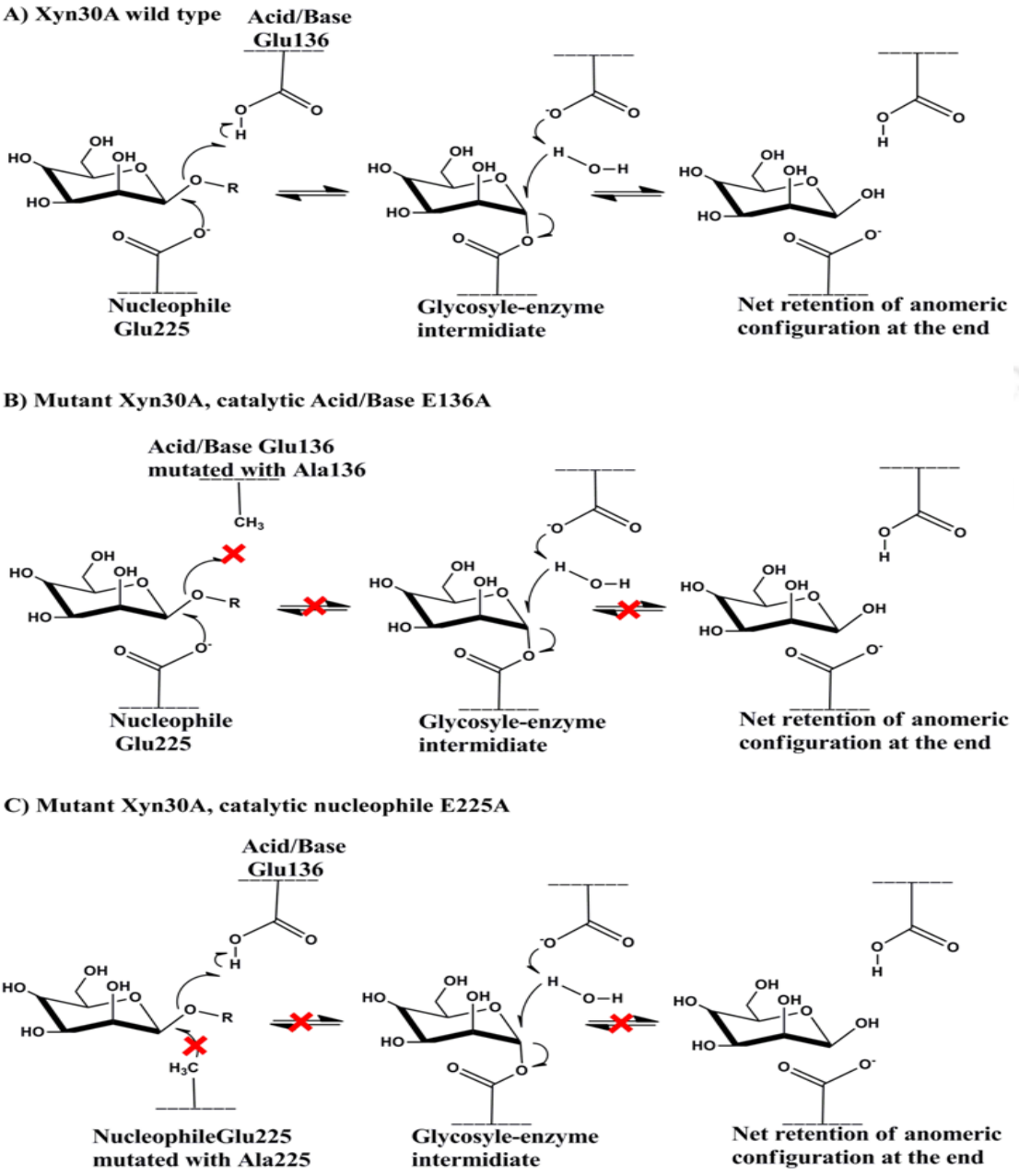


Fig. 3.3.10 Schematic representation of catalytic mechanism of *Ct*Xyn30A **A)** showing the retention type mechanism in wild type **B)** displaying the condition in mutant E136A of *Ct*Xyn30A **C)** displaying the condition in mutant E225A of *Ct*Xyn30A.

Conclusions

The enzymes identified and characterized in this study are the catalytic modules of full-length *Ct*XynGH30 and truncated derivative *Ct*Xyn30A of *Clostridium thermocellum* cellulosome and are members of family GH30-8 (or H). Biochemical analysis demonstrated that both *Ct*XynGH30 and *Ct*Xyn30A are active over broad range of pH from neutral to alkaline with an optimum pH at 6.0. The thermo-stability experiments showed that *Ct*XynGH30 and *Ct*Xyn30A are thermophilic xylanase displaying stability at a higher range of temperature (from 60 to 80ºC) with an optimum temperature of 70ºC. The substrate specificity analysis of *Ct*XynGH30 and *Ct*Xyn30A clearly showed that they act on xylan based polysaccharides with xylose as a monomeric unit. Moreover activity assays results on synthetic *p*NP derivatives did not show any activity eliminating the possibility of any exo-xylanase activity of the enzymes. Time dependent TLC analysis (beechwood xylan) hydrolysed by *Ct*XynGH30 displayed series of xylo-oligosaccharides confirming the endo-acting nature of the enzyme. The TLC analysis of linear xylo-oligosaccharides (xylohexaose) could not yield any further subsequent small xylo-oligosacharide signifying the requirement of the glucuronic acid substitution for the action of *Ct*XynGH30. The TLC profile with different polysaccharides displayed that *Ct*XynGH30 is capable of hydrolyzing not only xylan substrate substituted with glucuronic acid but also the polysaccharides decorated with arabinose as side chains. ^{1}H proton NMR study of hydrolyzed products from beechwood xylan generated by *Ct*XynGH30 showed the presence of acidic xylo-oligosaccharides having methyl glucuronic acid predominantly present penultimate to the reducing end terminal of xylose. The enzyme assays performed with the mutants of *Ct*Xyn30A (E136A and

E226A) displayed negligible activity towards the beechwood xylan. This indicated that both the residues (E136 and E226) are essential for the catalysis and further confirmed the results predicted by the *in silico* study that the glutamate 136 and 226 are the key catalytic residues of *Ct*Xyn30A.

References

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

Cavagna, F., H. Deger, and J. Puls. 1984. 2D-N.M.R. analysis of the structure of an aldotriouronic acid obtained from birch wood. Carbohydrate Research 129:1-8.

Chakraborty, S., Fernandes, V.O., Dias, F. M. V., Prates, J.A.M., Ferreira, L.M.A., Fontes, C.M.G.A., Goyal, A., Centeno, M.S.J. (2015) Role of pectinolytic enzymes identified in *Clostridium thermocellum* Cellulosome. PLOS One, 10-e0116787.

Bayer, E.A., Kenig, R., Lamed, R. (1983) Adherence of *Clostridium thermocellum* to cellulose. Journal of Bacteriology, 156, 818-827.

Dodd, D., Cann, I.K. (2009) Enzymatic deconstruction of xylan for biofuel production. Global Change Biology Bioenergy, 1, 2-17.

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

Ghosh, A., Verma, A. K., Rao, J. M. T., Shukla R., Goyal, A. (2014) Recovery and purification of oligosaccharides from coprameal by recombinant endo-β-mannanase and deciphering molecular mechanism involved and its role as potent therapeutic agent. Molecular Biotechnology, 57:111-127

Ghosh, A., Luis A. S., Bras, J.L.A., Fontes, C.M.G.A., Goyal, A. (2013) Thermostable recombinant β-(1→4)-mannanase from *Clostridium thermocellum*: biochemical characterization and manno-oligosaccharides production. Journal of Agricultural and Food Chemistry, 61, 12333-12344.

Larson, S.B., Day, J., Barba de la Rosa, A.P., Keen, N.T., McPherson, A. (2003) First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis. Biochemistry, 42, 8411-8422.

Kardosiova, A., Matulova, M., Anna, M. (1997) (4-O-Methyl-D-glucurono)-D-xylan from *Rudbeckia fulgida, var. sullivantii* (Boynton et Beadle) Carbohydrate Research, 308, 90-105

Nelson, N. (1944) A photometric adaptation of the Somogyi method for the determination of glucose. Journal of Biological Chemistry, 153, 375-380.

Rhee, M. S. Wei, L., Sawhney, N., Rice, J. D., St. John, F. J., Hurlbert J. C., Preston J. F. (2013) Engineering the xylan utilization system in *Bacillus subtilis* for production of acidic xylooligosaccharides, Applied and Environmental Microbiology, 3, 917-927

Sainz-Polo, M.A., Valenzuela, S.V., Gonzalez, B., Pastor, F.I., Sanz-Aparicio, J., (2014) Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity. Journal of Biological Chemistry, 289, 31088-31101.

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

Shallom, D., Shoham, Y. (2003) Microbial hemicellulases. Current Opinion in Microbiology 6, 219-228.

Songsiriritthigul C., Lapboonrueng S., Pechsrichuang P., Pesatcha P., Yamabhai M. (2010) Expression and characterization of *Bacillus licheniformis* chitinase (ChiA), suitable for bioconversion of chitin waste. Bioresource Technology, 101, 4096-4103

Somogyi, M., (1945) A New Reagent for the Determination of Sugars. Journal of Biological Chemistry, 160, 61-68.

St John, F.J., Dietrich, D., Crooks, C., Pozharski, E., Gonzalez, J.M., Bales, E., Smith, K., Hurlbert, J.C. (2014) A novel member of glycoside hydrolase family 30 subfamily 8 with altered substrate specificity. Acta Crystallographica Section D Biological Crystallography, 70, 2950-2958.

St John, F.J., Hurlbert, J.C., Rice, J.D., Preston, J.F., Pozharski, E. (2011) Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan-xylanohydrolase. Journal of Molecular Biology, 407, 92-109.

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. Journal of Bacteriology, 188, 8617-8626.

St John, F.J., Gonzalez, J.M., Pozharski, E. (2010) Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. FEBS Letters, 584, 4435-4441.

St John, F.J., Dietrich, D., Crooks, C., Pozharski, E., Gonzalez, J.M., Bales, E., Smith, K., Hurlbert, J.C. (2014) A novel member of glycoside hydrolase family 30 subfamily 8 with altered substrate specificity. Acta Crystallographica, Section D, Biological Crystallography, 70, 2950-2958.

Valenzuela, S.V., Diaz, P., Pastor, F.I. (2012) Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. Applied and Environmental Microbiology, 78, 3923-3931.

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*. Molecular Biology, 48, 278-286.

Vrsanska, M., Kolenova, K., Puchart, V., Biely, P. (2007) Mode of action of glycoside hydrolase family 5 glucuronoxylan xylanohydrolase from *Erwinia chrysanthemi*. FASEB Journal, 274, 1666-1677.

Urbanikova, L., Vrsanska, M., Kristian, B. R., Krogh, M., Hoff, T., Biely, P. (2011) Structural basis for substrate recognition by *Erwinia chrysanthemi* (GH) 30 glucuronoxylanase. FASEB Journal, 277, 2105-2116.

Chapter 4

Structural insights into native and ligand bound states of a family 30 glycosyl hydrolase glucuronoxylan-xylanohydrolase (*Ct*Xyn30A) of *Clostridium thermocellum*: Bioinformatics and X-Ray crystallographic structure analyses

4.1 Introduction

Xylan represents the second most abundant polysaccharides and major component of plant cell wall. It could constitute up to more than of 30% of its dry weight in the land plants. It is heterogeneous in nature made up of different monomeric units comprise of various side groups like D-xylose, D-mannose, L-arabinose, D-glucose, D-galactose, D-glucuronic acid, and D-galacturonic acid. The complex hemicellulosic polysaccharides classified according to released of monomeric unit after it get hydrolysed like if it is released xylose, it is a xylan; and likewise, hemicelluloses include mannans, arabinans, glucans, and galactans (Ebringerov, 2005; Whistler and Richards 1970). One of the major constituent of land plant is glucuronoxylan, mainly present in hardwood, soyabean hull, grasses, cereals and legumes plants. The glucuronoxylan from monocot (grasses and cereals) have higher degree of substitution of L-arabinose, p-coumaroyl and feruloyl groups as compared to glucuronoxylan from hardwood (Hurlbert and Preston, 2001).

The xylan main chain gets substituted at C-2 position with a 1, 2-linked 4-O-methyl-β-D-glucuronic acid residue, while 70% is acetylated at C-2, C-3, or both. On an average the birchwood xylan contains more than 1 mol of acetic acid per 2 mols of xylose. The presence of these acetyl groups is responsible for the partial solubility of xylan in water (Anna Ebringerova, 2000; Larson *et al*., 2003). Glucuronoxylan-xylanohydrolase requires this methyl glucuronic acid or glucuronic acid substitution to act on the main chain xylan, therefore, effectively degrade the acidic substituted xyalns. So far four entries of structurally characterized proteins from family GH30 subfamily 8 are available in the CAZy database. In brief, protein XynA from *Erwinia chrysanthemi* (Larson *et al*., 2003) and XynC from *Bacillus subtilis* (St John *et al*., 2009). Both preferably hydrolyse glucuronoxylan to branched xylo-oligosaccharides and showed negligible activity on linear β-1-4 linked xylo-oligosaccharides. Two more structure of xylanase of the same subfamily are recently solved that is from *Clostridium papyrosolvens* (*Cp*C71) which displayed low specificity for the glucuronoxylan (St John *et al*. 2014) and Xyn30D from *P. barcinonensis* acting on the branched (glucuronic and methyl glucuronic acid substituted) xylo-oligosaccharides but showed no catalytic activity on arabinose-substituted xylans. Regarding *Clostridium thermocellum*, various potential carbohydrate active enzymes *viz*. cellulase, xylanase, esterases, lyases and transferase were characterized time to time. So far as per our knowledge, there is no structural and biochemical report available on glucuronoxylan-xylanohydrolase enzyme from *clostridium thermocellum*. The present study describes the insight into the structural aspect of *Ct*Xyn30A representing the first glucuronoxylan-xylanohydrolase from *C. thermocellum* by means of preliminary bioinformatics analysis; which includes,

molecular modelling and its validation by various parameters, molecular dynamics simulation of modelled structure and molecular docking with xylo-oligosaccharides. The study also includes the high resolution 3-dimensional structure of *Ct*Xyn30A determination.

4.2 Materials and methods

4.2.1 computational and molecular docking study of *Ct*Xyn30A

4.2.1.1 Homology modeling

The catalytic module *Ct*Xyn30A was modelled by employing a computer program Modeller9v8 (Sali *et al*., 1995). It uses a knowledge based comparative protein structure modeling hence the first step is to align the query sequence with the best matched template of known protein structures. XynC belonging to family 30 glycoside hydrolase from *Bacillus subtilis*168 (PDB id: 3GTN) was selected as the template for structure modeling based on its sequence and functional homology. Modeller derives the restraints from the template and database of known structure. 10 independent models were generated after optimization of molecular probability density functions. Refinement of loops of the modeled structure was achieved with the help of loop optimization methods and after each cycle of loop refinement discrete optimized protein energy (DOPE) score was generated. This process was repeated in an iterative fashion until a negative DOPE score was attained (lower than -1) (Fiser *et al*., 2000). The model with the lowest discrete optimized protein energy score was chosen for further studies.

4.2.1.2 Model refinement and quality assessment

Energy minimization was carried out on modeled protein in order to relax any unfavorable bond angle and bad contacts before performing docking. Energy minimization was performed by a steepest descent algorithm with GROMOS96 43a1 force field and simple point charge (SPC) water model with GROMACS 4.0.7 package (http://www.gromacs.org/). After defining simulation box and solvent system (here water molecules) around protein, charge on the protein molecule was

neutralized by adding counter-ions (Na^+ was added) by replacing same number of water molecule which are at least 3.50 Å from the protein surface. Final structure after energy minimization was validated on structure analysis and verification server (SAVES) at NIH–MBI laboratory (http://nihserver.mbi.ucla.edu). Ramachandran plot which graphically represent the combination of possible and permitted torsional angles, backbone phi (ϕ) and psi (ψ) dihedral angles, was obtained using the PROCHECK program (Laskowski, *et al*., 1993). The Verify_3D program was used to determine a compatibility of an atomic model (3D) with its own amino acid sequence (1D). First it is categorized each residue into structural classes based on its location and environment (alpha, beta, loop, polar, nonpolar etc.) then it generates a score by comparing each residues with a collection of good structures as a reference (Bowie, *et al*., 1991). Stereo chemical quality and errors in protein structure were accessed at ProSA (Protein Structure Analysis) web Server (https://prosa.services.came.sbg.ac.at/prosa.php). It calculates a z-score which represents the overall model quality. Inaccurate protein structures are placed outside the range of values in the scatter plot (Sippl, 1993; Wiederstein and Sippl, 2007).

4.2.1.3 Molecular dynamics

The energy-minimized *Ct*Xyn30A model was placed inside a cubic box of single point charge (SPC) water molecule. The net charge of the system was neutralized with counter ions. The system was equilibrated for 50 pico second in NVT ensemble by restraining the solute atoms. Then the system was equilibrated for 50 pico second by NPT (number of particles (N), system pressure (P) and temperature (T) constant /conserved) ensemble twice, first with restraints followed by without restraints. Production run was carried out for 2 nano-seconds with NPT ensemble

using 2 femtosecond of integration time. The linear constraint solver (LINCS) algorithm was used to constrain the bonds involving hydrogen atoms (Hess, 2008). Conformations of *Ct*Xyn30A model at regular intervals throughout the production run were analyzed as time-dependent function to check whether the *Ct*Xyn30A model is energetically stable in the solvent system.

4.2.1.4 Prediction of active site

Active site involves key amino acid residues which participate in the reaction mechanism. Generally these residues are conserved within the family. Therefore, to identify the active site residues, structure superimposition of *Ct*Xyn30A with the template protein of family GH30 using PyMOL (0.99) was carried out. (The PyMOL Molecular Graphics System. San Carlos, 2002). Moreover, whether these active site residues are conserved among the GH30 family or not, we performed and analyzed multiple sequence alignment (MSA) taking all the structurally characterized representative members of GH30 subfamily H. The information on structurally related proteins of family GH30-8 was obtained from the CAZy database. The sequence and PDB coordinate files were retrieved from PDB (http://www.rcsb.org/pdb/). The alignment was generated by Clustal_X program (Thompson *et al*., 1997) and figure was produced by ESpript (http://espript.ibcp.fr).

4.2.1.5 Docking study on modeled CtXyn30A

Molecular docking studies were performed using Autodock 4.2.1. Autodock uses novel and robust docking method by implementing a new scoring function that estimates the free energy change during ligand binding (Goodsell and Olson, 1990). Linear xylo-oligosaccharides and branched glucuronoxylan were taken as ligand molecules. The following protocol was used to perform docking. Ligand PDBs were

obtained from pubchem (http://pubchem.ncbi.nlm.nih.gov) or from ligand bound protein complex (http://www.rcsb.org/pdb/). At first, the PDB files of protein and ligand were converted to PDBQT files by assigning gasteiger charges. Thereafter, non-polar hydrogens were merged and their charges were assigned to carbon atoms. Grid box was assigned around active site having dimension of 66, 46, 48 (*x*, *y*, *z* coordinates) with 0.375 Å grid point spacing. Lamarckian Genetic Algorithm (LGA) was implemented for docking simulation and conformational search (Morris, *et al*., 1998). Number of GA runs was set to be 30. Initial population size were 150, maximum energy evaluation per run 2500000 and maximum number of generations was set to be 27000. After successful completion of docking runs, docked conformation and interaction energies were saved and analyzed from the largest cluster having a minimum lowest free energy of binding (ΔG). 2D lig-plot (http://www.ebi.ac.uk/pdbsum/) of protein-ligand complex of best docked conformation was generated for the better understanding and representation of interaction.

4.2.2 Cloning, crystallization and structural analysis of *Ct*Xyn30A

4.2.2.1 Cloning, expression and purification of CtXyn30A

The complete methodology of cloning, expression and purification of *Ct*Xyn30A has been described in details in the Chapter 2. The additional exercises deployed to obtain highly purified and concentrated protein for seeding crystal drops are as follows:

The recombinant *Ct*Xyn30A containing a His_6 tag was purified from the cell-free extract by immobilized metal-ion affinity chromatography (IMAC) using a Sepharose column (HiTrap Chelating, GE Healthcare) and was eluted using 50 mM

HEPES buffer (pH 7.5) containing 300 mM imidazole, 1 M NaCl, 3 mM $CaCl_2.2H_2O$. Purified *Ct*Xyn30A was buffer-exchanged into 50 mM HEPES buffer (pH 7.5) containing 200 mM NaCl, 5 mM $CaCl_2$using a PD-10 column and was then subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml/min. Purified *Ct*Xyn30A was concentrated using a 10 kDa cutoff centrifugal concentrator (Millipore, USA) and washed three times with 0.5 mM $CaCl_2$. The purity and molecular mass of the recombinant protein was analysed by 13% (w/v) SDS–PAGE (Laemmli, 1970). The protein concentration was estimated using a molar extinction coefficient (ε) of 91853 per M/cm with a NanoDrop 2000c spectrophotometer (Thermo Scientific).

4.2.2.2 Crystallization conditions for CtXyn30A

The crystallization conditions were screened by the sitting-drop vapour-diffusion method using the commercial screens Crystal Screen, Crystal Screen 2 and PEG/Ion from Hampton Research (California, USA), JCSG from Molecular Dimensions (UK) and an in-house-prepared sparse-matrix screen of 80 conditions using a Nano drop robotic dispensing system Oryx8 (Douglas Instruments, UK). Drops consisting of 0.9 ml *Ct*Xyn30A solution at either 16 or 32 mg/ml and 0.9 ml reservoir solution were prepared at 292 K (Verma *et al.*, 2013). The crystals were harvetsted using harvesting buffer comprise the crystallization buffer containing 5% higher concentration of the precipitant. The crystals were cryocooled in liquid nitrogen after soaking in the appropriate cryoprotectant (30%, v/v glycerol added to the crystallization buffer for a few seconds, except for the crystal grown in 0.2 M potassium sulfate, 20%(w/v) PEG 3350, which was cryo-cooled in Paratone.

4.2.2.3 Data collection, structure determination and refinement of CtXyn30A

*Ct*Xyn30A was crystallized in several conditions and data from all the different crystals obtained were collected on beamline ID29 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) (de Sanctis *et al*., 2012), Diamond Light Source (Harwell, UK) and at PROXIMA-1 at SOLEIL (Orsay, France) using a PILATUS 6M detector with the crystals cooled to 100 K using a Cryostream (Oxford Cryosystems) (Verma *et al*., 2013). In addition, several *Ct*Xyn30A crystals, crystallized in different conditions, were soaked with xylohexaose or cellohexaose for 1 h, 2 h and 12 h. The data collected from the ESRF, Diamond and SOLEIL were processed using the programs iMOSFLM (Battye *et al*., 2011) and AIMLESS (Evans, 2011) from the CCP4 suite (Winn *et al*., 2011) or XDS (Kabsch, 2010). The statistics of data collection are given in Table 4.3.2. All of the crystals tested belonged to space group P1 with one molecule in the asymmetric unit and a solvent content ranging from 45 to 47% and had similar data-collection statistics, despite being grown in different conditions and having different crystal morphologies. The data collected at SOLEIL were from the best diffracting crystal (Fig. 4.3.8D) and were used to solve the *Ct*Xyn30A structure. 470° of data were collected with a Δφ of 0.2°. The crystal was broken/multiple and a second low-resolution data set was needed to index and to find the correct lattice before integrating the high-resolution data with XDS. The resolution was cut at a conservative 1.4 Å (the $CC_{1/2}$ value is below 0.5 for higher resolution data). The slightly high R_{merge} value at this resolution is due to radiation damage from a longer exposure for a P1 data collection. The Matthews coefficient of 2.33$Å^{-3}$ Da^{-1} (Matthews, 1968) indicated the presence of one molecule in the asymmetric unit with

a solvent content of 47.2%. The program Phaser (McCoy *et al*., 2007) was used to solve *Ct*Xyn30A structure by molecular replacement approach using, the crystal structure of XynC from *Bacillus subtilis* 168 (PDB id: 3GTN) as search model (St John *et al*., 2009) with a final solution giving LLG score of 181 and a TFZ of 24.5 as described in Verma *et al*. 2013. The structure was further refined and the refinement statistics are shown in Table 4.3.3. A model comprising 1 chain with 389 residues (native protein plus 3 residues in the N-terminal due to the cloning construct) was built from the initial map with program COOT (Emsley *et al*., 2010). Water molecules were added by visual inspection of the electron density maps, according to hydrogen bond criteria and as indicated by mFo-DFc maps to a total of 647 water molecules, and refined with REFMAC5 (Murshudov *at el*., 2011), as deemed appropriate from the r behaviour of the cross-validation (R_{free}) subset of reflections (5%). The final round of refinement was performed using the TLS/restrained refinement procedure using 3 segments (His-3 to Gln10, Val11 to Pro295 and Gly296 to Val386) as determined by the TLSMD server (Painter and Merritt, 2006), which gave final R and R_{free} factors of 0.148 and 0.172, respectively. The final model was verified with PROCHECK (Laskowski, *et al*., 1993) and checked and validated during submission to the PDB. The r.m.s. deviations of the bond lengths was 0.020 Å and bond angles 1.837 degrees. This structure of *Ct*Xyn30A (PDB id: 4CKQ) crystal figure 4.3.8D was used as a model for molecular replacement, to determine other *Ct*Xyn30A structures crystallized in different conditions (Fig. 4.3.9). The data were processed using XDS (Kabsch, 2010) and scaled with AIMLESS (Evans, 2011) from CCP4 suite (Winn *et al*., 2011). All data-collection statistics are indicated in Table 4.3.2. The three-dimensional structures were solved directly using the program Phaser

(McCoy *et al*., 2007) and refinement was carried out with REFMAC5 (Vagin *et al*., 2004) or using Phaser and AUTOBUILD (Terwilliger *et al*., 2008). The final round of refinement for structures with PDB codes 4uqe, 4uqd, 4uqc and 4uqb was performed in the PDBredo web server (Joosten *et al*., 2012). For the remaining structures, the final round of refinement was performed using TLS/restrained refinement using 3 segments (His-3-GLN10, Val11-Pro295, Gly296-Val386). In all *Ct*Xyn30A structures, both C- and N-terminal regions are well defined in the electron density maps as well as all the side chains. The refinement statistics are indicated in Table 4.3. Visualization and analysis of structures, as well as generation of figures, were conducted using both PyMOL (Schrodinger, 2010) and UCSF Chimera (Pettersen *et al*., 2004).

Data deposition: Coordinates and observed structure factor amplitudes for *Ct*Xyn30A structures have been deposited in the Protein Data Bank in Europe (PDBe), www.ebi.ac.uk/pdbe (PDB id code: 4CKQ, 4UQA, 4UQE, 4UQD, 4UQ9, 4UQC and 4UQB).

4.3 Results and Discussion

4.3.1 *In silico* structure analysis of *Ct*Xyn30A

4.3.1.1 Homology modeling and structure validation of CtXyn30A

Blastp of *Ct*Xyn30A in PDB database at NCBI (http://blast.ncbi.nlm.nih.gov/) for sequence similarity search showed a top hit for crystal structure of XynC family GH30 from *Bacillus subtilis*168 (PDB id: 3GTN) having a 72% sequence identity (St John *et al*., 2009). Before performing docking, modeled *Ct*Xyn30A was validated with different parameters like Ramachandran plot which showed 91% amino acid residues lies in most favourable region and 9% residues in additional allowed region, which implied that structure-backbone dihedral angles, phi (ϕ) and psi (ψ) occupied favourable positions (Fig. 4.3.1A). ProSA result as shown in the figure. 4.3.1B indicated that the modeled protein is error free and reside in the X-ray zone with scores of −9.58. Verify_3D (Luthy *et al*., 1992) result depicted that all 386 amino acid residues have average 3D–1D profile score greater than 0.22 which illustrated that no amino acid is in bad contact (Fig. 4.3.1C).

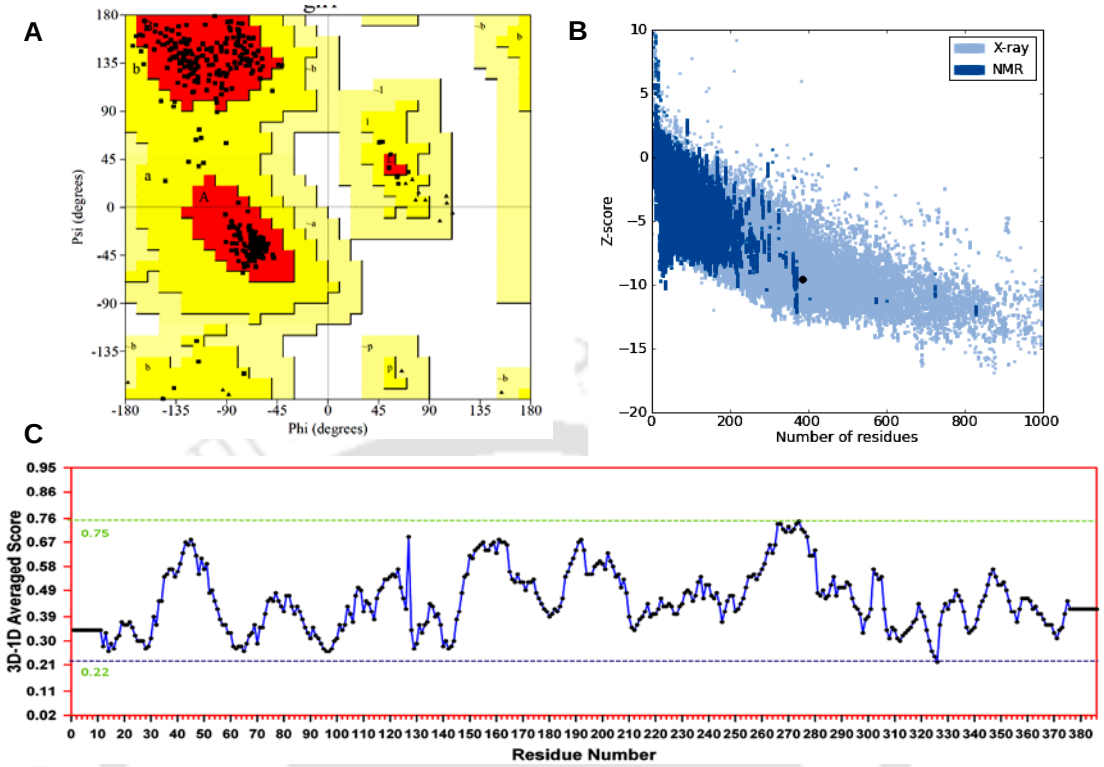


Fig. 4.3.1 Structure validation and quality assessment of *Ct*Xyn30A by (A) Ramachandran plot (B) ProSA web server (C) Verify_3D.

4.3.1.2 Molecular dynamics simulation of CtXyn30A

Molecular dynamics (MD) simulation was performed to ensure the "Compactness" of the protein throughout the simulation. Result suggested that *Ct*Xyn30A undergoes a significant change in the initial 1.0 nanosecond of simulations and converged with fluctuations less than 0.3 Å, which indicated the stable conformation of *Ct*Xyn30A model (Fig. 4.3.2).

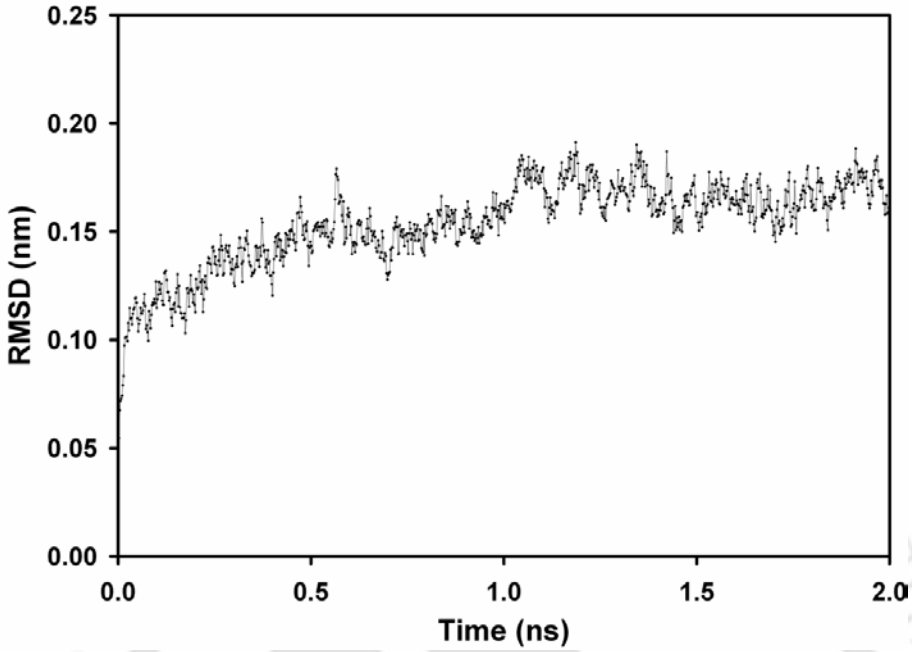


Fig. 4.3.2 MD simulation of modeled *Ct*Xyn30A shows that the energy-minimized structure is energetically stable under the simulation conditions.

The r.m.s. deviation between energy minimized model of *Ct*Xyn30A and final structure from MD was 1.7 Å. MD simulation suggested that all the parameters are well within the qualitative limit of compactness and stability during the simulation time signifying the energy minimized *Ct*Xyn30A model is satisfactory for molecular docking simulation.

4.3.1.3 Overall structural features of CtXyn30A

Modeled *Ct*Xyn30A showed common $(\beta/\alpha)_8$ barrel folding structure same as present in GH5 and GH10 families. The eight β-sheets surrounded by α-helix form a barrel shaped structure which makes a cavity to accommodate the substrate. However, there was a clear distinguishing structural feature in GH30 family observed that it had obligate 'side β-structure' (Fig. 4.3.3A) which is not associated with GH5 and GH10 families (Henrissat *et al*., 1996; Nishitani and Nevins, 1991; Pollet *et al*., 2010).

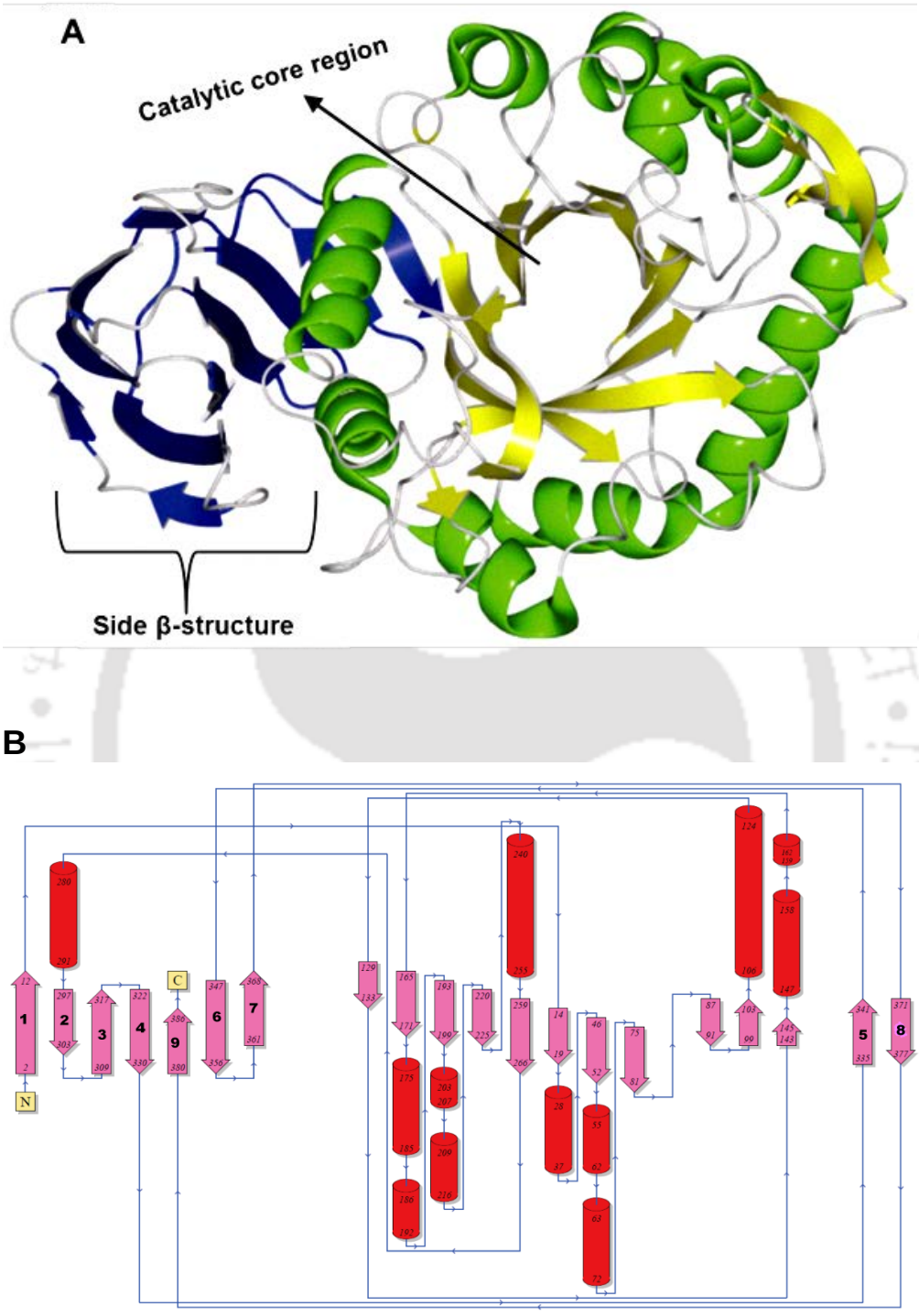


Fig. 4.3.3 **(A)** Overall 3-D cartoon representation of modeled *Ct*Xyn30A from front side (C-terminal) depicted the $(\beta/\alpha)_8$ barrel fold structure showing the catalytic core in which encircled β-sheets are surrounded by α-helix along with peculiar side β-structure. **(B)** Topology diagram of *Ct*Xyn30A represents a mutual orientation of secondary structure in 2D mentioning 1-9 arrows represent β-sheet of side β-structure; other arrows represent encircled β-sheets and cylinders representing the α-helices.

The structure analysis of the side β-structure depicted a β-strand rich structure having 9 β-strands (Fig. 4.3.3B) running anti-parallel to each other resembling a carbohydrate binding module (CBM). Moreover, it shows the first β-sheet β1 (written 1 in Fig. 4.3.3B) represent the N-terminal side as after this the amino acid chain enters $(\beta/\alpha)_8$ catalytic domain and β-strands from β2–β9 represent the C-terminal region. The same arrangement of side β-structure was observed in XynC (St John *et al*., 2010). Regarding the function and importance of side β-structure, it is demonstrated that it will act as a CBM thereby having the function to enhance the performance of the catalytic domain by keeping the substrate in close proximity (Boraston *et al*., 2004). The first evidence on the role of side β-structure was confirmed from the ligand 4-O-methyl glucuronoxylan ($MeGX_n$) bound structure of XynC which clearly indicated that this motif selectively targeting $MeGX_n$ as substrate (St John *et al*., 2011).

Concerning the importance of side β-structure, different truncated recombinant of the Xyn30D catalytic module was constructed which lacked full-length side β-structure. It was found that the truncated recombinant clone which lacked full-length side β-structure does not show any hydrolytic activity on xylan. It implied that the associated side β- domain is quite essential for catalysis (Valenzuela *et al*., 2012). In brief, an evolving function of family GH30 subgroup H is that it has inbuilt CBM along with the catalytic module which facilitates substrate binding and pivotal for activity of catalytic module.

***4.3.1.4 Catalytic core of Ct*Xyn30A**

The catalytic core of *Ct*Xyn30A formed by $(\beta/\alpha)_8$ barrel folding encircled β-sheets residues 13–291 aa (Fig. 4.3.3A). Close inspection of $(\beta/\alpha)_8$ barrel folding structure of *Ct*Xyn30A revealed that the loops connecting α-helix to the β-sheets are shorter toward N-terminal (or called back side) (Fig. 4.3.3A) than the loops at the C-terminus (or front side of the protein). These longer loops located at C-terminus probably facilitate the entrance of large and branched oligosaccharides (Larson *et al*., 2003). Aromatic amino acid residues of active site which shape the binding cleft and their probable role in catalysis of substrate was depicted by superimposing the *Ct*Xyn30A with XynC and XynA with backbone r.m.s. deviation of 0.3 Å and 1.2 Å, respectively (Holm and Rosenstrom, 2010). Analysis revealed that aromatic residues Trp23, Trp81, Tyr110, Trp143, Tyr200, Trp263 and Tyr264 of mature *Ct*Xyn30A protein (not included 30 aa of signal peptide) corresponding to Trp55, Trp113, Trp168, Tyr172, Tyr232, Trp289, and Tyr290 of XynA, which constitute the binding cleft (Larson *et al*., 2003). The analysis of the MSA results showed that the mentioned aromatic residues are highly conserved within the GH30 family (displayed as inverted triangle, Fig. 4.3.4).

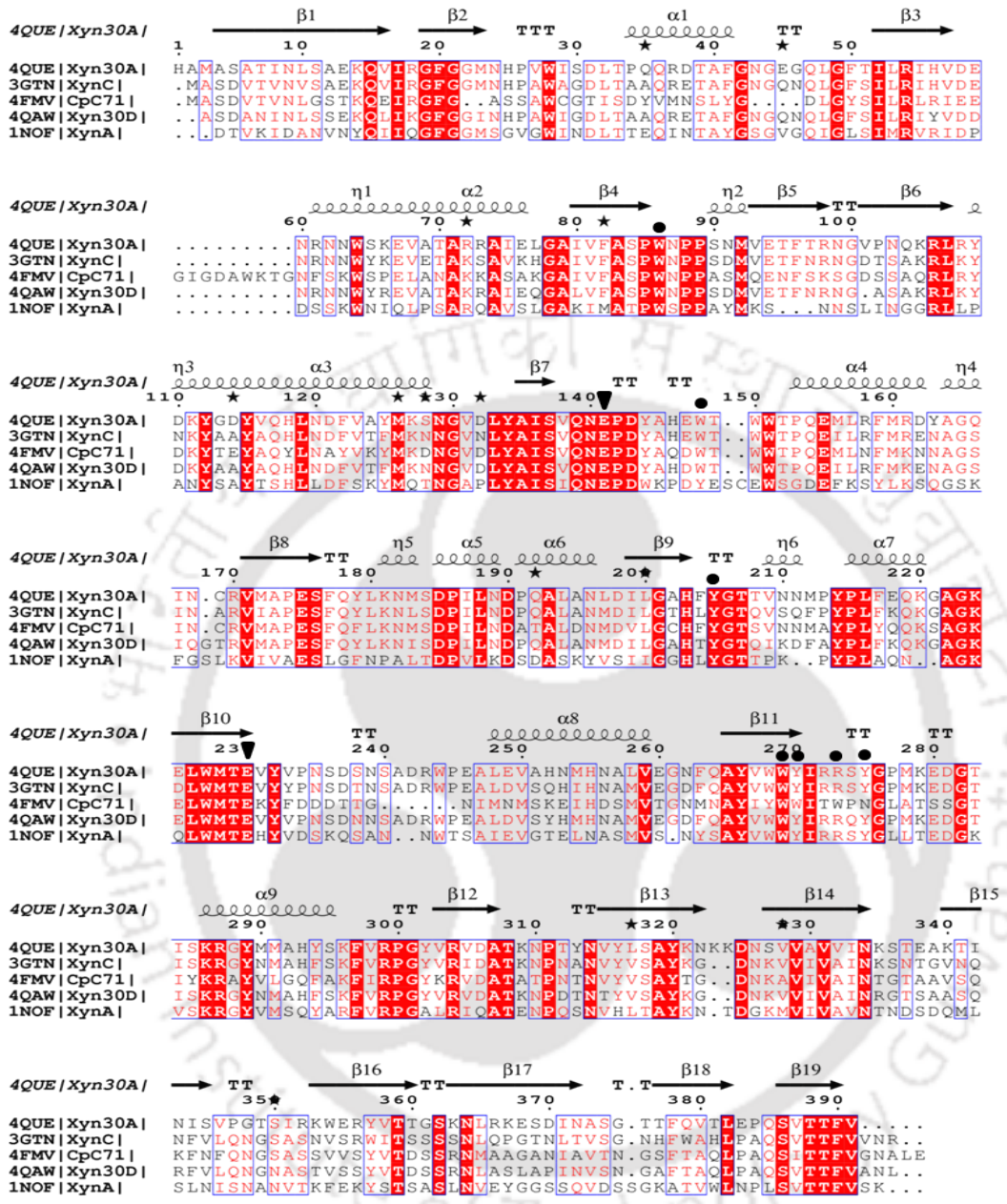


Fig. 4.3.4 Structure-based sequence alignment of all the structurally characterized members of GH30-8 family, with PDB entries, *Ct*Xyn30A (PDB id, 4QUE) from *C. thermocellum*, XynC (PDB id, 3GTN) from *B. subtilis* 168, *Cp*C71 (PDB id, 4FMV) from *C. papyrosolvens* C71, Xyn30D (PDB id, 4QAW) from *P. barcinonensis* and XynA (PDB id, 1NOF) from *E.chrysanthemi*. AYVW Showing conserved and SNS semi-conserved amino acid residues. and represents the corresponding secondary structure (alpha helix and beta sheet respectively) of *Ct*Xyn30A

4.3.1.5 Probable mechanism of catalysis

Hydrolysis of a glycosidic bond occurs through either the retention or by inversion of anomeric configuration. The similarity between these two reactions mechanism is both form an oxacarbenium-ion-like transition states and needed a pair of carboxylic acids at catalytic site. Whilst there is subtle difference in both the mechanisms like in inverting glycoside the average distance between two carboxylic acid residues which acts as a proton donor /acid and the other as a nucleophile/base is of ~10 Å and the reaction occurs via a double-displacement mechanism wherein retaining mechanism the distance would be ~5.5 Å and reaction occurs via a single-displacement mechanism(Davies and Henrissat, 1995; McCarter and Withers, 1994; Rye and Withers, 2000). Catalysis of glycosidic bond in family GH30 follows the retention mechanism (http://www.cazy.org).

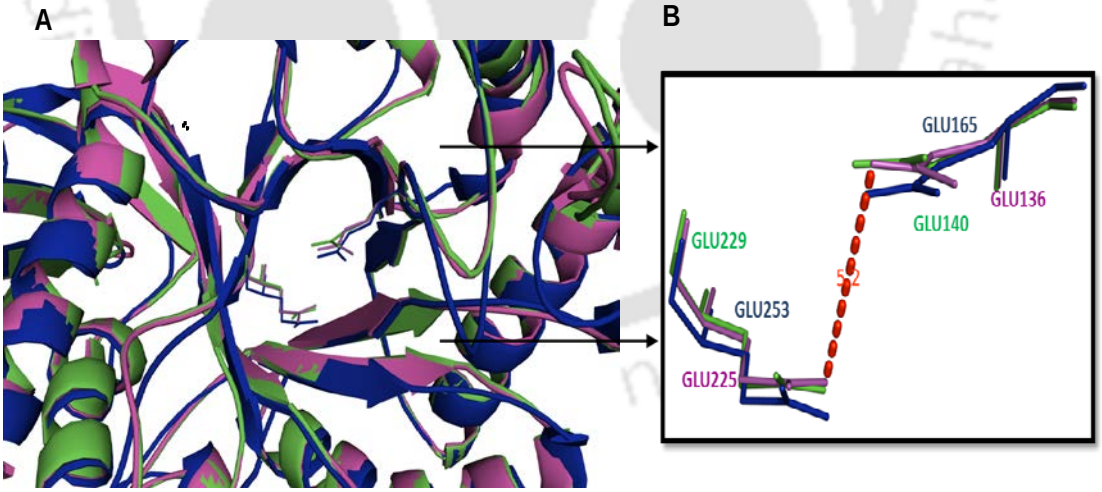


Fig. 4.3.5 **(A)** Ribbon representation of superimposed *Ct*Xyn30A (purpule), XynC (blue) and XynA (green colour) proteins. **(B)** Enlarged view of superimposed active site residues of template and target (XynC and XynA) proteins. The figure was generated using the PyMol program (http://www.pymol.org).

Study of superimposed structures (Fig. 4.3.5A and 4.3.5B) of *Ct*Xyn30A with XynC and XynA for finding the proper orientation and distance between the key catalytic residues showed that, Glu136 of *Ct*Xyn30A corresponds to Glu140 of XynC and Glu165 of XynA, which acted as acid/base while Glu225 of *Ct*Xyn30A matched with Glu229 of XynC and Glu253 of XynA, which acted as a nucleophile. Moreover, the distance between the carboxyl groups of catalytic glutamate was found to be 5.2 Å which is appropriate for retention type mechanism (Larson *et al*., 2003; St John *et al*., 2011). The MSA results confirmed that these active site catalytic residues are conserved within the family (marked as star in Fig. 4.3.4) and protein-ligand docking analysis showed these residues are directly involved in making polar contact with ligand molecule (Fig. 4.3.7).

4.3.1.6 Docking of substrate at the active site

The crystal structures of GH5 and GH10 bound with xylo-oligomer ligands having the catalytic core formed by $(\beta/\alpha)_8$ barrel folding structure have been reported (Correia *et al*., 2011; Leggio *et al*., 2000). It gives valuable information about residues interacting with the xylan backbone forming hydrogen bonds and hydrophobic interactions. After knowing the location of catalytic core and conserved active site residues of *Ct*Xyn30A, the docking study was carried out to obtain best fit conformation of different ligands. The results of docking were summarized in Table 4.3.1.

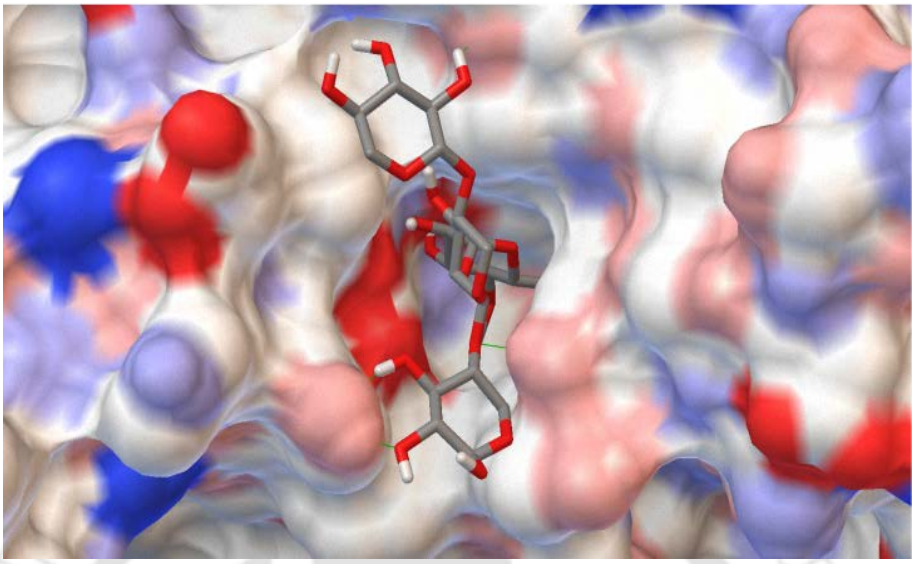


Fig. 4.3.6 Surface view of the active site cleft of *Ct*Xyn30A representing the binding mode with ligand 4-O-methyl glucoronoxylan ($MeGlcA^2Xyl_3$). Different color illustrated electrostatic potential of the enzyme surface negatively charged regions (red colour) and positively charged regions (blue). This figure was generated with the python molecular viewer program.

The ligand xylopentaose showed highest affinity with the free energy of binding of –9.39 kcal/mol. *Ct*Xyn30A in binding mode with ligand 4-O-methylglucuronoxylan ($MeGlc^2Xyl_3$) depicted the cavity formed by barrel shaped catalytic core which allowed ligands to access active site residues (Fig. 4.3.6). Enzyme-ligand complexes (Fig. 4.3.7) of the same inferred that many significant polar interactions are taking place between active site amino acids and ligand molecules like O2–OE2 Glu136 (3.02 Å), O2B–OH Tyr139 (3.07 Å), O3B–OH Tyr139 (2.31 Å), O6B–OH Tyr227 (2.83 Å), O5–OH Tyr227 (2.92 Å), O2B–ND2 Asn135 (3.17 Å), O2B–OE1 Glu225 (2.6 Å) and O3B–OE1 Glu225 (3.04 Å) while His141, Glu142, Trp143, His198, Tyr200, Val226 and Trp264 were found to be in hydrophobic interaction.

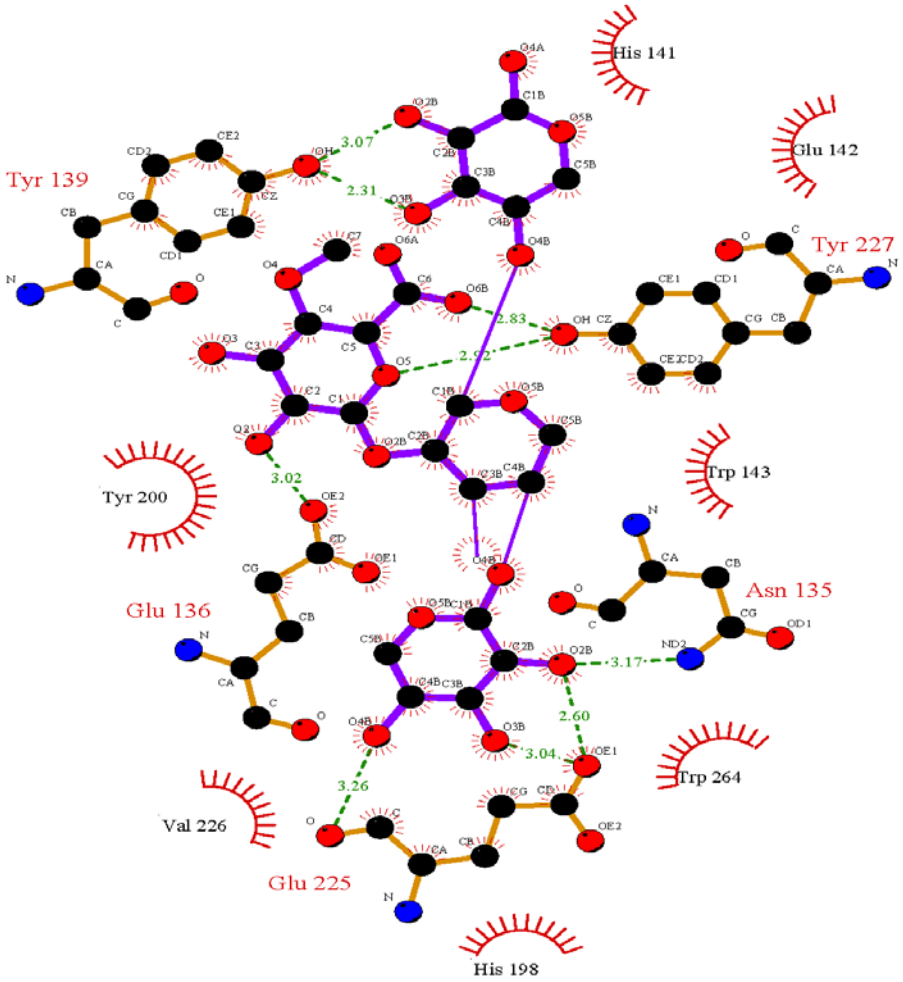


Fig. 4.3.7 Schematic 2D depiction of ligand 4-O-methyl glucoronoxylan ($MeGlcA^2Xyl_3$) interacting with active site amino acid residues of *Ct*Xyn30A protein. (▬▬) Dashed lines shows hydrogen bonds with labelled bond length, residues displaying arc with spokes making hydrophobic interaction, contacted atoms are shown with spokes radiating back.

Table 4.3.1 Amino acid residues making polar and hydrophobic interaction (underlined) on docking.

Ligand	Predicted free energy of binding region (ΔG) (Kcal/mol)	Polar interactions	Residue within 4Å
Xylobiose	-5.67	Tyr227, Glu225, Tyr139, Glu136	Phe172, Tyr200, His198, Trp264, Trp143, Asn135, Trp81
Xylotriose	-5.42	Tyr139, Glu136, Tyr200, Phe172, Tyr227	Trp81, Trp143, Trp264, Glu225, Asn135, His198, Val226, Phe199
Xylotetraose	-7.13	Asn135,Trp143 Glu54,Glu225, His20	Asn82, Trp81, Glu136, His198, Val52, His51, Asp53, Asn55 Glu142, Tyr227, Trp264, Tyr265, Trp23
Xylopentaose	-9.39	Glu136, Glu142, Trp81, Val52, His20	Tyr227, Tyr139, Trp264, Tyr265, Trp143, Pro80, Asn82, Asp53, Glu54, Trp23
Glucuronoxylan ($MeGlcA^2Xyl_3$)	-7.19	Tyr139, Tyr227, Glu225, Asn135, Glu136	Glu142, Trp264, Val226, Tyr200, His198, His141, Trp143

4.3.2 Crystallization and 3-dimensional structure analysis of native and ligand bound form of *Ct*Xyn30A by X-ray crystallography

4.3.2.1 Crystals of CtXyn30A under different conditions

The crystals of *Ct*Xyn30A appeared within 1 day, reaching their maximum size within a few days in many different conditions as displayed in figure 4.3.8.

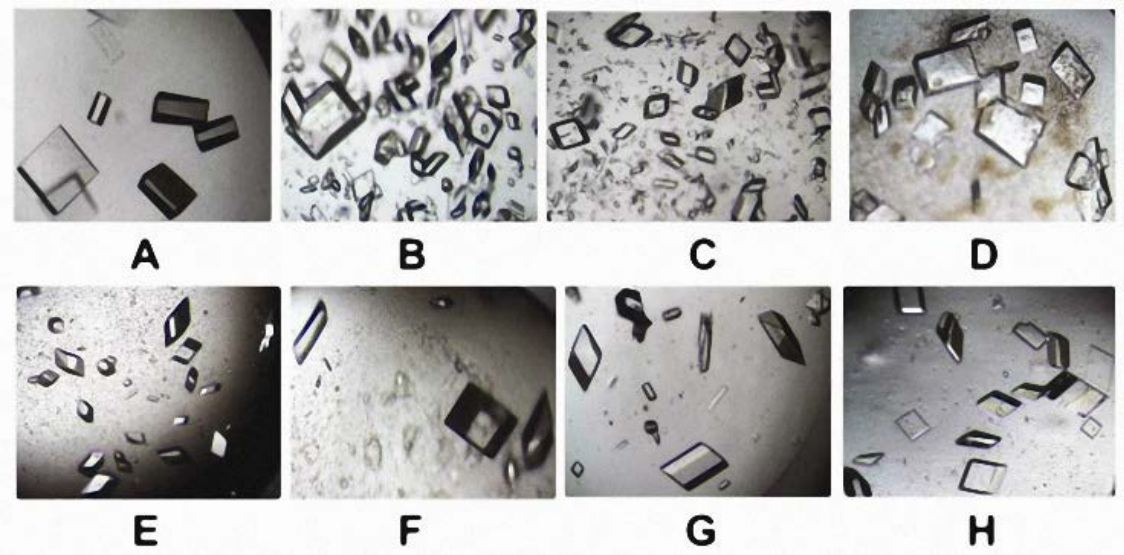


Fig. 4.3.8 *Ct*Xyn30A crystallized in different conditions. Crystals of *Ct*Xyn30A obtained by sitting-drop vapour diffusion method. **(A)** 2M ammonium sulfate; **(B)** 0.1 M HEPES pH 7.5, 10% polyethylene glycol (PEG) 6000, 5% (v/v) (±)-2- methyl-2,4-pentanediol; **(C)** 0.2 M sodium fluoride, 20% (w/v) PEG 3350; **(D)** 0.2 M potassium sulfate, 20% (w/v) PEG 3350; (PDB code 4ckq) **(E)** 0.2 M sodium acetate trihydrate pH 7.0, 20% (w/v) PEG 3350; **(F)** 2% (v/v) Tacsimate pH 4.0, 0.1 M sodium acetate trihydrate 4.6, 16% (w/v) PEG 3350; **(G)** 0.1M sodium formate, 12% (w/v) PEG-3350; **(H)** 0.15 M cesium chloride, 15% (w/v) PEG 3350. The largest crystals were approximately, 200x120x60 µm in size.

4.3.2.1 *Crystals of ligand bound CtXyn30A at different condition*

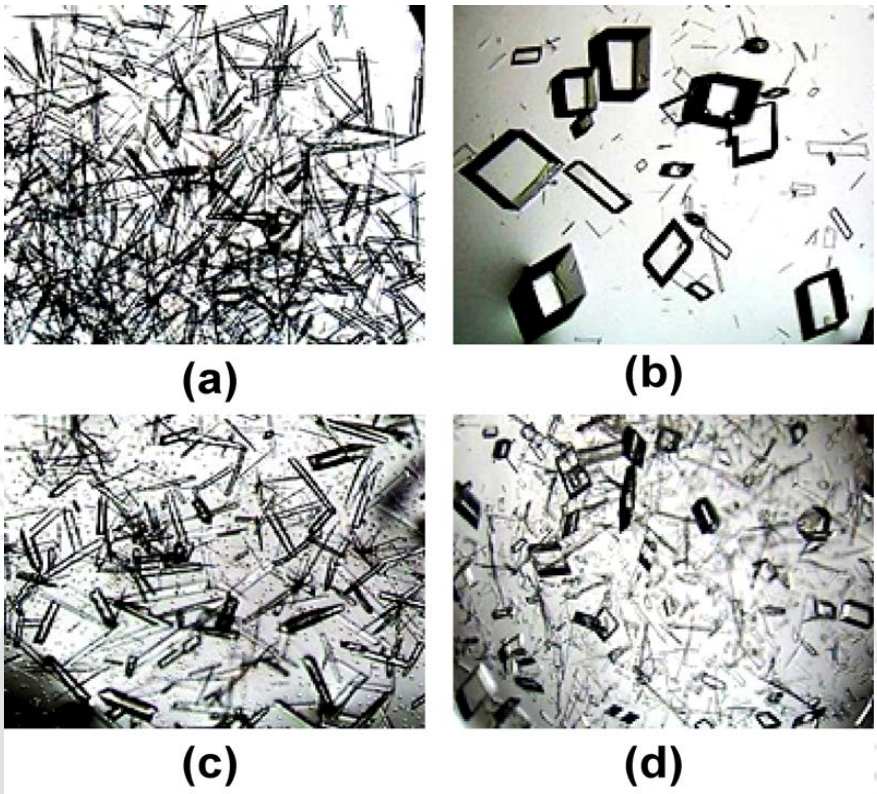


Fig. 4.3.9 The best results were obtained for crystals grown in the following conditions: **(a)** 0.2 M $CaCl_2$, 0.1 M HEPES 7.5, 25% (v/v) PEG 3350, drop of 2 µl of protein at 18 g/L + 2 µl of precipitant, hanging-drop vapour diffusion method (PDB codes 4uqe and 4uqd); **(b)** 0.1 M Tris-HCl 8.5, 8% (v/v) PEG 8000, drop of 0.9 µl of protein at 16.9 g/L + 0.9 µl of precipitant, sitting drop vapour diffusion method (PDB code 4uq9); **(c)** 0.4 M sodium/potassium tartrate, drop of 0.9 µl of protein at 16.9 g/L + 0.9 µl of precipitant, sitting drop vapour diffusion method; **(d)** 0.1 M HEPES 7.5, 10% (v/v) PEG 8000, 8% (v/v) ethylene glycol, drop of 0.9 µl of protein at 16.9 g/L + 0.9 µl of precipitant, sitting drop vapour diffusion method

The crystal diffracted to a resolution of at 1.1 Å (Fig. 4.3.10). All the important parameters like beamline used, wavelength, space group, unit cell parameters, resolution limits and other information of native and ligand bound form of *Ct*Xyn30A along with their PDB id were collected and shown in the Table 4.3.2

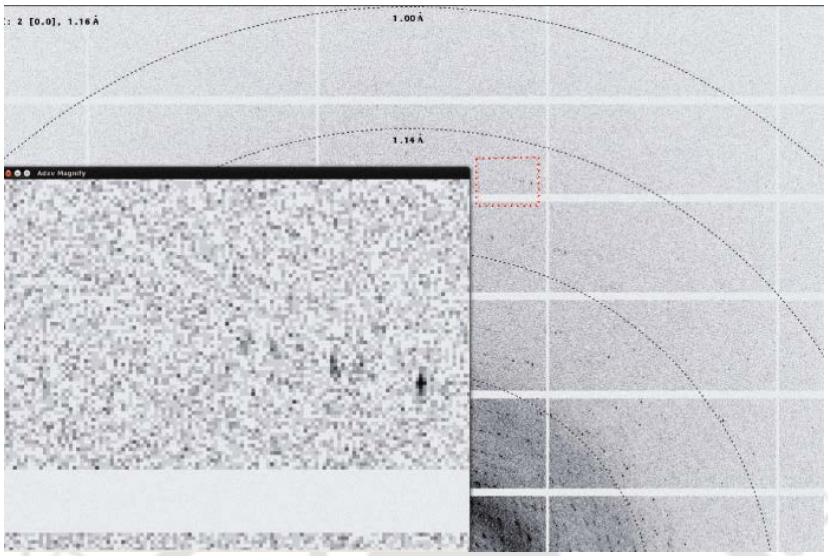


Fig. 4.3.10 Representative diffraction pattern of a *Ct*Xyn30A crystal with resolution rings shown as dashed circles. The inset shows an enlargement of the contents of the red box, showing spots at ~1.2 Å resolution

Table 4.3.2. Data collection statistics.

PDB code	4ckq	4uqa	4uqe	4uqd	4uq9	4uqc	4uqb
Beamline	PROXIMA-1, SOLEIL	IO3-4 DIAMOND	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL
Wavelength (Å)	0.8260	0.9795	0.9537	0.9537	0.9537	0.9537	0.9537
Space group	P1	P1	P1	P1	P1	P1	P1
Unit cell parameters (Å,°)	*a* = 46.09, *b* = 47.46, *c* = 53.62, *α* = 83.15, *β* = 73.41, *γ* = 65.87	*a* = 46.73, *b* = 50.37, *c* = 58.87, *α* = 65.00, *β* = 67.28, *γ* = 76.84	*a* = 45.94, *b* = 46.07, *c* = 59.05, *α* = 68.11, *β* = 80.19, *γ* = 60.99	*a* = 45.64, *b* = 46.74, *c* = 53.07, *α* = 83.37, *β* = 73.53, *γ* = 65.79	*a* = 46.62, *b* = 50.33, *c* = 58.71, *α*= 65.16, *β*= 67.56, *γ* = 76.99	*a* = 50.39, *b* = 50.44, *c* = 58.19, *α*=111.90, *β*=110.80, *γ* = 7.67	*a* = 46.98, *b* = 50.59, *c* = 59.31, *α*=114,77, *β*=101.43, *γ*= 103.85
Resolution limits (Å)	42.93 – 1.40 (1.48 – 1.40)	50.39–1.52(1.57 -1.52)	40.18–1.28 (1.30–1.28)	50.9–1.25 (1.29-1.25)	45.53-1.77 (1.84-1.77)	44.84–1.30 (1.32–1.30)	50.68–1.68 (1.74–1.68)
Total number of reflections	347015 (11412)	182666 (11567)	410803 (17816)	393936 (33211)	173410 (1554)	463345 (19528)	214085 (18208)
Number of unique reflections	76088 (2464)	66458 (2618)	96062 /4507)	96003 (8944)	40534 (3836)	109732 (4770)	47651 (4525)
Multiplicity	4.6 (4.6)	4.3 (4.4)	4.3 (4.0)	4.0 (3.7)	4.1 (4.0)	4.2 (4.1)	4.3 (4.0)
Completeness(%)	97.2 (95.6)	96.51 (94.56)	95.1 (90.1)	89.80 (75.83)	93.44 (79.54)	94.4 (82.8)	92.29 (77.25)
<I/σ(I)>	3.5 (1.3)	19.48 (10.22)	23.0 (10.7)	15.21 (1.58)	14.15 (1.91)	20.8 (7.9)	13.63 (1.77)
$CC_{1/2}$	0.971 (0.536)	0.999 (0.996)	0.998 (0.926	0.998 (0.564)	0.991 (0.764)	0.994 (0.973)	0.997 (0.837)
R_{merge}[a]	0.196 (0.860)	0.050 (0.142)	0.056 (0.202)	0.066 (0.827)	0.132 (0.952)	0.055 (0.177)	0.085 (0.827)
$R_{p.i.m.}$[b]	0.101 (0.449)	0.063 (0.778)	0.030 (0.115)	0.069 (1.075)	0.175 (2.080)	0.031 (0.098)	0.101 (1.438)

Values in parentheses are for the lowest/highest resolution shells

[a], $R_{merge} = \sum_{hkl}\sum_i(I_i(hkl) - \langle I(hkl)\rangle)/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the i[th] intensity measurement of reflection *hkl*, including symmetry-related reflections, and $\langle I(hkl)\rangle$ is its average.

[b], $R_{p.i.m.} = \left(\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_{j=1}^{n}|I_{hkl,j} - \langle I_{hkl}\rangle|\right)\Big/\left(\sum_{hkl}\sum_j I_{hkl,j}\right)$, where $\langle I_{hkl}\rangle$ is the average of symmetry-related observations of a unique reflection.

$CC_{1/2}$ is the half-data-set correlation coefficient (Diederichs and Karplus, 2013)

4.3.2.1 Three dimensional structure analysis of *Ct*Xyn30A

*Ct*Xyn30A structure consists of the expected $(\beta/\alpha)_8$ TIM barrel classified as belonging to the carbohydrate-active enzyme database clan A enzymes (4/7 hydrolases). A C-terminal β-sheet domain is appended to this barrel through a hydrophobic patch on the outer side of α-helices 7 and 8. The TIM barrel domain contains residues from Val11 to Pro295, with a core of 8 β-strands surrounded by 8 α helices. The side associated β-domain comprises residues His3 to Gln10 and Gly296 to Val386 (Fig. 4.3.11).

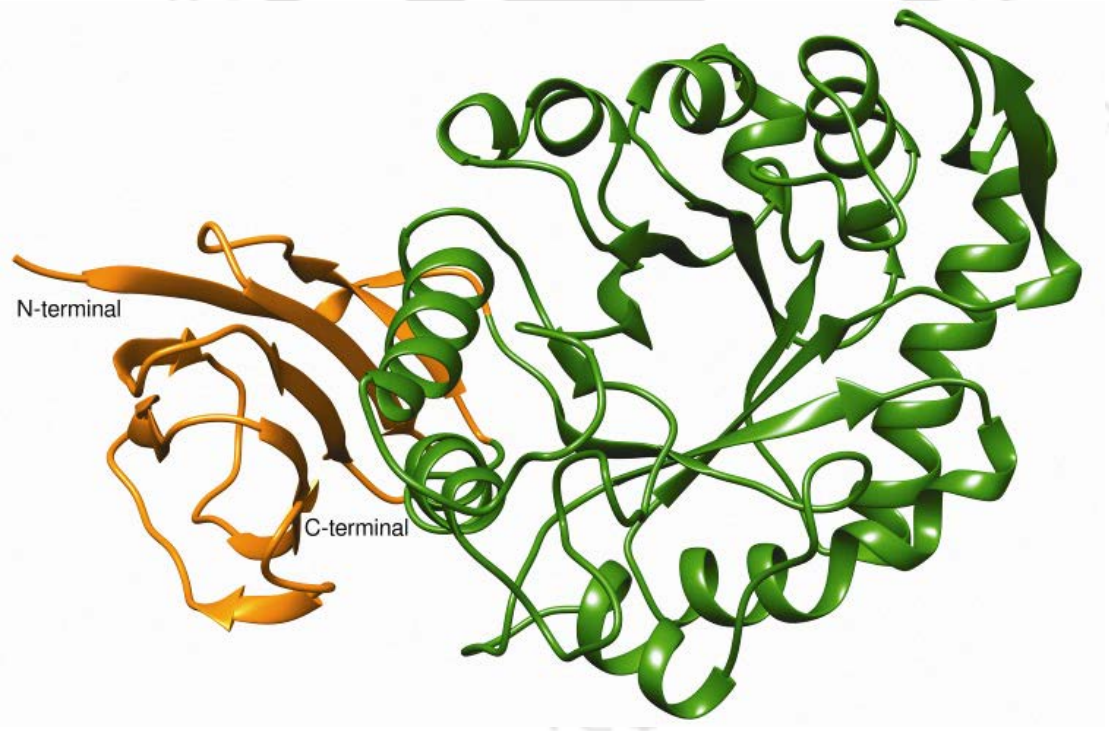


Fig. 4.3.11 Overall structure of *Ct*Xyn30A. The TIM barrel domain is represented in forest green and consists of residues V11 to P295, and the side β-domain (represented in orange) comprising residues H3 to Q10 and G296 to V386.

The modeling study allowed the prediction of *Ct*Xyn30A structure as reported in Section 4.3.1.3 (Verma and Goyal, 2014). The r.m.s. deviation between the

predicted 3-D structure of *Ct*Xyn30A and the 3-D crystallographic structure of *Ct*Xyn30A (PDB code 4uqe) was 0.429 for 302 Cα atoms, confirming the high structural similarity (Fig. 4.3.12).

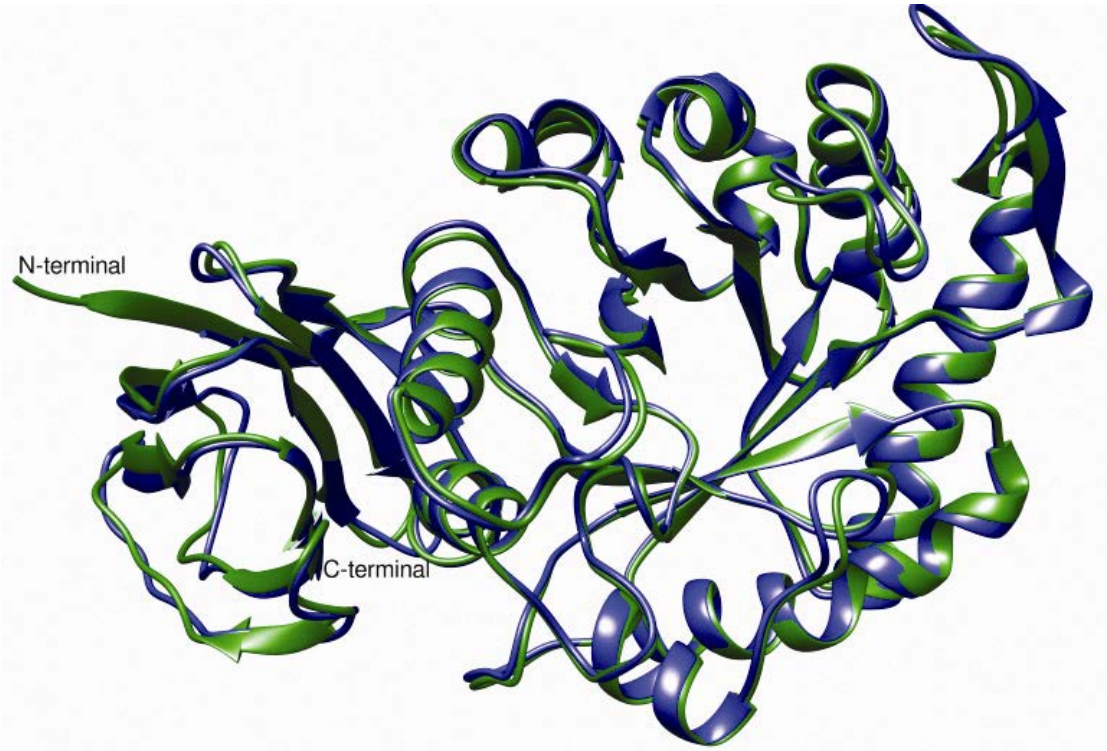


Fig. 4.3.12 *Ct*Xyn30A structures overlay. Superposition of *Ct*Xyn30A structure determined by X-ray crystallography (forest green) PDB id: 4uqe, and the structure obtained by molecular modeling (blue) as described by Verma *et al*. (Verma *et al*., 2013), showing an r.m.s. deviation of 0.429 for 302 Cα atoms.

The structure comparison of members of GH30 subfamily 8, from different organisms showed that *Ct*Xyn30A structure is the most similar to XynC structure (PDB code 3GTN) from gram positive bacterium *B. subtilis* (r.m.s.d. between the two structures of 0.442 Å for 307 Cα atoms) and XynA (PDB code 1NOF) from the gram negative bacterium *Erwinia chrysanthemi* (r.m.s.d. of 0.695 Å for 275 Cα atoms). Small differences between the three structures are shown in figure 4.3.13.

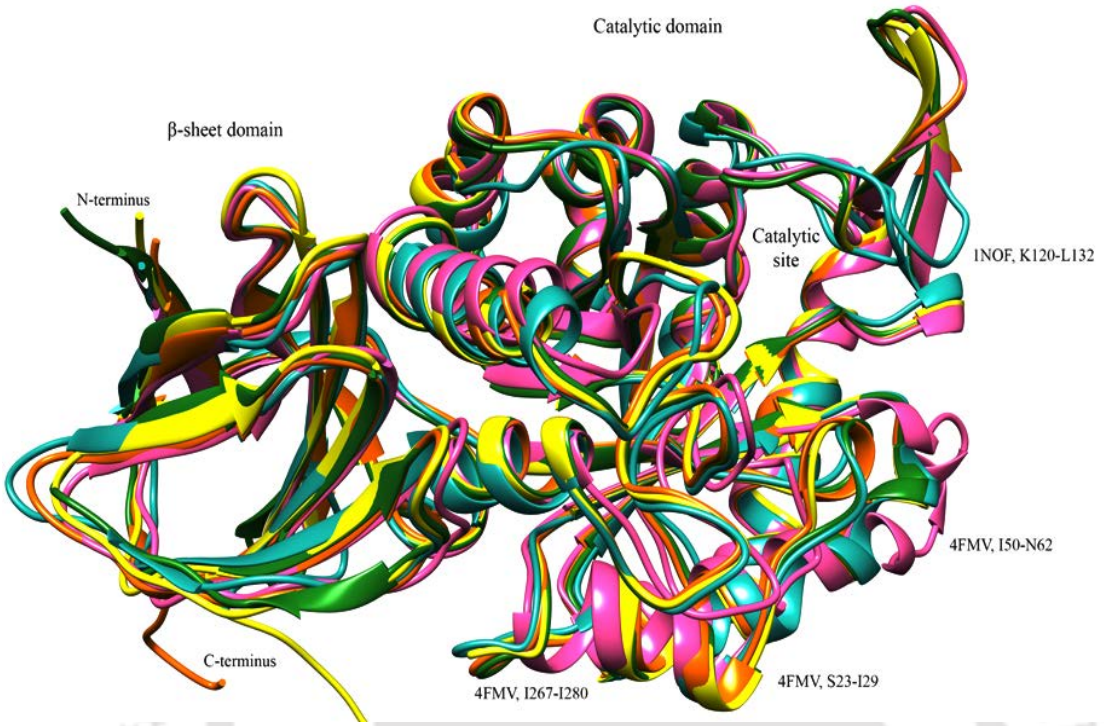


Fig. 4.3.13 Superposition of XynC (orange, PDB code 3GTN) from *B. Subtilis*, XynA (blue, PDB id, 1NOF) from *Erwinia chrysanthemi*, *Cp*C71 (pink, PDB id, 4FMV) from *Clostridium papyrosolvens* C71, *Ct*Xyn30A (green, PDB id, 4QUE) from *Clostridium thermocellum* and Xyn30D from *P. barcinonesis* (yellow, PDB id, 4QAW), excluding the CBM35 domain. The main differences among the four structures were observed in *Cp*C71, regions Ser23-Ile29, Ile50-Asn62and Ile267-Ile280 and in XynA, the region from Lys120 to Leu132.

However, comparison of *Ct*Xyn30A (PDB code 4CKQ) with *C. papyrosolvens* C71 (*Cp*C71, PDB code 4FMV) showed a lower structural similarity (r.m.s.d. of 0.836 Å for 2043 atoms) (Fig. 4.3.14).

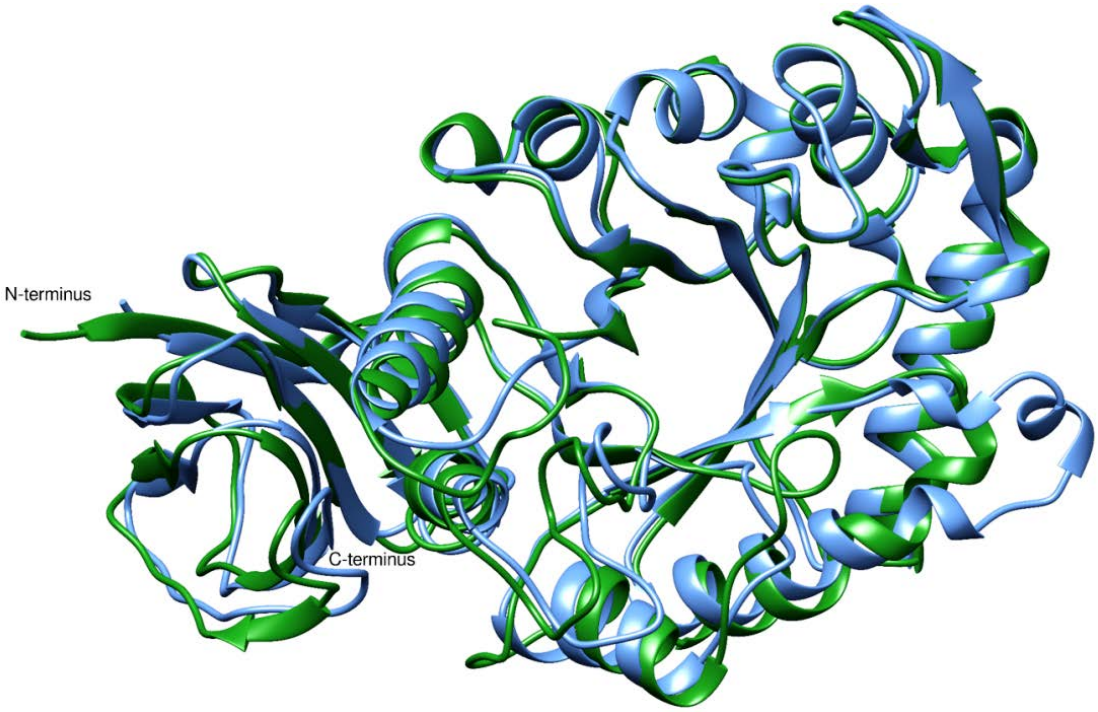


Fig. 4.3.14 Superposition of *C. thermocellum Ct*Xyn30A (green, PDB id, 4QUE) and *C. papyrosolvens*, *Cp*C71 (blue, PDB id, 4 FMV) X-ray structures.

The main structural difference was observed in the loop comprising residues Ile50 to Asn62 in *Cp*C71, in which an α-helix (Glu51-Gly61) can be observed in *Cp*C71 structure. This secondary structure element was not observed in *Ct*Xyn30A. Other differences were observed in the solvent exposed loops comprising residues Ser23 to Ile29 and Ile267 to Ile280 in *Cp*C71 structure. These differences were probably the result of crystal packing. Interestingly the β8-α8 loop, which defines the function of GH30-8 subfamily of enzymes and its role in GlcA recognition, is longer in *Ct*Xyn30A when compared to *Cp*C71. This loop in *Ct*Xyn30A is similar in XynC and *Ct*Xyn30A, which is in agreement with the conservation of the residues observed in this region in the three structures (Fig. 4.3.13). The two catalytic residues, Glu136 and Glu225 in *Ct*Xyn30A and Glu143 and Glu232 in *Cp*C71 are also located in the same region (Fig. 4.3.15).

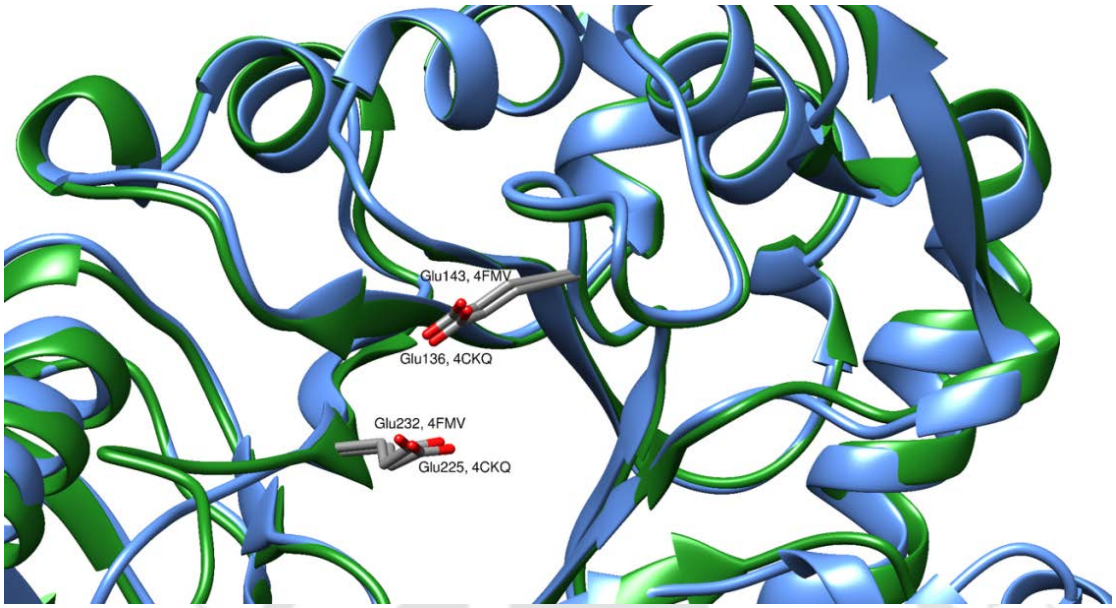


Fig. 4.3.15 Closer view of the catalytic site of *Ct*Xyn30A and *Cp*C71 displaying the side chains of the catalytic residues of both enzymes, Glu136 and Glu225 and Glu143 and Glu232, respectively.

The full length *Ct*XynGH30 and *Cp*C71, contain a non-catalytic CBM6 domain. This domain is connected to *Ct*Xyn30A by a 14 residues linker (419-432). Recently Sainz Polo *et al* (Sainz-Polo *et al*., 2014) reported the crystal structure of *P. barcinonesis* Xyn30D comprising the catalytic domain with the associated β-structure with an attached CBM35. The catalytic domain, like in *Ct*Xyn30A, has an $(\alpha/\beta)_8$ TIM barrel fold with an associated β-structure having a hydrophobic patch in between.

4.3.2.2 *Ct*Xyn30A catalytic site analysis

Multiple sequence analysis of all available structures from family GH30-8 show that the catalytic residues Glu136 and Glu225 in *Ct*Xyn30A participating in the catalysis are absolutely conserved (Fig. 4.3.4, filled inverted triangle). The distance between Glu136 and Glu225, the residues involved in the enzyme catalytic activity varies in the different *Ct*Xyn30A structures from 4.1 Å to 4.5 Å. The docking studies predicted a distance between these catalytic residues of 5.2 Å (Verma and Goyal,

2014). The distance observed between these residues in the different *Ct*Xyn30A structures confirms the presence of retention type mechanism in *Ct*Xyn30A (Davies and Henrissat, 1995). Since a large number of crystals were available in several crystallization conditions attempts were made to bind the ligands, xylohexaose or cellohexaose, by soaking experiments and a large number of data were subsequently collected. All tested crystals belong to space group P1 and display good diffraction in a resolution range from 1.25 Å to 1.77 Å, with one molecule in the asymmetric unit. Despite the different conditions in which *Ct*Xyn30A crystallized, namely pH values and precipitant agents, the overall three dimensional structure is very similar for all of them (with r.m.s.d. ranging from 0.1 Å to 0.3 Å). No xylohexaose or cellohexaose was observed bound in any of the structures. This may be due to the larger size of these ligands, which may hinder their diffusion into the crystal. However, part of the proteolysed his-tag peptide or molecules that are present in the crystallization condition or cryo protectant solution such as glycerol, tartaric acid and malonic acid were identified in the catalytic binding site of the various structures (Fig. 4.3.16) and Table 4.3.2).

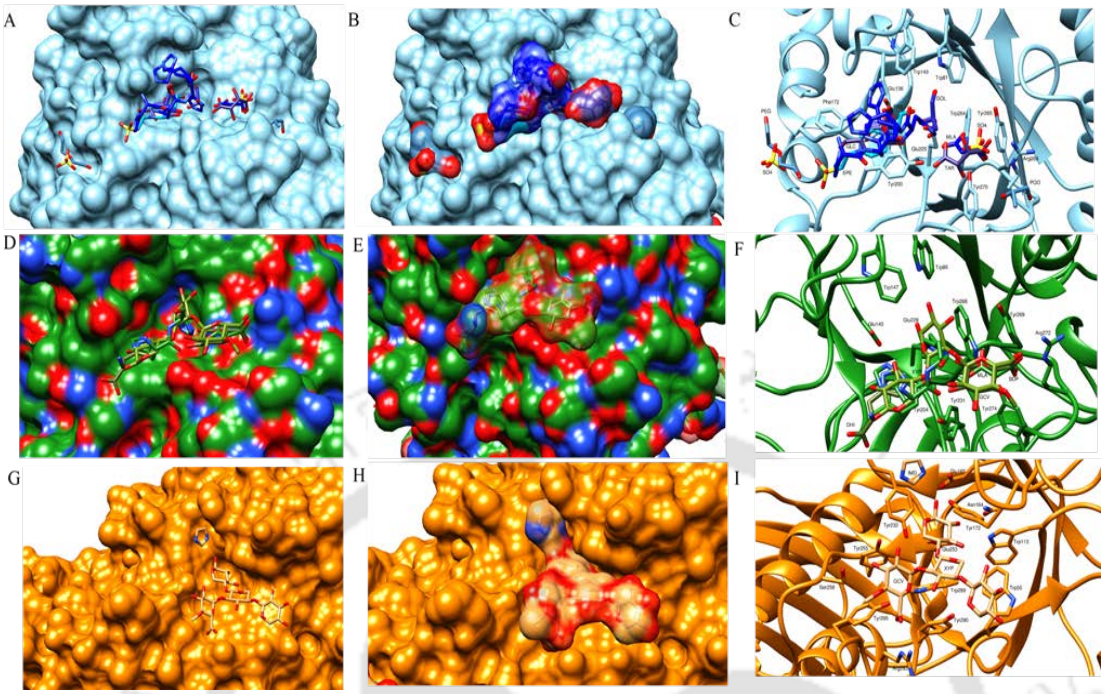


Fig. 4.3.16 *Ct*Xyn30A, XynC and XynA catalytic site structure organization. Surface representation of *Ct*Xyn30A **(A, D)**, XynC **(B, E)** and XynA (**C, F)** catalytic site (PDBs id, 4UQD, 3KL0 and 2Y24, respectively) with the several ligands observed, represented in sticks (A, B and C, respectively), (PDBs id, 4CKQ, 4UQB, 4UQC, 4UQD for *Ct*Xyn30, 3KL0, 3KL3 and 3KL5 for XynC, and 2Y24 for XynA) and in surface representation (D, E and F, respectively); *Ct*Xyn30A (G), XynC (H) and XynA (I) catalytic site with the several ligands observed and the side chains of the residues involved in ligand binding represented in sticks. (TAR– (D)-tartaric acid; MLA– malonic acid; IMD– imidazole; GOL– glycerol; GLC– D-glucose; XYP– β-D-xylopyranose; TRS-2-amino-2-hydroxymethyl-propane-1,3-diol; GCV– 4-O-methyl-α-D-glucoronic acid; EPE– HEPES; PGO-1,2-propanediol; BDP– β-D-glucopyranuronic acid; PGE - ; DHI – (D)-histidine).

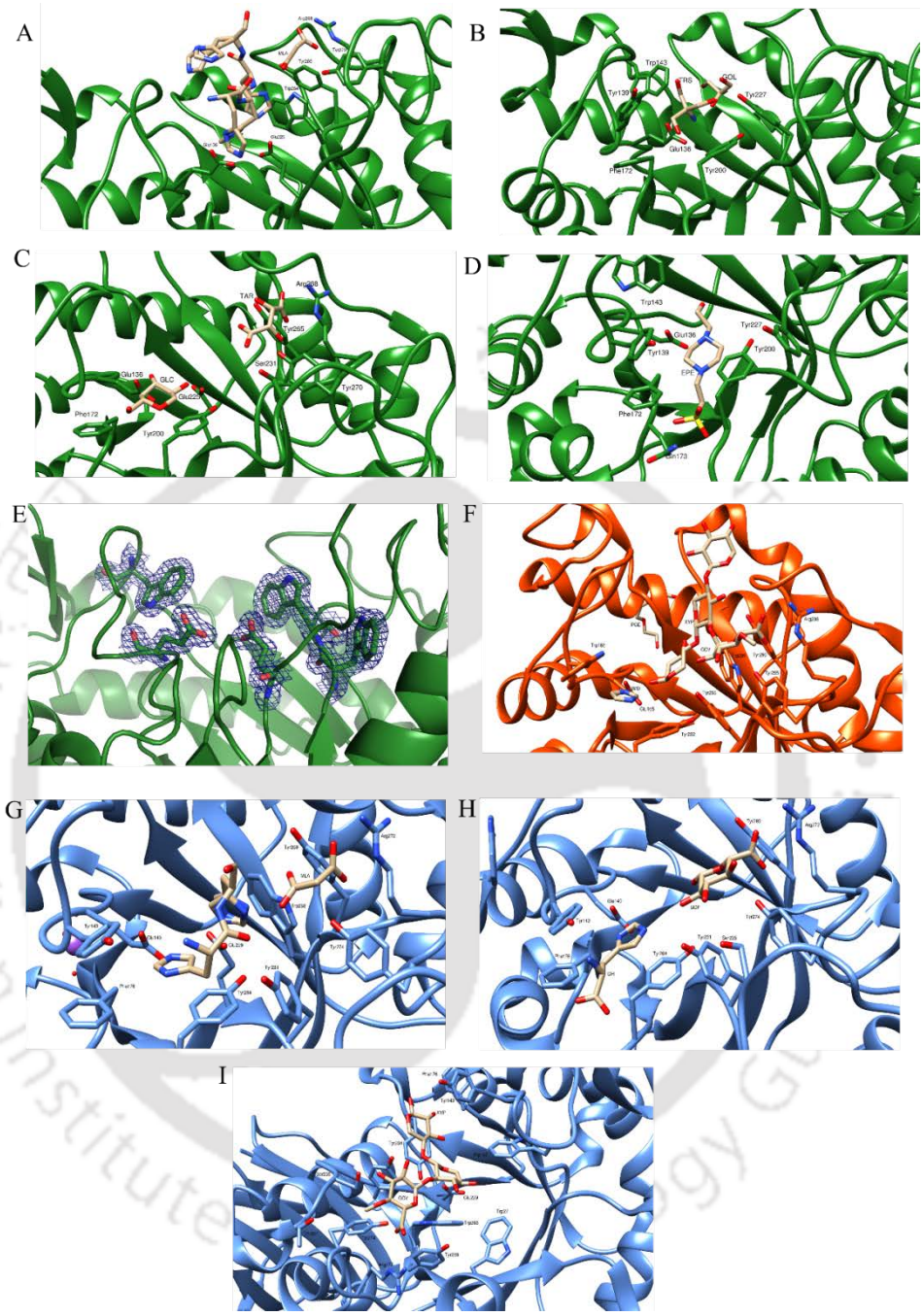


Fig. 4.3.17 Structure organization of *Ct*Xyn30A, XynC and XynA catalytic site displaying **(A – D)** closer view of *Ct*Xyn30A catalytic site, PDB codes 4CKQ, 4UQD, 4UQC and 4UQB, respectively; **(E)** *Ct*Xyn30A catalytic site (PDB code 4UQE) showing 2mFo-DFc difference Fourier map of Glu136, Trp143, Glu225, Trp264 and Trp265 residues in blue and contoured at 1 σ; **(F–H)** XynC catalytic site (PDB codes 3KL0, 3KL3 and 3KL5, respectively; **(I)** XynA catalytic site (PDB code 2Y24). The side-chain of the residues involved in ligand binding is represented. (TAR – (D)-tartaric acid ; MLA – malonic acid; IMD – imidazole; GOL– glycerol; GLC– D-glucose; XYP– D-xylopyranose; TRS - 2-amino-2-hydroxymethyl-propane-1,3-diol; GCV – 4-O-methyl-α-D-glucoronic acid ;EPE– HEPES; PGO- 1,2-propanediol; BDP – β-D-glucopyranuronic acid; PGE - ; DHI – (D)-histidine).

Moreover, the aromatic residues present in the catalytic site of *Ct*Xyn30A (Fig. 4.3.3, in filled circle) were also found to be conserved and are listed in the Table 4.3.3.

Table 4.3.3 Important residues in the binding pocket.

***Ct*Xyn30A**	**XynC**	**XynA**	***Cp*C71**	**Xyn30D**
Trp81	Trp85	Trp113	Trp88	Trp84
Trp143	Trp147	Tyr172	Trp150	Trp145
Tyr200	Tyr204	Tyr232	Tyr207	Tyr203
Trp264	Trp268	Trp289	Trp265	Trp267
Tyr265	Tyr269	Tyr290	Trp266	Tyr268
Arg268	Arg272	Arg293	Trp269	Arg271
Tyr270	Tyr274	Tyr295	Asn271	Tyr273

Docking studies performed on *Ct*Xyn30A showed that these aromatic residues (Trp81, Tyr139, Trp143, Phe172, Tyr200, Tyr227, Trp264 and Tyr265) are involved in substrate binding (Verma and Goyal, 2014). They are important in the binding of the different small molecules as shown in figures 4.3.16 and 4.3.17. In *Ct*Xyn30A catalytic site, PDB code 4CKQ (Fig. 4.3.16 and 4.3.17), a malonic acid molecule was observed, stabilized by a hydrophobic interaction with Tyr265 residue and by hydrogen bond with Trp264 Nε1, Arg268 NH_2 group and Nε, Tyr270 OH group and Nδ1 from His3 residue from the proteolysed his-tag peptide. In the catalytic site of *Ct*Xyn30A, PDB code 4UQD (Fig. 4.3.16 and 4.3.17B), a glycerol molecule was observed establishing hydrophobic contacts with Tyr139 and hydrogen bonding with a 2-amino-2-hydroxymethyl-propane-1,3-diol molecule, which is itself interacting with Trp143 and Tyr200 by hydrophobic contacts and with Tyr227 by hydrogen bonding. In *Ct*Xyn30A structure, PDB code 4UQC (Fig. 4.3.16 and 4.3.17C), a tartaric acid molecule is seen, stabilized by a hydrophobic interaction with Tyr265

residue and hydrogen bonding with Ser231 Oγ, Tyr270 OH group and Arg268 NH_2 group and Nε. A 4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid (HEPES) molecule was identified in catalytic site of *Ct*Xyn30A, PDB code 4UQB (Fig. 4.3.16 and 4.3.17D). This molecule is stabilized by hydrophobic interactions with Glu136, Tyr139, Trp143, Phe172, Tyr200, Gly201 and Tyr 227 residues, and hydrogen bonding with Gln173 Nε2 group. The different chemical nature of the ligands observed in *Ct*Xyn30A catalytic site may suggest the possible binding of different decorations of xylose subunits of GX_n at this site.

Similar binding pocket sites were also observed in three dimensional structures of XynC (St John *et al*., 2011; St John *et al*., 2006), XynA (Larson *et al*., 2003), *Cp*C71 (St John *et al*., 2014) and Xyn30D (Sainz-Polo *et al*., 2014). The majority of these residues are conserved in the five structures (Table 4.3.3). Comparison of the catalytic site of *Ct*Xyn30A with the several ligands observed, superimposed with XynC shows that the HEPES molecule observed in *Ct*Xyn30A (PDB, id 4UQB) catalytic site is positioned in a similar manner to the β-D-xylopyranose molecule in XynC (PDB id, 3KL5) catalytic site. In addition, the tartaric acid molecule in *Ct*Xyn30A catalytic site seems to be mimicking the 4-O-methyl-α-D glucoronic acid molecule (MeGlcA) found in Xyn30C catalytic site (PDB id, 3KL5). The tartaric acid molecule is stabilized by hydrogen bonding to Ser231, Arg268 and Tyr270 corresponding to residues Ser235, Arg272 and Tyr274 in XynC, which binds MeGlcA. The same residues are also involved in the binding of a β-D-glucopyranuronic acid molecule to XynC (PDB id, 3KL3). Interestingly, a malonic acid molecule is observed in *Ct*Xyn30A (PDB id, 4CKQ) and XynC (PDB id, 3KL0) structures. This molecule is bound to *Ct*Xyn30A by hydrogen bonding to Trp 264,

Arg268 and Tyr270 and also by a hydrophobic interaction with Tyr265 and similarly, by the conserved residues Trp268, Arg272, Tyr274 and Tyr269 in XynC. In the catalytic site of XynA (PDB id, 2Y24, Fig. 4.4) a MeGlcA molecule (stabilized by hydrogen bonding with Tyr255, Arg293 and Tyr295, and hydrophobic contacts with Trp289 and Tyr290) occupies the malonic acid site found in *Ct*Xyn30A and XynC. A tartaric acid molecule is also observed in *Ct*Xyn30A (PDB id, 4UQC) catalytic site, stabilized by Tyr 265, Arg268 and Tyr270 residues. Besides MeGlcA, an imidazole molecule and a β-D-xylopyranose molecule are observed in XynA (PDB id, 2Y24) catalytic site. The imidazole makes hydrogen bond with Glu165 and β-D-xylopyranose molecule and is further stabilised by hydrophobic contacts with Trp268. Likewise, a glycerol molecule establishing a hydrogen bonding to a 2-amino-2-hydroxymethyl-propane-1,3-diol molecule is observed in *Ct*Xyn30A (PDB id, 4UQD). In this case, the glycerol molecule is stabilized by hydrophobic interactions with Tyr139 (Trp168 in XynA) and Phe172 (Phe193 in XynA). The 2-amino-2-hydroxymethyl-propane-1,3-diol is making hydrogen bonding interactions with Glu136 (Glu165 in XynA), and Tyr227 and an hydrophobic interaction with Trp143. In *Ct*Xyn30A catalytic site (PDB id, 4UQE), a glycerol molecule hydrogen bonding with a 2-amino-2-hydroxymethyl-propane-1,3-diol molecule is observed. In this case, a slight change in orientation is observed. The 2-amino-2-hydroxymethyl-propane-1,3-diol molecule is hydrogen bonding with Glu136, glycerol and Tyr227. The 2-amino-2-hydroxymethyl-propane-1,3-diol molecule is interacting through hydrophobic contacts with Tyr139, Trp143, Phe172 and Tyr200. Thus, the ligand interactions involve similar residues in *Ct*Xyn30A and XynA, despite the different nature of the ligands. The residues involved in ligand binding to *Ct*Xyn30A, XynC

and XynA (Fig. 4.3.16 and 4.3.17 and Table 4) are located in the aglycon moiety and are generally conserved in the structure of *Cp*C71. However, in *Cp*C71, the residues Arg268 and Tyr270 are replaced by Trp269 and Asn271. *Cp*C71 also lacks the β8–α8 loop region which may explain its low specificity for the GlcA appendage.

Conclusions

The identification of the active site residues and understanding of their role in hydrolysis of substrate was studied. *Ct*Xyn30A showed sequence and structural similarity with XynC from *B.subtilis*, and both were reassigned in the family GH30 glucuronoxylan-xylanohydrolase; subfamily H. The 3-D structure of *Ct*Xyn30A was modeled and validated. Molecular dynamic simulation ensured the structural integrity of energy minimized *Ct*Xyn30A. Structure analysis of *Ct*Xyn30A displayed that it has a $(\beta/\alpha)_8$ barrel fold structure along with an associated side β-structure resembling to CBM, attached to the main catalytic core. The side β-structure helps in substrate selection and crucial for the activity of catalytic core. The findings inferred that carboxylic group of amino acid residue Glu136 act as a catalytic acid/base whereas Glu225 act as a nucleophile, moreover these residues are 5.2 Å apart from each other which confirmed the presence of retention type mechanism in *Ct*Xyn30A. These key active site residues were found to be conserved within GH30 family. The catalytic core of *Ct*Xyn30A consisted of β-barrel and connecting loops toward the C-terminal side, formed by aromatic amino acid residues Trp53, Trp173, Tyr139, Tyr230, Trp111, Trp293 and Tyr294 which found to be conserved. The three dimensional structure of *Ct*Xyn30A contained a central $(\beta/\alpha)_8$ TIM barrel to which a side β-sheet domain is tightly connected. Glu136 and Glu225 are the key residues for catalytic activity of *Ct*Xyn30A. Trp81, Tyr139, Trp143, Gln173, Tyr200, Tyr227, Ser231, Trp264, Tyr265, Arg268 and Tyr270 are important residues for ligand binding. A comparison of *Ct*Xyn30A with XynA and XynC structures with various bound ligands confirmed that equivalent residues are involved in substrate recognition providing a common consensus in the catalytic site. Thus, GH30-8 contain specificity

pockets near the active site, enabling GH30-8 members to bind to a number of ligands that may mimic the different decorations of xylose in the GX_n chains of their substrates. The present study described the complete structural features and the conserved active site residues as well as the conserved aromatic residues which overall architect the catalytic core of glucuronoxylan-xylanohydrolase *Ct*Xyn30A from *Clostridium thermocellum*.

References

Anna Ebringerová, T.H. (2000) Xylan and xylan derivatives-biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties Macromolecular Rapid Communications, 21, 542-556.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallographica. Section D, Biological Crystallography, 67, 271-281.

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

Bowie, J., Luthy, R. Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science, 253, 164-170.

Correia, M.A., Mazumder, K., Bras, J.L., Firbank, S.J., Zhu, Y., Lewis, R.J., York, W.S., Fontes, C.M., Gilbert, H.J. (2011) Structure and function of an arabinoxylan-specific xylanase. The Journal of Biological Chemistry, 286, 22510-22520.

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

De Sanctis, D., Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., Mairs, T., McSweeney, S., Petitdemange, S., Rey-Bakaikoa, V., Surr, J., Theveneau, P., Leonard, G.A., Mueller-Dieckmann, C. (2012) ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering. Journal of Synchrotron Radiation, 19, 455-461.

Diederichs, K., Karplus, P.A. (2013) Better models by discarding data Acta Crystallographica Section D Biological Crystallography, 69, 1215-1222.

Ebringerov, A. (2005) Structural Diversity and Application Potential of Hemicelluloses. Macromolecular Symposia, 232, 1-12.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010) Features and development of Coot. Acta Crystallographica Section D Biological Crystallography, 66, 486-501.

Evans, P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta crystallographica. Section D, Biological crystallography, 67, 282-292.

Fiser, A., Do, R.K., Sali, A. (2000) Modeling of loops in protein structures. Protein Science 9, 1753-1773.

Goodsell, D.S., Olson, A.J. (1990) Automated docking of substrates to proteins by simulated annealing. Proteins, 8, 195-202.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., Davies, G. (1996) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proceedings of the National Academy of Sciences of the United States of America, 93, 5674.

Hess, B. (2008) P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. Journal of Chemical Theory and Computation, 4, 116-122.

Holm, L., Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Research, 38, W545-549.

Hurlbert, J.C., Preston, J.F.,3rd (2001) Functional characterization of a novel xylanase from a corn strain of *Erwinia chrysanthemi*. Journal of Bacteriology, 183, 2093-2100.

Joosten, R.P., Joosten, K., Murshudov, G.N., Perrakis, A. (2012) PDB_REDO: constructive validation, more than just looking for errors. Acta Crystallographica. Section D, Biological Crystallography, 68, 484-496.

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

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227, 680-685.

Larson, S.B., Day, J., Barba de la Rosa, A.P., Keen, N.T., McPherson, A. (2003) First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis. Biochemistry, 42, 8411-8422.

Laskowski R. A., Macarthur, M.W., Moss, D. S., Thornton J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26, 283-291.

Leggio, L.L., Jenkins, J., Harris, G.W., Pickersgill, R.W. (2000) X-ray crystallographic study of xylopentaose binding to *Pseudomonas fluorescens* xylanase A. Proteins, 41, 362-373.

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

Matthews, B.W. (1968) Solvent content of protein crystals. Journal of Molecular Biology 33, 491-497.

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

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

Morris G. M., Goodshell, D.S., Halliday, R.S., Huey R., Hart, W. E., Belew, R. K., Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry, 19, 1639-1662.

Murshudov, G. N., Skub P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn,M. D., Long, F. Vagin, A. A. (2011) REFMAC 5 for the refinement of macromolecular crystal structures. Acta Crystalography Section D, 67, 355-367.

Nishitani, K., Nevins, D.J. (1991) Glucuronoxylan xylanohydrolase. A unique xylanase with the requirement for appendant glucuronosyl units. The Journal of Biological Chemistry, 266, 6539-6543.

Painter, J., Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallographica. Section D, Biological Crystallography 62, 439-450.

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. Journal of Computational Chemistry, 25, 1605-1612.

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

Rye, C.S., Withers, S.G. (2000) Glycosidase mechanisms. Current Opinion in Chemical Biology, 4, 573-580.

Sainz-Polo, M.A., Valenzuela, S.V., Gonzalez, B., Pastor, F.I., Sanz-Aparicio, J., (2014) Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity. Journal of Biological Chemistry, 289, 31088-31101.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., Karplus, M. (1995) Evaluation of comparative protein modeling by MODELLER. Proteins, 23, 318-326.

Schrodinger, L., (2010) The PyMOL Molecular Graphics System. Version 1.6.X ed.

Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. Proteins, 17, 355-362.

St John, F.J., Dietrich, D., Crooks, C., Pozharski, E., Gonzalez, J.M., Bales, E., Smith, K., Hurlbert, J.C. (2014) A novel member of glycoside hydrolase family 30 subfamily 8 with altered substrate specificity. Acta Crystallographica Section D Biological Crystallography, 70, 2950-2958.

St John, F.J., Godwin, D.K., Preston, J.F., Pozharski, E., Hurlbert, J.C. (2009) Crystallization and crystallographic analysis of *Bacillus subtilis* xylanase C. Acta crystallographica. Section F, Structural Biology and Crystallization Communications, 65, 499-503.

St John, F.J., Gonzalez, J.M., Pozharski, E. (2010) Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. FEBS Letters, 584, 4435-4441.

St John, F.J., Hurlbert, J.C., Rice, J.D., Preston, J.F., Pozharski, E. (2011) Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase. Journal of Molecular Biology, 407, 92-109.

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. Journal of Bacteriology, 188, 8617-8626.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.W., Read, R.J., Adams, P.D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallographica, Section D, Biological Crystallography, 64, 61-69.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F.,Higgins D.J. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 24, 4876–4882.

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., Murshudov, G.N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallographica. Section D, Biological Crystallography, 60, 2184-2195.

Valenzuela, S.V., Diaz, P., Pastor, F.I. (2012) Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. Applied and Environmental Microbiology, 78, 3923-3931.

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*. Molecular Biology, 48, 278-286.

Verma, A.K., Goyal, A., Freire, F., Bule, P., Venditto, I., Bras, J.L., Santos, H., Cardoso, V., Bonifacio, C., Thompson, A., Romao, M.J., Prates, J.A., Ferreira, L.M., Fontes, C.M., Najmudin, S., (2013) Overexpression, crystallization and preliminary X-ray crystallographic analysis of glucuronoxylan xylanohydrolase (Xyn30A) from *Clostridium thermocellum*. Acta Crystallographica, Section F, Structure Biology Crystal Communications 69, 1440-1442.

Whistler, R.L., Richards, E.L, (1970) Hemicelluloses, In pigman, W., Horton, D. (Eds,), The Carbohydrats second ed. Academic press. pp. 447-469.

Wiederstein, M., Sippl, M.J., (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research 35, W407-410.

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., 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 Crystallographica, Section D, Biological Crystallography, 67, 235-242.

Chapter 5

Structural, biochemical and *in silico* determinants of ligand binding specificity of family 6 carbohydrate binding module (*Ct*CBM6) from *Clostridium thermocellum*

5.1 Introduction

Carbohydrate-binding modules (CBMs) are important components of modular carbohydrate-active enzymes. They are non-catalytic modules having the ability to express and fold independently of their associated module. By definition CBMs are non-catalytic modules which can target the catalytic module and facilitate its prolong and intimate association with the substrate molecule and by doing so they increase the effective concentration of substrate in the vicinity of associated catalytic module and improves its performance (Guillen *et al*., 2010). The specificity of protein-carbohydrates interaction is vital for understanding numerous biological processes, including cell signaling and host-pathogen interaction and these microbial CBMs interact with plant cell wall complex carbohydrates. Thus, for investigating insights of protein-carbohydrate interactions, CBM can be an excellent model system (Boraston *et al.*, 2004; Hashimoto, 2006; Brett and Waldron, 1996). The CBM-polysaccharide binding relationship can be studied by four popular and well-established

methodologies; (1) the solid state depletion assay, (2) affinity gel electrophoresis, (3) UV difference and fluorescence spectroscopy and (4) isothermal titration calorimetry (ITC). Although, affinity gel electrophoresis have been the most popular method for the last several decades (Takeo, 1995), now-a-days, isothermal titration calorimetry analysis is gaining importance mainly due to its high sensitivity and less time consumption.

Based on sequence similarities, CBMs are grouped into 71 families as listed in the CAZy database (http://www.cazy.org) till date. Presently 946 bacterial, 74 archaea and 14 eukaryotic (protein entries) CBMs have been reported in family 6 (http://www.cazy.org/CBM6.html). Interestingly, family 6 CBM shows diverse binding specificity towards different polysaccharides *viz.* β-agarose having 3,6-anhydro-L-galactoses and D-galactoses alternate linked by α-(1→3) and β-(1→4) linkages (Henshaw et al., 2006), xylan (β-1→4-linked xylose) (Pires et al., 2004), cellulose (β-1→4-linked glucose homopolymer) (Correia *et al*., 2009; Henshaw *et al*., 2004), laminarin (β-1→3-linked glucose homopolymer), lichenan (β-1→4- and β-1→3-mixed linked β-glucans) (Boraston *et al*., 2003; van Bueren *et al*., 2005) and oat spelt xylan (β-1→4-linked xylose main chain with β-1→2 or β-1→3-linked arabinose side chain)) (Ahmed *et al*., 2013a; Ahmed *et al*., 2013b). CBM6 not only exhibited affinity towards soluble substrates, but also showed affinity towards insoluble substrates and thereby helping in enhancing the activity of cognate catalytic modules. (Bae *et al*., 2003; Fernandes *et al*., 1999). The CBMs of family 6 showed two potential binding sites termed as cleft A and cleft B. Cleft A is formed by the loops that connect the inner and the outer β-sheets of a protein while cleft B is located on the concave surface of the β sandwich fold (Henshaw *et al*., 2004).

In the present study, the structural and functional relationship aspect of CBM6 from *Clostridium thermocellum* was studied. *Ct*CBM6 was modelled and molecular docking analysis was performed in order to understand the binding cleft architecture and to know the important residues participating in binding the carbohydrates. *Ct*CBM6 was cloned and expressed successfully which is described in Chapter 2. Affinity gel electrophoresis and isothermal titration calorimety technique (ITC) were employed for qualitative and quantitative analysis of binding affinity of *Ct*CBM6 with various ligands. ITC is the only technique that can directly measure the ligand binding energetics of proteins. It is one of the most significant technique used in molecular glycobiology (Dam and Brewer, 2002). Biochemical characterization of *Ct*CBM6 was carried out in order to study the structure and functional relationships.

5.2 Materials and Methods

5.2.1 Reagents, chemicals and substrates

The reagents for polyacrylamide gel electrophoresis *viz.* acrylamide, bisacrylamide, tris free base, glycine, bromophenol blue, β-mercaptoethanol and Coomassie Brilliant Blue R250 were obtained from Sigma-Aldrich Pvt. Ltd. USA. Chemicals and other ingredient like glycerol, methanol, concentrated hydrochloric acid, glacial acetic acid and $CaCl_2$ were supplied by Merck, India. Natural polysaccharides like rye arabinoxylan, wheat arabinoxylan (soluble and insoluble), arabinogalactan, sugar beet arabinan, rhamnogalactouronan, konjac glucomannan and locus bean galactomannan were procured from Megazyme International, Ireland. Oat spelt xylan, birchwood xylan, beechwood xylan, 4-O-methyl glucuronoxylan, barley β-D-glucan, carboxy methylcellulose (CMC), carboxy ethylcellulose (CEC), curdlan, pullulan, pectic galactan (from apple, citrus and lupin), xyloglucan and galactan (lupin) were purchased from Sigma-Aldrich, USA.

5.2.2 Cloning, expression and purification of WT *Ct*CBM6

The complete methodology of cloning, expression and purification of *Ct*CBM6 is described in Chapter 2.

5.2.3 Ligand binding analysis of *Ct*CBM6

5.2.3.1 CtCBM6 on native-PAGE with soluble ligands

Affinity gel electrophoresis (AGE) of *Ct*CBM6 using Native-PAGE was carried out against soluble ligands (polysaccharides) following the method described elsewhere (Takeo, 1995; Tomme *et al*., 2000). The 6.5% resolving (native) gel was prepared using stock solutions *viz.* 30% (w/v) acrylamide, 1.0% (w/v) polysaccharide, 1M Tris-HCl buffer (pH 8.8), 50% (v/v) glycerol, 10% (w/v) ammonium per sulfate

and TEMED as described for SDS-PAGE in Chapter 2, Section 2.2.20. There were certain differences which were followed to run the native-PAGE to perform affinity gel electrophoresis. No SDS in the resolving gel (there was no stacking gel) as well as in the native-PAGE running buffer was added. 1.0% (w/v) stock solution of all the mentioned ligands in Section 5.2.1 were made by dissolving in sterile deionized water until a clear solution was obtained. Thereafter, necessary amount of ligand solution (from the stock) was added to the resolving gel prior to the polymerization to achieve 0.008 to 0.2% (w/v) final concentrations. The remaining components of native-PAGE, described above were added and the gel was polymerized. The native-PAGE was carried out at 4°C rather than 25°C as in case of SDS-PAGE. Bovine serum albumin, BSA (10 μg) was run in the native gel as an internal reference to check for any nonspecific binding interactions. The electrophoresis was carried out at constant current of 2 mA per lane and at 4ºC using a Mini PROTEAN Tetra pack (Bio-Rad, USA). The gels were stained with Coomassie Brilliant Blue R250 (Sigma) and de-stained as described in Chapter 2, Section 2.2.20.6 to detect the protein bands.

5.2.3.2 Preparation of native-PAGE running buffer

The following composition (Table 5.2.1) was used to make a 5x native-PAGE running buffer of pH 8.3 without SDS.

Table 5.2.1 Composition of 5x native-PAGE running buffer.

Components	Final concentration (5x buffer)
Tris base	0.125 M
Glycine	1.25 M

5.2.3.3 Preparation of protein sample loading buffer

A 5x sample loading buffer was prepared by dissolving the components as described in Table 5.2.2 and pH of the buffer was adjusted to 6.8. Final concentration of sample buffer brings to 1x before using, by taking 4 volume of sample (protein) and 1 volume of 5x sample buffer.

Table 5.2.2 Composition of 5x sample loading buffer (Laemmli, 1970).

Components	Final concentration (**5x buffer**)
Tris-HCl buffer (pH 6.8)	0.2 M
Glycerol	20.0 (%,v/v)
Bromophenol Blue	0.05 (%, w/v)

5.2.3.4 Binding analysis of CtCBM6 with soluble polysaccharides

Affinity gel electrophoresis (AE) using Native-PAGE was carried out in order to determine the equilibrium association constant (K_a) of *Ct*CBM6 against soluble polysaccharides following the method described earlier (Takeo, 1995; Tomme *et al*., 2000). Native-PAGE using 6.5% gel containing soluble polysaccharides (birchwood, beechwood, 4-O-methyl glucuronoxylan, oat spelt xylan, wheat arabinoxylan or rye arabinoxylan) at concentrations varying from 0.008 to 0.2% (w v^{-1}), were added before gel polymerization. A 6.5% gel without any soluble polysaccharide was also run as a control. The electrophoresis was carried out at constant current of 2 mA per lane and at 4ºC using a Mini PROTEAN Tetra pack (Bio-Rad, USA). Purified *Ct*CBM6 (5.0 µg) was run on native gels with or without soluble polysaccharides. Bovine serum albumin, BSA (10 µg) was also run as a reference to check for any non-specific binding interactions. After staining the gel with coomassie brilliant blue, the migration of *Ct*CBM6 was measured directly on the gel from the edge of the loading well to the middle of the stained protein spot and the distances were assigned as migration without polysaccharide (R_0, mm), migration in the presence of

polysaccharide (R, mm) and distance travelled by dye front (R_d, mm). The relative mobility (r) was calculated as

$$r = R/R_d \quad \text{or} \quad r = R_0/R_d$$

The equilibrium binding constant (K_a) was determined by plotting 1/r (y-axis) against the concentration of polysaccharide on x-axis (%, w v^{-1}). The regression line of the plot was extrapolated to the x-axis and the point where the regression line intersected the x-axis gave equilibrium dissociation constant (K_d). The reciprocal of K_d gives K_a as reported earlier by Tomme *et al*. 2000.

$$K_d = 1/K_a$$

The effect of Ca^{2+} ions on ligand binding affinity of *Ct*CBM6 was also studied by incorporating 5 mM Ca^{2+} ions and polysaccharide into the gel prior to its polymerization. The gels were analysed by measuring the retardation in *Ct*CBM6 migration with respect to the control.

5.2.3.5 *Polysaccharides binding analysis of CtCBM6 by Isothermal titration calorimetry*

Isothermal titration calorimetry (ITC) was employed in order to determine the thermodynamic parameters *viz.* Gibb's free energy (ΔG), enthalpy (ΔH), entropy (ΔS) and stoichiometry (N_0) of *Ct*CBM6 upon ligand binding. ITC experiments were carried out essentially as described previously (Carvalho *et al*., 2007; Carvalho *et al*., 2004), except that the titrations were carried out at 55ºC, as the *Ct*CBM6 is from *Clostridium thermocellum. Ct*CBM6 was taken in 50 mM Na-HEPES buffer, pH 7.5, containing 2 mM $CaCl_2$. During titration *Ct*CBM6 (60 µM) was stirred at 300 rpm in the reaction cell, which was injected with 28 successive 2 µl aliquots of ligands at (1.5 mM) 180s intervals. Integrated heat effects, after correction for the heats of dilution, were analysed by non-linear regression using a single site-binding model (Microcal

ORIGIN, Version 5.0; Microcal Software). The fitted data yielded the association constant (K_a) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated by using the standard thermodynamic equation, $-RT\ln K_a = \Delta G = \Delta H - T\Delta S$.

5.2.3.6 *Binding affinity of CtCBM6 with insoluble polysaccharides*

Qualitative binding affinity analysis of *Ct*CBM6 with insoluble polysaccharides was performed as described earlier (Boraston *et al*., 2000). 20 µg of purified *Ct*CBM6 was incubated with 1 mg insoluble wheat arabinoxylan (WAXI) or 1 mg avicel in 200 µl, 20 mM Tris-HCl, pH 7.0 with vigorous shaking at 4ºC for 2h. The reaction mixture was centrifuged at 13000***g*** at 4°C for 5 min. The polysaccharide pellet was washed 3 times with 200 µl of buffer (20.0 mM Tris-HCl, pH 7.0). Finally, the polysaccharide pellet was re-suspended in 200 µl of 10% (w/v) SDS containing 10% (w/v) β-mercaptoethanol and boiled for 10 min. The supernatant containing the unbound protein and the resuspended washed pellet containing bound protein (each 30 µl) were analyzed by SDS-PAGE using 14% gel. The quantitative binding affinity assessment by adsorption isotherm was carried out by the method of Gilkes *et al* (Gilkes *et al*., 1992). The reaction was carried out in a 200 µl volume containing varying concentrations (1µM-25µM) of *Ct*CBM6 and 1 mg/ml insoluble polysaccharide using the same reaction conditions as described above. The reaction was set up in triplicate with the respective controls, without the polysaccharide run in parallel. The supernatant represents the unbound or free protein [F], which was measured spectro-photometrically at A_{280}. The concentration of bound protein [B] for a particular concentration was calculated by subtracting the amount of free protein [F] from the initial protein concentration.

Based on this, equilibrium association constant (K_a) was obtained from the depletion isotherms (plot of [B] versus [F]) after fitting (nonlinear regression) of the raw data to an equilibrium binding type or adsorption model [39].

$$K_a = [B]/N\,[F]$$

where, N is the concentration of available binding sites on the substrate (the number of polysaccharide lattice units) [moles g^{-1} of CBM], [B] is the concentration of bound CBM (mol g^{-1} of substrate), [F] is the molar concentration of free CBM and K_a is the equilibrium association constant [M^{-1}].

The relative equilibrium association constant, K_r (litres/g of substrate) was calculated as,

$$K_r = K_a\, N_o$$

K_a = is the equilibrium association constant [M^{-1}]

N_o = is the concentration of available binding sites on the substrate (the number of polysaccharide lattice units) [moles g^{-1} of CBM]

N_o were calculated from a non-linear regression plot against µ mole of protein bound per gram of wheat arabinoxylan insoluble (µ mole g^{-1}) versus the free protein *Ct*CBM6 (µM) (Gilkes *et al*., 1992). Data were analyzed by GraphPad (Prism version 5) software using non-linear regression using one binding site equation.

5.2.4 Melting curve of *Ct*CBM6

The protein melting curve for *Ct*CBM6 was generated by incubating 20 µg/ml protein in 20 mM Tris-HCl buffer, pH 7.0 at temperature varying from 40-100°C using a peltier temperature controller. The protein melting experiments were performed on a UV-visible spectrophotometer (Varian, Carry 100-Bio). The absorbance at 280 nm (A_{280}) (tryptophan absorption maximum) was measured with the increase in the temperature following the method, described earlier (Dvortsov *et*

al., 2009). The temperature of the cuvette block was raised to 3°C/min and data were collected after every 3°C rise in the temperature. Measurements were repeated three times to achieve a 5% margin of error. The melting point of unfolded *Ct*CBM6 was graphically determined by plotting UV absorption units *versus* temperature (Abou-Hachem *et al*., 2002). The effect of Ca^{2+} ions on *Ct*CBM6 was also analyzed by generating the melting curve in the presence of 5 mM Ca^{2+} ions.

5.2.5 *In silico* characterization of *Ct*CBM6

5.2.5.1 Homology modeling, model refinement and quality assessment

The protein sequence (Gene accession number ABN54208) search for a closest structural homologue template to model *Ct*CBM6 was achieved by 'BLASTp' program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) from the Protein Data Bank (PDB) database. 3-D modeling was accomplished by Modeller9v8 using the template of crystal structure of family 6 CBM from *C. thermocellum* (PDB id: 1GMM) (Sali *et al*., 1995). Query sequence was aligned with the template sequence of known structure which acted as an input for the Modeller. Initially, twenty independent models were generated after optimization of molecular probable density functions (MOLPDF score). The modeled structure was further subjected to loop refinement by loop optimization and after each cycle of loop refinement, a discrete optimized protein energy (DOPE) score was generated. This process was repeated in an iterative fashion until DOPE score attained a negative value (Fiser *et al*., 2000). The best model having reasonable DOPE and MOLPDF scores was chosen for further refinement. Energy minimization steps were performed by a steepest descent algorithm with GROMOS96 43a1 force field and simple point charge (SPC) water model using GROMACS4.0.7 package (http://www.gromacs.org/). A simulation box

was created by introducing the water molecules as a solvent system for carrying out *Ct*CBM6 stability study. Overall charge on the protein molecule was neutralized by adding counter-ions (here 3 Na^+ ions) and the same number of water molecules which were at least 3.50 Å apart was replaced from the protein surface. The final energy minimized model was validated through various parameters on the structure analysis and verification server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES/) such as Ramachandran plot (Ramachandran *et al*., 1963), VERIFY3D (Eisenberg *et al*., 1997) and ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) (Sippl, 1993; Wiederstein and Sippl, 2007)

5.2.5.2 Molecular dynamic simulation of CtCBM6 model

Molecular dynamics simulations of energy-minimized *Ct*CBM6 model was performed in the defined cubic box of single point charge (SPC) water molecules. The net charge of the system was neutralized with counter ions (Na^+ ions). The whole system was equilibrated for 500 ps by NVT (N; amount of substance, V; volume, T; temperature) ensemble by restraining the solute atoms through a harmonic force constant of 1000 kJ nm^{-2}. Then the system was equilibrated for 500 ps by NPT (N; amount of substance, P; pressure, T; temperature) ensemble twice, first with restraints followed by without restraints. Production run was carried out for 20 ns with NPT ensemble using 2 fs of integration time. The linear constraint solver (LINCS) algorithm was used to constrain the bonds involving hydrogen atoms (Goodsell and Olson, 1990). Conformations of *Ct*CBM6 model at regular intervals throughout the production run were analyzed as time-dependent function to check whether the *Ct*CBM6 model is energetically stable in the solvent system.

5.2.5.3 Binding site prediction

Multiple sequence alignment and structural superposition of *Ct*CBM6 were performed with known structures of family 6 CBMs *viz. Ct*CBM6 (PDB id: 1GMM *Clostridium thermocellum*), *Cs*CBM6 (PDB id: 1NAE from *Clostridium stercorarium*), *Cm*CBM6-2 (PDB id: 1UY1 from *Cellvibrio mixtus*), *Bh*CBM6 (PDB id: 1W9S from *B. halodurans*) and *Cc*CBM6 (PDB id: 2V4V from *C. cellulolytium*) to identify the key residues and their role in ligand binding. Structure superposition was done using PyMOL (DeLano W.L. (2002) The PyMOL Molecular Graphics System. San Carlos). Related protein sequence were retrieved from PDB (http://www.rcsb.org/pdb/), saved in FASTA format. Alignment was performed using the Clustal X program and the final alignment was generated by ESpript (http://espript.ibcp.fr).

5.2.5.4 Molecular docking of modeled CtCBM6

Molecular docking studies were carried out using Autodock 4.2.1. Autodock uses a novel and robust docking method by implementing a new scoring function that estimates the free energy change during ligand binding (Goodsell and Olson, 1990). Required input files were prepared by assigning gasteiger charges, merging non-polar hydrogens and save in PDBQT format using AutoDock Tools (ADT) 1.5.4. Grid box was set to 40, 52, 40 (x,y,z) with 0.375 spacing covering the entire binding site cavity. Lamarckian Genetic Algorithm was chosen for conformational search and numbers of evaluation were set to be 30 (Morris *et al*., 1998). Best docked conformation from the largest cluster having a minimum lowest free energy of binding (ΔG), were saved. Results were analyzed using PyMOL (0.99) for possible polar and hydrophobic interaction. Representation of carbohydrate-protein complex was generated via

Ligplot (http://www.ebi.ac.uk/pdbsum/) for a better understanding of protein ligand interaction.

5.2.6 Secondary structure analysis of *Ct*CBM6

Analysis of *Ct*CBM6 by Circular dichroism (CD) was carried out at 25°C on a spectropolarimeter (Jasco Corporation, Tokyo, JASCO J-810), equipped with a peltier system for temperature control. The protein concentration (0.09 mg/ml) in 20 mM TrisHCl buffer (pH 7.4) was used. The average of six scans was carried out from 195-250 nm wavelengths, with scan a speed of 50 nm min^{-1}, using a bandwidth of 1 nm and resolution of 1 nm. The far-UV spectrum was plotted in terms of the mean residue ellipticity (MRE, expressed as deg cm^2 $dmol^{-1}$) as a function of wavelength, calculated by the procedure described earlier (Kelly *et al*., 2005). The percentage of secondary structure, content was calculated using web based K2d3 program (http://k2d3.ogic.ca/). It uses different algorithms or neural network approach to estimate the percentage of secondary structure by comparing the similarity between the CD spectrum of unknown protein with the database of known CD spectra (reference set) as described by earlier (Louis-Jeune *et al*., 2011). The content of secondary structure elements of *Ct*CBM6 were also verified by using PSIPRED web server (http://bioinf.cs.ucl.ac.uk/psipred/). It uses the two-stage neural network approach based on the position specific scoring matrices generated by PSI-BLAST (Jones, 1999).

5.2.7 Construction of *Ct*CBM6 mutants by site-directed mutagenesis

Site-directed mutagenesis was carried out using the PCR-based NZY mutagenesis kit (NZY Tech Ltd., Portugal) according to the manufacturer's instructions, using recombinant plasmid containing the gene encoding *Ct*CBM6 (as described in Chapter 2) as template. Three mutants of *Ct*CBM6, Y24A, F84A and double mutant (Y24A/F84A) were created. The sequences of the primers used to generate these mutants are displayed in Table 5.2.3.

Table 5.2.3 Oligonucleotide primers for site-specific mutant of *Ct*BM6.

Construct	Primer sequence
WT	Fwd:5'-GGAAGCGGTCTCGGATATATCGAAAACGGCAAC-3'
Tyr-24-Ala	Fwd:5'- GGAAGCGGTCTCGGAGCTATCGAAAACGGCAAC-3'
	Rev: 5'- GTTGCCGTTTTCGATAGCTCCGAGACCGCTTCC-3'
WT	Fwd:5'-GCTGCAACCGGAGGCTTTAATGCCTATGAGG-3'
Phe-84-Ala	Fwd:5'- GCTGCAACCGGAGGCGCTAATGCCTATGAGG-3'
	Rev: 5'-CCTCATAGGCATTAGCGCCTCCGGTTGCAGC-3'

5.2.7.1 RCR reaction

The PCR reaction was setup by following the manufacturer's instructions. The different PCR components and their respective final concentration are mentioned in the Table 5.2.4. The components of PCR reaction were mixed gently and the tube was short spin for 10 second before placing for amplification.

Table 5.2.4 PCR mixture reagents for site-directed mutagenesis of *Ct*CBM6.

PCR components	Volume (µl)	Final concentration
10x reaction buffer	5.0	1x
25 mM dNTP mix	0.5	0.25 mM
Forward primer (100 mM)	1.0	2.5 ng
Reverse primer (100 mM)	1.0	2.5 ng
Sigma water, pH 8.0	40.5	--
WT *Ct*CBM6 DNA (50 ng/µl)	1.0	1.0 ng
Pfu DNA polymerase (2.5 U/µl)	1.0	2.5 U
Total	50.0	--

5.2.7.2 *PCR conditions*

The PCR conditions were chosen according to the manufacturer's instructions. The PCR amplification was performed using a thermal cycler (Applied Biosystems, GeneAmp PCR System 9700). PCR cycles for amplification is shown in Table 5.2.5

Table 5.2.5 PCR cycles for generating site-specific mutant of *Ct*CBM6.

Steps		Time (min)
I. Denaturation at 95°C		1
II. 18 cycles of	i) Denaturation at 95°C	1
	ii) Annealing at 60°C	1
	iii) Extension at 68°C	11
III. Final extension at 68°C		10

5.2.7.3 *DpnI endonuclease digestion of PCR amplified product*

The restriction digestion of PCR amplified product by *DpnI* endonuclease was carried out as described earlier in the Chapter 3, Section 3.2.11.3.

5.2.7.4 *Transformation of resultant product*

DpnI digested reaction product was transformed using the *E. coli* BL-21 competent cell following the protocol as descried earlier in Chapter 3, Section 3.2.11.4.

5.2.7.5 *Screening of positive mutants*

The mutant plasmid DNAs were send for sequencing to GATC Biotech AG**,** European Custom Sequencing Centre, Cologne, Germany to confirm if the appropriate mutations were incorporated.

5.3 Results and Discussion

5.3.1 Cloning, expression and purification of recombinant *Ct*CBM6

*Ct*CBM6 was cloned, expressed and purified as a soluble ~14 kDa protein as describes in the chapter 2.

5.3.2 Qualitative binding analysis of *Ct*CBM6 with soluble polysaccharides by affinity gel electrophoresis

Affinity gel electrophoresis (AE) of *Ct*CBM6 with 24 different soluble natural polysaccharides clearly suggested that it predominantly binds xylan backbone (β-1→4-xylopyranose) containing polysaccharides (Fig. 5.3.1). The screening results of binding affinity are summarized in Table 5.3.1. *Ct*CBM6 showed negligible affinity towards celluloses, pectins and other polysaccharides. The ligand specificity of *Ct*CBM6 is also in the agreement with the fact that it displayed affinity to same polysaccharides to which its cognate catalytic module, *Ct*Xyn30A showed activity as described in the Chapter 3, Section 3.3.1. After elucidating the specificity of *Ct*CBM6 towards xylan main chain substrate, the equilibrium binding constant and other thermodynamic parameters were calculated for respective natural polysaccharides.

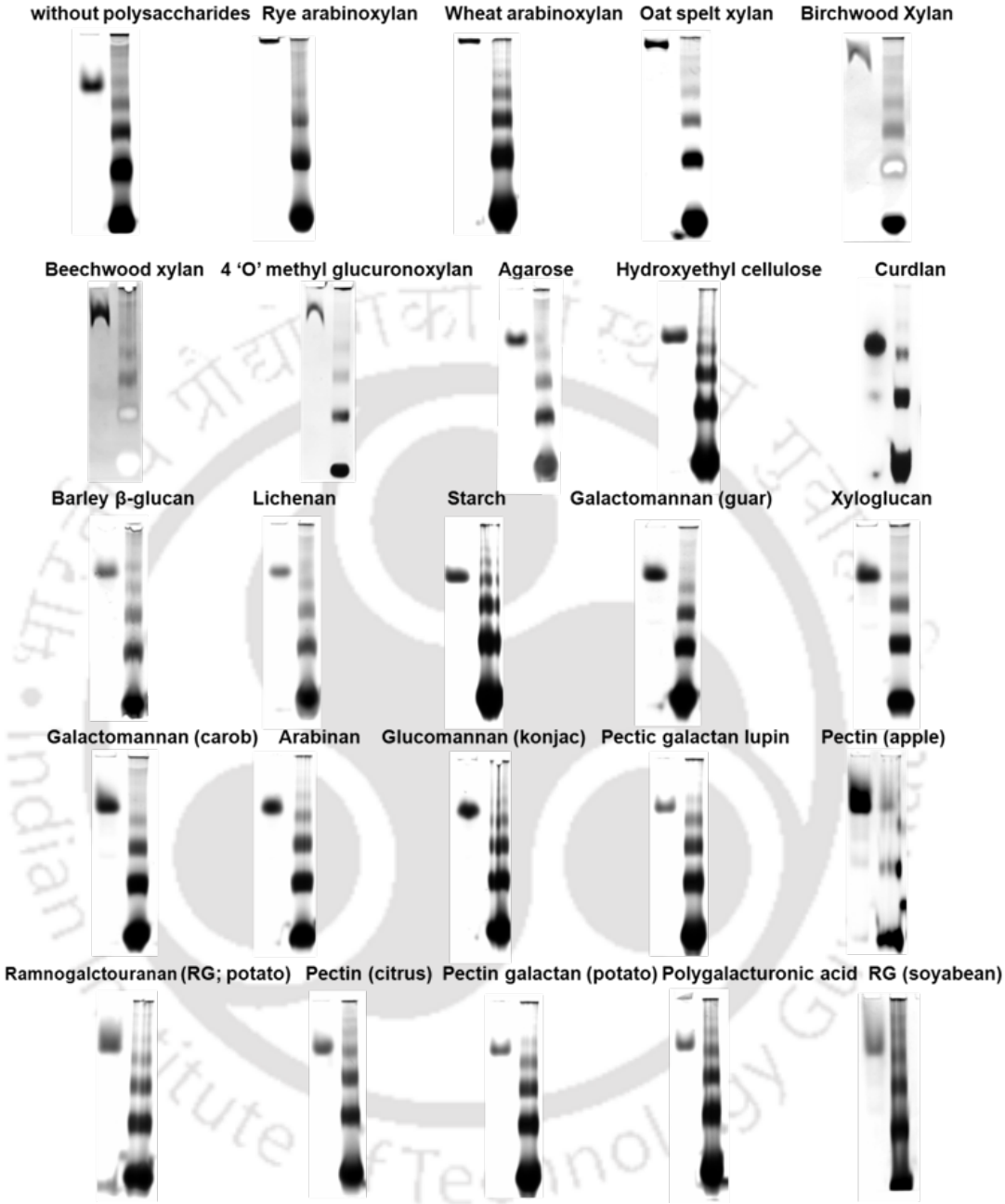


Fig. 5.3.1 Screening of various polysaccharides substrate binding with *Ct*CBM6 by native-PAGE using 6.5% gels run at 4°C. Bovine Serum Albumin (BSA) was used as an internal reference Final concentration 0.1% (w/v) of ligand was used.

Table 5.3.1 Binding affinity of *Ct*CBM6 towards various polysaccharides.

Ligand[a]	Binding[b]
Xylans	
Beechwood xylan	++
Birchwood xylan	++
Glucuronoxylan	++
Oat-spelt Xylan	+++
Rye arabinoxylan	+++
Wheat arabinoxylan	+++
Cellulosic substrates*	–
Pectins**	–
Hemicelluloses***	–

* *Barley β-glucan, Carboxymethyl cellulose (CMC), Curdlan, Hydroxyl ethyl cellulose (HEC), Lichenan (icelandic moss) and Starch*

** *Pecticgalactan (lupin), Pecticgalactan (Potato), Pectin from apple, Polygalacturonic acid (PGA), Rhamnogalacturonan (Potato) and Rhamnogalacturonan (soyabean)*

****Arabinan (sugar beet), Galactomannan (carob), Galactomannan (guar), Glucomannan (Konjac), Xyloglucan (Tamarind)*

[a] Ligand concentration, 1 mg/ml.

[b] Symbols used: –no detectable binding; ++Significant binding; +++ Strong binding.

5.3.2.1 Quantitative binding analysis of CtCBM6 by affinity gel electrophoresis

Quantitative assessment of the binding affinity of *Ct*CBM6 towards soluble polysaccharides in terms of the equilibrium association constant (K_a) was carried out by AE. The increase in retardation of *Ct*CBM6 migration with an increase in the concentration of soluble polysaccharides was observed (Fig. 5.3.2 A, i-F, i) and the plot of 1/*r* versus polysaccharide concentration for determination of K_a by regression analysis for each polysaccharide is shown in figure 5.3.2 (A, ii – F, ii). The K_a values indicated that *Ct*CBM6 shows higher affinity towards wheat- and rye- arabinoxylans (highly decorated with arabinose moiety) than poorly decorated xylan (substituted with glucuronic acid or arabinose) such as beechwood-, birchwood- and oat spelt-

xylan (Table 5.3.2). The presence of Ca^{2+} ions (5 mM) along with the polysaccharides only slightly reduced the migration of *Ct*CBM6 henceforth; it helps to increase insignificantly the binding affinity towards the ligands (Fig. 5.3.3).

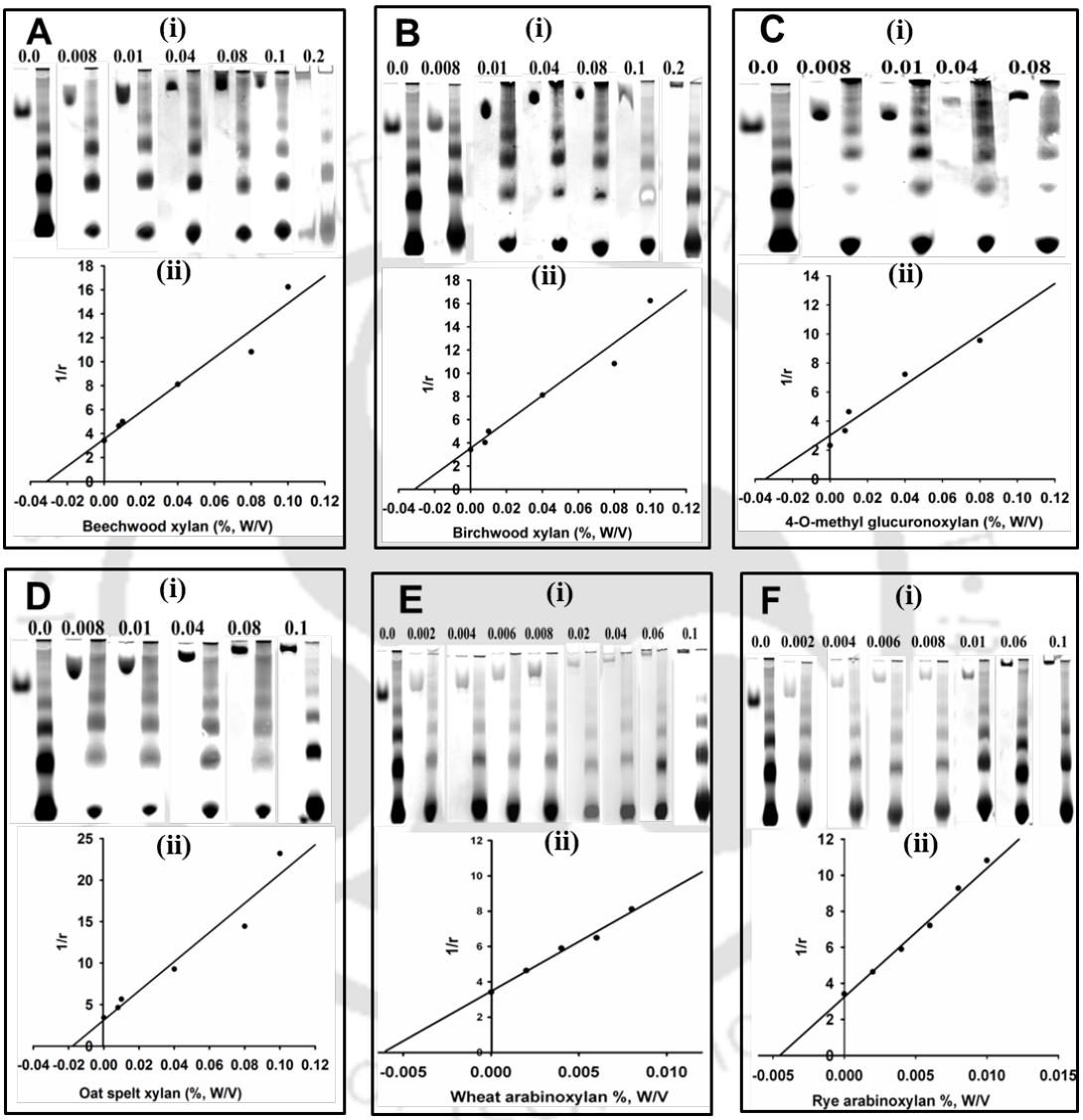


Fig. 5.3.2 Quantitative binding analysis by affinity gel electrophoresis using Native-PAGE (6.5%, w/v) gels run at 4°C with different xylan based substrates, **(A, i & ii**) Beechwood xylan **(B, i & ii)** Birchwood xylan **(C, i & ii)** 4-O-methyl glucuronoxylan **(D, i & ii)** Oat spelt xylan **(E, i & ii)** Wheat arabinoxylan and **(F, i & ii)** Rye arabinoxylan. Bovine Serum Albumin (BSA) was used as a reference protein.

Table 5.3.2 Quantitative binding analysis of *Ct*CBM6 with soluble polysaccharides by AE and ITC.

Ligand	$K_a \times 10^4$ (M^{-1}) AE (4°C)	$K_a \times 10^4$ (M^{-1}) ITC (55°C)	ΔG kcal/mol	ΔH kcal/mol	$T\Delta S$ kcal/mol	N_0
Beechwood xylan	0.53	3.33±0.33	-6.8	-10.4±0.58	-3.6	1.18±0.05
Birchwood xylan	0.52	3.22±0.37	-6.8	-7.9±0.48	-1.1	1.30±0.06
Oat spelt xylan	0.61	3.20±0.4	-6.7	-2.9±0.17	3.8	1.50±0.06
4-O-methyl Glucuronoxylan	0.50	3.25±0.35	-6.8	-9.2±0.56	-2.4	1.17±0.06
Wheat arabinoxylan	2.75	3.39±0.45	-6.8	-2.9±0.15	3.9	1.68±0.06
Rye arabinoxylan	3.71	1.40±0.37	-6.3	-12.5±0.51	-6.2	0.77±0.27

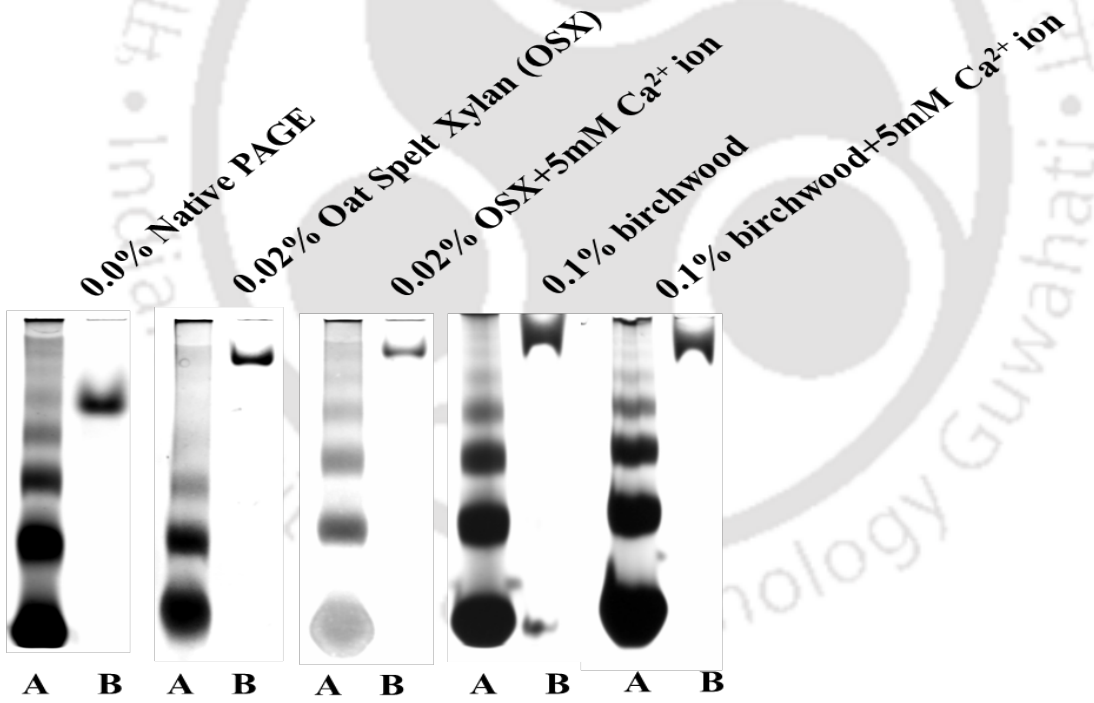


Fig. 5.3.3 Affinity electrophoresis of *Ct*CBM6 using 6.5% native gel to demonstrate the effect of Ca^{2+} ion on binding. Lane A represent the Bovine serum albumin (BSA), Lane B represent the *Ct*CBM6.

5.3.3 Determination of binding constant and thermodynamic parameters of *Ct*CBM6 by ITC

To investigate, in more detail, the ligand specificity of *Ct*CBM6, the capacity of the protein to bind to xylan constituents was assessed by ITC. The data, reported in Table 5.3.2 with example titrations displayed in Fig. 5.3.4, show that *Ct*CBM6 displays high binding affinity with different xylans through an exothermic mechanism. Thus, the thermodynamic data show that the binding of *Ct*CBM6 to all its ligands is enthalpically driven, while the change in entropy makes a negative contribution to overall affinity, as observed for the majority of CBMs studied to date. The equilibrium association constant (K_a) determined by ITC at 55°C showed higher values than those obtained by affinity gel electrophoresis at 4°C for all polysaccharides (Table 5.3.2). The difference in K_a values between AE and ITC might be due to the fact that *Ct*CBM6 is thermophilic and thus optimally functional at higher temperature. The limitation with AE is that, it cannot be performed at higher temperature nonetheless both the methods exhibited similar binding patterns displaying a relatively higher affinity against arabinoxylans (Table 5.3.2).

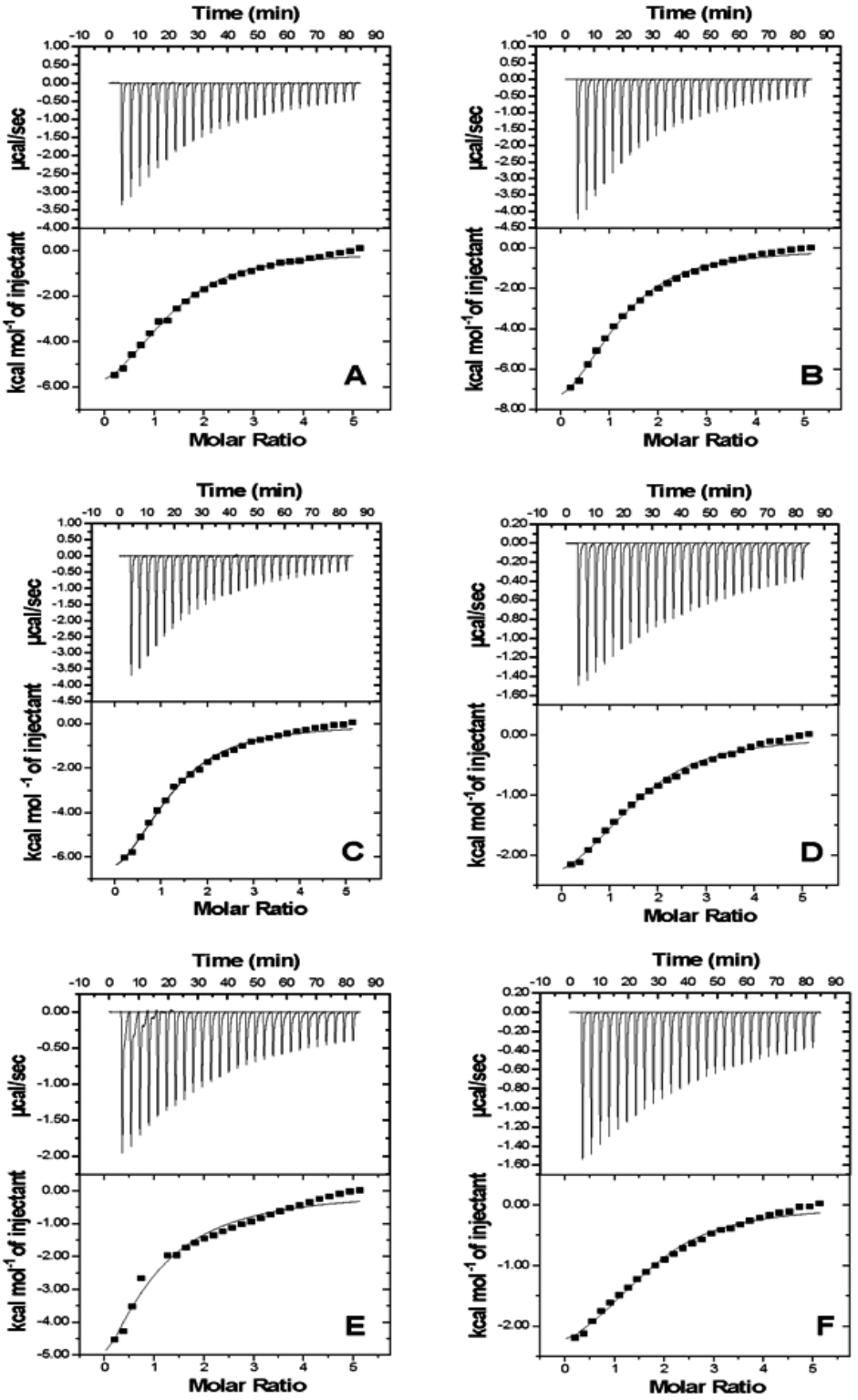


Fig. 5.3.4 Isothermal titration calorimetry (ITC) analysis of *Ct*CBM6 bonding with xylan-based polysaccharides at 55°C. **(A)** Birchwood xylan **(B)** Beechwood xylan **(C)** 4-O-methyl glucuronoxylan **(D)** Oat spelt xylan **(E)** Rye arabinoxylan and **(F)** Wheat arabinoxylan.

5.3.4 Binding analysis of *Ct*CBM6 with insoluble polysaccharide

Qualitative binding affinity experiments with insoluble ligand clearly indicated that *Ct*CBM6 binds insoluble wheat arabinoxylan, but under the same conditions, no significant binding was observed with insoluble avicel (Fig. 5.3.5).

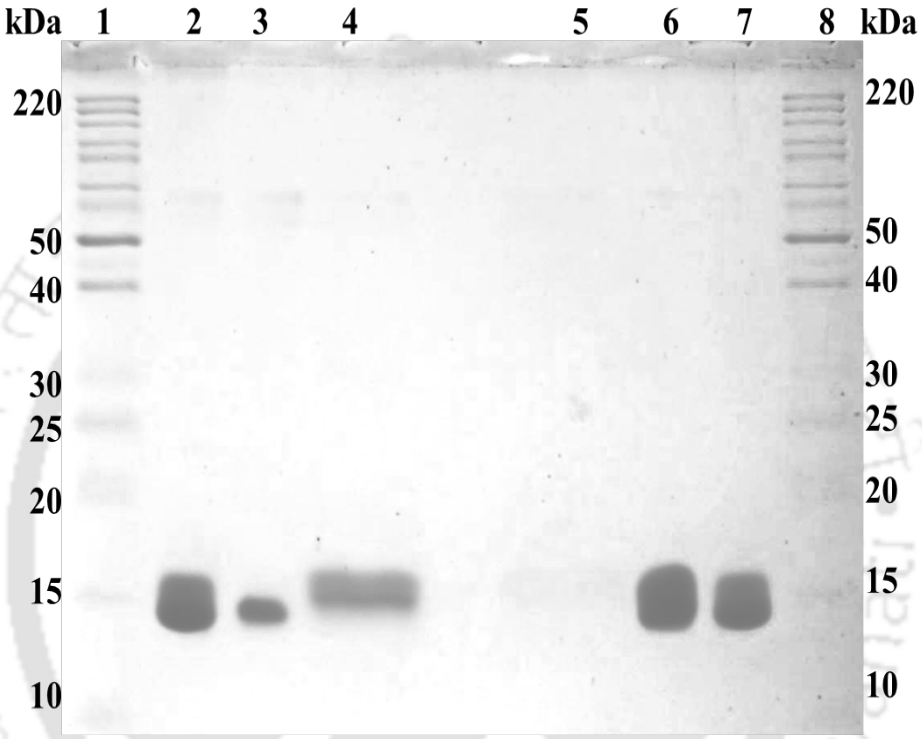


Fig. 5.3.5 Qualitative binding analysis of *Ct*CBM6 against insoluble wheat arabinoxylan and avicel by SDS-PAGE. Lane 1 and 8: protein marker, Lane 2 and 7: purified *Ct*CBM6 (20 µg), Lane 3: bound *Ct*CBM6 fraction (wheat arabinoxylan), Lane 4: unbound *Ct*CBM6 fraction (wheat arabinoxylan), Lane 5: bound *Ct*CBM6 fraction (avicel), Lane 6: unbound *Ct*CBM6 fraction (avicel).

The quantitative analysis of binding carried out with insoluble wheat arabinoxylan using adsorption isotherm (Fig. 5.3.6) showed the association constant (K_a) value of 6.0×10^7 M^{-1}. The estimated values of relative equilibrium constant K_r and concentration of binding sites [N_0] were 0.485 ± 0.11lg^{-1} and 0.287 ± 0.08 µmole g^{-1}, respectively.

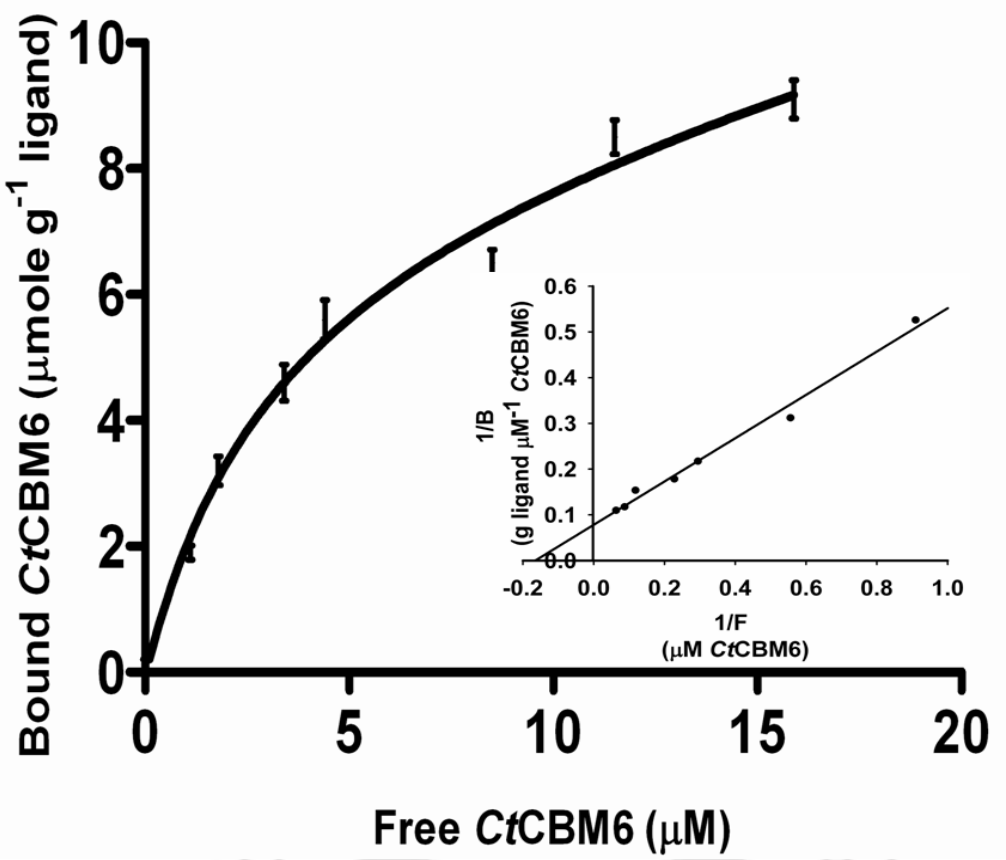


Fig. 5.3.6 Adsorption of *Ct*CBM6 to insoluble wheat arabinoxylan. The figure shows the equilibrium adsorption isotherm ([B] versus [F]) for *Ct*CBM6. Initial protein concentrations of *Ct*CBM6 were 1–25 mM. Inset shows the linear regression plot of 1/[B] versus 1/[F], to derive the association constant (*K*a).

5.3.5 Protein melting-curve of *Ct*CBM6

The protein melting curve of *Ct*CBM6 showed a single melting peak at 82°C confirming that this CBM is highly thermostable. The presence of 5 mM Ca^{2+} ions slightly shifted the peak towards the higher temperature 85°C, indicating the insignificant role of Ca^{2+} ions in the structural stabilization of *Ct*CBM6 (Fig. 5.3.7).

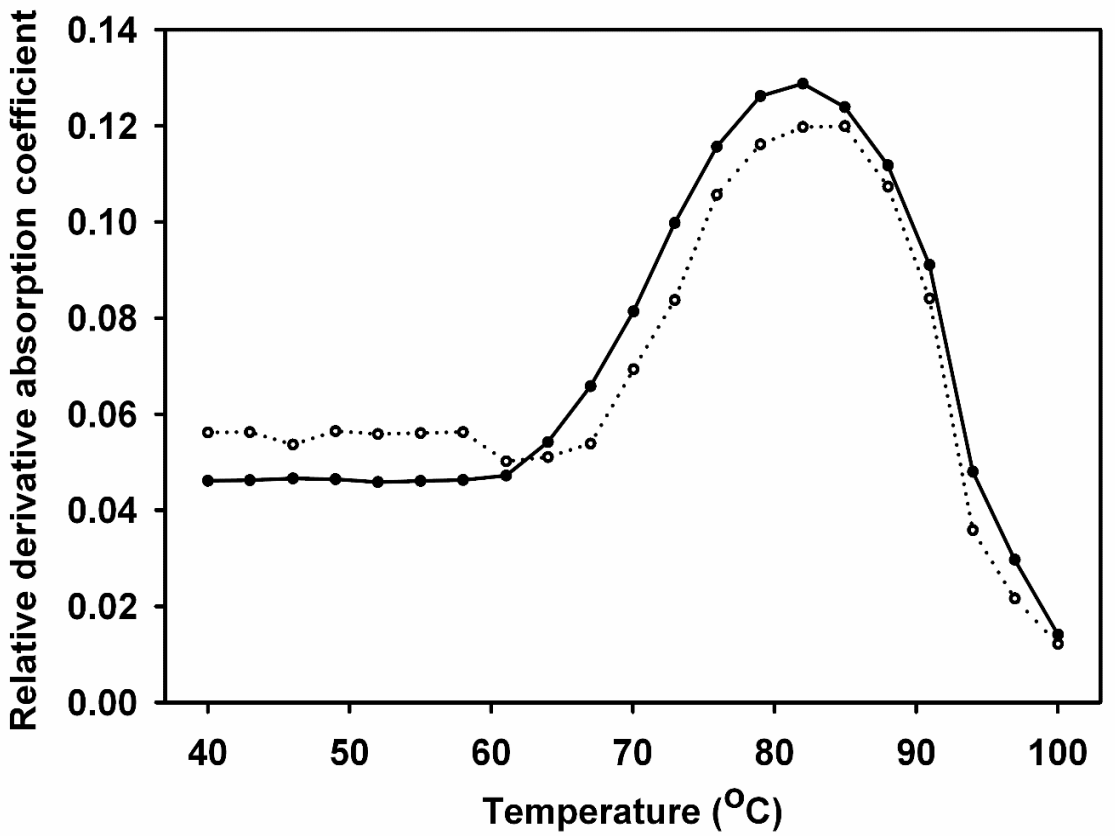


Fig. 5.3.7 Protein melting curve of *Ct*CBM6 with no additive, black filled circle (—●—) and in the presence of 5 mM Ca^{2+} ions represented with open circle (—○—).

5.3.6 Structure characterization of *Ct*CBM6

5.3.6.1 Homology modeling and structure validation

Homology modeling of *Ct*CBM6 having GenBank Accession No: ABN54208 and uniprot ID A3DJS9 was accomplished by using the closest structural homologue as a template. The crystal structure of family 6 CBM from *C. thermocellum* (PDB id: 1GMM) having 63% sequence identity helped to model the *Ct*CBM6 satisfactorily in terms of folds and functional residues. The final model of *Ct*CBM6 was validated with different quality assessment parameters. Ramachandran plot which implied that structure-backbone of dihedral angles, phi (φ) and psi (ψ) occupied acceptable statistics of amino acids with 96.1% in the core and 3.9% in additional allowed region (Fig. 5.3.8A). ProSA result indicated that the modeled protein was error free and resides in the NMR zone with Z-score of -6.17 as in figure 5.3.8B. VERIFY3D result depicted that all 123 amino acid residues have average 3D-1D profile score greater than 0.22 which illustrated that no amino acid is in bad contact (Fig. 5.3.8C). The quality assessment results were satisfactory and allowed using the structural information to support further studies on *Ct*CBM6.

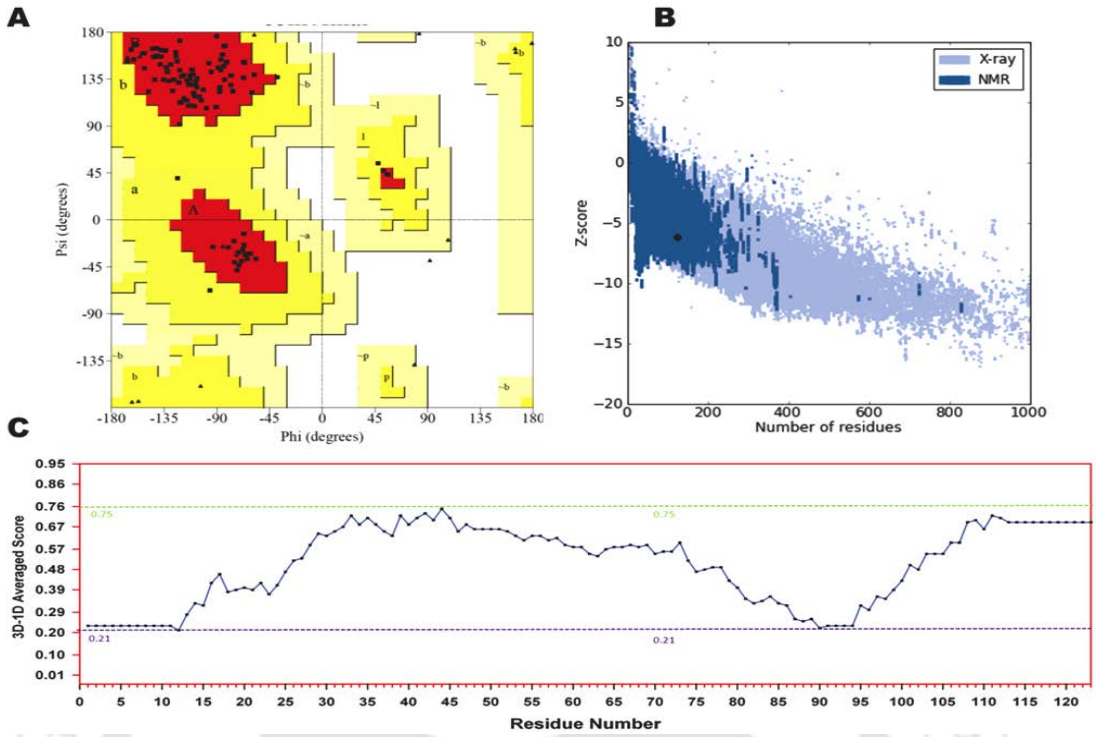


Fig. 5.3.8 Structure validation and quality assessment of *Ct*CBM6 by **(A)** Ramachandran plot **(B)** ProSA web server **(C)** Verify 3D.

5.3.6.2 Molecular dynamics study of CtCBM6

Molecular dynamics (MD) simulation was performed to ensure the stability of the protein throughout the simulation run. The results suggested that *Ct*CBM6 undergoes a significant change in the RMSD of backbone atom between 12 to 14 nano second (ns) of simulations revealing the possibility of conformational changes in *Ct*CBM6. The RMSD value converged from 12 to 20 ns, with fluctuations of less than 0.07 Å indicating the most stable conformation of *Ct*CBM6 model (Fig. 5.3.9). The RMSD between an energy minimized model of *Ct*CBM6 and the final structure from MD was 1.4 Å, represented the stable conformation. The MD plot signified that all the parameters are well within the qualitative limit of compactness and stability during the simulation. Therefore, the energy minimized model of *Ct*CBM6 is satisfactory for molecular docking simulation.

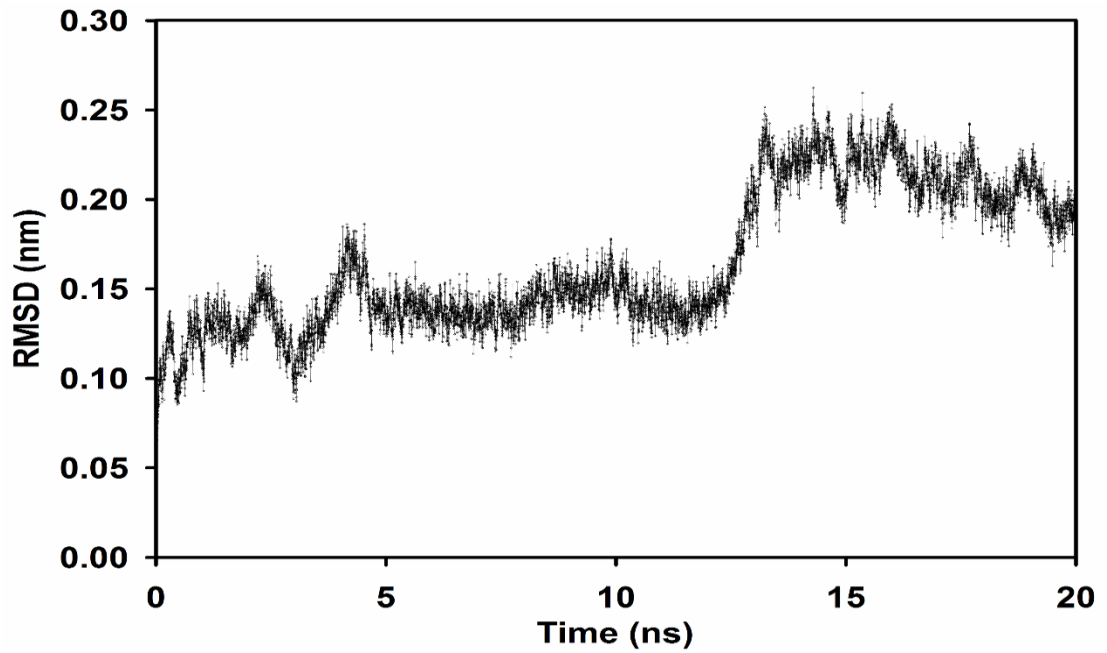


Fig. 5.3.9 MD simulation of modeled *Ct*CBM6 shows that the energy-minimized structure is energetically stable under the simulation conditions.

5.3.6.3 Overall structure features of CtCBM6

Modeled *Ct*CBM6 was rich in β strands and showed β-sandwich fold also known as β-jelly roll fold (Fig. 5.3.10A). The overall structure of *Ct*CBM6 has two anti-parallel β-sheets each consisting of five β-strands in which four β-strands from each sheet are opposite to each other, creating a β-sandwich fold. The remaining two β-strands extend away from β-sandwich core (β-3 and β-4 strand, Fig. 5.3.10B), making a finger-like structure similar to as reported for *Cs*CBM6 of *Clostridium stercorarium*, constituting a part of carbohydrate-binding site (Lammerts van Bueren and Boraston, 2004). The closest structure homologue of *Ct*CBM6 was *Cs*CBM6 (PDB id: 1NAE) with r.m.s. deviation of 0.5 Å over 110 matched C^{α} atoms on DALI server (Holm and Rosenstrom, 2010). The other CBMs which shared the similar fold were *Ct*CBM6 (*Clostridium thermocellum*; PDB id: 1GMM) with 0.7 Å r.m.s. deviation and *Cc*CBM6 (*Clostridium cellulolyticum*; PDB id: 2V4V) with 1.1 Å r.m.s. deviation.

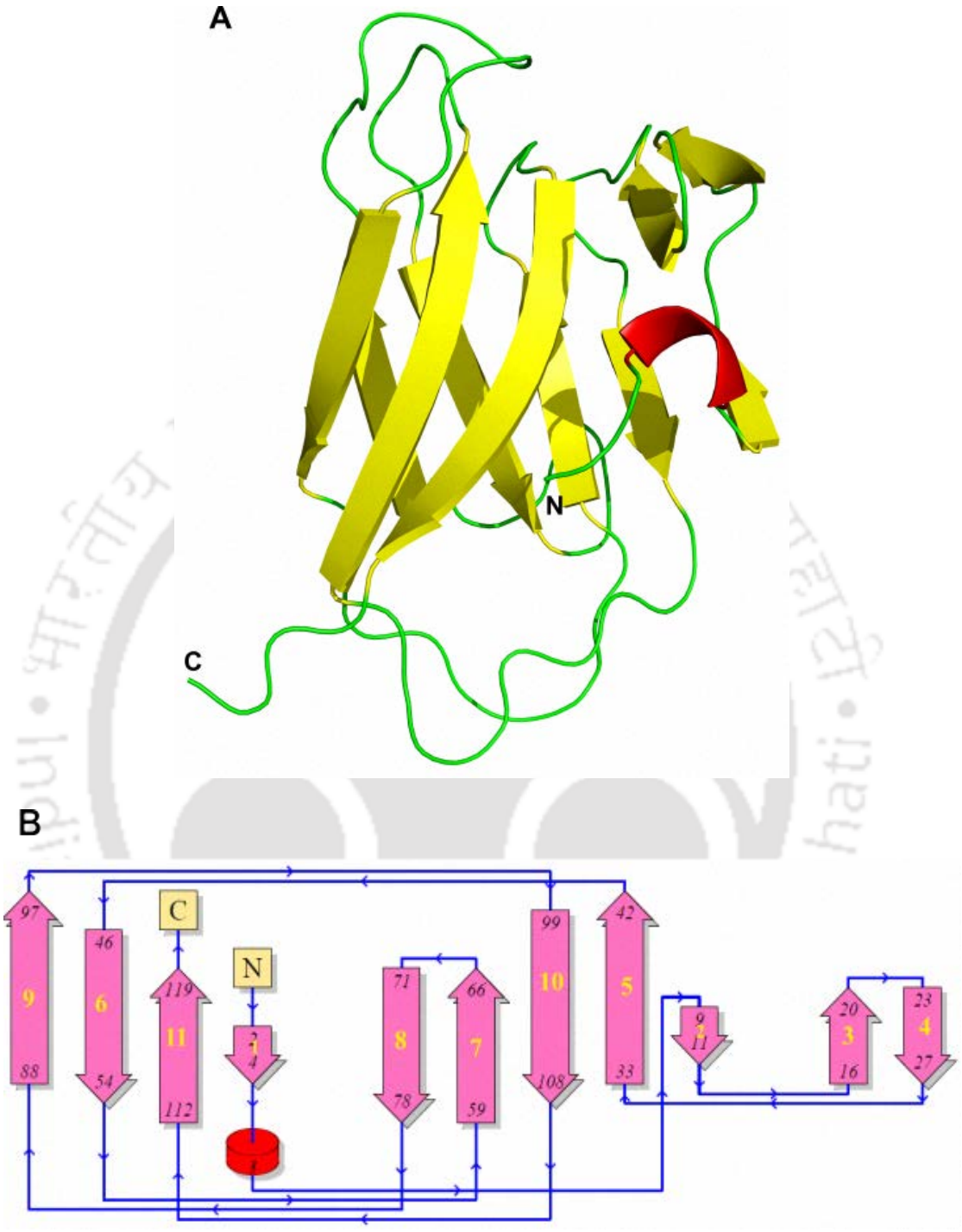


Fig. 5.3.10 **(A)** Three-dimensional cartoon representation of modeled *Ct*CBM6 showing β-Jelly roll containing predominantly anti-parallel β-sheets, **(B)** Topology diagram of *Ct*CBM6 represents mutual orientation of 1-11 β-strands shown by arrows (pink color) and cylinder (red color) representing the α-helix.

5.3.6.4 Secondary structure analysis of CtCBM6

The secondary structure predicted by PSIPRED inferred that *Ct*CBM6 has 47% β-stands and 53% random coil (Fig. 5.3.10). The determination of *Ct*CBM6 secondary structure from CD spectrum suggested the abundance of β-stands containing 44% β-stands and 55% random coil (Fig. 5.3.11). The secondary structure elements of *Ct*CBM6 determined by both the methods are summarized in the Table 5.3.3 display that the results of CD analysis corroborated those of the predicted structure.

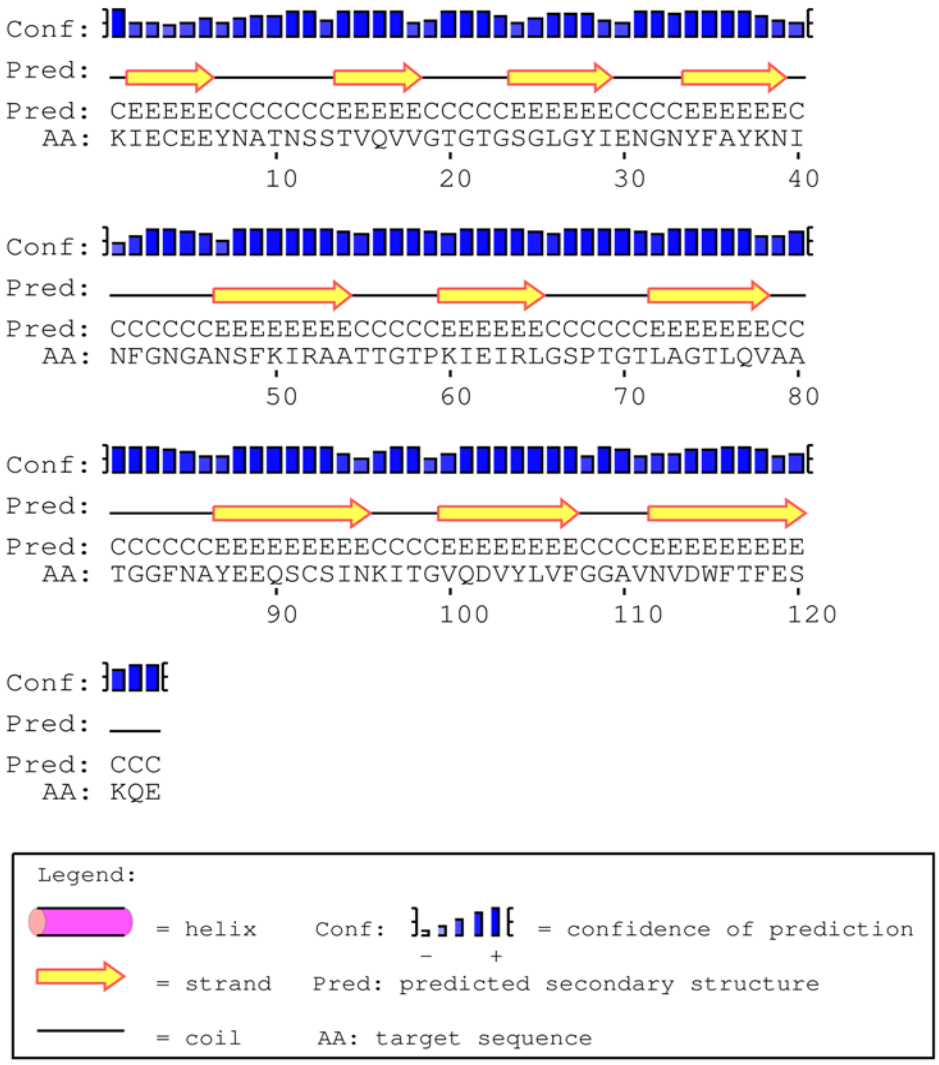


Fig. 5.3.10 Secondary structure prediction of *Ct*CBM6 by PSIPRED VIEW.

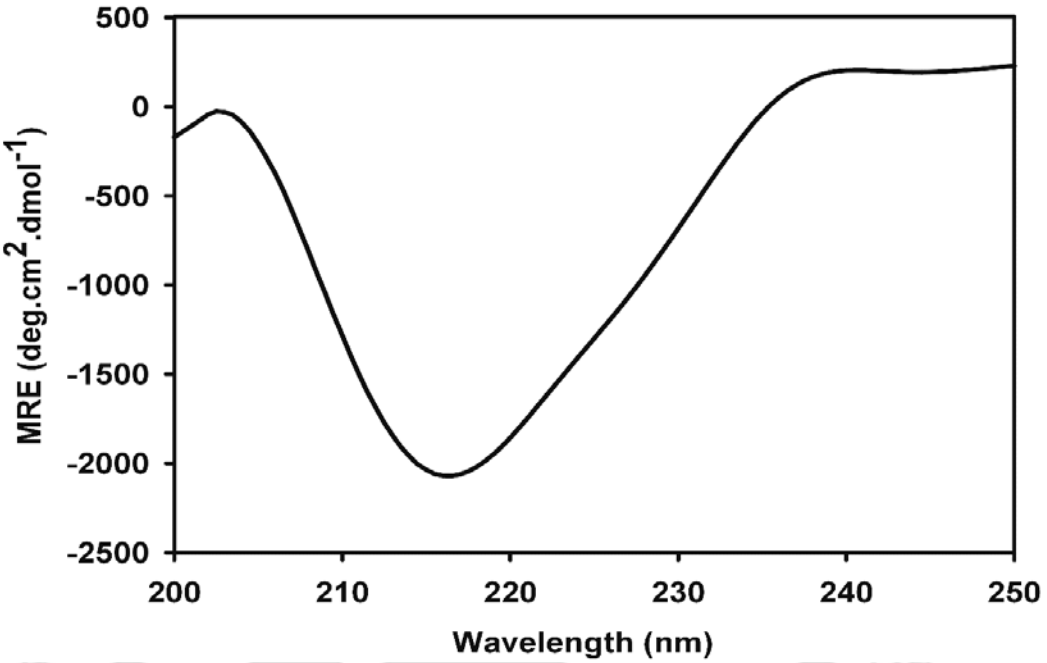


Fig. 5.3.11 Far UV-CD spectra of *Ct*CBM6 from *C. thermocellum*.

Table 5.3.3 Secondary structure elements of *Ct*CBM6 analyzed by far-UV CD spectrum and PSIPRED.

Secondary structure content of *Ct*CBM6	Percentage (%) by CD analysis	Percentage (%) by PSIPRED VIEW
α-Helix	0.86	0.00
β-stand	44.49	47.2
Random Coil	54.65	52.8

5.3.6.5 Binding clefts and key residues of CtCBM6

The family 6 carbohydrate binding modules display more than one binding cleft. The cleft A is situated in the region connecting the inner and outer β-sheets of the jelly roll fold, whereas the cleft B is formed by one of the concave surfaces of β-sheet (Pires *et al*., 2004). Multiple sequence alignment and structure superimposition with known structures of CBM6 depicted five important regions termed A to E, in cleft A, which play key roles in ligand selection and affinity (Fig. 5.3.12 A&B). The universal mechanism of ligand selectivity of CBM6 was described earlier that gave appropriate framework to employ bioinformatics tools to determine the location of

potential binding sites, the orientation of conserved binding site residues and important regions responsible for specificity and binding (Abbott *et al*., 2009). The structurally aligned *Ct*CBM6 (in green color) with *Ct*CBM6 (PDB id: 1GMM) in cyan color is shown in figure 5.3.12, B. Region A, B and C are spatially and functionally conserved, and are placed in a one group while region D and E are "hot spot" for primary and tertiary structure variations which lead to determination of the functional specificity of CBM6s (Abbott *et al*., 2009). This A-E_{LOOP} model was adopted in order to understand the role of key residues in determining the substrate specificity and binding cleft topology. In *Ct*CBM6, region A and B are made up of Phe84 and Tyr28, respectively.

These conserved aromatic residues form a flat platform to accommodate sugar ring. In addition, conserved Asn112 (region C) just beneath the region A and B is making the floor of the side platform created by aromatic rings of Phe84 and Tyr28. Region D is associated with selectivity of ligands and also determines the binding mode with carbohydrate molecule i.e. whether it binds an internal region or terminal end of the carbohydrate. Region D generally has an aliphatic isoleucine or an aromatic phenylalanine (Fig. 5.3.12A) but in case of *Ct*CBM6, region D contains an aliphatic valine (Val18 in purple color; Fig. 5.3.12B) which is positioned same as Ile23 of *Ct*CBM6 (1GMM) i.e. in a parallel trajectory to the binding site made by Phe84 and Tyr28. Hence, it facilities the entry of sugar rings to cross over the small aliphatic side chain (as compared to Ile) and interact with internal region of sugar rings, creating an open cleft (Fig. 5.3.12B).

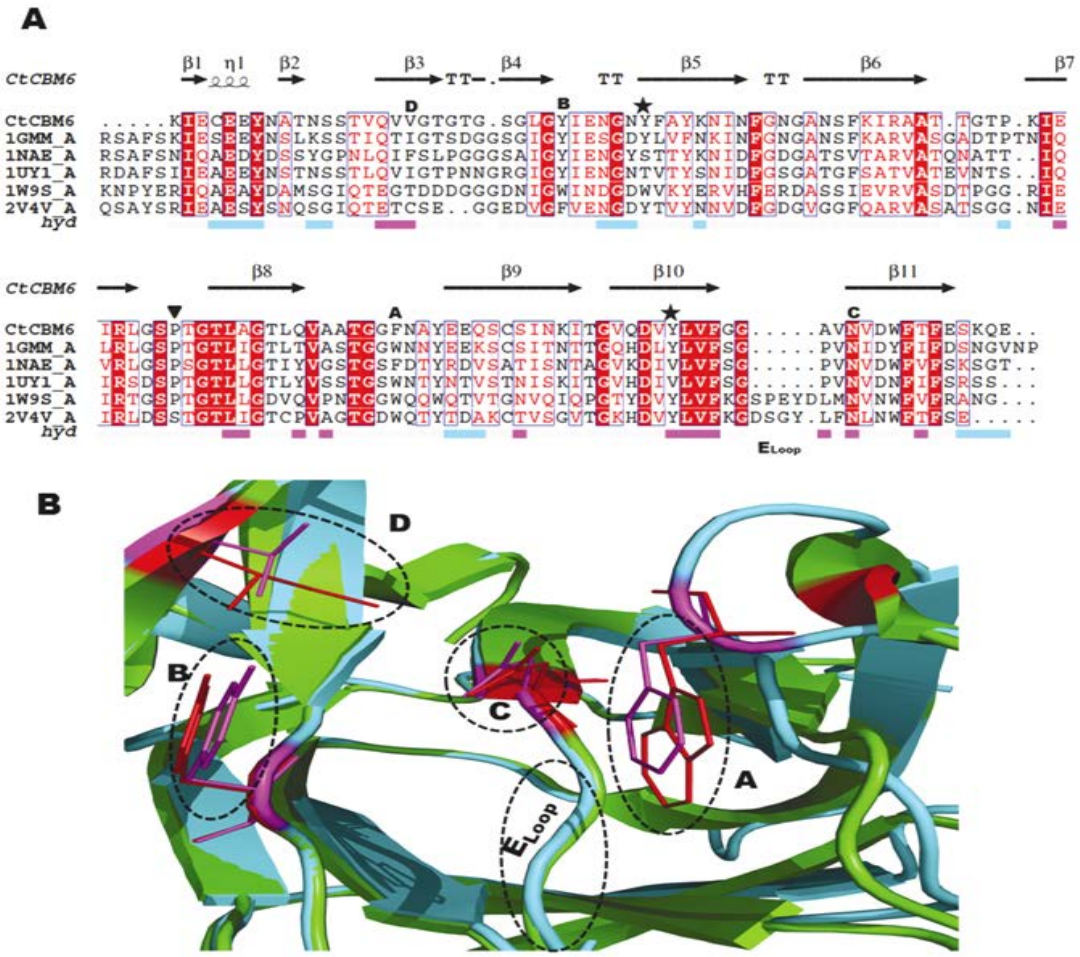


Fig. 5.3.12 **(A)** Multiple sequence alignment of *Ct*CBM6 with representative members of family 6 CBM. *Ct*CBM6 (PDB id: 1GMM from *Clostridium thermocellum*), *Cs*CBM6 (PDB id: 1NAE from *Clostridium stercorarium*), *Cm*CBM6-2 (PDB id: 1UY1 from *Cellvibrio mixtus*), *Bh*CBM6 (PDB id: 1W9S from *B. halodurans*) and *Cc*CBM6 (PDB id: 2V4V from *C. cellulolytium*). AYVW showing conserved and SNS semi-conserved amino acid residues; qqq and ⟶ represent the corresponding secondary structure (alpha helix and beta sheet, respectively) of *Ct*CBM6; ▬ represents the relative hydrophobicity (Pink-hydrophobic, grey-intermediate and cyan-hydrophilic); ▼ Represent conserved proline residues; ★ Represent key residues of cleft B; the sequence shown as A-E_{LOOP} displays the important residues for ligand binding of cleft A. **(B)** Structure superimposition of *Ct*CBM6 in green color with *Ct*CBM6 (PDB id: 1GMM) in cyan color showing the structurally conserved regions. A-E_{LOOP} are indicated by black circles dashed lines while key residues are in stick form (*Ct*CBM6 in purple color and *Ct*CBM6 (PDB id: 1GMM) in red color) responsible for ligand binding and selectivity within cleft A.

While, in some CBM6s like *Sd*CBM6 (PDB id: 2CDP; *Saccharophagus degradans*), *Cm*CBM6 (PDB id: 1UY1; *Cellvibrio mixtus*) and *Bh*CBM6 (PDB id: 1W9S; *B. halodurans*) region D amino acid residue (conserved glutamate) orients itself in such a way that, it blocks the passage of sugar rings by extending out over it forming a closed cleft, therefore, allowing only terminal sugar to interact (Henshaw *et al*., 2006; Henshaw *et al*., 2004b; van Bueren *et al*., 2005). The region E is very crucial for ligand selectivity, but structurally it is poorly defined (Fig. 5.3.12A). It is made up of a loop (E_{Loop}) that connects β10 and β11 strands. In *Ct*CBM6 this loop is formed by Gly108, Gly109, Ala110 and Val111. This short loop facilitates the accommodation of large oligosaccharides via internal sugar ring binding mode.

*Ct*CBM6 exhibited a putative ligand binding site, located at the concave surface of a jelly roll fold, termed cleft B (Fig. 5.3.13), apart from the structurally conserved A, B and C regions which create cleft A. The cleft B of a *Ct*CBM6 binding site resembles CBMs of families 4, 15, 17 and 22 (Boraston *et al*., 2002; Notenboom *et al*., 2001; Szabo *et al*., 2001). Cleft B is formed by two aromatic residues present close to each other in the cavity which is formed by the concave face. Here, in *Ct*CBM6 the aromatic residues Tyr34 and Tyr104 are forming the same groove to accommodate sugar ring, as in cleft A (Fig. 5.3.13). However, the accessibility of these surface aromatic residues might be occluded by a conserved proline (Pro68) residue situated in the loop connecting β-7 to β-8 strand. The capacity of clefts A and B to recognize polysaccharides will be explored below.

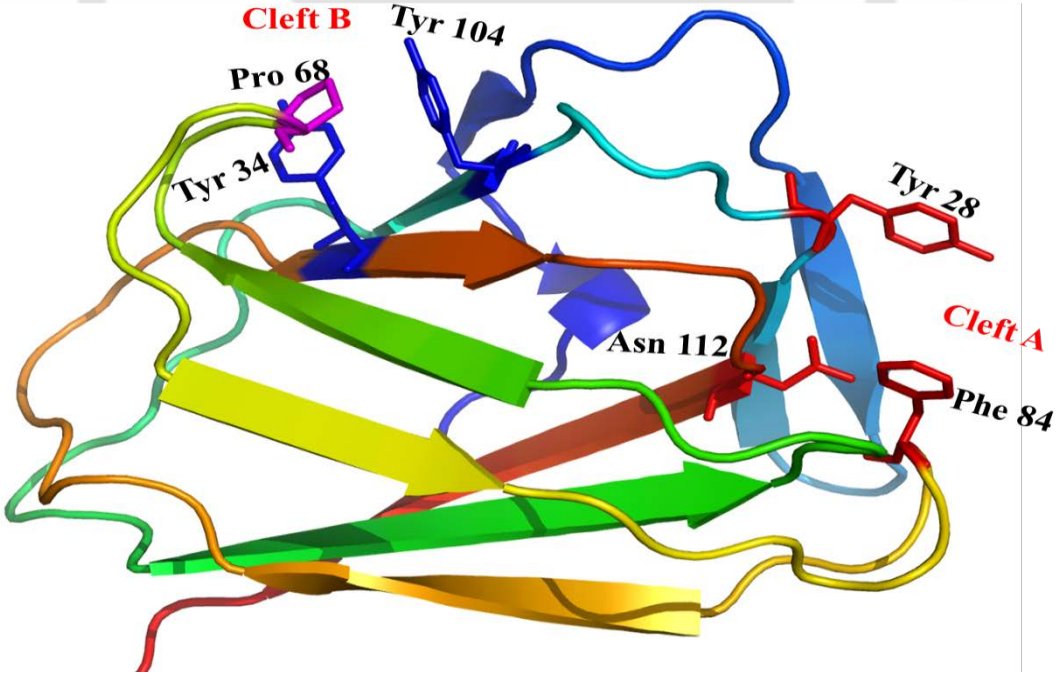


Fig. 5.3.13 Ribbon representation of *Ct*CBM6 depicted the location of two different binding cleft in *Ct*CBM6. Residues Phe84, Tyr28 and Arg112 architect cleft A while Tyr34 and Tyr104 making the same plane hydrophobic surface at cleft B which is occluded by Pro68.

5.3.6.6 Docking analysis of CtCBM6 at binding cleft A

The ligand docking analysis of *Ct*CBM6 at binding cleft A illustrated the amino acid residues that are involved in the interaction (Table 5.3.4). The predicted free binding energy (ΔG) with different ligands mentioned in the Table 5.3.4 clearly suggested that as the size of ligand molecule increases the free binding energy (ΔG) of interaction decreases and it is least for xylopentaose and for xylohexaose it was a positive value (not mentioned in Table 5.3.4). The ligand (xylobiose) bound complex structure of *Ct*CBM6 depicted the orientation of the terminal sugar ring which is parallel to aromatic amino acid residues Tyr28 and Phe84 present at cleft A (Fig. 5.3.14, A). Amino acid residues Asn85, Asn112 and Tyr28 are involved in forming the hydrogen bond while residues Phe84, Gly27 and Val18 making a hydrophobic interaction with ligand molecules (Fig. 5.3.14, B).

Table 5.3.4 Cleft A binding site residues displaying interaction with ligands.

Ligand	Predicted free binding energy (ΔG) (kcal/mol)	Binding site residues present in 4Å region of cleft A
Xylobiose	-3.97	VAL18, GLY27, TYR28, PHE84, ASN85, ASN112.
Xylotriose	-3.27	VAL18, THR20, TYR28, ILE29, PHE84, ASN85, TYR87, ASN112, VAL111, ALA110.
Xylotetraose	-2.5	VAL18, THR20, TYR28, ILE29, PHE84, ASN85, ALA86, TYR87, ALA110, VAL111, ASN112.
Xylopentaose	-1.95	VAL18, THR20, TYR28, ARG52, PHE84, ASN85, ALA86, TYR87, GLU88, ASN112.
Glucuronoxylan ($MeGlcA^2Xyl_3$)	-2.21	VAL18, GLU19, THR20, TYR28, PHE84, ASN85, TYR87. ASN112.

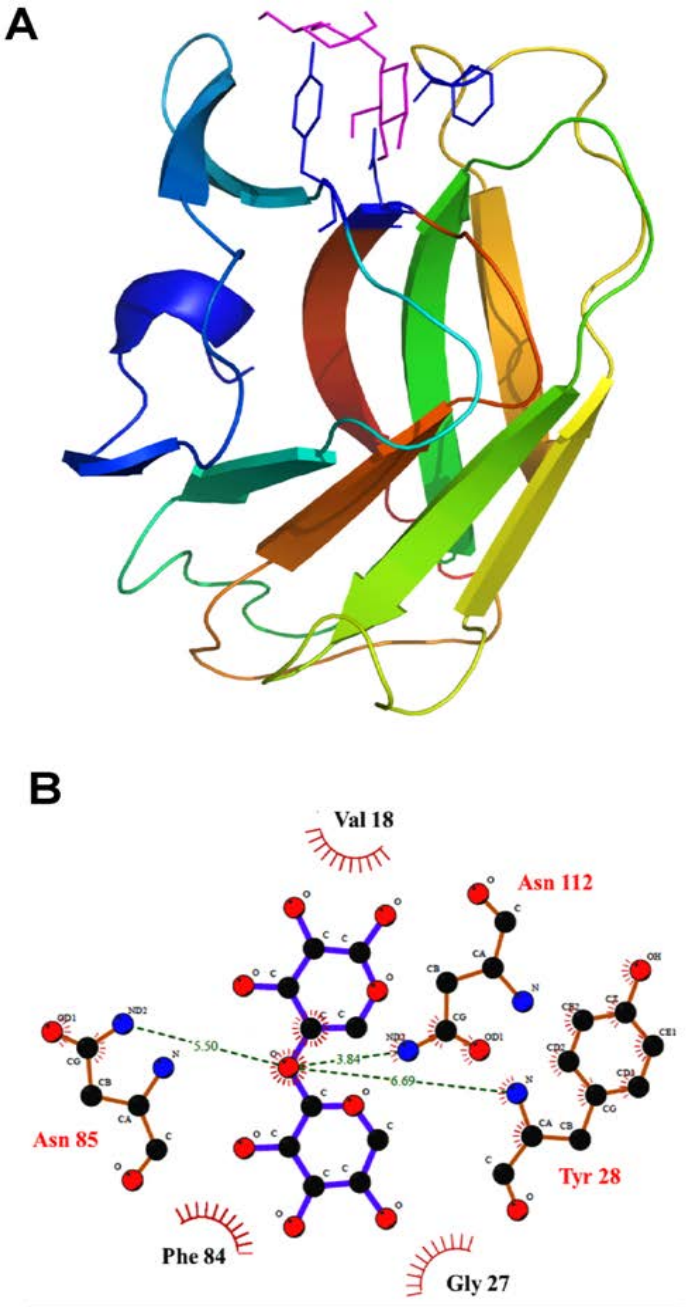


Fig. 5.3.14 **(A)** Ligand bond docked structure of *Ct*CBM6 with xylobiose (purple) showing orientation of aromatic residues making a parallel platform to accommodate flat pyranose ring of ligand molecule at cleft A **(B)** Schematic 2D depiction of ligand xylobiose interacting with amino acid residues at binding cleft A. (▬▬▬) Dashed lines shows hydrogen bonds with labelled bond length, residues displaying arc with spokes making hydrophobic interaction, contacted atoms are shown with spokes radiating back.

Surface overview of best docked pose of xylotetraose showed solvent-accessible surface and depth of the binding groove at cleft A (Fig. 5.3.15A) while residues Val18, Tyr28, Ile29, Phe84, Asn85, Ala86, Tyr87, Ala110, Val111 and Asn112 were found making hydrogen bond or hydrophobic interactions with ligand molecules (Fig. 5.3.15B)

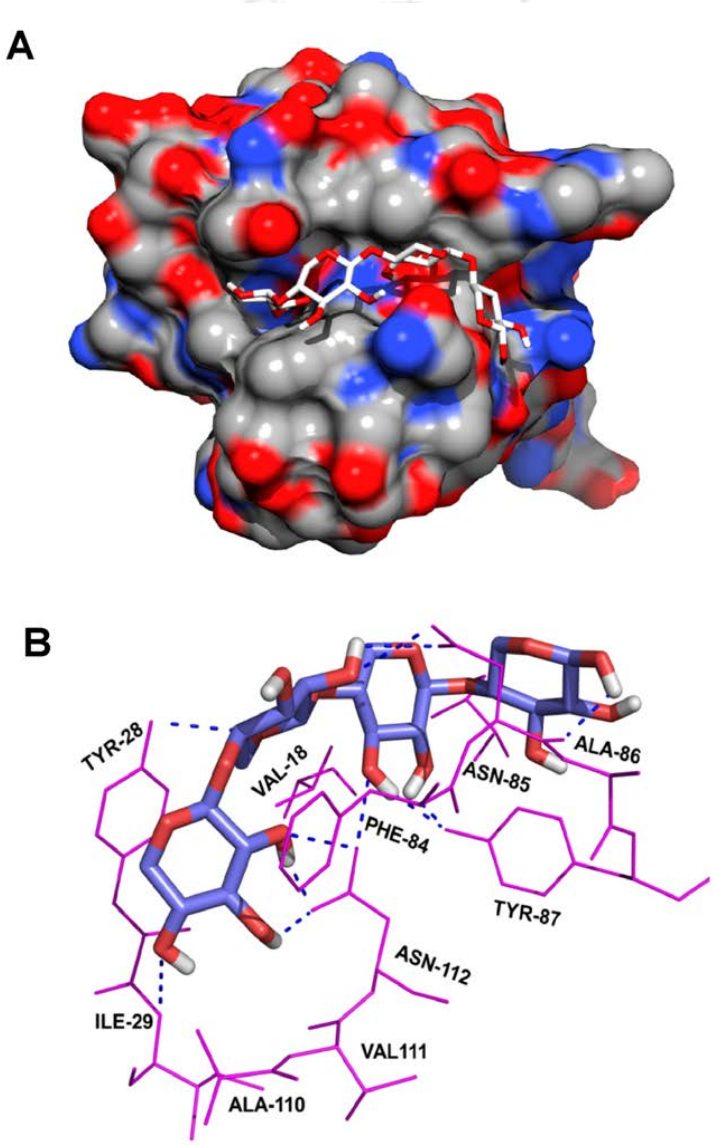


Fig. 5.3.15 **(A)** Surface view of the binding cleft A of *Ct*CBM6 showing solvent-accessible surface and depth of the binding groove accommodating ligand xylotatraose, **(B)** Binding mode interactions of *Ct*CBM6 with ligand xylotatraose best hit obtained by AutoDock.

5.3.7 Assessment of involvement of binding residues of *Ct*CBM6 by site-directed mutagenesis

The structure of *Ct*CBM6 suggests that the protein contain two putative clefts, termed cleft A and B. Structure analysis suggest that only cleft A is functional as cleft is not available for carbohydrate recognition due to steric hindrance. When *Ct*CBM6 was superposed with other CBM6 structures (Fig. 5.3.12, B) as well as docking studies with xylo-oligosacharides (Table 5.3.4, Fig 5.3.14, A&B) suggest that in cleft A Tyr28 and Phe84 might constitute the two key residues involved in ligand recognition. Thus, *Ct*CBM6 mutant derivatives Y28A, F84A and the double mutant Y28A/F84A were produced. ITC analyses of the mutants could provide only qualitative information on their binding to ligands because the affinities were too low to obtain proper titration curves (Fig. 5.3.16, A-F). The results revealed that F84A and Y28A/F84A completely lost their binding affinity towards both glucuronic acid as well as arabinose substituted xylans. However, the binding affinity of Y28A was significantly reduced but not entirely as could be observed by ITC analysis and affinity electrophoresis using 4-O-methyl glucuronoxylan and wheat arabinoxylan (Fig. 5.3.16 A&B) and (Fig. 5.3.17B&C) as also summarized in the Table 5.3.5. The results confirmed that Tyr 28 and Phe84 are the key amino acid residues playing a critical role in the ligand binding. Therefore, cleft A is the carbohydrate binding platform of *Ct*CBM6.

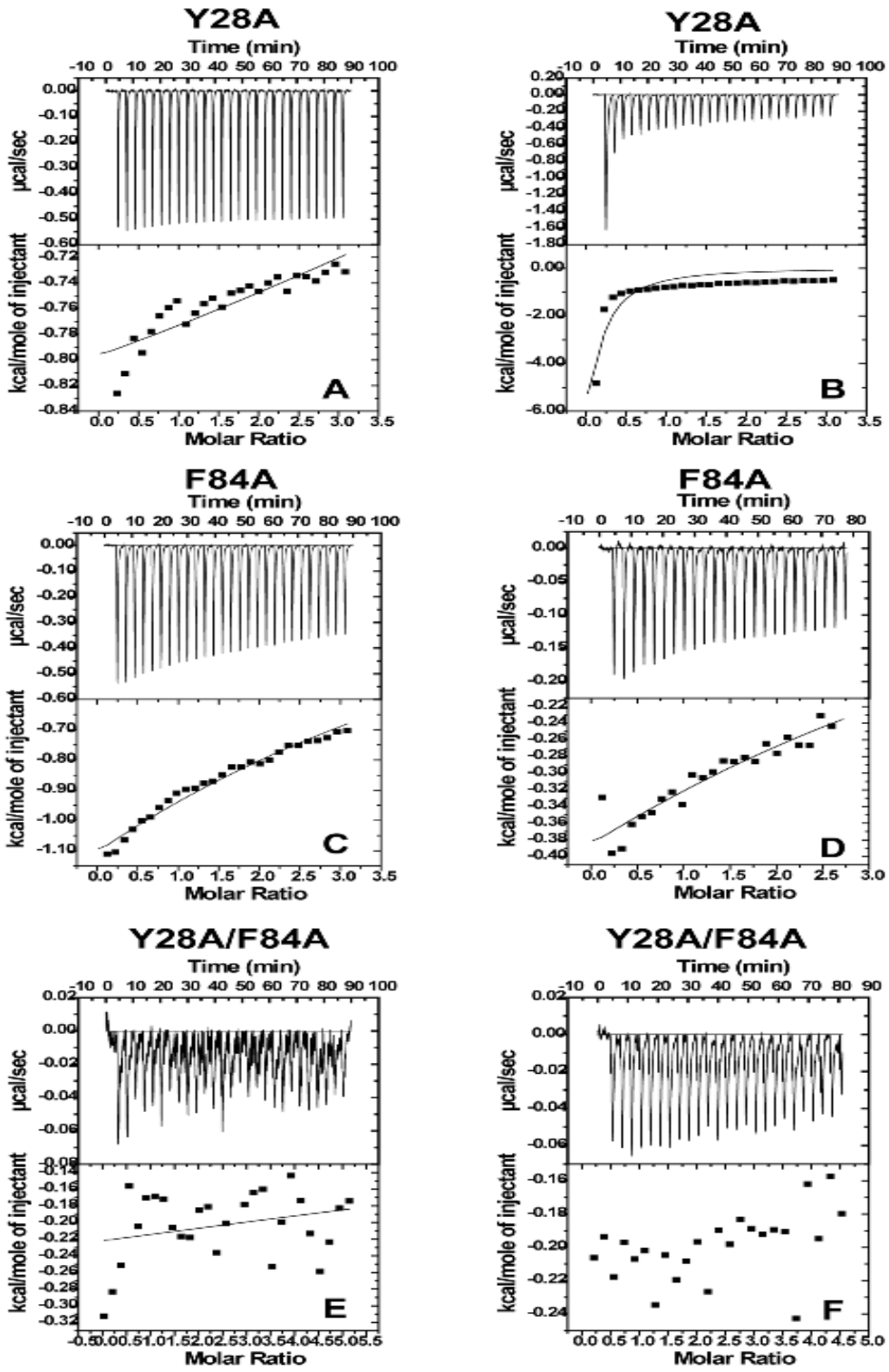


Fig. 5.3.16 Isothermal titration calorimetry analysis of mutants of *Ct*CBM6 binding with xylan-based polysaccharides at 55°C. The ITC profiles of *Ct*CBM6 mutants Y28A (**A**), F84A (**C**) and Y28A/F84A (**E**) with ligand 4-O-methyl glucuronoxylan; Y28A (**B**), F84A (**D**) and Y28A/F84A **(F**) with Wheat arabinoxylan.

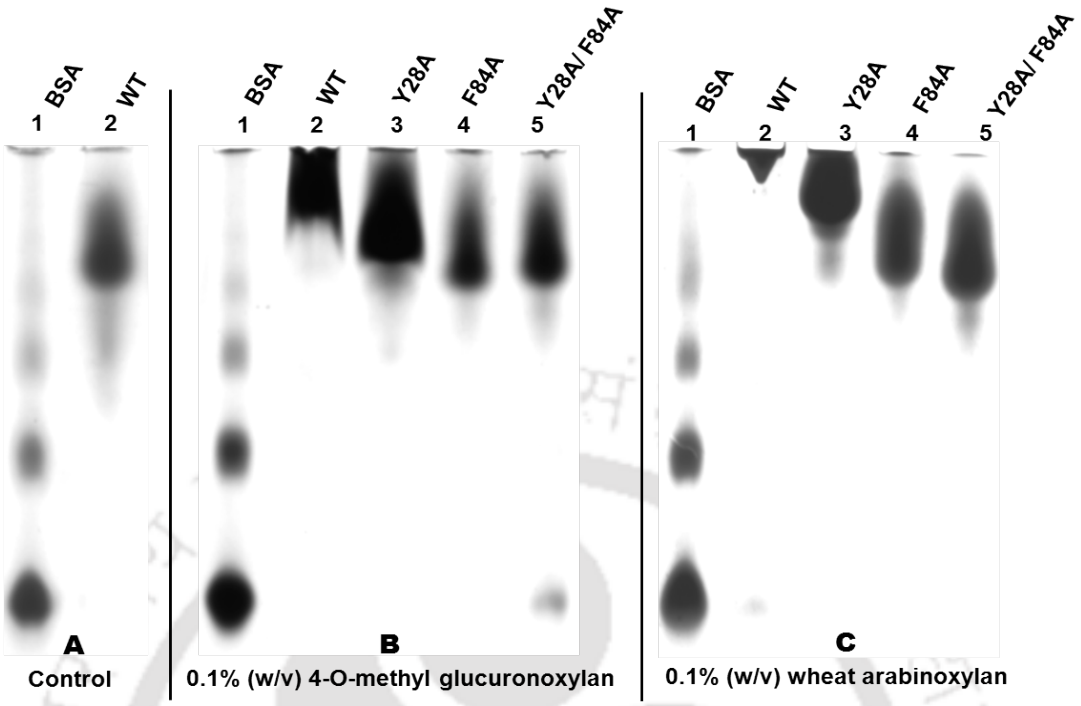


Fig. 5.3.17 Qualitative binding analysis of *Ct*CBM6 by affinity electrophoresis using native-PAGE (6.5%, w/v) gels run at 4°C with and without substrate, **(A)** Native-PAGE gel showing the migration of Bovine serum albumin (BSA), used as a reference protein (lane 1), wild-type (WT) *Ct*CBM6 (lane 2), **(B)** Native-PAGE gel containing 4-O-methyl glucuronoxylan (0.1%, w/v) showing the retardation and hence binding of WT *Ct*CBM6 (lane 2), reduced binding of mutant Y28A (lane3), no binding of mutants F84A (lane 4) and Y28A/F84A of *Ct*CBM6 (lane 5), **(C)** Native-PAGE gel containing wheat arabinoxylan (0.1%, w/v) showing binding of WT *Ct*CBM6 (lane 2), reduced binding of mutant Y28A (lane3) and no binding of F84A (lane 4) and Y28A/ F84A (lane 5).

Table 5.3.5 Qualitative binding analysis of *Ct*CBM6 mutants by ITC and affinity electrophoresis (AE).

Protein	Ligand	ITC	AE
Y28A	Wheat arabinoxylan	++[a]	++[a]
Y28A	4-O-methyl glucuronoxylan	+[b]	+[b]
F84A	Wheat arabinoxylan	-[c]	-[c]
F84A	4-O-methyl glucuronoxylan	-	-
Y28A/F84A	Wheat arabinoxylan	-	-
Y28A/F84A	4-O-methyl glucuronoxylan	-	-

[a] (++) indicates weak binding. [b] (+) indicates very weak binding. [C] (-) indicates no binding. Although the affinity was too low to accurately convolute by isotherm but the binding was clearly detectable by AE.

Conclusions

*Ct*CBM6 demonstrated affinity towards α-L-arabinosyl and glucuronic acid substituted xylans. The equilibrium association constant (K_a) determined by affinity electrophoresis and isothermal titration calorimetry suggests that *Ct*CBM6 displays similar affinities against highly substituted xylans (wheat- and rye- arabinoxylans) and poorly substituted xylans (birch-, beech- and oat spelt- xylan). *Ct*CBM6 is unable to interact with other plant cell wall carbohydrates. Data presented here revealed that *Ct*CBM6 display β-jelly roll β-sandwich fold with two potential binding sites; cleft A and B. Cleft A, located on the loops connecting the two β-sheets is decorated by the side chains of Phe-84, Tyr-28 and Asn-112 whereas cleft B is located in the concave surface of the protein where residues Tyr-34 and Tyr-104 might contribute to ligand recognition. The functional importance of the two putative clefts for carbohydrate recognition was investigated by mutagenesis analysis. The data revealed that two cleft A protein mutant derivatives, the single mutant F84A and the double mutant Y28A/F84A, completely lost the ability to recognize xylans, whereas, the single mutant Y28A was able to recognize glucuronic acid and arabinose decorated xylan polysaccharides but with significantly reduced affinity. These results confirm that cleft A, through the action of Tyr28 and Phe84, constitutes the functional carbohydrate binding platform of *Ct*CBM6. The protein melting of *Ct*CBM6 at 83°C conveyed the thermostable nature of the protein. The presence of Ca^{2+} ions increased only marginally the structure stability of *Ct*CBM6 by shifting the melting peak by 2°C to 85°C. Overall the data confirms that within the CBM6 family a subset of proteins have evolved the capacity to recognize xylans through cleft A, while in this cases cleft B remain non-functional.

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, 10395-10404.

Abou-Hachem, M., Karlsson, E.N., Simpson, P.J., Linse, S., Sellers, P., Williamson, M.P., Jamieson, S.J., Gilbert, H.J., Bolam, D.N., Holst, O. (2002) Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from *Rhodothermus marinus*. Biochemistry, 41, 5720-5729.

Ahmed, S., Luis, A.S., Bras, J.L., Fontes, C.M., Goyal, A. (2013a) Functional and structural characterization of family 6 carbohydrate-binding module (*Ct*CBM6A) of *Clostridium thermocellum* alpha-L-arabinofuranosidase. Biochemistry, (Moscow) 78, 1272-1279.

Ahmed, S., Luis, A.S., Bras, J.L., Fontes, C.M., Goyal, A. (2013b) The family 6 carbohydrate-binding module (*Ct*CBM6B) of *Clostridium thermocellum* alpha-L-arabinofuranosidasebinds xylans and thermally stabilized by Ca^{2+} ions Biocatalysis and Biotransformation, 31, 217-225

Bae, H.J., Turcotte, G., Chamberland, H., Karita, S., Vezina, L.P. (2003) A comparative study between an endoglucanase IV and its fused protein complex Cel5-CBM6. FEMS Microbiology Letters, 227, 175-181.

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

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

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

Boraston, A.B., Tomme, P., Amandoron, E.A., Kilburn, D.G., (2000) A novel mechanism of xylan binding by a lectin-like module from Streptomyces lividans xylanase 10A. The Biochemical Journal, 350, 933-941.

Brett, C., Waldron, K. (1996) Physiology and Biochemistry of Plant Cell Walls. Springer Netherlands.

Carvalho, A.L., Dias, F.M., Nagy, T., Prates, J.A., Proctor, M.R., Smith, N., Bayer, E.A., Davies, G.J., Ferreira, L.M., Romao, M.J., Fontes, C.M., Gilbert, H.J. (2007) Evidence for a dual binding mode of dockerin modules to cohesins. Proceedings of the National Academy of Sciences, USA, 104, 3089-3094.

Carvalho, A.L., Goyal, A., Prates, J.A., Bolam, D.N., Gilbert, H.J., Pires, V.M., Ferreira, L.M., Planas, A., Romao, M.J., Fontes, C.M. (2004) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. Jornal of Biological Chemistry, 279, 34785-34793.

Correia, M.A., Pires, V.M., Gilbert, H.J., Bolam, D.N., Fernandes, V.O., Alves, V.D., Prates, J.A., Ferreira, L.M., Fontes, C.M. (2009) Family 6 carbohydrate-binding modules display multiple beta1,3-linked glucan-specific binding interfaces. FEMS Microbiology Letters, 300, 48-57.

Dam, T.K., Brewer, C.F. (2002) Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. Chemical Reviews, 102, 387-429.

DeLano W.L. (2002) The PyMOL Molecular Graphics System. San Carlos, C.D.S.

Dvortsov, I.A., Lunina, N.A., Chekanovskaya, L.A., Schwarz, W.H., Zverlov, V.V., Velikodvorskaya, G.A. (2009) Carbohydrate-binding properties of a separately folding protein module from beta-1,3-glucanase Lic16A of *Clostridium thermocellum*. Microbiology, 155, 2442-2449.

Eisenberg, D., Luthy, R., Bowie, J.U. (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods in Enzymology, 277, 396-404.

Fernandes, A.C., Fontes, C.M., Gilbert, H.J., Hazlewood, G.P., Fernandes, T.H., Ferreira, L.M. (1999) Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. The Biochemical Journal, 342, 105-110.

Fiser, A., Do, R.K., Sali, A. (2000) Modeling of loops in protein structures. Protein science : a publication of the Protein Society, 9, 1753-1773.

Gilkes, N.R., Jervis, E., Henrissat, B., Tekant, B., Miller, R.C., Jr., Warren, R.A., Kilburn, D.G. (1992) The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. Journal of Biological Chemistry, 267, 6743-6749.

Goodsell, D.S., Olson, A.J., (1990) Automated docking of substrates to proteins by simulated annealing. Proteins, 8, 195-202.

Guillen, D., Sanchez, S., Rodriguez-Sanoja, R. (2010) Carbohydrate-binding domains: multiplicity of biological roles. Applied Microbiology and Biotechnology, 85, 1241-1249.

Hashimoto, H. (2006) Recent structural studies of carbohydrate-binding modules. Cellular and Molecular Life Sciences, 63, 2954-2967.

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

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

Henshaw, J.L., Bolam, D.N., Pires, V.M.R., Czjzek, M., Henrissat, B., Ferreira, L.M.A., Fontes, C.M.G.A., Gilbert, H.J. (2004) The family 6 carbohydrate binding module *Cm*CBM6-2 contains two ligand-binding sites with distinct specificities. Journal of Biological Chemistry, 279, 21552-21559.

Holm, L., Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Research, 38, 545-549.

Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology, 292, 195-202.

Kelly, S.M., Jess, T.J., Price, N.C. (2005) How to study proteins by circular dichroism. Biochima et Biophysica Acta 1751, 119-139.

Lammerts van Bueren, A., Boraston, A.B. (2004) Binding sub-site dissection of a carbohydrate-binding module reveals the contribution of entropy to oligosaccharide recognition at "non-primary" binding subsites. Journal of Molecular Biology, 340, 869-879.

Louis-Jeune, C., Andrade-Navarro, M.A., Perez-Iratxeta, C. (2011) Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. Proteins, 80, 374-381.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry, 19, 1639-1662.

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

Pires, V.M., Henshaw, J.L., Prates, J.A., Bolam, D.N., Ferreira, L.M., Fontes, C.M., Henrissat, B., Planas, A., Gilbert, H.J., Czjzek, M. (2004) The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5a in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. The Journal of Biological Chemistry, 279, 21560-21568.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. Journal of Molecular Biology, 7, 95-99.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., Karplus, M. (1995) Evaluation of comparative protein modeling by MODELLER. Proteins, 23, 318-326.

Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. Proteins, 17, 355-362.

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

Takeo, K. (1995) Advances in affinity electrophoresis. Journal of Chromatography A, 698, 89-105.

Tomme, P., Boraston, A., Kormos, J.M., Warren, R.A., Kilburn, D.G. (2000) Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules. Enzyme and Microbial Technology, 27, 453-458.

van Bueren, A.L., Morland, C., Gilbert, H.J., Boraston, A.B. (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of beta-1,3-linked glucans by presenting a unique ligand binding surface. The Journal of Biological Chemistry, 280, 530-537.

Wiederstein, M., Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35, 407-410.

Future prospects

The family 30 glycoside hydrolase, *Ct*XynGH30 a glucuronoxylan-xylanohydrolase from *Clostridium thermocellum* displayed activity towards xylan main chain polysaccharides decorated with glucuronic acid (or 4-O-methyl glucuronic acid) and arabinose side chain. Therefore *Ct*XynGH30 can be used in various industrial processes involving degradation of xylan based plant polysaccharides *viz*. pulp and paper industries for reducing the use chlorine, clarification of fruit juices and wine, animal food for improving digestibility. The thermostability of *Ct*XynGH30 makes it a good candidate for use in high temperature industrial operations.

*Ct*XynGH30 an endo-xylanase (glucuronoxylan-xylanohydrolase) mainly generated acid xylo-oligaosachharides, which are widely used as food additives, nutraceuticals and therapeutic purposes. These glucuronic acid-containing (acidic) xylo-oligaosachharides are known to be potential anti-cancerous, immunostimulating and anti-inflammatory agents.

The non-catalytic module described in this study, *Ct*CBM6 showed high affinity towards xylan main chain polysaccharides decorated with glucuronic acid (or 4-O-methyl glucuronic acid) or with arabinose side chains. The thermostable nature of *Ct*CBM6 makes it a potential candidate for the purification of biomolecules by adapting immobilized affinity ligand technology. The high thermo-stability and specificity of *Ct*CBM6 towards the acidic xylo-oligomers (glucuronic acid or arabinose substituted xylo-oligaosachharides) will allows it to separate from the rest by retaining them at higher temperatures.

Published/accepted (* corresponding author, † equal contribution)

1. **Anil K. Verma**, Pedro Bule, Teresa Ribeiro, Joana L. A. Brás, Joyeeta Mukherjee, Munishwar N. Gupta, Carlos M.G.A. Fontes and Arun Goyal* (**2015**) The family 6 Carbohydrate Binding Module (*Ct*CBM6) of glucuronoxylanase (*Ct*XynGH30) of *Clostridium thermocellum* binds decorated and undecorated xylans through Cleft A. Archives of Biochemistry and Biophysics, Elesvier 575, 8-21.
2. **Anil K.Verma** and Arun Goyal* (**2014**) *In silico* structural characterization and molecular docking studies of first glucuronoxylan-xylanohydrolase (Xyn30A) from family 30 glycosyl hydrolase (GH30) from *Clostridium thermocellum*, Molecular Biology 48, 278-286.
3. **Anil K.Verma**, Arun Goyal*, Freire F., Bule P., Venditto I., Joana L. A. Bras, Santos H., Cardoso V., Bonifacio C., Thompson A., Romao M. J., Prates J. A. M., Ferreira L. M. A., Carlos M.G.A. Fontes, and Shabir Najmudin (**2013**) Overexpression, crystallization and preliminary X-ray crystallographic analysis of glucuronoxylan-xylanohydrolase (Xyn30A) from *Clostridium thermocellum*, Acta Crystallografica F69 1440-1442.
4. Arabinda Ghosh†, **Anil K. Verma**†, Ana Sofia Luis, Joana L. A. Bras, Carlos M. G. A. Fontes and Arun Goyal* (**2014**) Mannan specific family 35 carbohydrate binding module (*Ct*CBM35) of *Clostridium thermocellum*: structure analysis and ligand binding. Biologia (Section Cellular and Molecular Biology) 69(10), 1271-1282.
5. Arabinda Ghosh†, **Anil K. Verma**†, Saurabh Gautam, Munishwar N. Gupta and Arun Goyal* (**2014**) Structure and functional investigation of ligand binding by a family 35 carbohydrate binding module (*Ct*CBM35) of β-mannanase of family 26 Glycoside Hydrolase from *Clostridium thermocellum*. Biochemistry (Moscow) 79, 672-86.
6. Arabinda Ghosh†, **Anil K. Verma**†, Jagan Mohan Rao T., Rishikesh Shukla and Arun Goyal* (**2014**) Recovery and purification of oligosaccharides from coprameal by recombinant endo-β-mannanase and deciphering molecular mechanism involved and its role as potent therapeutic agent. Molecular Biotechnology 57, 111-127.

To be submitted

1. Filipe Freire†, **Anil K. Verma**†, Arun Goyal*, Carlos M.G.A. Fontes and Shabir Najmudin (**2015**) Conservation in the catalytic machinery of glucuronoxylan endo-β-1,4-xylanases revealed by the structure of the thermostable *Ct*Xyn30A from *Clostridium thermocellum*. Journal of Structural Biology. (Submitted after revision)

Manuscripts under preparation

1. **Anil K. Verma** and Arun Goyal* (**2015**) Cloning, expression and biochemical characterization of a modular glucuronoxylan–xylanohydrolase (*Ct*XynGH30) from *Clostridium thermocellum*.
2. Shraddha Shukla, **Anil K Verma**, Rviwoo Barua, Kati Katina, Rikka Juvonen, MaijaTenkanen and Arun Goyal* (**2015**) *Weissella confusa* Cab3 from sauerkraut: Molecular characterization, expression and in silico analysis of dextransucrase in *Lactococcus lactis*.
3. Kedar Sharma, **Anil K. Verma** and Arun Goyal* **(2015)** Molecular cloning, expression and characterization of novel endo-β-1, 4-mannanase of a family 10 glycoside hydrolase from *Pedobacter saltans* DSM12145.

Conferences/Symposia/Meetings

International

1. **Anil K. Verma**, Pedro Bule, Teresa Ribeiro, Joana L. A. Brás, Carlos M.G.A. Fontes and Arun Goyal (**2015**) Insight into structural, biochemical and *in silico* determinants of ligand binding specificity of family 6 carbohydrate binding module (*Ct*CBM6) from *Clostridium thermocellum.* 11th Carbohydrate Bioengineering Meeting, May 10-13, 2015 Espoo, Finland.
2. **Anil K.Verma**, Arun Goyal, Filipe Freire, Carlos M.G.A. Fontes and Shabir Najmudin (**2015**) Insights into the mechanism of glucuronoxylan hydrolysis revealed by the 3-dimensional crystal structures of glucuronoxylan-xylanohydrolase (*Ct*Xyn30A) from *Clostridium thermocellum.* 11th Carbohydrate Bioengineering Meeting, May 10-13, 2015 Espoo, Finland.
3. Kedar Sharma, **Anil K. Verma** and Arun Goyal (**2015**) Molecular cloning, expression and characterization of novel endo-β-1, 4-mannanase of a family 10 glycoside hydrolase from *Pedobacter saltans* DSM12145 (2015). 11th Carbohydrate Bioengineering Meeting, May 10-13, 2015 Espoo, Finland.
4. **Anil K. Verma,** Filipe Freire, Arun Goyal, Carlos M.G.A. Fontes and Shabir Najmudin (**2014**) Structural and biochemical characterization of glucuronoxylan-xylanohydrolase (Xyn30A) from *Clostridium thermocellum*. 27th International Carbohydrate Symposium, Jan 12-17, 2014, Indian Institute of Science, Bangalore, India.
5. Arabinda Ghosh, **Anil K. Verma** and Arun Goyal (**2014**) Novel thermostable recombinant endo-β-mannanase of *Clostridium thermocellum* for manno-oilgosaccharides production. National Seminar on Microbiology (Microbial World 2014). March 14, 2014, University of North Bengal, Siliguri, India.
6. **Anil K. Verma,** Arun Goyal, Filipe Freire, Joyeeta Mukherjee, Munishwar N. Gupta, Carlos M.G.A. Fontes and Shabir Najmudin (**2014**) Structure and functional analyses of recombinant glucuronoxylan-xylanohydrolase (*Ct*XynGH30), its truncated derivative Xyn30A and associated family 6 carbohydrate binding module *Ct*CBM6 from *Clostridium thermocellum*. International Conference on Emerging Trends in Biotechnology (ICETB-2014)

and XI Biotech Research Society India (BRSI) Convention, November 6-9, 2014, Jawaharlal Nehru University, New Delhi, India.

7. **Anil K. Verma,** Carlos M.G.A. Fontes and Arun Goyal (**2013**) Cloning, expression and binding analysis of carbohydrate binding module family 6 (*Ct*CBM6) from *Clostridium thermocellum*. 10th Carbohydrate Bioengineering meeting, April 21-24 2013, Institute of microbiology, Academy of Science of the Czech Republic, prague Czech Republic.
8. Arabinda Ghosh**, Anil K. Verma,** Neeta Pathaw, Nikhil K. Chrungoo, Saurabh Gautam, Munishwar N. Gupta and Arun Goyal (**2013**) Conformational change upon ligand binding of a manno-configured substrate specific family 35 Carbohydrate Binding Module (CBM35) from *Clostridium thermocellum*. 10th Carbohydrate Bioengineering Meeting, April 21-24, 2013, Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic.
9. Rajan Choudhary, S. P. Das, **Anil K. Verma,** D. Das and Arun Goyal (**2013**) Efficient bioethanol production from lignocellulosic leafy biomass of poplar (Populas nigra). International Conference on Conserving Biodiversity for Sustainable Development, Aug 16-18, 2013, National Institute of Technology, Rourkela, Odisha, India.
10. Arabinda Ghosh**, Anil K. Verma,** Ana Sofia Luis, Joana L. A. Bras, Carlos M. G. A. Fontes and Arun Goyal (**2013**) 3-Dimensional structure and ligand binding of family 35 carbohydrate binding module (*Ct*CBM35) of *Clostridium thermocellum* by in silico and affinity electrophoresis studies. 10th Convention of Biotech Research Society and International Conference on Advances in Biotechnology and Bioinformatics, Nov 25-27, 2013, D.Y. Patil Institute of Biotechnology and Bioinformatics, Pune, India.
11. Arabinda Ghosh, **Anil K. Verma,** Carlos M.G.A. Fontes and Arun Goyal (**2012**) Manno-configured substrate specific family 35 carbohydrate binding module (*Ct*CBM35) from *Clostridium thermocellum*. International conference of industrial biotechnology and IX convention of the biotech research society, India (ICIB 2012). Nov 21-23, 2012, Punjabi University, Patiala, Punjab.
12. **Anil K. Verma** and Arun Goyal (**2012**) Structural characterization and docking analysis of modeled *Ct*XynGH30 protein family 30 glucoronoxylan-xylanohydrolase of *Clostridium thermocellum*. International Symposium on "Bioengineering 2012" (ISBE 2012). December 10, Biotech Hub, Centre for the Environment, Indian Institute of Technology Guwahati, India.
13. **Anil K. Verma**, Arabinda Ghosh and Arun Goyal (**2011**) *In silico* structure and substrate binding analyses of family 35 Carbohydrate binding Module (CBM35) from cellulosome of *Clostridium thermocellum*. World Congress in Biotechnology, OMICS Group Conferences, March 21-23, 2011 Hyderabad, India.
14. Arabinda Ghosh, Shadab Ahmed, **Anil K. Verma,** Carlos M.G.A. Fontes and Arun Goyal (**2011**) Cloning, expression and Biochemical characterization of family 26 glycoside hydrolase (GH26-CBM35) and Carbohydrate binding

Module (CBM35) from *Clostridium thermocellum*. 9th Carbohydrate Bioengineering Meeting, May 15-18, 2011, Technical University of Lisbon, Portugal.

15. **Anil K. Verma** and Arun Goyal (**2011**) Investigating the 3-dimensional structure of family 5 glycoside hydrolase (*Ct*GH5), from cellulosome of *Clostridium thermocellum*. International Conference on New Horizons in Biotechnology (8th Annual Convention of the Biotech Research Society, India), November 21-24, 2011 National Institute for Interdisciplinary Science and Technology (NIIST), CSIR, Trivandrum.

National

1. **Anil K. Verma** and Arun Goyal (**2014**) Biotech Hub Symposium-2014, Dec 2, 2014, Biotech Hub, Centre for the Environment, Indian Institute of Technology Guwahati.
2. **Anil K. Verma,** Arun Goyal, Filipe Freire, Carlos M.G.A. Fontes and S. Najmudin (**2013**) Cloning, hyper-expression, crystallization and X-ray crystallographic structure analysis of glucuronoxylan-xylanohydrolase (Xyn30A) from *Clostridium thermocellum*. 42nd National Seminar on Crystallography and International Workshop on Application of X-Ray Diffraction for Drug Discovery, Nov 21-23, 2013, Jawaharlal Nehru University, New Delhi, India.
3. **Anil K. Verma and** Arun Goyal (**2011**) Cloning of a family 5 glycoside hydrolase (GH5-CBM6) and carbohydrate binding module (CBM6) from *Clostridium thermocellum*. Carbo XXVI, Symposium, 2011 on Carbohydrates At The Interface of Chemistry and Biology, November 23-25, 2011 Indian Institute of Chemical Biology (IICB), Kolkata.

Oral Presentation

1. **Anil K. Verma** and Arun Goyal (**2013**) Structure characterization and molecular docking analysis of modeled family 6 carbohydrate binding module (*Ct*CBM6) of *Clostridium thermocellum*. 42nd National Seminar on Crystallography and International Workshop on Application of X-Ray Diffraction for Drug Discovery. Nov 21-23, 2013, Jawaharlal Nehru University, New Delhi, India.

VITAE

The author was born on September 10, 1985 in Indore, Madhya Pradesh, India. He passed the Secondary School Examination conducted by Central Board of Secondary Education, New Delhi, in 2000 and later passed Higher Secondary School Examination conducted by Central Board of Secondary Education, New Delhi in 2002. He completed B.Sc. Biotechnology from Holkar Science College affiliated to Devi Ahilya Vishwavidyalaya, Madhya Pradesh, India, in 2007. He completed M. Sc. Industrial Microbiology from School of Life Sciences at Devi Ahilya Vishwavidyalaya, Madhya Pradesh, India, 2009.

Mr. Anil Kumar Verma joined Ph.D. programme in July 2009 at Department of Biotechnology (then), Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India. He successfully completed the course work with 7.6/10 Cumulative Point Index (CPI). He received Institute Fellowship (IIT Guwahati) from July 2009 to July 2014, under the scheme run by the Ministry of Human Resource and Development (MHRD), New Delhi. From September 2014 onwards he received fellowship as Senior Research Fellowship from a Indo-Finland DBT sponsored project at Department of Biosciences & Bioengineering under Prof. Arun Goyal. He delivered the Open (PhD Synopsis) Seminar on January 27, 2015 and presented his thesis work before the Doctoral Committee and his performance was satisfactory. He submitted the PhD thesis in March, 2015. Presently, he is continuing as Senior Research Fellow under the same project at the same Department.

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Overexpression, crystallization and preliminary X-ray crystallographic analysis of glucuronoxylan xylanohydrolase (Xyn30A) from *Clostridium thermocellum*

Anil Kumar Verma,[a] Arun Goyal,[a]* Filipe Freire,[b] Pedro Bule,[b] Immacolata Venditto,[b] Joana L. A. Brás,[b] Helena Santos,[b] Vânia Cardoso,[b] Cecília Bonifácio,[c] Andrew Thompson,[d] Maria João Romão,[c] José A. M. Prates,[b] Luís M. A. Ferreira,[b] Carlos M. G. A. Fontes[b] and Shabir Najmudin[b]*

[a]Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati 781 039, India, [b]CIISA–Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal, [c]REQUIMTE–CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal, and [d]Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin, 91192 Gif-sur-Yvette, France

Correspondence e-mail: arungoyl@iitg.ernet.in, shabir@fmv.utl.pt

Received 6 August 2013
Accepted 9 September 2013

The modular carbohydrate-active enzyme belonging to glycoside hydrolase family 30 (GH30) from *Clostridium thermocellum* (*Ct*XynGH30) is a cellulosomal protein which plays an important role in plant cell-wall degradation. The full-length *Ct*XynGH30 contains an N-terminal catalytic module (Xyn30A) followed by a family 6 carbohydrate-binding module (CBM6) and a dockerin at the C-terminus. The recombinant protein has a molecular mass of 45 kDa. Preliminary structural characterization was carried out on Xyn30A crystallized in different conditions. All tested crystals belonged to space group *P*1 with one molecule in the asymmetric unit. Molecular replacement has been used to solve the Xyn30A structure.

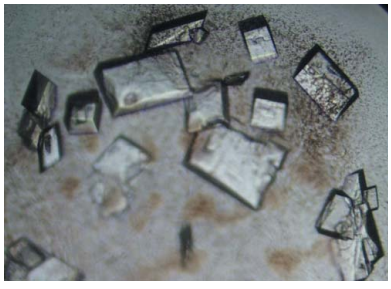
1. Introduction

The modular carbohydrate-active enzyme belonging to glycoside hydrolase family 30 (GH30) from *Clostridium thermocellum* (*Ct*XynGH30) is a cellulosomal protein which plays an important role in plant cell-wall degradation. It contains an N-terminal signal peptide followed by a catalytic module (Xyn30A), a C-terminal carbohydrate-binding module belonging to family 6 (CBM6) and a dockerin module. *Ct*XynGH30 was originally classified as a glycoside hydrolase (GH) family 5 member in the CAZy database. It has now been reclassified as a member of the GH family 30 glucuronoxylan xylanohydrolases, subfamily H (in the CAZy database it is in subfamily 8; Cantarel *et al.*, 2009; St John *et al.*, 2010; Aspeborg *et al.*, 2012). Structurally characterized proteins of this subfamily include XynA from *Erwinia chrysanthemi* (Hurlbert & Preston, 2001; Vrsanská *et al.*, 2007) and XynC from *Bacillus subtilis* (St John *et al.*, 2009). These display preferential hydrolysis of glucuronoxylan to branched xylo-oligosaccharides and negligible activity towards linear β-1–4-linked xylo-oligosaccharides (St John *et al.*, 2006, 2011). Sequence alignment, protein-modelling and molecular-docking studies of Xyn30A showed that the modelled structure is close to the XynC family GH30 from *B. subtilis* 168, presenting a common $(\beta/\alpha)_8$-barrel fold with distinguishing structural features of family GH30 xylanases such as the obligate 'side β-structure' similar to CBM (St John *et al.*, 2006). Determining the three-dimensional structure of *C. thermocellum* Xyn30A is an essential step for the further biochemical understanding of the protein activity towards its different substrates. In the present study, recombinant protein Xyn30A was cloned, overexpressed, purified and crystallized. X-ray diffraction data were collected for preliminary structure analysis.

2. Materials and methods

2.1. Bacterial strains and plasmids

Escherichia coli DH5α cells were used for cloning and *E. coli* BL21 (DE3) cells were used as the expression host. The plasmid used for cloning and expression was pET-28a(+). All of the above-mentioned items were procured from Novagen (Madison, Wisconsin, USA). pGEM-T Easy vector was purchased from Promega Corporation (Madison, Wisconsin, USA).



2.2. Gene amplification and cloning

Oligonucleotide primers containing *Nhe*I and *Xho*I restriction sites were designed. The primers used for amplifying Xyn30A were the forward primer 5′-CTCT**GCTAGC**GCAACAATCAACTTGTCG-3′ and the reverse primer 5′-CACA**CTCGAG**TTATACAAAAGTT-GTAACGC-3′. PCR was carried out using denaturation at 367 K for 5 min and 30 cycles of (i) denaturation at 367 K for 30 s, (ii) annealing at 328 K for 60 s and (iii) extension at 345 K for 2 min followed by extension at 345 K for 5 min. The PCR product was run on 0.8% agarose gel, purified using a QIAquick gel-extraction kit, ligated into pGEM-T Easy vector and positive clones were screened (Suzuki *et al.*, 1997). The Xyn30A gene fragment was extracted and cloned into pET-28a(+) vector using *Nhe*I and *Xho*I restriction enzymes. *E. coli* DH5α cells were transformed by the heat-shock method using cloned DNA. The transformed cells were grown on LB–agar plates supplemented with kanamycin (50 µg ml^{-1}) at 310 K. Positive clones were confirmed by restriction digestion of the recombinant plasmid (pXyn30A) purified using a QIAprep Spin Miniprep Kit.

2.3. Expression and purification of recombinant Xyn30A

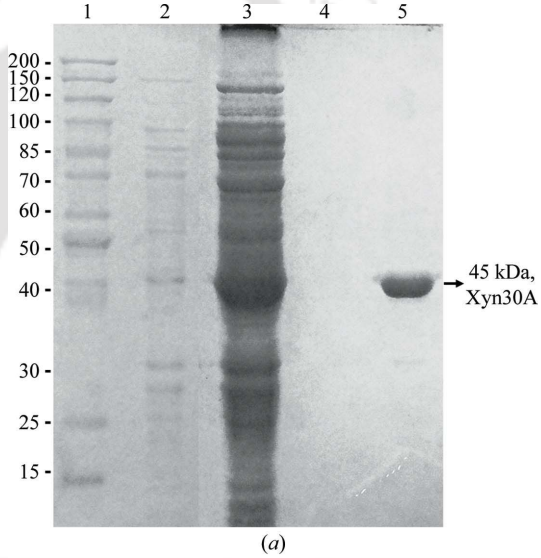
E. coli BL21(DE3) cells were transformed with the recombinant plasmid pXyn30A. Cells harbouring pXyn30A were grown in LB containing kanamycin (50 µg ml^{-1}) at 310 K at 180 rev min^{-1} until the mid-exponential phase ($A_{550} \simeq 0.6$), recombinant protein expression was induced with 1.0 m*M* isopropyl β-D-1-thiogalactopyranoside and the cells were left for an additional 16 h at 292 K and 180 rev min^{-1} (Ichinose *et al.*, 2008). The cells were harvested by centrifugation at 8000*g* at 277 K for 15 min and the resulting cell pellet was resuspended in 50 m*M* HEPES buffer containing 10 m*M* imidazole, 1 *M* NaCl, 3 m*M* $CaCl_2.2H_2O$ pH 7.5. The cells were then sonicated on ice for 15 min (8 s on/16 s off pulse, 33% amplitude) and centrifuged again at 17 000*g* at 277 K for 40 min to obtain the cell-free extract. The recombinant Xyn30A containing a His_6 tag was purified from the cell-free extract by immobilized metal-ion affinity chromatography (IMAC) using a Sepharose column (HiTrap Chelating, GE Healthcare) and was eluted using 50 m*M* HEPES buffer containing 300 m*M* imidazole, 1 *M* NaCl, 3 m*M* $CaCl_2{\cdot}2H_2O$ pH 7.5. Purified Xyn30A was buffer-exchanged into 50 m*M* HEPES buffer pH 7.5 containing 200 m*M* NaCl, 5 m*M* $CaCl_2$ using a PD-10 column and was then subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml min^{-1}. Purified Xyn30A was concentrated using a 10 kDa cutoff centrifugal concentrator (Millipore, USA) and washed three times with 0.5 m*M* $CaCl_2$. The purity and molecular mass of the recombinant protein were analysed by 13%(*w*/*v*) SDS–PAGE (Laemmli, 1970; Fig. 1*a*). The purified recombinant Xyn30A contains an N-terminal His_6 tag (MGSSHHHHHHSSGLVPRGSHMAS), giving a total of 409 amino-acid residues with a molecular mass of 45.18 kDa. The protein concentration was estimated using a molar extinction coefficient (ε) of 91 853 M^{-1} cm^{-1} with a NanoDrop 2000c spectrophotometer (Thermo Scientific).

2.4. Crystallization

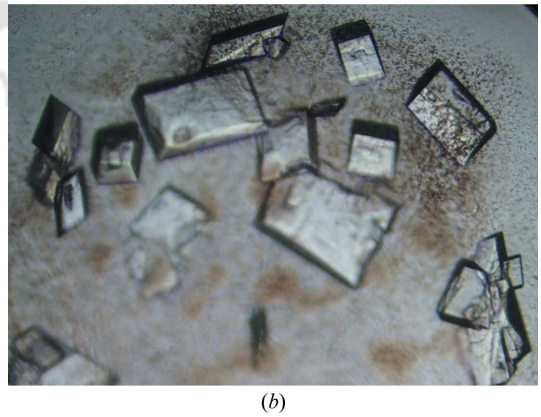
The crystallization conditions were screened by the sitting-drop vapour-diffusion method using the commercial screens Crystal Screen, Crystal Screen 2 and PEG/Ion from Hampton Research (California, USA), JCSG from Molecular Dimensions and an in-house-prepared sparse-matrix screen of 80 conditions using a NanoDrop robotic dispensing system Oryx8 (Douglas Instruments). Drops consisting of 0.9 µl Xyn30A solution at either 16 or 32 mg ml^{-1} and 0.9 µl reservoir solution were prepared at 292 K. Crystals appeared within 1 d, reaching their maximum size within a few days in many different conditions: (i) 2 *M* ammonium sulfate; (ii) 0.1 *M* HEPES pH 7.5, 10% polyethylene glycol (PEG) 6000, 5%(*v*/*v*) (±)-2-methyl-2,4-pentanediol; (iii) 0.2 *M* sodium fluoride, 20%(*w*/*v*) PEG 3350; (iv) 0.2 *M* potassium sulfate, 20%(*w*/*v*) PEG 3350 (Fig. 1*b*); (v) 0.2 *M* sodium acetate trihydrate pH 7.0, 20% PEG 3350; and (vi) 2%(*v*/*v*) Tacsimate pH 4.0, 0.1 *M* sodium acetate trihydrate 4.6, 16% PEG 3350. The crystals were cryocooled in liquid nitrogen after soaking in the appropriate cryoprotectant [30%(*v*/*v*) glycerol added to the crystallization buffer] for a few seconds, except for the crystal grown in 0.2 *M* potassium sulfate, 20%(*w*/*v*) PEG 3350, which was cryocooled in Paratone.

2.5. Data collection and processing

Data from all the different crystals obtained were collected on beamline ID29 at the European Synchrotron Radiation Facility (ESRF), Grenoble (de Sanctis *et al.*, 2012) and at PROXIMA-1 at SOLEIL (Orsay, France) using a PILATUS 6M detector with the crystals cooled to 100 K using a Cryostream (Oxford Cryosystems). Data from the ESRF and SOLEIL were processed using the



(*a*)



(*b*)

Figure 1
(*a*) SDS–PAGE [13%(*w*/*v*)] showing overexpression and purification of Xyn30A. Lane 1, Page Ruler protein marker (labelled in kDa); lane 2, uninduced Xyn30A cells; lane 3, cell-free extract; lane 4, last wash from column; lane 5, purified Xyn30A (~45 kDa). (*b*) Crystals of Xyn30A obtained by sitting-drop vapour diffusion. The largest crystals are approximately 200 × 120 × 60 µm in size.

Table 1
Data-collection statistics.

Values in parentheses are for the lowest/highest resolution shells.

Data set	Xyn30A
Beamline	PROXIMA-1, SOLEIL
Wavelength (Å)	0.826
Space group	*P*1
Unit-cell parameters (Å, °)	a = 46.09, b = 47.46, c = 53.62, α = 83.15, β = 73.41, γ = 65.87
Resolution limits (Å)	42.93–1.40 (42.93–4.43/1.48–1.40)
Mosaicity (°)	0.39
No. of observations	347015 (11412/50059)
No. of unique observations	76088 (2464/10874)
Multiplicity	4.6 (4.6/4.6)
Completeness (%)	97.2 (99.3/95.6)
$\langle I/\sigma(I)\rangle$	3.5 (7.3/1.3)
$CC_{1/2}$†	0.971 (0.974/0.536)
R_{merge}‡	0.196 (0.100/0.860)
$R_{p.i.m.}$§	0.101 (0.052/0.449)

† $CC_{1/2}$ is the half-data-set correlation coefficient (Diederichs & Karplus, 2013). ‡ $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith intensity measurement of reflection hkl, including symmetry-related reflections and $\langle I(hkl)\rangle$ is its average. § $R_{p.i.m.} = \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the average of symmetry-related observations of a unique reflection.

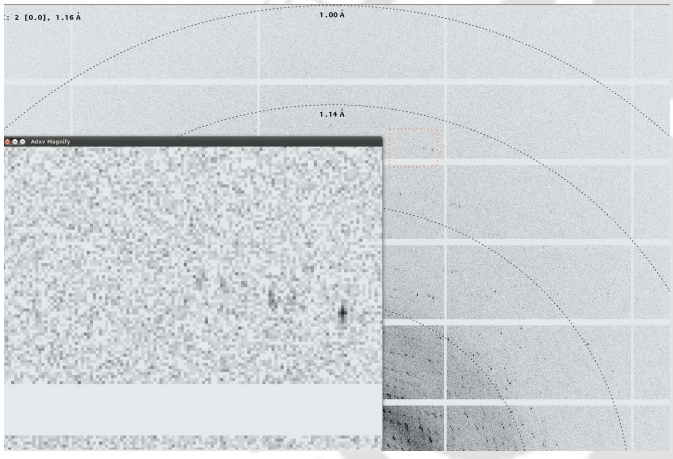


Figure 2
Representative diffraction pattern of a Xyn30A crystal with resolution rings shown as dashed circles. The inset shows an enlargement of the contents of the red box, showing spots at ∼1.2 Å resolution.

programs *iMOSFLM* (Battye *et al.*, 2011) and *AIMLESS* (Evans, 2011) from the *CCP*4 suite (Winn *et al.*, 2011) or *XDS* (Kabsch, 2010). All of the crystals tested belonged to space group *P*1 with one molecule in the asymmetric unit and a solvent content ranging from 45 to 47% and had similar data-collection statistics, despite being grown in different conditions and having different crystal morphologies.

The data collected at SOLEIL were from the best diffracting crystal (Fig. 1*b*) and were used to solve the Xyn30A structure. 470° of data were collected with a $\Delta\varphi$ of 0.2°. The crystal diffracted to a resolution of 1.1 Å (Fig. 2). The crystal was broken/multiple and a second low-resolution data set was needed to index and to find the correct lattice before integrating the high-resolution data with *XDS*. The resolution was cut at a conservative 1.4 Å (the $CC_{1/2}$ value is below 0.5 for higher resolution data). The slightly high R_{merge} value at this resolution is due to radiation damage from a longer exposure for a *P*1 data collection. The Matthews coefficient of 2.33 $Å^3$ Da^{-1} (Matthews, 1968) indicates the presence of one molecule in the asymmetric unit with a solvent content of 47.2%. Data-collection statistics are given in Table 1.

The three-dimensional structure was solved by molecular replacement using the program *Phaser* (McCoy *et al.*, 2007) as implemented in the *PHENIX* software package (Adams *et al.*, 2010) and using the crystal structure of XynC from *B. subtilis* 168 (PDB code 3gtn; St John *et al.*, 2009) as a search model, with the final solution giving an LLG score of 181 and a TFZ of 24.5. Structure refinement and analysis are ongoing.

This research work was supported in part by a Cutting-edge Research Enhancement and Scientific Training (CREST) Fellowship from the Department of Biotechnology, Ministry of Science and Technology to AG. A scholarship from IIT Guwahati to AKV is gratefully acknowledged. The authors also acknowledge financial support from Fundacão para a Ciência e a Tecnologia, Portugal through projects PTDC/BIA-PRO/103980/2008 and PTDC/BIA-PRO/100359/2008. We acknowledge SOLEIL for provision of synchrotron-radiation facilities (proposal ID 20120372) and Dr Victor D. Alves for help with data collection. We would also like to thank the ESRF, Grenoble (beamline ID29) for access and technical support during data collection.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H. III & Henrissat, B. (2012). *BMC Evol. Biol.* **12**, 186.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* D**69**, 1215–1222.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Hurlbert, J. C. & Preston, J. F. III (2001). *J. Bacteriol.* **183**, 2093–2100.
Ichinose, H., Yoshida, M., Fujimoto, Z. & Kaneko, S. (2008). *Appl. Microbiol. Biotechnol.* **80**, 399–408.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Sanctis, D. de *et al.* (2012). *J. Synchrotron Rad.* **19**, 455–461.
St John, F. J., Godwin, D. K., Preston, J. F., Pozharski, E. & Hurlbert, J. C. (2009). *Acta Cryst.* F**65**, 499–503.
St John, F. J., González, J. M. & Pozharski, E. (2010). *FEBS Lett.* **584**, 4435–4441.
St John, F. J., Hurlbert, J. C., Rice, J. D., Preston, J. F. & Pozharski, E. (2011). *J. Mol. Biol.* **407**, 92–109.
St John, F. J., Rice, J. D. & Preston, J. F. (2006). *J. Bacteriol.* **188**, 8617–8626.
Suzuki, T., Ibata, K., Hatsu, M., Takamizawa, K. & Kawai, K. (1997). *J. Ferment. Bioeng.* **84**, 86–89.
Vrsanská, M., Kolenová, K., Puchart, V. & Biely, P. (2007). *FEBS J.* **274**, 1666–1677.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

ISSN 0026-8933, Molecular Biology, 2014, Vol. 48, No. 2, pp. 278–286. © Pleiades Publishing, Inc., 2014.
Published in Russian in Molekulyarnaya Biologiya, 2014, Vol. 48, No. 2, pp. 322–332.

STRUCTURAL AND FUNCTIONAL ANALYSIS OF BIOPOLYMERS AND BIOPOLYMER COMPLEXES

UDC 577.152.351

In silico Structural Characterization and Molecular Docking Studies of First Glucuronoxylan-Xylanohydrolase (Xyn30A) from Family 30 Glycosyl Hydrolase (GH30) from *Clostridium thermocellum*[1]

Anil Kumar Verma and Arun Goyal

Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati, 781039, Assam, India
e-mail: arungoyl@iitg.ernet.in

Received June 13, 2013; in final form, August 22, 2013

Abstract—*Ct*XynGH30 is a carbohydrate active modular enzyme and component of cellulosome of *Clostridium thermocellum*. The full length *Ct*XynGH30 contains an N-terminal catalytic module named as Xyn30A and a family 6 carbohydrate binding module (CBM6) at C-terminus. Xyn30A was modeled by computer program Modeller9v8 taking crystal structure of XynC from *B. subtilis* as a template to generate the molecular model. Model refinement was done using energy minimization by implementing steepest descent algorithm with GROMOS96 43a1 force field. Quality assessment by Ramachandran plot showed that 91% amino acids lie in most favourable region and 9% in additional allowed region. Structural analysis depicted that Xyn30A has a $(\beta/\alpha)_8$ barrel fold. Additionally, it had a β-strand rich structure called 'side β-structure' attached with main catalytic core. Structural superimposition reflected that Glu136 act as a catalytic acid/base while Glu225 act as a catalytic nucleophile. Multiple sequence alignment showed that these catalytic residues are conserved within the family. The docking results showed that these residues display polar interaction with linear and substituted xylo-oligosaccharides. The binding interaction of ligands depicted that aromatic amino acids Trp81, Tyr139, Trp143, Phe172, His198, Tyr200, Tyr227, Trp264 and Tyr265 create binding site pocket around the active site. We report overall structural feature, conserved active site residues and enzyme-ligand docking of first glucuronoxylan-xylanohydrolase (Xyn30A) of family 30 glycosyl hydrolase (GH30) from *Clostridium thermocellum*.

DOI: 10.1134/S0026893314020022

Keywords: glycoside hydrolase family 30, *Clostridium thermocellum*, Xyn30A, homology modeling, heteroxylans, ligand binding

INTRODUCTION

Plant polysaccharides are chemically complex structures whose main constituents are cellulose, hemicellulose and lignin [1]. Hemicellulose is a heteropolysaccharide that mainly constitute xylan, the second most abundant polymer on earth along with glucans, mannans and arabinans. Due to heterogeneity and complexity of xylan structure, it required action of different enzyme to complete degradation into simple sugars [2, 3]. Therefore, some anaerobic microbes and fungi are evolved with multi-enzyme complex systems called cellulosomes. *Clostridium thermocellum* is a gram positive thermophilic anaerobic bacterium known to have one of the highest rates of cellulose utilization [4, 5]. *Ct*XynGH30 is one of the proteins from the cellulosome of *Clostridium thermocellum*. CAZy is a continuously updated database of carbohydrate-active enzymes classified on the basis of amino acid sequence similarity and hydrophobic cluster analysis [6]. Initially *Ct*XynGH30 was placed in Glycoside Hydrolase (GH) family 5 in CAZy database but after amino acid sequence alignment and phylogenetic analysis it has been reclassified into GH family 30 glucuronoxylan xylanohydrolase, sub-family H (in CAZy database it is in sub-family 8) [7]. In this sub-family, only two proteins are structurally characterized till now viz. XynC from *B. subtilis* a Gram-positive bacterium (PDB id: 3GTN) and XynA (PDB id: 1NOF) from Gram-negative phytopathogenic *E. chrysanthemi* and both of proteins were studied in their native as well as ligand bound crystal structures [8, 9]. Four more proteins belonging to family GH30 are biochemically characterized viz. Xyn5B from *Bacillus* sp. strain BP-7 showing activity on neutral, non-substituted xylo-oligosaccharides [10] and Xyn30D from *Paenibacillus barcinonensis* which efficiently hydrolyzed glucuronoxylans and methyl-glucuronic acid branched xylo-oligosaccharides [11]. Xyn30A from *Bacillus licheniformis* SVD1 was characterized as glucuronoarabinoxylan endo-1,4-β-xylanase [12] and XynD as xylanase from *Aeromonas caviae* ME-1[13].

Chemically, glucuronoxylan (GX_n) is composed of a linear backbone of β-(1-4)-linked D-xylopyranose units having a substitute of α-D-glucuronic acid

[1] The article is published in the original.

(GlcA) or 4-*O*-methyl α-D-glucuronic acid (MeGlcA) residue at O2 position by α(1-2) glycosidic bond. GX_n is the major component of secondary cell wall of hardwood plant and crop residues hence constitutes a large part of plant biomass [14]. The xylanases of GH30 subfamily H are specific for hydrolysis of particular type of xylan such as glucuronoxylan hence they differ from other xylanases belonging to families GH5, GH10 and GH11 which cleave internal xylosidic bonds of mainly linear xylo-oligosaccharides displaying endo-1,4-β-xylanase activity [15]. Glucuronoxylan xylanohydrolase differs from xylanases of family GH39 and GH43 which showed β-xylosidase activity hence producing mainly xylose as the main product [16, 17]. Glucuronoxylan xylanohydrolase also differs from family GH 67 α-glucuronidase that cleaves α(1-2) glycosidic bond of GlcA or MeGlcA residues, substituted on main chain of xylan [18]. Xylanases of family GH30 required a xylan decorated with GlcA or MeGlcA moiety. It cleaves the β-1,4 xylosidic bonds position along with α(1-2) substituted GlcA or MeGlcA moiety [19]. The complete degradation of glucuronoxylan required cumulative action of endo-1,4-β-xylanase, β-xylosidase and α-glucuronidase enzymes. The important applications of microbial xylanases have been reported in paper and pulp industry, bioconversion of plant polysaccharides for bioethanol production, improvement in the digestibility of animal feed stock, baking and brewing industries [20–23]. Current work, focused on structural features, conserved catalytic residues and probable mechanism of catalysis of first glucuronoxylan-xylanohydrolase (named as Xyn30A) from *Clostridium thermocellum*. We also analyzed the enzyme-ligand complex obtained by docking which gave valuable information about the conserved aromatic residues present in catalytic core and their possible roles in binding of the substrates.

EXPERIMENTAL

Retrieval of sequence and defining boundaries of modular *Ct*XynGH30. Sequences were retrieved from the CAZy Database (http://www.cazy.org/) having gene accession number ABN54208 and uniprot ID A3DJS9. Domain boundaries of modular *Ct*XynGH30 were defined by analyzing the result of PSI-Blast [24] and InterProScan [25]. N-terminal signal peptide was detected by signaIP [26]. Final adjustment in domain borders was carried out by sequence alignment using Clustal X program [27].

Homology modeling. The catalytic module Xyn30A was modelled by employing computer program Modeller9v8 [28]. It uses a knowledge based comparative protein structure modeling hence the first step is to align the query sequence with best matched template of known protein structure. XynC family GH30 from *Bacillus subtilis* 168 (PDB id: 3GTN) selected as the template for structure modeling based on its sequence and functional homology. Modeller derives the restraints from the template and database of known structure. 10 independent models were generated after optimization of molecular probability density functions. Refinement of loops of modeled structure was achieved with the help of loop optimization methods and after each cycle of loop refinement discrete optimized protein energy (DOPE) Score was generated this process was repeated in an iterative fashion until DOPE Score attained to be negative (lower than −1) [29]. The model with the lowest discrete optimized protein energy (DOPE) was chosen for further studies.

Model refinement and quality assessment. Energy minimization was carried out on modeled protein in order to relax any unfavorable bond angle and bad contacts before performing docking. Energy minimization was performed by steepest descent algorithm with GROMOS96 43a1 force field and simple point charge (SPC) water model with GROMACS 4.0.7 package (http://www.gromacs.org/). After defining simulation box and solvent system (here water molecule) around protein, charge on the protein molecule was neutralized by adding counter-ions (Na^+ was added) by replacing same number of water molecule which are at least 3.50 Å from the protein surface. Final structure after energy minimization was validated on structure analysis and verification server (SAVES) at NIH–MBI laboratory (http://nihserver.mbi.ucla.edu). Ramachandran plot which graphically represent the combination of possible and permitted torsional angles, backbone phi (φ) and psi (ψ) dihedral angles, was obtain using PROCHECK program [30]. The Verify_3D program was used to determine a compatibility of an atomic model (3D) with its own amino acid sequence (1D). First it is categorized each residue into structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc.) then it generates a score by comparing each residues with a collection of good structures as a reference [31]. Stereochemical quality and errors in protein structure were accessed at ProSA (Protein Structure Analysis) web Server (https://prosa.services.came.sbg.ac.at/prosa.php). It calculates a Z-score which represents the overall model quality. Inaccurate protein structures are placed outside the range of values in the scatterplot [32, 33].

Molecular dynamics. The energy-minimized Xyn30A model was placed inside a cubic box of single point charge (SPC) water molecules. The net charge of system was neutralized with counter ions. The system was equilibrated for 50 ps in NVT ensemble by restraining the solute atoms. Then the system was equilibrated for 50 ps by NPT ensemble twice, first with restraints followed by without restraints. Production run was carried out for 2 ns with NPT ensemble using 2 fs of integration time. The linear constraint solver (LINCS) algorithm was used to constrain the bonds involving hydrogen atoms [34]. Conformations of Xyn30A model at regular intervals throughout the

production run were analyzed as time-dependent function to check whether the Xyn30A model is energetically stable in the solvent system.

Prediction of active site. Active site involves key amino acid residues which participate in the reaction mechanism. Generally these residues are conserved within the family. Therefore, to identify the active site residues, we performed structure alignment of Xyn30A with template protein of family GH30 using PyMOL (0.99). Moreover, whether these active site residues are conserved among the GH30 family or not, we performed and analyzed multiple sequence alignment (MSA) taking all the representative members of GH30 subfamily H, sequences whose structural and biochemical characterization are available. Related protein sequence was retrieved from PDB (http://www.rcsb.org/pdb/) or CAZy Database, saved in FASTA format. Alignment was performed using Clustal X program and final figure was generated by ESpript (http://espript.ibcp.fr), for better understanding of conserved residues.

Docking study on modeled Xyn30A. Molecular docking studies were performed using Autodock 4.2.1. Autodock uses novel and robust docking method by implementing new scoring function that estimates the free energy change during ligand binding [35]. Linear xylo-oligosaccharides and branched glucuronoxylan were taken as ligand molecules. The following protocol was used to perform docking. Ligand PDBs were obtained from pubchem (http://pubchem.ncbi.nlm.nih.gov) or from ligand bound protein complex (http://www.rcsb.org/pdb/). At first, the PDB files of protein and ligand were converted to PDBQT files by assigning gasteiger charges. Thereafter, non-polar hydrogens were merged and there charges were assigned to carbon atoms. Grid box were assigned around active site having dimension of 66, 46, 48 (x, y, z coordinates) with 0.375 Å grid point spacing. Lamarckian Genetic Algorithm (LGA) was implemented for docking simulation and conformational search [36]. Number of GA runs was set to be 30. Initial population size were (150), maximum energy evaluation per run (2500000) and maximum number of generation was set to be 27000. After successful completion of docking runs, docked conformation and interaction energies were saved and analyzed from the largest cluster having a minimum lowest free energy of binding (ΔG). 2D lig-plot (http://www.ebi.ac.uk/pdbsum/) of protein-ligand complex of best docked conformation was generated for the better understanding and representation of interaction.

RESULTS AND DISCUSSION

Molecular Architecture of *Ct*XynGH30

The molecular architecture of modular carbohydrate active enzyme *Ct*XynGH30 showed N-terminal catalytic module Xyn30A of 1158 bp and a C-terminal carbohydrate binding module belonging to family 6 (CBM6) of 369 bp. *Ct*XynGH30 is a 630 aa protein. Full length protein comprises, 1–30 aa signal peptide, catalytic module Xyn30A of 391 aa, carbohydrate binding module of 127 aa, dockerin type I of 71 aa and a linker of 6 aa which connect CBM to catalytic module.

Homology Modeling and Structure Validation

Blastp of Xyn30A in PDB database at NCBI (http://blast.ncbi.nlm.nih.gov/) for sequence similarity search showed a top hit of crystal structure of XynC family GH30 from *Bacillus subtilis* 168 (PDB id: 3GTN) having a 72% sequence identity [8]. Before performing docking, modeled Xyn30A was validated with different parameters like Ramachandran plot which showed 91% amino acid residues lies in most favourable region and 9% residues in additional allowed region, which implied that structure-backbone dihedral angles, phi (φ) and psi (ψ) occupied favourable positions (Fig. 1a). ProSA result as shown in the Fig. 1b indicated that the modeled protein is error free and reside in the x-ray zone with scores of –9.58. Verify_3D [37] result depicted that all 386 amino acid residues has average 3D–1D profile score greater than 0.22 which illustrated that no amino acid is in bad contact (Fig. 1c).

Molecular Dynamics Simulation of Xyn30A

Molecular dynamics (MD) simulation was performed to ensure the "Compactness" of the protein throughout the simulation. Result suggested that Xyn30A undergoes a significant change in the initial 1.0 ns of simulations and converged with fluctuations less than 0.3 Å, which indicated the stable conformation of Xyn30A model (Fig. 2). The RMSD between energy minimized model of Xyn30A and final structure from MD was 1.7 Å. MD simulation suggested that all the parameters are well within the qualitative limit of compactness and stability during the simulation time signifying the energy minimized Xyn30A model is satisfactory for molecular docking simulation.

Overall Structural Feature of Xyn30A

Modeled Xyn30A showed common $(\beta/\alpha)_8$ barrel folding structure same as present in GH5 and GH10 families. The eight β-sheets surrounded by α-helix form a barrel shaped structure which makes a cavity to accommodate the substrate. However, there was a clear distinguishing structural feature in GH30 family observed that it had obligate 'side β-structure' (Fig. 3a) which is not associated with GH5 and GH10 families [38–40]. The structure analysis of side β-structure depicted a β-strand rich structure having 9-strands (Fig. 3b) running anti-parallel to each other resembling a carbohydrate binding module (CBM). Moreover, it shows the first β-sheet β1 (written 1 in Fig. 3b)

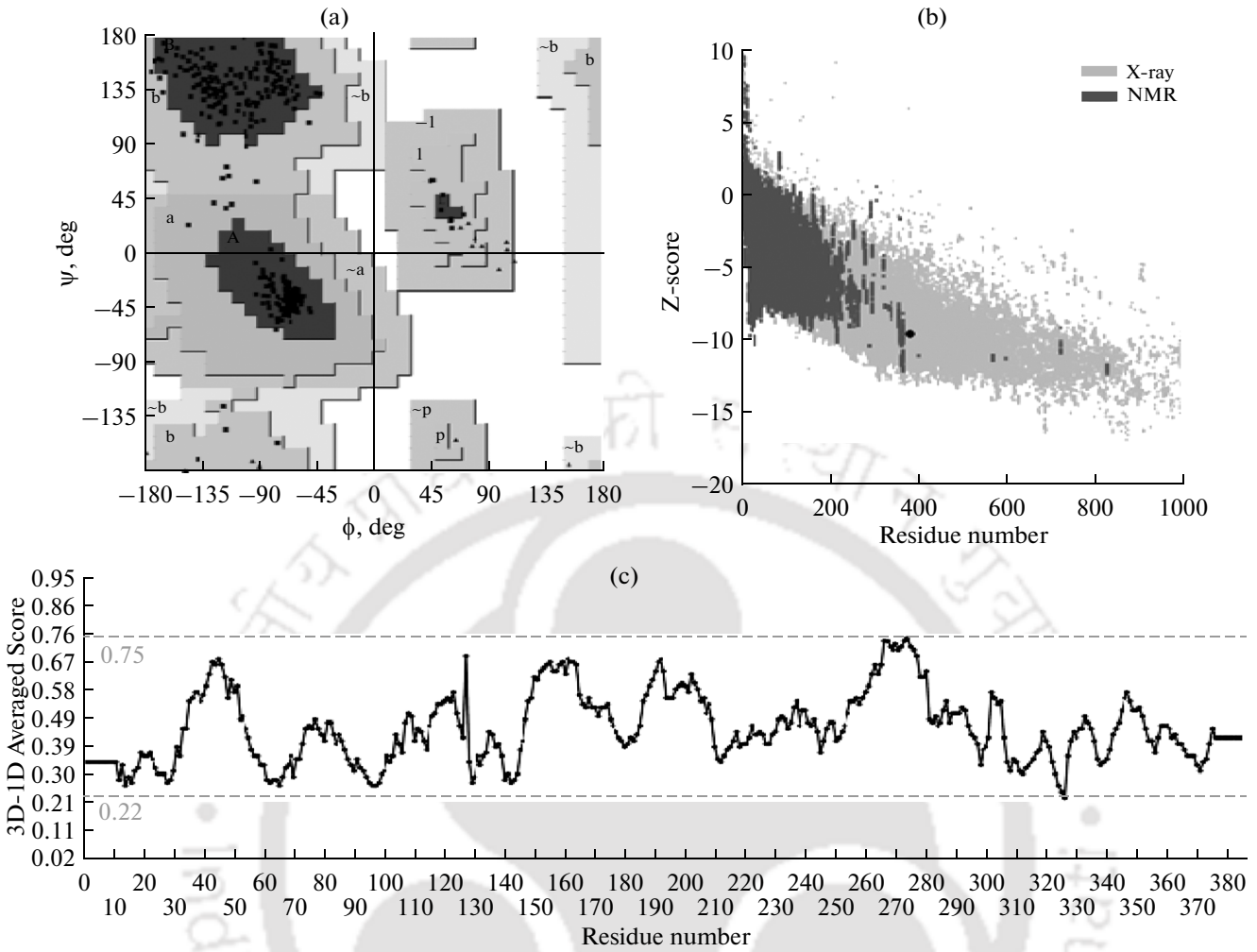


Fig. 1. Structure validation and quality assessment of Xyn30A by (a) Ramachandran plot, (b) ProSA web server, and (c) Verify 3D.

represent the N-terminal side as after this the amino acid the chain enters $(\beta/\alpha)_8$ catalytic domain and β-strands from β2–β9 represent the C-terminal region. The same arrangement of side β-structure was observed in XynC [7]. Regarding the functionality and importance of side β-structure, it is demonstrated that it will act as a CBM thereby having the function to enhance the performance of catalytic domain by keeping the substrate in close proximity [41]. The first evidence about the role of side β-structure was confirmed from the ligand 4-*O*-methyl glucuronoxylan ($MeGX_n$) bound structure of XynC which clearly indicated that this motif selectively targeting $MeGX_n$ as substrate [42]. Concerning the importance of side β-structure, different truncated recombinant of Xyn30D catalytic module was constructed which lacked full-length side β-structure. It was found that the truncated recombinant clone which lacked full-length side β-structure does not show any hydrolytic activity on xylan. It implied that the associated side β-domain is quite essential for catalysis [11]. In brief, an evolving func-

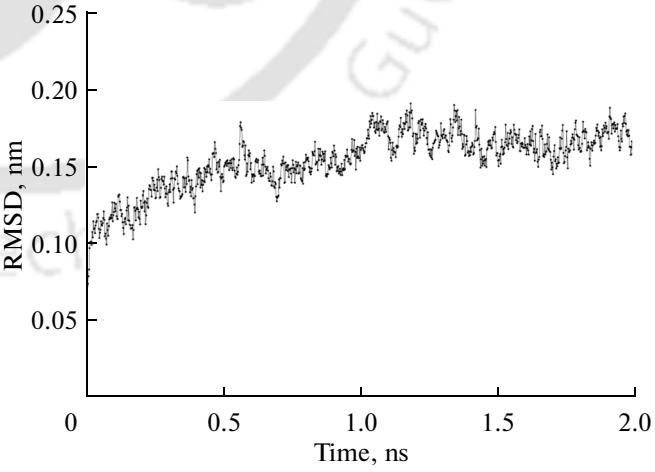


Fig. 2. MD simulation of modeled Xyn30A shows that the energy-minimized structure is energetically stable under the simulation conditions.

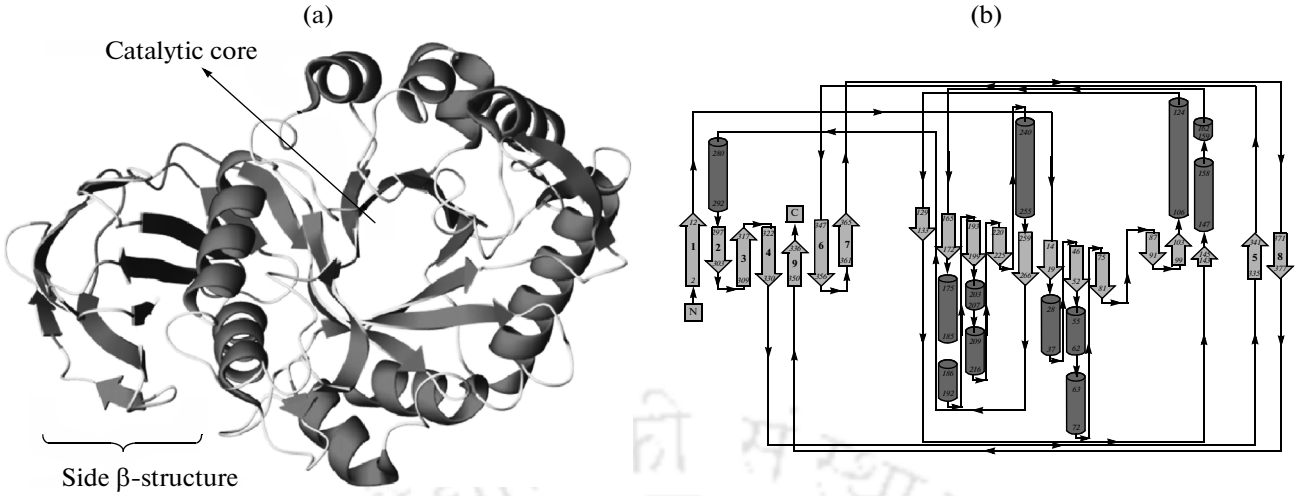


Fig. 3. (a) Overall 3-D cartoon representation of modeled Xyn30A from front side (C-terminal) depicted the $(\beta/\alpha)_8$ barrel fold structure showing the catalytic core in which encircled β-sheets are surrounded by α-helix along with peculiar side β-structure. (b) Topology diagram of Xyn30A represents a mutual orientation of secondary structure in 2D mentioning 1–9 arrows represent β-sheet of side β-structure, other arrows represent encircled β-sheets and cylinders representing the α-helices.

tion of family GH30 subfamily H is that it has inbuilt CBM along with catalytic module which facilitates substrate binding and pivotal for activity of catalytic module.

Catalytic Core of Xyn30A

The catalytic core of Xyn30A formed by $(\beta/\alpha)_8$ barrel folding encircled β-sheets residues 13–291 aa (Fig. 3a). Close inspection of $(\beta/\alpha)_8$ barrel folding structure of Xyn30A revealed that the loops connecting α-helix to the -sheets are shorter toward N-terminal (or called back side) (Fig. 3a) than the loops at the C-terminus (or front side of the protein). These longer loop located at C-terminus probably facilitate the entrance of large and branched oligosaccharides [9]. Aromatic amino acid residues of active site which shape the binding cleft and their probable role in catalysis of substrate was depicted by superimposing the Xyn30A with XynC and XynA with backbone RMSD of 0.3 Å and 1.2 Å, respectively [43]. Analysis revealed that aromatic residues Trp23, Trp81, Tyr110, Trp143, Tyr200, Trp263 and Tyr264 of mature Xyn30A protein (not included 30 aa of signal peptide) corresponding to Trp55, Trp113, Trp168, Tyr172, Tyr232, Trp289, and Tyr290 of XynA, which constitute the binding cleft [9]. The analysis of the MSA results showed that the above mentioned aromatic residues are highly conserved within the GH30 family (displayed as inverted solid triangle, ▼ Fig. 4).

Probable Mechanism of Catalysis

Hydrolysis of a glycosidic bond occurs through either the retention or by inversion of anomeric configuration. The similarity between these two reactions mechanism is both form an oxacarbenium-ion-like transition states and needed a pair of carboxylic acids at catalytic site. Whilst there is subtle difference in both the mechanisms like in inverting glycoside the average distance between two carboxylic acid residues which acts as a proton donor /acid and the other as a nucleophile/base is of ~10 Å and the reaction occurs via a double-displacement mechanism wherein retaining mechanism the distance would be ~5.5 Å and reaction occurs via a single-displacement mechanism [44–46]. Catalysis of glycosidic bond in family GH30 follows the retention mechanism (http://www.cazy.org). Study of superimposed structures (Figs. 5a, 5b) of Xyn30A with XynC and XynA for finding the proper orientation and distance between the key catalytic residues showed that, Glu136 of Xyn30A corresponds to Glu140 of XynC and Glu165 of XynA, which acted as acid/base while Glu225 of Xyn30A matched with Glu229 of XynC and Glu253 of XynA, which acted as a nucleophile. Moreover, the distance between the carboxyl groups of catalytic glutamate was found to be 5.2 Å which is appropriate for retention type mechanism [9, 42]. The MSA results confirmed that these active site catalytic residues are conserved within the family (marked as star in Fig. 4) and protein-ligand docking analysis showed these residues are directly involved in making polar contact with ligand molecule (Fig. 6b).

Docking of Substrate at the Active Site

The crystal structures of GH5 and GH10 bound with xylo-oligomer ligands having the catalytic core formed by $(\beta/\alpha)_8$ barrel folding structure have been reported [47, 48]. It gives valuable information about residues interacting with the xylan backbone forming hydrogen bonds and hydrophobic interactions. After

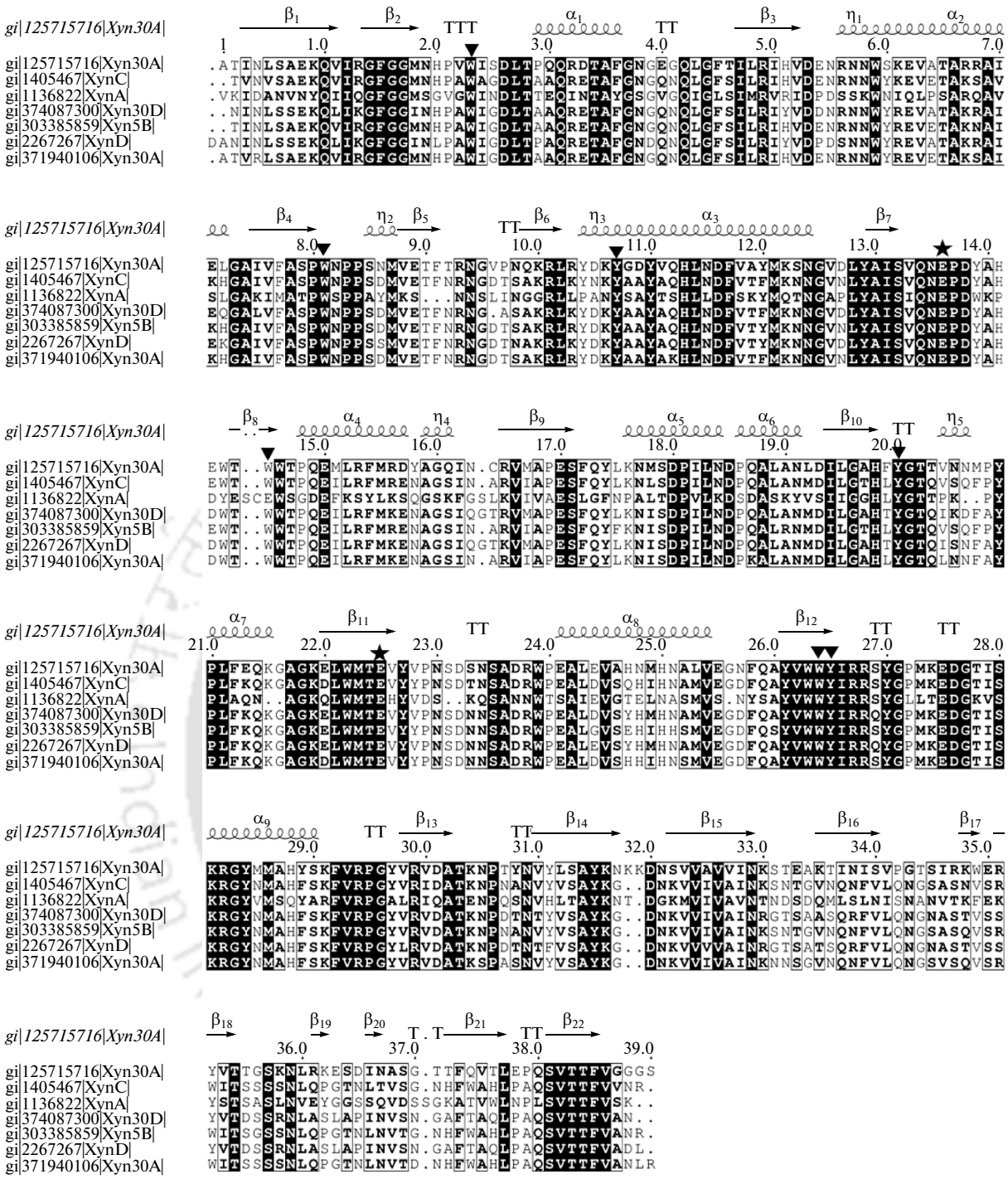


Fig. 4. Multiple sequence alignment of target (Xyn30A) with representative member of family GH30 subfamily 8 viz. XynA (PDB id: 1NOF), XynC (PDB id: 3GTN), Xyn30D, Xyn5B, Xyn30A and XynD. AYVW showing conserved and SNS semi-conserved amino acid residues. and → represent the corresponding secondary structure (α-helix and β-sheet, respectively) of Xyn30A. ▼ Represent conserved aromatic residues and ★ represent conserved catalytic residues.

knowing the location of catalytic core and conserved active site residues of Xyn30A, we performed docking to obtain best fit conformation of different ligands. The results of docking were summarized in Table. The ligand xylopentaose showed highest affinity with free energy of binding of –9.39 Kcal/mol. Xyn30A in binding mode with ligand 4-*O*-methylglucuronoxylan ($MeGlc^2Xyl_3$) depicted the cavity formed by barrel-shaped catalytic core which allowed ligands to access active site residues (Fig. 6a). Enzyme-ligand com-

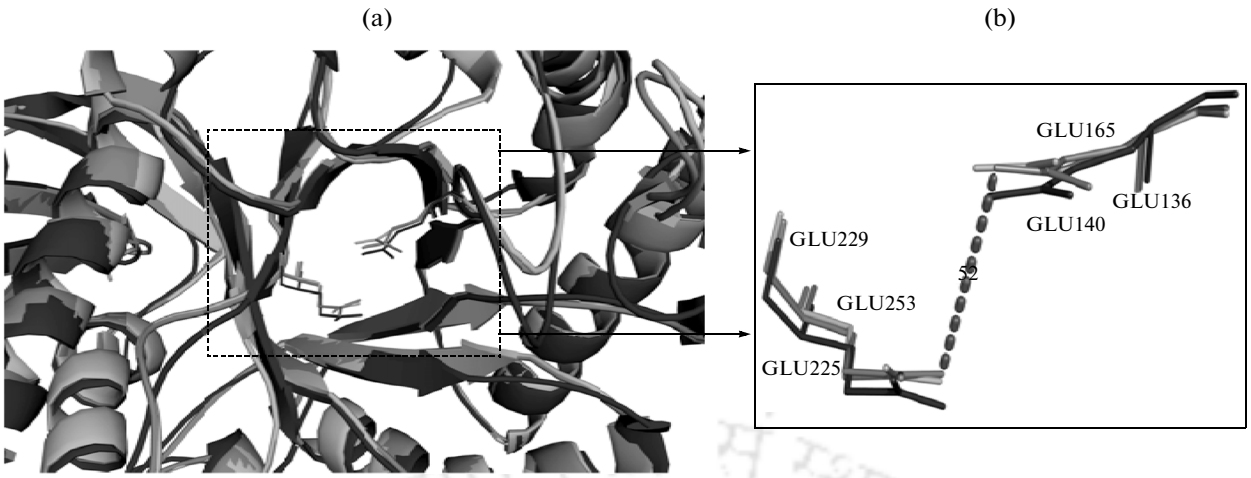


Fig. 5. (a) Ribbon representation of superimposed Xyn30A (grey). XynC (black) and XynA (light grey) proteins. (b) Enlarged view of superimposed active site residues of template, Xyn30A showing Glu136 and Glu225 (grey), the targets, XynC showing Glu165 and Glu253 (Black) and XynA showing Glu140 and Glu229 (light grey). The figure was generated using the PyMol program (http://www.pymol.org).

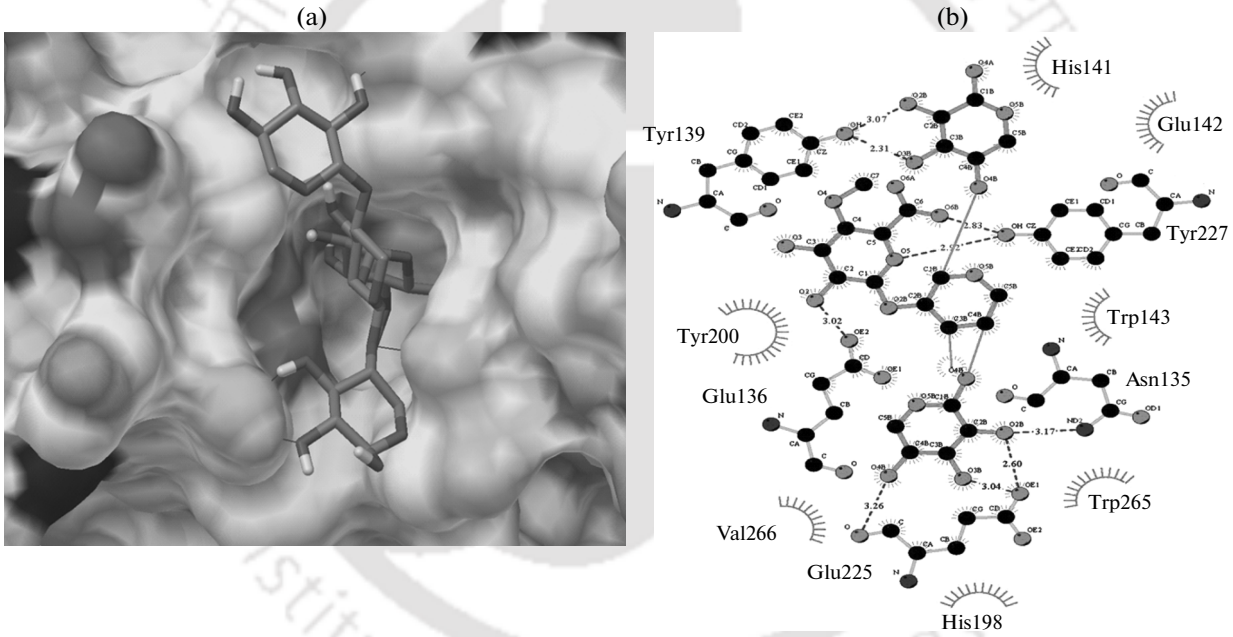


Fig. 6. (a) Surface view of the active site cleft of Xyn30A representing the binding mode with ligand 4-O-methyl glucoronoxylan ($MeGlcA^2Xyl_3$). This figure was generated with the python molecular viewer program. (b) Schematic 2D depiction of ligand 4-*O*-methyl glucoronoxylan ($MeGlcA^2Xyl_3$) interacting with active site amino acid residues of Xyn30A protein. (----) Dashed lines shows hydrogen bonds with labelled bond length, residues displaying arc with spokes making hydrophobic interaction, contacted atoms are shown with spokes radiating back.

plexes (Fig. 6b) of the same inferred that many significant polar interaction are taking place between active site amino acids and ligand molecules like O2–OE2 Glu136 (3.02 Å), O2B–OH Tyr139 (3.07 Å), O3B–OH Tyr139 (2.31 Å), O6B–OH Tyr227 (2.83 Å), O5–OH Tyr227 (2.92 Å), O2B–ND2 Asn135 (3.17 Å), O2B–OE1 Glu225 (2.6 Å) and O3B–OE1 Glu225 (3.04 Å) while His141, Glu142, Trp143, His198, Tyr200, Val226 and Trp264 were found to be in hydrophobic interaction.

CONCLUSIONS

The identification of the active site residues and understanding of their role in hydrolysis of substrate was studied. Xyn30A showed sequence and structural

Amino acid residues making polar and hydrophobic interaction (underlined) on docking

Ligand	Predicted free energy of binding by docking, ΔG, Kcal/mol	Polar interactions	Residue within 4 Å region
Xylobiose	–5.67	Tyr227, Glu225, Tyr139, Glu136	Phe172, Tyr200, His198, Trp264, Trp143, Asn135, Trp81
Xylotriose	–5.42	Tyr139, Glu136, Tyr200, Phe172, Tyr227	Trp81, Trp143, Trp264, Glu225, Asn135, His198, Val226, Phe199
Xylotetraose	–7.13	Asn135, Trp143, Glu54, Glu225, His20	Asn82, Trp81, Glu136, His198, Val52, His51, Asp53, Asn55, Glu142, Tyr227, Trp264, Tyr265, Trp23
Xylopentaose	–9.39	Glu136, Glu142, Trp81,Val52, His20	Tyr227, Tyr139, Trp264, Tyr265, Trp143, Pro80, Asn82, Asp53, Glu54, Trp23
Glucuronoxylan, $MeGlcA^2Xyl_3$	–7.19	Tyr139, Tyr227, Glu225, Asn135, Glu136	Glu142, Trp264, Val226, Tyr200, His198, His141, Trp143

similarity with XynC from *B. subtilis*, and both were reassigned in the family GH30 glucuronoxylan xylanohydrolase subfamily H. The 3-D structure of Xyn30A was modeled and validated. Molecular dynamic simulation ensured the structural integrity of energy minimized Xyn30A. Structure analysis of Xyn30A displayed that it has a $(\beta/\alpha)_8$ barrel fold structure along with an associated side β-structure resembling to CBM, attached to the main catalytic core. The side β-structure helps in substrate selection and crucial for the activity of catalytic core. Our findings inferred that carboxylic group of amino acid residue Glu136 act as a catalytic acid/base whereas Glu225 act as a nucleophile, moreover these residues are 5.2 Å apart from each other which confirmed the presence of retention type mechanism in Xyn30A. These key active site residues were found to be conserved within GH30 family. The catalytic core of Xyn30A consisted of β-barrel and connecting loops toward the C-terminal side, formed by aromatic amino acid residues Trp53, Trp173, Tyr139, Tyr230, Trp111, Trp293 and Tyr294 which found to be conserved. So the present study enlightened the complete structural features and the conserved active site residues as well as the conserved aromatic residues which over-all architect the catalytic core of glucuronoxylan-xylanohydrolase Xyn30A from *Clostridium thermocellum*.

ACKNOWLEDGMENTS

The research work in part was supported by Cutting-edge Research Enhancement and Scientific Training (CREST) Fellowship from Department of Biotechnology, Ministry of Science and Technology to Arun Goyal. The scholarship from IIT Guwahati to AKV is gratefully acknowledged. The authors thank Mr. P. Saravanan for help in molecular dynamics studies.

REFERENCES

1. Gilbert H.J. 2010. The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol.* **153**, 444–455.
2. Shallom D., Shoham Y. 2003. Microbial hemicellulases. *Curr. Opin. Microbiol.* **6**, 219–228.
3. Kulkarni N., Shendye A., Rao M. 1999. Molecular and biotechnological aspects of xylanases. *FEMS Microbiol. Rev.* **23**, 411–456.
4. Fontes C.M., Gilbert H.J. 2010. Cellulosomes: Highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* **79**, 655–681.
5. Shoham Y., Lamed R., Bayer E.A. 1999. The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends Microbiol.* **7**, 275–281.
6. Cantarel B.L., Coutinho P.M., Rancurel C., Bernard T., et al. 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* **37**, 233–238.
7. St John F.J., González J.M., Pozharski E. 2010. Consolidation of glycosyl hydrolase family 30: A dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *FEBS Lett.* **584**, 4435–4441.
8. St John F.J., Godwin D.K., Preston J.F., Pozharski E., Hurlbert J.C. 2009. Crystallization and crystallographic analysis of *Bacillus subtilis* xylanase C. *Acta Crystallogr. F: Struct. Biol. Cryst. Commun.* **65**, 499–503.
9. Larson S.B., Day J., Barba dela Rosa A.P., Keen N.T., McPherson A. 2003. First crystallographic structure of a xylanase from glycoside hydrolase family 5: Implications for catalysis. *Biochemistry*. **42**, 8411–8422.
10. Gallardo O., Fernandez M.F., Valls C., Valenzuela S.V., et al. 2010. Characterization of a family GH5 xylanase with activity on neutral oligosaccharides and evaluation as a pulp bleaching aid. *Appl. Environ. Microbiol.* **76**, 6290–6394.
11. Valenzuela S.V., Diaz P., Pastor F.I. 2012. Modular glucuronoxylan-specific xylanase with a Family CBM35 carbohydrate-binding module. *Appl. Environ. Microbiol.* **78**, 3923–3931.
12. Keppler B.D., Showalter A.M. 2010. IRX14 and IRX14-LIKE, two glycosyl transferases involved in glu-

curonoxylan biosynthesis and drought tolerance in *Arabidopsis*. *Mol. Plant*. **5**, 834–841.
13. Sakkaa M., Tachinoa S., Katsuzakia H., van Dyk J.S., Pletschke B.I., Kimuraa T., Sakkaa K. 2012. Characterization of Xyn30A and Axh43A of *Bacillus licheniformis* SVD1 identified by its genomic analysis. *Enzyme Microb. Tech*. **51**, 193–199.
14. Suzuki T., Ibata K., Hatsu M., Takamizawa K., Kawai K. 1997. Cloning and expression of a 58 kDa xylanase VI gene (xynD) of *Aeromonas caviae* ME-l in *Escherichia coli* which is not categorized as a family F or family G Xylanase. *J. Ferment. Bioeng*. **84**, 86–89.
15. Biely P., Vrsanska M., Tenkanen M., Kluepfel D. 1997. Endo-β-1,4-xylanase families: Differences in catalytic properties. *J. Biotechnol*. **57**, 151–166.
16. Collins T., Gerday C., Feller G. 2005. Xylanases, xylanase families, and extremophilic xylanases. *FEMS Microbiol. Rev*. **29**, 3–23.
17. Ahmed S., Charan R., Ghosh A., Goyal A. 2012. Comparative modeling and ligand binding site prediction of a family 43 glycoside hydrolase from *Clostridium thermocellum*. *J. Proteins Proteomics*. **3**, 31–38.
18. Zaide G., Shallom D., Shulami S., Zolotnitsky G., et al. 2001. Biochemical characterization and identification of catalytic residues in α-glucuronidase from *Bacillus stearothermophilus* T-6. *Eur. J. Biochem*. **268**, 3006–3016.
19. Preston J.F., Hurlbert J.C., Rice J.D., Ragunathan A., St. John F.J. 2003. Microbial strategies for the depolymerization of glucuronoxylan: Leads to biotechnological applications of endoxylanases. In: *Applications of Enzymes to Lignocellulosics*. Eds. Mansfield S.D., Saddler J.N. Washington, DC.: Am. Chemical Soc., pp. 191–210.
20. Qiua Z., Shia P., Luoa H., Baia Y., et al. 2010. A xylanase with broad pH and temperature adaptability from *Streptomyces megasporus* DSM 41476, and its potential application in brewing industry. *Enzyme Microb. Technol*. **46**, 506–512.
21. Subramaniyan S., Prema P. 2002. Biotechnology of microbial xylanases: Enzymology, molecular biology, and application. *Crit. Rev. Biotechnol*. **22**, 33–46.
22. Bajpai P. 1999. Application of enzymes in the pulp and paper industry. *Biotechnol. Prog*. **15**, 147–157.
23. Jiang Z., Li X., Yang S., Li L., Tan S. 2005. Improvement of the bread making quality of wheat flour by the hyperthermophilic xylanase B from *Thermotoga maritima*. *Food Res. Int*. **38**, 37–43.
24. 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.
25. Zdobnov E.M., Apweiler R. 2001. InterProScan: An integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. **17**, 847–848.
26. Nielsen H., Engelbrecht J., Brunak S., Heijne G.V. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*. **10**, 1–6.
27. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.J. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **24**, 4876–4882.
28. Sali A., Potterton L., Feng Y., Herman V., Martin K. 1995. Evaluation of comparative protein modeling by Modeller. *Protein Struct. Funct. Genet*. **23**, 318–326.
29. Fiser A., Do R.K., Sali A. 2000. Modeling of loops in protein structures. *Prot. Sci*. **9**, 1753–1773.
30. Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structure. *J. Appl. Cryst*. **26**, 283–291.
31. Bowie J.U., Luthy R., Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. **253**, 164–170.
32. 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.
33. Sippl M.J. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins*. **17**, 355–362.
34. Hess B. 2008. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput*. **4**, 116–122.
35. Goodsell D.S., Olson A.J. 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins: Struct. Funct. Genet*. **8**, 195–202.
36. Morris G.M., Goodsell D.S., Halliday R.S., Huey R., et al. 1998. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J. Comput. Chem*. **19**, 1639–1662.
37. Luthy R., Bowie J.U., Eisenberg D. 1992. Assessment of protein models with three-dimensional profiles. *Nature*. **356**, 83–85.
38. Pollet A., Delcour J.A., Courtin C.M. 2010. Structure determinants of the substrate specificities of xylanases from different glycoside hydrolase families. *Crit. Rev. Biotechnol*. **30**, 176–191.
39. Henrissat B., Callebaut I., Fabrega S., Lehn P., et al. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. U. S. A*. **92**, 7090–7094.
40. Nishitani K., Nevins D.J. 1991. Glucuronoxylan xylanohydrolase. A unique xylanase with the requirement for appendant glucuronosyl units. *J. Biol. Chem*. **266**, 6539–6543.
41. Boraston A.B., Bolam D.N., Gilbert H.J., Davies G.J. 2004. Carbohydrate binding modules: Fine-tuning polysaccharide recognition. *Biochem. J*. **382**, 769–781.
42. St. John F.J., Hurlbert J.C., Rice J.D., Preston J.F., Pozharski E. 2011. Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase. *J. Mol. Biol*. **407**, 92–109.
43. Holm L., Rosenstrom P. 2010. Dali server: Conservation mapping in 3D. *Nucleic Acids Res*. **38**, 545–549.
44. Davies G., Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. *Curr. Biol*. **3**, 853–859.
45. Rye C.S., Withers S.G. 2000. Glycosidase mechanisms. *Curr. Opin. Chem. Biol*. **4**, 573–580.
46. McCarter J., Withers S.G. 1994. Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol*. **4**, 885–892.
47. Leggio L.L., Jenkins J., Harris G.W., Pickersgill R.W. 2000. X-ray crystallographic study of xylopentaose binding to *Pseudomonas fluorescens* xylanase A. *Proteins*. **41**, 362–373.
48. Correia M.A., Mazumder K., Bras J.L., Firbank S.J., et al. 2011. Structure and function of an arabinoxylan-specific xylanase. *J. Biol. Chem*. **286**, 22510–22520.

Archives of Biochemistry and Biophysics 575 (2015) 8–21

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

The family 6 Carbohydrate Binding Module (*Ct*CBM6) of glucuronoxylanase (*Ct*XynGH30) of *Clostridium thermocellum* binds decorated and undecorated xylans through cleft A

Anil Kumar Verma [a], Pedro Bule [b], Teresa Ribeiro [b], Joana L.A. Brás [b], Joyeeta Mukherjee [c], Munishwar N. Gupta [c], Carlos M.G.A. Fontes [b], Arun Goyal [a,*]

[a] Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India
[b] CIISA-Faculdade de Medicina Veterinária, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal
[c] Department of Chemistry, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India

ARTICLE INFO

Article history:
Received 17 January 2015
and in revised form 24 February 2015
Available online 7 April 2015

Keywords:
Clostridium thermocellum
Binding cleft
Ligand binding affinity and isothermal titration calorimetry

ABSTRACT

*Ct*CBM6 of glucuronoxylan-xylanohydrolase (*Ct*XynGH30) from *Clostridium thermocellum* was cloned, expressed and purified as a soluble ~14 kDa protein. Quantitative binding analysis with soluble polysaccharides by affinity electrophoresis and ITC revealed that *Ct*CBM6 displays similar affinity towards decorated and undecorated xylans by binding wheat- and rye-arabinoxylans, beechwood-, birchwood- and oatspelt-xylan. Protein melting studies confirmed thermostable nature of *Ct*CBM6 and that Ca^{2+} ions did not affect its structure stability and binding affinity significantly. The *Ct*CBM6 structure was modeled and refined and CD spectrum displayed 44% β-strands supporting the predicted structure. *Ct*CBM6 displays a jelly roll β-sandwich fold presenting two potential carbohydrate binding clefts, A and B. The cleft A, is located between two loops connecting β4–β5 and β8–β9 strands. Tyr28 and Phe84 present on these loops make a planar hydrophobic binding surface to accommodate sugar ring of ligand. The cleft B, is located on concave surface of β-sandwich fold. Tyr34 and Tyr104 make a planar hydrophobic platform, which may be inaccessible to ligand due to hindrance by Pro68. Site-directed mutagenesis revealed Tyr28 and Phe84 in cleft A, playing a major role in ligand binding. The results suggest that *Ct*CBM6 interacts with carbohydrates through cleft A, which recognizes equally well both decorated and un-decorated xylans.

Introduction

Plant polysaccharides constitute one of the major renewable sources of carbon and energy on earth albeit being recalcitrant to degradation. Microbes have evolved a remarkable complex enzyme machinery to utilize these abundant carbon sources. Plant cell wall degrading enzymes are modular in nature and comprise catalytic module(s) appended to one or more non-catalytic Carbohydrate Binding Modules (CBMs[1]) [1–3]. The function of CBM is to increase the effective concentration of enzymes on the surface of the recalcitrant and sometimes insoluble substrates, thus potentiating catalysis [4]. Moreover, CBMs may also be involved in substrate targeting and disruption of insoluble crystalline plant polysaccharides [5]. The ability of CBMs to fold and express independently of the associated module makes them attractive candidates for numerous biotechnological applications, such as expression and purification tags [6]. For example, CBMs are used in immobilized affinity ligand technology for purification of biomolecules [7,8]. Highly thermostable CBMs are effective in separating different carbohydrate analyte (cello- and xylo-oligomers) by retaining them at higher temperatures for better separation of oligosaccharides [9]. CBMs are also used as biosensors, such as the CBM of *Cellulomonas fimi* (*C. fimi*) that was used for glucose sensing in bioreactors [10,11]. Whole-cell immobilization by cellulosic material was first demonstrated when an *Escherichia coli* (*E. coli*) surface anchored CBM, derived from *C. fimi*, was attached to cellulose [12,13]. The cell immobilization technique using CBM, was also explored for bioremediation purpose such as detoxification of nerve gas by immobilizing the organo-phosphorus

* Corresponding author. Fax: +91 (361) 2690762.
E-mail address: arungoyl@iitg.ernet.in (A. Goyal).

[1] *Abbreviations used:* CBMs, Carbohydrate Binding Modules; CMC, carboxy methylcellulose; CEC, carboxy ethylcellulose; AE, affinity gel electrophoresis; WAXI, wheat arabinoxylan insoluble; PDB, Protein Data Bank; DOPE, discrete optimized protein energy; SPC, simple point charge; SAVES, structure analysis and verification server; LINCS, linear constraint solver; ADT, AutoDock Tools; CD, Circular Dichroism; MD, molecular dynamics.

http://dx.doi.org/10.1016/j.abb.2015.03.026
0003-9861/
TH-1369_09610621

hydrolase as well as removal of heavy metal contamination from polluted air [14,15].

Based on sequence similarities, CBMs are grouped into 71 families as listed in the CAZy database (http://www.cazy.org) till date. Presently 946 bacterial, 74 archaea and 14 eukaryotic CBMs have been reported in family 6 (http://www.cazy.org/CBM6.html). Interestingly, family 6 CBMs display a diverse binding specificity and thus different CBM6s were shown to recognize different polysaccharides *viz.* β-agarose having 3,6-anhydro-L-galactoses and D-galactoses alternatively linked by α-(1,3) and β-(1,4) linkages [16], xylan (β-1,4-linked xylose) [17], cellulose (β-1,4-linked glucose homopolymer) [18,19], laminarin (β-1,3-linked glucose homopolymer), lichenan (β-1,4- and β-1,3-mixed linked β-glucans) [20,21] and oat spelt xylan (β-1,4-linked xylose main chain with β-1,2 or β-1,3-linked arabinose side chain) [22,23]. CBM6 not only exhibited affinity on soluble ligands but also showed capacity to bind insoluble polysaccharides thereby potentiating the capacity of cognate catalytic modules to depolymerize insoluble substrates [24,25]. Previous structural studies revealed that family 6 CBMs display two potential binding sites termed as cleft A and cleft B. Cleft A is formed in the loops that connect the inner and the outer β-sheets of the protein while cleft B is located on the concave surface of the β sandwich [26].

Protein–carbohydrate interactions are vital for numerous biological processes, including cell signaling, host–pathogen interaction and the attachment of microbial enzymes to plant cell wall polysaccharides. Thus, CBMs are an excellent model to reveal the mechanisms of protein–carbohydrate recognition [27–29]. In the present study, we show that *Ct*CBM6 allows *Ct*XynGH30 to recognize decorated and undecorated xylans. *Ct*XynGH30 contains an N-terminal catalytic module *Ct*Xyn30A and is part of *Clostridium thermocellum* cellulosome. As it is characteristic of other CBM6s, the structure of *Ct*CBM6 displays a β-sandwich fold and contains two putative carbohydrate interacting surfaces, defined as clefts A and B. Mutagenesis studies confirm that recognition of the backbone of xylan molecules is mediated by the side chains of two aromatic residues located in cleft A.

Materials and methods

Bacterial strains and plasmids

E. coli DH5α cells were used for cloning and *E. coli* BL21 (DE3) cells were used for expression. The plasmid used for cloning and expression were pGEM-T Easy and pET-28a(+), respectively. *E. coli* BL21 (DE3) cells and pET-28a(+) vector were procured from Novagen (Madison, USA) while pGEM-T Easy vector was purchased from Promega Co. (Madison, WI, USA).

Fine chemicals and polysaccharides

*Pfu*DNA polymerase was procured from Stratagene, USA. The dNTPs and $MgCl_2$ were purchased from Bioline, UK. Restriction enzymes *Nhe*I and *Xho*I were purchased from Fermentas, Germany. Natural polysaccharides like rye arabinoxylan, wheat arabinoxylan (soluble and insoluble), arabinogalactan, sugar beet arabinan, rhamnogalactouronan, konjac glucomannan and locus bean galactomannan were procured from Megazyme International, Ireland. Oat spelt xylan, birchwood xylan, beechwood xylan, 4-O-methyl glucuronoxylan, barley β-D-glucan, carboxy methylcellulose (CMC), carboxy ethylcellulose (CEC), curdlan, pullulan, pectic galactan (from apple, citrus and lupin), xyloglucan and galactan (Lupin) were purchased from Sigma–Aldrich, USA.

Retrieval of sequence and defining boundaries of CtXynGH30 and its truncated modules

The protein sequence of *Ct*XynGH30 was retrieved from the NCBI protein sequence database with accession number ABN54208 (http://www.ncbi.nlm.nih.gov/protein) and uniprot ID A3DJS9. Domain boundaries of full length modular protein designated as *Ct*XynGH30 were defined by analyzing the result of PSI-Blast (http://blast.ncbi.nlm.nih.gov/) and InterProScan (http://www.ebi.ac.uk/interpro/). N-terminal signal peptide was detected by signalP (http://www.cbs.dtu.dk/services/SignalP-4.0/). A final adjustment in domain borders was carried out by sequence alignment using the Clustal X program [30].

Amplification and cloning of gene encoding CtCBM6

Oligonucleotide primers containing *Nhe*I and *Xho*I restriction sites were designed. The primers used for the amplification of *Ct*CBM6 were, forward 5′-CTCT**GCTAGC**AAGATAGAATGCGAAGAA-3′ and reverse primer 5′-CACA**CTCGAG**TTACTCCTGTTTTGACTC-3′. The 50 μl PCR reaction mixture contained $MgCl_2$ (2.5 mM), dNTPs (1.6 mM), primers (0.45 μM each), 1.0 μl of Taq DNA polymerase (1 Unit μl^{-1}) and 0.5 μl (7.7 ng) of genomic DNA of *C. thermocellum.* The PCR amplification conditions used were denaturation at 94 °C for 5 min followed by 30 cycles of (i) denaturation at 94 °C for 30 s, (ii) annealing at 55 °C for 60 s (iii) extension at 72 °C for 1 min and final extension, 72 °C for 5 min. The PCR amplified product was separated on 1.0% agarose gel and extracted from the gel using QIAquick gel extraction kit (Qiagen, USA). The purified PCR DNA was ligated to pGEM-Teasy vector. The positive clones were screened by blue–white colony selection followed by restriction digestion [31]. The pGEM-Teasy cloned DNA and the pET-28a(+) expression vector were digested with *Nhe*I and *Xho*I restriction enzymes. A ligation reaction was set up containing the insert releases after the restriction digestion, the digested pET-28a(+) vector and T4 DNA ligase and incubated at 16 °C for overnight. The ligation product was transformed using *E. coli* DH5α competent cells by heat-shock method. The transformed cells were grown on LB-agar plates supplemented with kanamycin (50 μg ml^{-1}) by incubation at 37 °C for overnight. Positive clones were confirmed by restriction digestion analysis of the isolated recombinant plasmid DNA from the *E. coli* DH5α cells by QIAprep spin miniprep kit (Qiagen, USA).

Expression and purification of recombinant CtCBM6

E. coli BL-21 (DE3) cells were used for of transformation of recombinant plasmid containing the gene encoding *Ct*CBM6. The cells harboring the recombinant plasmid were grown in 100 ml LB medium containing kanamycin (50 μg ml^{-1}) by incubating at 37 °C with shaking at 180 rpm till the cell growth reached mid-exponential phase having an absorbance at 550 nm (A_{550}) = 0.6. The cells were induced with isopropyl-1-thio-β-D-galactopyranoside at 1.0 mM final concentration and further grown at 24 °C and 180 rpm for 16 h for expression of recombinant *Ct*CBM6 [23]. The cells were harvested by centrifugation at 10,000*g* and 4 °C for 15 min. The resulting cell pellet was re-suspended in 5 ml lysis buffer (20 mM phosphate buffer, pH 7.0, 300 mM NaCl, 0.2 mg/ml lysozyme and 1.0 mM phenylmethylsulfonyl fluoride, PMSF). Then the cells were sonicated (Sonics, Vibra cell) on ice for 10 min (8 s on/16 s off pulse; 33% amplitude) and centrifuged at 17,000*g* at 4 °C for 45 min to get the cell free extract. The recombinant protein from the cell free extract was purified by immobilized metal ion affinity chromatography (IMAC) using Sepharose columns (GE Healthcare, 1 ml HiTrap

chelating) and further dialyzed against 20 mM sodium phosphate buffer pH 7.0. The purity and molecular mass of recombinant proteins was verified by 14% SDS–PAGE [32]. The concentration of the purified protein was determined by measuring the absorbance at 280 nm (A_{280}) using the molar absorption coefficient, 14565 M^{-1} cm^{-1} for *Ct*CBM6 [33].

Binding analysis of CtCBM6 with soluble polysaccharides

Affinity gel electrophoresis (AE) using Native–PAGE was carried out in order to determine the equilibrium association constant (K_a) of *Ct*CBM6 against soluble polysaccharides following the method described earlier [34,35]. Native–PAGE using 6.5% gel containing soluble polysaccharides (birchwood, beechwood, 4-O-methyl glucuronoxylan, oat spelt xylan, wheat arabinoxylan or rye arabinoxylan) at concentrations varying from 0.008% to 0.2% (w v^{-1}), were added before gel polymerization. A 6.5% gel without any soluble polysaccharide was also run as a control. The electrophoresis was carried out at constant current of 2 mA per lane and at 4 °C using a Mini PROTEAN Tetra pack (Bio-Rad, USA). Purified *Ct*CBM6 (5.0 μg) was run on native gels with or without soluble polysaccharides. Bovine serum albumin, BSA (10 μg) was also run as a reference to check for any non-specific binding interactions. After staining the gel with coomassie brilliant blue, the migration of *Ct*CBM6 was measured directly on the gel from the edge of the loading well to the middle of the stained protein spot and the distances were assigned as migration without polysaccharide (R_0, mm), migration in the presence of polysaccharide (R, mm) and distance traveled by dye front (R_d, mm). The relative mobility (r) was calculated as

$$r = R/R_d \quad \text{or} \quad r = R_0/R_d$$

The equilibrium binding constant (K_a) was determined by plotting $1/r$ (y-axis) against the concentration of polysaccharide on x-axis (%, w v^{-1}). The regression line of the plot was extrapolated to the x-axis and the point where the regression line intersected the x-axis gave equilibrium dissociation constant (K_d). The reciprocal of K_d gives K_a as reported earlier by Tomme et al. [34].

$$K_d = 1/K_a$$

The effect of Ca^{2+} ions on ligand binding affinity of *Ct*CBM6 was also studied by incorporating 5 mM Ca^{2+} ions and polysaccharide into the gel prior to its polymerization. The gels were analyzed by measuring the retardation in *Ct*CBM6 migration with respect to the control.

Polysaccharides binding analysis of CtCBM6 by Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was employed to study the thermodynamic parameters *viz.* Gibb's free energy (ΔG), enthalpy (ΔH), entropy (ΔS) and stoichiometry (N_0) of *Ct*CBM6 upon polysaccharide binding. ITC experiments were carried out essentially as described previously [36,37], except that the titrations were carried out at 55 °C, as the *Ct*CBM6 is from *C. thermocellum.* The protein was in 50 mM Na–HEPES buffer, pH 7.5, containing 2 mM $CaCl_2$. During titration *Ct*CBM6 (60 μM) was stirred at 300 rpm in the reaction cell, which was injected with 28 successive 2 μl aliquots of ligands at (1.5 mM) 180 s intervals. Integrated heat effects, after correction for the heats of dilution, were analyzed by non-linear regression using a single site-binding model (Microcal ORIGIN, Version 5.0; Microcal Software). The fitted data yielded the association constant (K_a) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated by using the standard thermodynamic equation,

Binding affinity of CtCBM6 with insoluble polysaccharides

Qualitative binding analysis of *Ct*CBM6 with insoluble polysaccharides was performed as described by Boraston et al. [38]. 20 μg of purified *Ct*CBM6 was incubated with 1 mg insoluble wheat arabinoxylan (WAXI) or 1 mg avicel in 200 μl, 20 mM Tris–HCl, pH 7.0 with gentle shaking at 4 °C for 2 h. The reaction mixture was centrifuged at 13,000*g* at 4 °C for 5 min. The polysaccharide pellet was washed 3 times with 200 μl of buffer (20.0 mM Tris–HCl, pH 7.0). Finally, the polysaccharide pellet was re-suspended in 200 μl of 10% (w v^{-1}) SDS containing 10% (w v^{-1}) β-mercaptoethanol and boiled for 10 min. The supernatant containing the unbound protein and the resuspended washed pellet containing bound protein (each 30 μl) were analyzed by SDS–PAGE using 14% gel. The quantitative assessment by adsorption isotherm was carried out by the method of Gilkes et al. [39]. The reaction condition was kept same as describe above in a 200 μl reaction volume but with varying concentrations (1–25 μM) of *Ct*CBM6 along with 1 mg ml^{-1} insoluble polysaccharide. The reaction was set up in triplicate with the respective controls, in parallel without the polysaccharide. The supernatant represents the unbound or free protein [F], which was measured at A_{280}. The concentration of bound protein [B] for a particular concentration was calculated by subtracting the amount of free protein [F] from the initial protein concentration.

Based on this, equilibrium association constant (K_a) was obtained from the depletion isotherms (plot of [B] versus [F]) after fitting (nonlinear regression) of the raw data to an equilibrium binding type or adsorption model [39].

$$K_a = [B]/N[F]$$

where, N is the concentration of available binding sites on the substrate (the number of polysaccharide lattice units) [moles g^{-1} of CBM], [B] is the concentration of bound CBM (mol g^{-1} of substrate), [F] is the molar concentration of free CBM and K_a is the equilibrium association constant [M^{-1}].

The relative equilibrium association constant, K_r (l g^{-1} of substrate) was calculated as,

$$K_r = K_a N_o$$

K_a = is the equilibrium association constant [M^{-1}]
N_o = is the concentration of available binding sites on the substrate (the number of polysaccharide lattice units) [moles g^{-1} of CBM]

N_o were calculated from a non-linear regression plot against μ mole of protein bound per gram of wheat arabinoxylan insoluble (μ mole g^{-1}) versus the free protein *Ct*CBM6 (μM) [39]. Data were analyzed by GraphPad (Prism version 5) software using non-linear regression using one binding site equation.

Thermal denaturation of CtCBM6

The protein melting curve for *Ct*CBM6 was generated by incubating 20 μg ml^{-1} final concentration in 20 mM Tris–HCl buffer, pH 7.0 at temperature varying from 40 to 100 °C using a peltier temperature controller attached to a UV–visible spectrophotometer (Varian, Carry 100-Bio). UV spectrum was recorded by measuring the absorbance at 280 nm (tryptophan absorption maximum) following the method, described by Dvortsov et al [40]. The temperature of the cuvette block was raised to 3 °C/min and data were collected after every 3 °C rise in the temperature. Measurements were repeated three times to achieve a 5% margin of error. The melting point of unfolded *Ct*CBM6 was determined by plotting absorption units versus temperature [41]. The effect of Ca^{2+} ions

on *Ct*CBM6 was also analyzed by generating the melting curve in the presence of 5 mM Ca^{2+} ion.

Homology modeling, model refinement and quality assessment

The protein sequence search for a closest structural homolog template to model *Ct*CBM6 was achieved by 'BLASTp' program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) from the Protein Data Bank (PDB) database. 3-D modeling was accomplished by Modeller9v8 using the template of crystal structure of family 6 CBM from *C. thermocellum* (PDB id: 1GMM) [42]. Query sequence was aligned with the template sequence of known structure which acted as an input for the Modeller. Initially, twenty independent models were generated after optimization of molecular probable density functions (MOLPDF score). The modeled structure was further subjected to loop refinement by loop optimization and after each cycle of loop refinement, a discrete optimized protein energy (DOPE) score was generated. This process was repeated in an iterative fashion until DOPE score attained a negative value [43]. The best model having reasonable DOPE and MOLPDF scores was chosen for further refinement. Energy minimization steps were performed by a steepest descent algorithm with GROMOS96 43a1 force field and simple point charge (SPC) water model using GROMACS4.0.7 package (http://www.gromacs.org/). A simulation box was created by introducing the water molecules as a solvent system for carrying out *Ct*CBM6 stability study. Overall charge on the protein molecule was neutralized by adding counter-ions (here 3 Na^+ ions) and the same number of water molecules which were at least 3.50 Å apart was replaced from the protein surface. The final energy minimized model was validated through various parameters on the structure analysis and verification server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES/) such as Ramachandran plot [44], VERIFY3D [45] and ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) [46,47].

Molecular dynamic analysis of CtCBM6 model

Molecular dynamics stimulations of energy-minimized *Ct*CBM6 model was performed in the defined cubic box of single point charge (SPC) water molecules. The net charge of the system was neutralized with counter ions (Na^+ ions). The whole system was equilibrated for 500 ps by NVT ensemble by restraining the solute atoms through a harmonic force constant of 1000 kJ nm^{-2}. Then the system was equilibrated for 500 ps by NPT ensemble twice, first with restraints followed by without restraints. Production run was carried out for 20 ns with NPT ensemble using 2 fs of integration time. The linear constraint solver (LINCS) algorithm was used to constrain the bonds involving hydrogen atoms [48]. Conformations of *Ct*CBM6 model at regular intervals throughout the production run were analyzed as time-dependent function to check whether the *Ct*CBM6 model is energetically stable in the solvent system.

Binding site prediction

Structural superimposition and multiple sequence alignment of *Ct*CBM6 were performed with known structure of CBMs of family 6 to identify the key residues and their role in substrate binding. Structural superimposition was done using PyMOL [49]. Related protein sequence was retrieved from PDB (http://www.rcsb.org/pdb/), saved in FASTA format. Alignment was performed using the Clustal X program and the final alignment was generated by ESPript (http://espript.ibcp.fr).

Molecular docking of modeled CtCBM6

Molecular docking studies were carried out using Autodock 4.2.1. Autodock uses a novel and robust docking method by implementing a new scoring function that estimates the free energy change during ligand binding [48]. Required input files were prepared by assigning gasteiger charges, merging non-polar hydrogens and save in PDBQT format using AutoDock Tools (ADT) 1.5.4. Grid box was set to 40, 52, 40 (x, y, z) with 0.375 spacing covering entire binding site cavity. Lamarckian Genetic Algorithm was chosen for conformational search and numbers of evaluation were set to be 30 [50]. Best docked conformation from the largest cluster having a minimum lowest free energy of binding (ΔG), were saved. Results were analyzed using PyMOL (0.99) for possible polar and hydrophobic interaction. Representation of carbohydrate–protein complex was generated via Ligplot (http://www.ebi.ac.uk/pdbsum/) for a better understanding of protein ligand interaction.

Secondary structure analysis of CtCBM6

The secondary structure elements of *Ct*CBM6 were determined by PSIPRED on web server (http://bioinf.cs.ucl.ac.uk/psipred/). It uses two-stage neural network approach based on the position specific, scoring matrices generated by PSI-BLAST [51]. The secondary structure of *Ct*CBM6 was also confirmed by Circular Dichroism (CD). The CD spectrum analysis was carried out at 25 °C on a spectropolarimeter (Jasco Corporation, Tokyo, JASCO J-810), equipped with a peltier system for temperature control. The protein concentration (0.09 mg ml^{-1}) in 20 mM Tris–HCl (pH 7.4) was used. The average of six scans was carried out from 195–250 nm wavelengths, with a scan speed of 50 nm min^{-1}, using a bandwidth of 1 nm and resolution of 1 nm. The far-UV spectrum was plotted in terms of the mean residue ellipticity (MRE, expressed as deg cm^2 $dmol^{-1}$) as a function of wavelength, calculated by the procedure described earlier [52]. The percentage of secondary structure, content was calculated using web based K2d3 program (http://k2d3.ogic.ca/). It uses different algorithms or neural network approach to estimate the percentage of secondary structure by comparing the similarity between the CD spectrum of unknown protein with the database of known CD spectra (reference set) as described earlier [53].

Construction of CtCBM6 mutants by site-directed mutagenesis

Site-directed mutagenesis was carried out using the PCR-based NZY mutagenesis kit (NZY Tech Ltd., Portugal) according to the manufacturer's instructions, using plasmid pCBM6 as template. The sequences of the primers used for generating of these mutants are displayed in Table 1. The mutation in the plasmid DNA was assessed by sequencing to ensure that only the appropriate mutations were incorporated.

Results and discussion

CtCBM6 is truncated derivative of modular enzyme CtXynGH30

InterProScan and Blast analysis using the full length primary sequence of *Ct*XynGH30 from *C. thermocellum* confirmed the presence of different protein modules. Thus, *Ct*XynGH30 contains an N-terminal catalytic module from family 30 glycoside hydrolase (GH30), named *Ct*Xyn30A, an internal family 6 carbohydrate binding module (*Ct*CBM6), which is connected via a flexible linker of 6 amino acid residues to the catalytic domain, and a C-terminal dockerin that anchors the enzyme into *C. thermocellum* cellulosome.

Table 1
Oligonucleotide primers used for generating site-specific variant of *Ct*CBM6.

Construct	Primer sequence
Mutation Tyr-24-Ala	Fwd:5′-GGAAGCGGTCTCGGATATATCGAAAACGGCAAC-3′ (WT)
	Fwd:5′-GGAAGCGGTCTCGGAGCTATCGAAAACGGCAAC-3′
	Rev: 5′-GTTGCCGTTTTCGATAGCTCCGAGACCGCTTCC-3′
Mutation Phe-84-Ala	Fwd:5′-GCTGCAACCGGAGGCTTTAATGCCTATGAGG-3′ (WT)
	Fwd:5′-GCTGCAACCGGAGGCGCTAATGCCTATGAGG -3′
	Rev: 5′-CCTCATAGGCATTAGCGCCTCCGGTTGCAGC -3′

Cloning, expression and purification of recombinant CtCBM6

The gene encoding *Ct*CBM6 was amplified using oligonucleotide primers containing desired restriction sites. *Ct*CBM6 was hyper-expressed using *E. coli* (BL21) and purified as displayed in Fig. 1. Recombinant *Ct*CBM6 displayed molecular mass of approximately, 14 kDa. The molecular size of *Ct*CBM6 predicted by sequence analysis using pepstats (http://emboss.bioinformatics.nl/) server was 13.8 kDa including the Histidine tag (His_6 tag). However, this did not include a few amino acids from N-terminal which are present in the protein from to the pET-28a(+) vector. The theoretical molecular size of *Ct*CBM6 was in the close agreement with the size displayed by SDS–PAGE.

Qualitative binding analysis of CtCBM6 with soluble polysaccharides by affinity gel electrophoresis

Affinity electrophoresis (AE) of *Ct*CBM6 with 24 different soluble natural polysaccharides clearly suggested that it predominantly interacts with ligands containing a xylan backbone (β-1,4-xylopyranose) (Fig. 1S). *Ct*CBM6 showed negligible affinity towards celluloses, pectins and other polysaccharides. The screening results of binding affinity are summarized in Table 2. The ligand specificity of *Ct*CBM6 reflects the substrate specificity of the cognate *Ct*Xyn30A catalytic module [unpublished data]. After elucidating the specificity of *Ct*CBM6 towards xylan main chain backbone, the equilibrium binding constant and other thermodynamics parameter were calculated for respective natural polysaccharides.

Quantitative binding analysis of CtCBM6 by affinity gel electrophoresis

Quantitative assessment of the affinity towards soluble polysaccharides in terms of the equilibrium association constant (K_a) was carried out by AE. The increase in retardation of *Ct*CBM6 migration with an increase in the concentration of soluble polysaccharides was observed (Fig. 2(A, i–F, i)) and the plot of $1/r$ versus polysaccharide concentration for determination of K_a by regression analysis for each polysaccharide is shown in Fig. 2(A, ii–F, ii). The K_a values indicated that *Ct*CBM6 shows higher affinity towards wheat- and rye-arabinoxylans (highly substituted) than poorly substituted xylan such as beechwood-, birchwood- and oat spelt-xylan (Table 3). The presence of Ca^{2+} ions (5 mM) only slightly reduced the migration of *Ct*CBM6 henceforth it helps to increase insignificantly the binding affinity towards ligands (Fig. 2S)

Determination of binding constant and thermodynamic parameters of CtCBM6 by ITC

To investigate, in more detail, the ligand specificity of *Ct*CBM6, the capacity of the protein to bind to xylan constituents was assessed by ITC. The data, reported in Table 3 with example titration-displayed in Fig. 3 show that *Ct*CBM6 displays high binding affinity with different xylans through an exothermic mechanism. Thus, the thermodynamic data show that the binding of *Ct*CBM6 to all its ligands is enthalpically driven, while the change in entropy makes a negative contribution to overall affinity, as observed for the majority of CBMs studied to date. The equilibrium association constant (K_a) determined by ITC at 55 °C showed significantly higher values than those obtained by affinity gel electrophoresis at 4 °C for all polysaccharides (Table 3). The difference in K_a values between AE and ITC might be due to the fact that *Ct*CBM6 is more functional at higher temperatures. The limitation of AE is that, it cannot be performed at higher temperatures; nonetheless both the methods exhibited similar binding patterns displaying a relatively higher affinity against arabinoxylans (Table 3).

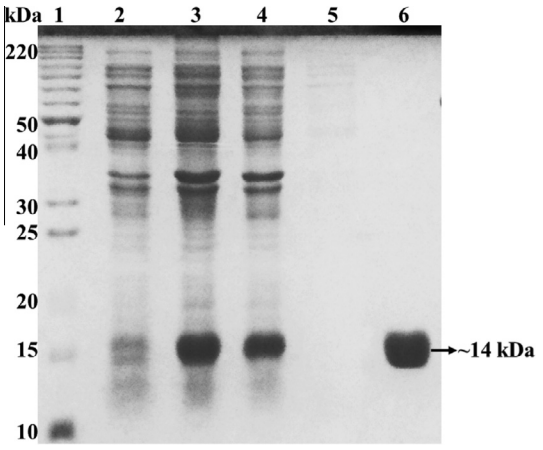


Fig. 1. SDS–PAGE (14%) showing over-expression and purification profile of recombinant *Ct*CBM6, Lane 1: protein marker, Lane 2: uninduced *E. coli* BL-21 cells, Lane 3: induced *E. coli* BL-21 cells, Lane 4: cell pellet, Lane 5: column last wash and Lane 6: purified protein.

Table 2
Binding affinity of *Ct*CBM6 towards various polysaccharides.

Ligand[a]	Binding[b]
Xylans	
Beechwoodxylan	++
Birchwood xylan	++
Glucuronoxylan	++
Oat-spelt Xylan	+++
Rye arabinoxylan	+++
Wheat arabinoxylan	+++
Cellulosic substrates*	–
Pectins**	–
Hemicelluloses***	–

* Barley β-glucan, carboxymethyl cellulose (CMC), curdlan, hydroxyl ethyl cellulose (HEC), lichenan (icelandic moss) and starch.
** Pecticgalactan (lupin), pecticgalactan (Potato), pectin from apple, polygalacturonic acid (PGA), rhamnogalacturonan (Potato) and rhamnogalacturonan (soyabean).
*** Arabinan (sugar beet), galactomannan (carob), galactomannan (guar), glucomannan (Konjac), xyloglucan (tamarind).
[a] Ligand concentration, 1 mg ml^{-1}.
[b] Symbols used: –, no detectable binding; ++, significant binding; +++, strong binding.

Binding of CtCBM6 to insoluble polysaccharide

Qualitative binding affinity experiments with insoluble ligands clearly indicated that *Ct*CBM6 binds insoluble wheat arabinoxylan, but under the same conditions, no significant binding was observed with insoluble avicel (Fig. 4A). The quantitative analysis

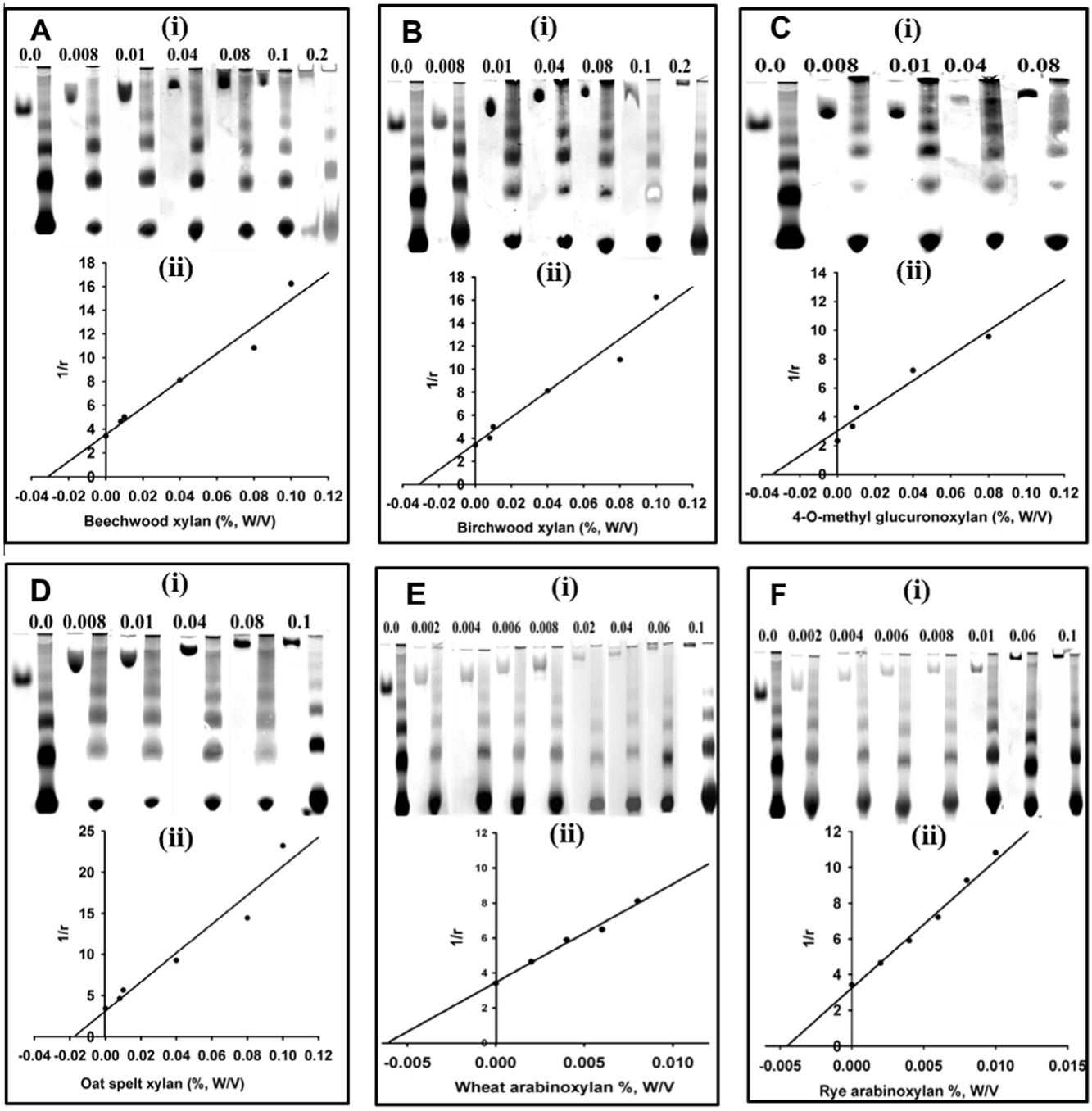


Fig. 2. Quantitative binding analysis by affinity gel electrophoresis using Native–PAGE (6.5%, w/v) gels run at 4 °C with different xylan based substrates (A, i & ii) beechwood xylan, (B, i & ii) birchwood xylan, (C, i & ii) 4-O-methylglucuronoxylan, (D, i & ii) oat spelt xylan, (E, i & ii) wheat arabinoxylan and (F, i & ii) rye arabinoxylan. Bovine Serum Albumin (BSA) was used as a reference protein.

Table 3
Quantitative binding analysis of *Ct*CBM6 with soluble polysaccharides by AE and ITC.

Ligand	$K_a \times 10^4$ (M^{-1}) AE (4 °C)	$K_a \times 10^4$ (M^{-1}) ITC (55 °C)	ΔG kcal/mol	ΔH kcal/mol	$T\Delta S$ kcal/mol	N_0
Beechwoodxylan	0.53	3.33 ± 0.33	−6.8	−10.4 ± 0.58	−3.6	1.18 ± 0.05
Birchwood xylan	0.52	3.22 ± 0.37	−6.8	−7.9 ± 0.48	−1.1	1.30 ± 0.06
Oat spelt xylan	0.61	3.20 ± 0.4	−6.7	−2.9 ± 0.17	3.8	1.50 ± 0.06
4-O-methyl Glucuronoxylan	0.50	3.25 ± 0.35	−6.8	−9.2 ± 0.56	−2.4	1.17 ± 0.06
Wheat arabinoxylan	2.75	3.39 ± 0.45	−6.8	−2.9 ± 0.15	3.9	1.68 ± 0.06
Rye arabinoxylan	3.71	1.40 ± 0.37	−6.3	−12.5 ± 0.51	−6.2	0.77 ± 0.27

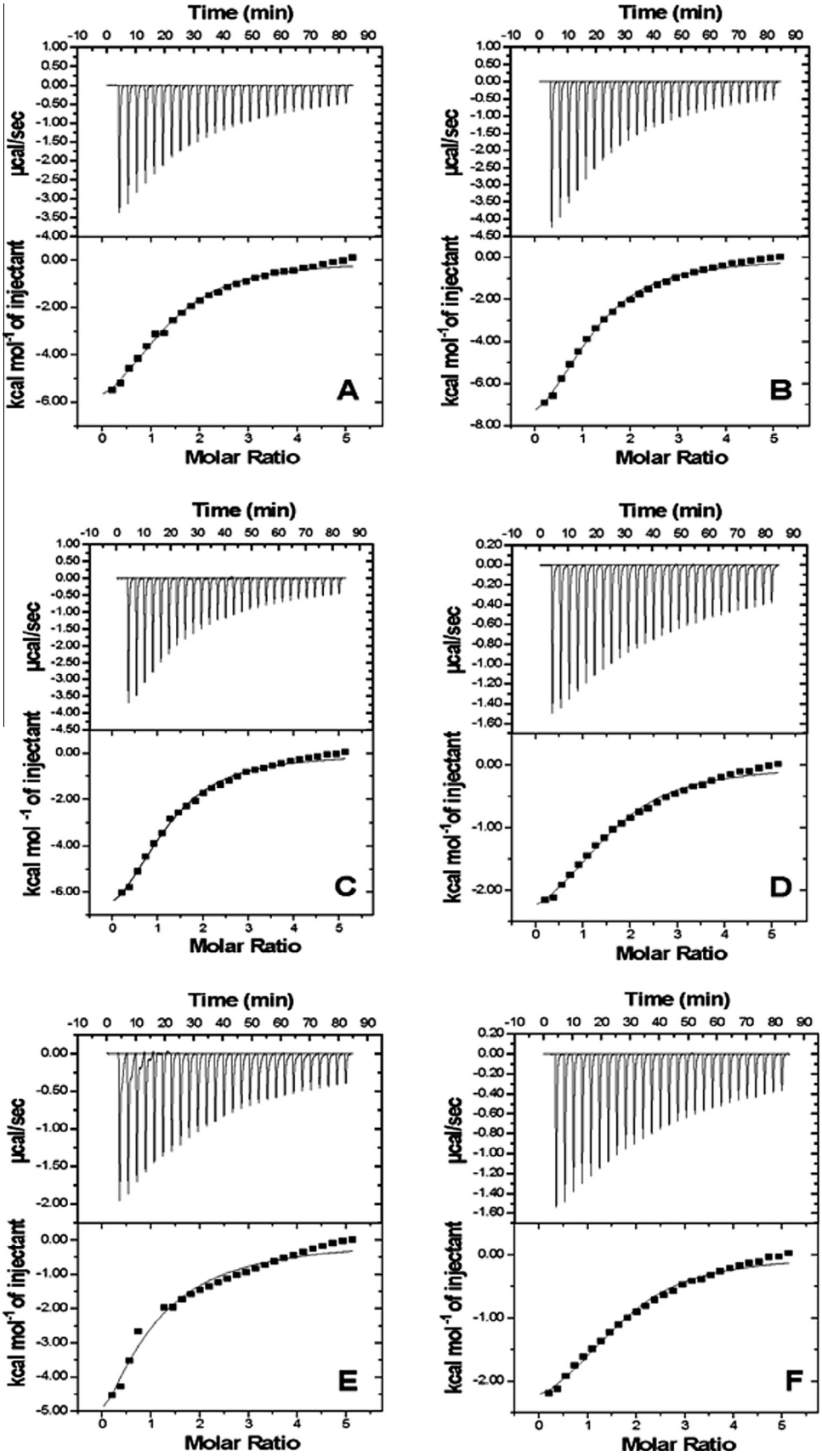


Fig. 3. Isothermal titration calorimetry analysis of *Ct*CBM6 with xylan-based polysaccharides at 55 °C. (A) Birchwood xylan, (B) beechwood xylan, (C) 4-O-methylglucuronoxylan, (D) oat spelt xylan, (E) rye arabinoxylan and (F) wheat arabinoxylan.

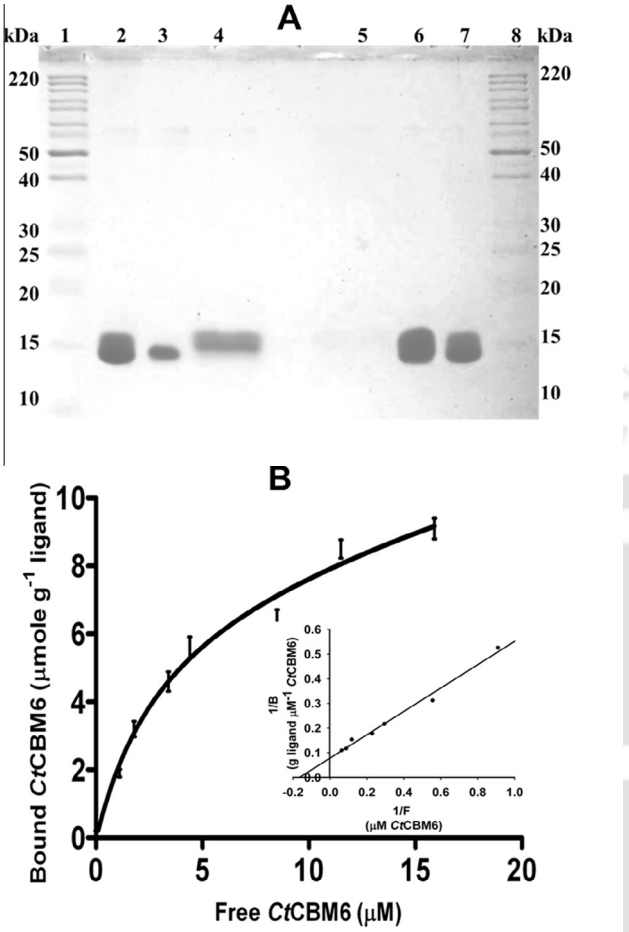


Fig. 4. (A) Qualitative binding analysis of *Ct*CBM6 against insoluble wheat arabinoxylan and avicel by SDS–PAGE. Lanes 1 & 8: protein marker, Lanes 2 & 7: purified *Ct*CBM6 (20 μg), Lane 3: bound *Ct*CBM6 fraction (wheat arabinoxylan), Lane 4: unbound *Ct*CBM6 fraction (wheat arabinoxylan), Lane 5: bound *Ct*CBM6 fraction (avicel), Lane 6: unbound *Ct*CBM6B fraction (avicel). (B) Adsorption of *Ct*CBM6 to insoluble wheat arabinoxylan. The figure shows the equilibrium adsorption isotherm ([B] versus [F]) for *Ct*CBM6. Initial protein concentrations of *Ct*CBM6 were 1–25 mM. Inset shows the linear regression plot of 1/[B] versus 1/[F], to derive the association constant (K_a).

of binding carried out with insoluble wheat arabinoxylan using adsorption isotherm (Fig. 4B) showed the association constant (K_a) value of 6.0×10^7 M^{-1}. The estimated values of relative equilibrium constant K_r and concentration of binding sites [N_0] were 0.485 ± 0.11 l g^{-1} and 0.287 ± 0.08 μmole g^{-1}, respectively.

Protein melting-curve of CtCBM6

The protein melting curve of *Ct*CBM6 showed a single melting peak at 82 °C confirming that this CBM is highly thermostable. The presence of 5 mM Ca^{2+} ions slightly shifted the peak towards the higher temperature 85 °C, indicating the insignificant role of Ca^{2+} ions in the structural stabilization of *Ct*CBM6 (Fig. 5).

Homology modeling and structure validation of CtCBM6

The crystal structure of family 6 CBM from *C. thermocellum* (PDB id: 1GMM) having 63% sequence identity helped to model the *Ct*CBM6 satisfactorily in terms of folds and functional residues. The final model of *Ct*CBM6 was validated with different quality assessment parameters. Ramachandran plot showed that structure 86.0% of the dihedral angles, phi (φ) and psi (ψ) occupied

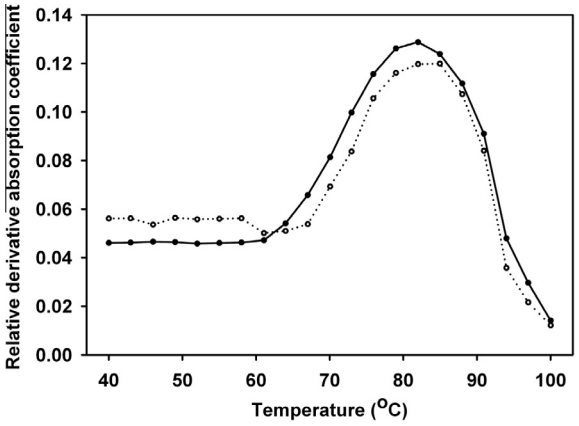


Fig. 5. Protein melting curve of *Ct*CBM6 with no additive, black filled circle (—●—) and in the presence of 5 mM Ca^{2+} ions represented with open circle (—○—).

acceptable statistics of amino acids with 96.1% in the allowed region and 3.9% in additional allowed region (Fig. 3S, A). ProSA result indicated that the modeled protein was error free and resides in the NMR zone with *Z*-score of −6.17 as (Fig. 3S, B). VERIFY3D result depicted that all 123 amino acid residues have average 3D–1D profile score greater than 0.22 which illustrated that no amino acid is in bad contact (Fig. 3S, C). The quality assessment results were satisfactory and allowed using the structural information to support further studies on *Ct*CBM6.

Molecular dynamics study of CtCBM6

Molecular dynamics (MD) simulation was performed to ensure the stability of the protein throughout the simulation run. The results suggested that *Ct*CBM6 undergoes a significant change in the RMSD of backbone atom between 12 and 14 ns of simulations revealing the possibility of conformational changes in *Ct*CBM6. The RMSD converged from 12 to 20 ns with fluctuations, of less than 0.07 Å, which indicated the most stable conformation of *Ct*CBM6 model (Fig. 4S). The RMSD between an energy minimized model of *Ct*CBM6 and the final structure from MD was 1.4 Å, represented the stable conformation. The MD plot signified that all the parameters are well within the qualitative limit of compactness and stability during the simulation. Therefore, the energy minimized model of *Ct*CBM6 is satisfactory for molecular docking simulation.

Structure of CtCBM6

Modeled *Ct*CBM6 was rich in β strands and displayed a β-sandwich fold (Fig. 6A). *Ct*CBM6 has two anti-parallel β-sheets each consisting of five β-strands in which four β-strands from each sheet is opposite to each other. The remaining two β-strands extend away from β-sandwich core (β-3 and β-4 strand, Fig. 6B), making a finger-like structure similar to as reported for *Cs*CBM6 of *Clostridium stercorarium* [54]. The closest structure homolog of *Ct*CBM6 was *Cs*CBM6 (PDB id: 1NAE) with RMSD of 0.5 Å over 110 matched C^{α} atoms on DALI server [55]. The other CBMs which shared the similar fold were *Ct*CBM6 (*C. thermocellum*, PDB id: 1GMM) with 0.7 Å RMSD and *Cc*BM6 (*Clostridium cellulolyticum*; PDB id: 2V4V) with 1.1 Å RMSD. The secondary structure predicted by PSIPRED inferred that *Ct*CBM6 has 47% β-stands and 53% random coil (Fig. 5S). The determination of *Ct*CBM6 secondary structure from CD spectrum suggested the abundance of β-stands containing 44% β-stands and 55% random coil (Fig. 6C). The secondary structure elements of *Ct*CBM6 determined by both the

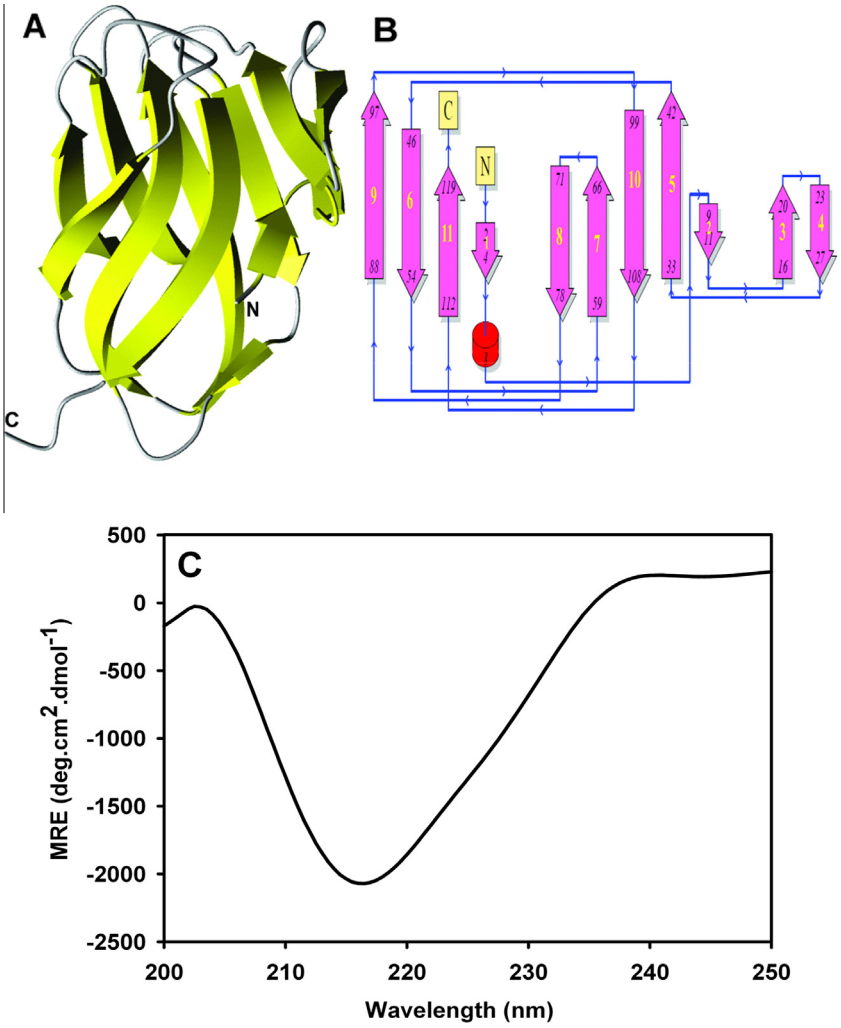


Fig. 6. (A) Three-dimensional pictorial representation of modeled *Ct*CBM6 showing predominantly anti-parallel β-sheets, (B) Topology diagram of *Ct*CBM6 represents mutual orientation of 1–11 β-strands shown by arrows (pink color) and cylinder (red color) representing the α-helix. (C) Far UV-CD spectrum of *Ct*CBM6 from *C. thermocellum*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Secondary structure elements analysis of *Ct*CBM6 by CD and PSIPRED.

Secondary structure content of *Ct*CBM6	Percentage (%) by CD analysis	Percentage (%) by PSIPRED
α-Helix	0.86	0.00
β-Stand	44.49	47.2
Random coil	54.65	52.8

methods are summarized in the Table 4 display that results of CD analysis corroborated those of the predicted structure.

Binding clefts and key residues of CtCBM6

Multiple sequence alignment and structure superimposition with known CBM6 structures depicted five important regions termed A to E, which play key roles in ligand selection and affinity [56] (Fig. 7A and B). The universal mechanism of ligand selectivity by CBM6 was described earlier and provided an appropriate framework to employ bioinformatics tools to determine the location of potential binding sites, the orientation of conserved binding site residues and important regions responsible for specificity and binding [57]. *Ct*CBM6 (in green color) was superimposed with *Cm*CBM6 (PDB id: 1GMM in cyan color Fig. 7B). Region A, B and C are spatially and functionally conserved, and are placed in a one group while region D and E are "hot spot" for primary and tertiary structure variations which lead to determination of the functional specificity of CBM6s [57]. This A–E_{LOOP} model was adopted in order to understand the role of key residues in ligand specificity and binding cleft topology. In *Ct*CBM6, regions A and B are made up of Phe84 and Tyr28, respectively. These conserved aromatic residues form a flat platform that may accommodate sugar ring. In addition, conserved Asn112 (region C) just beneath of regions A and B, is making the floor of the side platform created by aromatic rings of Phe84 and Tyr28. Region D is associated with selectivity of ligands and also determines the binding mode with carbohydrate molecule i.e. whether it binds an internal region or terminal end of the carbohydrate. Region D generally has an aliphatic isoleucine or an aromatic phenylalanine (Fig. 7A) but in case of *Ct*CBM6, region D contains an aliphatic valine (Val18 in purple color; Fig. 7B) which is positioned similarly to Ile23 of *Ct*CBM6 (1GMM) i.e. in a parallel trajectory to the binding site made by Phe84 and Tyr28. Hence, it facilities the entry of sugar rings to cross over the small aliphatic side chain (as compared to Ile) and interact with internal region of sugar rings, creating an open cleft (Fig. 7B). While, in some CBM6s like *Sd*CBM6 (PDB id: 2CDP; *Saccharophagus degradans*), *Cm*CBM6 (PDB id: 1UY1; *Cellvibrio mixtus*) and *Bh*CBM6 (PDB id: 1W9S; *Bacillus halodurans*) region D amino acid residue (conserved glutamate) orients itself in such a

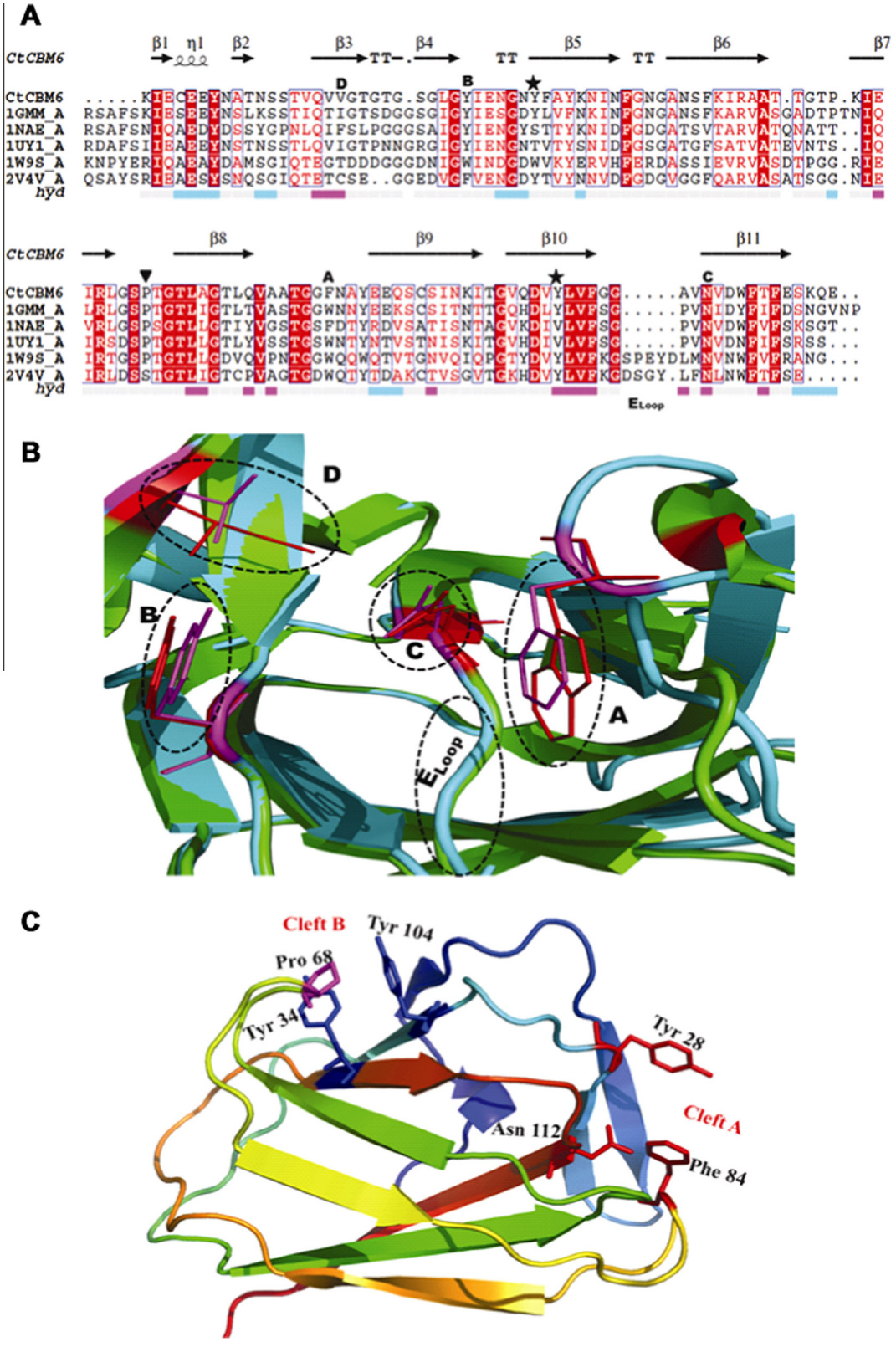


Fig. 7. (A) Multiple sequence alignment of *Ct*CBM6 with representative members of family 6 CBM. *Ct*CBM6 (PDB id: 1GMM from *C. thermocellum*), *Cs*CBM6 (PDB id: 1NAE from *Clostridium stercorarium*), *Cm*CBM6-2 (PDB id: 1UY1 from *Cellvibrio mixtus*), *Bh*CBM6 (PDB id: 1W9S from *B. halodurans*) and *Cc*CBM6 (PDB id: 2V4V from *C. cellulolytium*). AVW showing conserved and SNS semi-conserved amino acid residues; and represent the corresponding secondary structure (alpha helix and beta sheet, respectively) of *Ct*CBM6; represents the relative hydrophobicity (Pink-hydrophobic, gray-intermediate and cyan-hydrophilic); represent conserved proline residues; represent key residues of cleft B; the sequence shown as A–E_{LOOP} displays the important residues for ligand binding of cleft A. (B) Structure superimposition of *Ct*CBM6 in green color with *Ct*CBM6 (PDB id: 1GMM) in cyan color showing the structurally conserved regions. A–E_{LOOP} are indicated by black circles dashed lines while key residues are in stick form (*Ct*CBM6 in purple color and *Ct*CBM6 (PDB id: 1GMM) in red color) responsible for ligand binding and selectivity within cleft A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

way that, it blocks the passage of sugar rings by extending out over it forming a closed cleft, therefore, allowing only terminal sugar to interact [16,18,20]. The region E is crucial for ligand selectivity, but structurally it is poorly defined (Fig. 7A). It is made up of a loop (E_{Loop}) that connects β10 and β11 strands. In *Ct*CBM6 this loop is formed by Gly108, Gly109, Ala110 and Val111. This short loop facilitates the accommodation of large oligosaccharides via internal sugar ring binding mode.

*Ct*CBM6 exhibited a putative ligand binding site, located at the concave surface of a jelly roll fold, termed cleft B (Fig. 7C), apart from the structurally conserved A, B and C regions which create cleft A. The cleft B of a *Ct*CBM6 binding site resembles CBMs of families 4, 15, 17 and 22 [58–60]. Cleft B is formed by two aromatic residues present close to each other in the cavity which is formed by the concave face. In *Ct*CBM6 the aromatic residues Tyr34 and Tyr104 are forming the same groove to accommodate sugar ring, as in cleft A (Fig. 7C). However, the accessibility of these surface aromatic residues might be occluded by a conserved proline (Pro68) residue situated in the loop connecting β-7 to β-8 strand. The capacity of clefts A and B to recognize polysaccharides will be explored below.

Docking analysis of CtCBM6 at binding cleft A

The ligand docking analysis of *Ct*CBM6 at binding cleft A illustrated the amino acid residues involved in the interaction (Table 1S). The predicted free binding energy (ΔG) with different ligands mentioned in Table 1S clearly suggested that as the size of ligand molecule increases the free binding energy (Δ*G*) of interaction decreases and it is least for xylopentaose and for xylohexaose it was a positive value (not mentioned in Table 1S).The ligand (xylobiose) bound complex structure of *Ct*CBM6 depicted the orientation of the terminal sugar ring which is parallel to aromatic amino acid residues Tyr28 and Phe84 present at cleft A (Fig. 8A). Amino acid residues Asn85, Asn112 and Tyr28 are involved in forming the hydrogen bond while residues Phe84, Gly27 and Val18 making a hydrophobic interaction with ligand molecules (Fig. 8B). Surface overview of best docked pose of xylotetraose showed solvent-accessible surface and depth of the binding groove at cleft A (Fig. 8C) while residues Val18, Tyr28, Ile29, Phe84, Asn85, Ala86, Tyr87, Ala110, Val111 and Asn112 were found making hydrogen bond or hydrophobic interactions with ligand molecules (Fig. 8D).

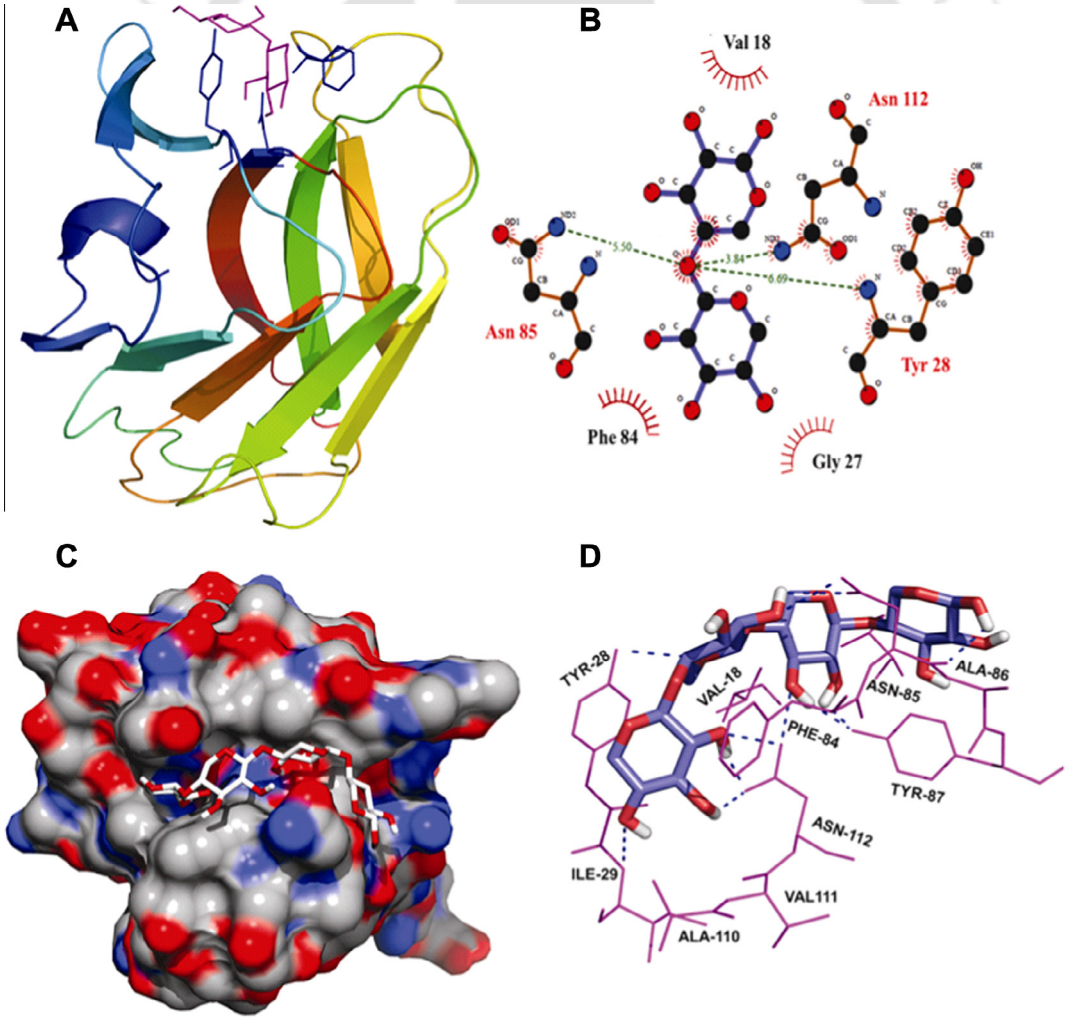


Fig. 8. (A) Ligand bond docked structure of *Ct*CBM6 with xylobiose showing orientation of aromatic residues making a parallel platform to accommodate flat pyranose ring of ligand molecule at cleft A (B) Schematic 2D depiction of ligand xylobiose interacting with amino acid residues at binding cleft A. (▬ ▬) Dashed lines shows hydrogen bonds with labeled bond length, residues displaying arc with spokes ▫ making hydrophobic interaction, contacted atoms are shown with ▪ spokes radiating back. (C) Surface view of the binding cleft A of *Ct*CBM6 showing solvent-accessible surface and depth of the binding groove accommodating ligand xylotatraose. (D) Binding mode interactions of *Ct*CBM6 with ligand xylotatraose best hit obtained by AutoDock. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

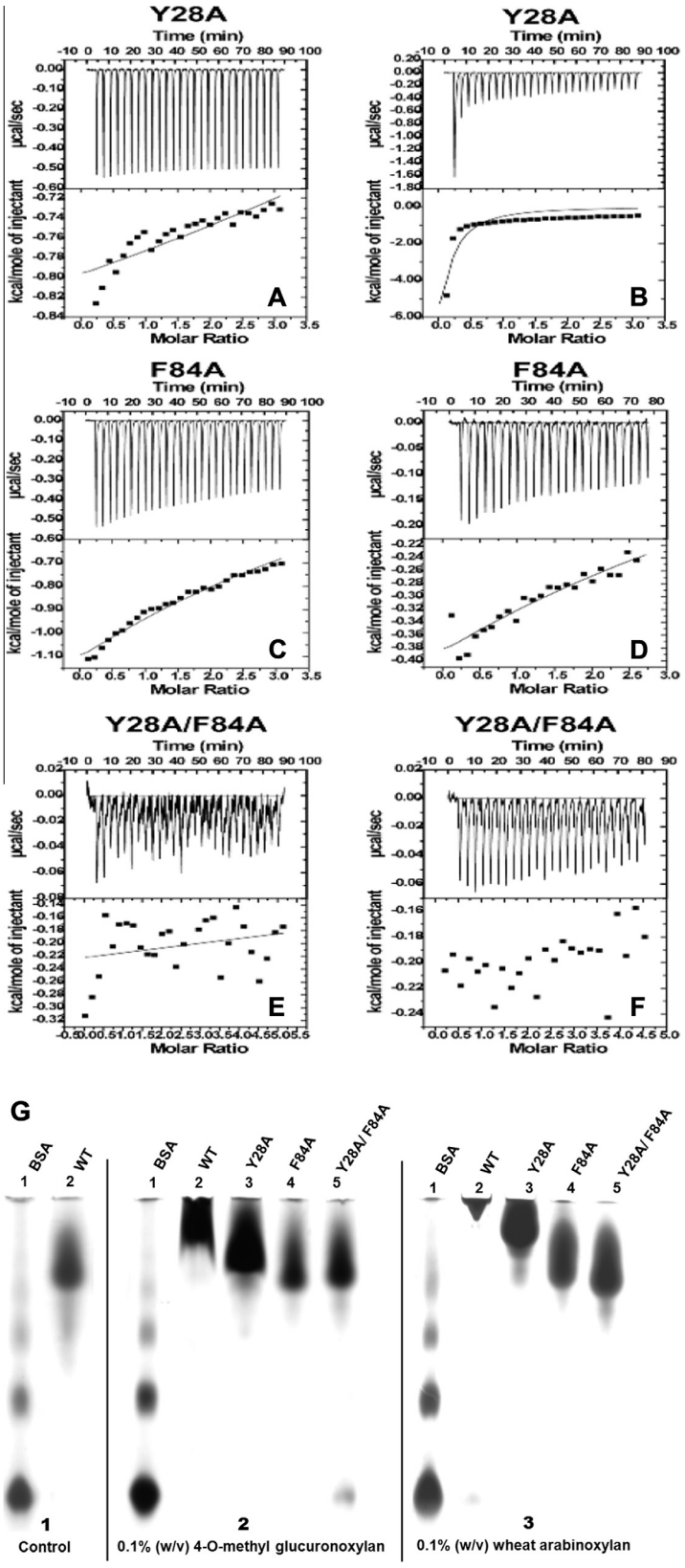


Fig. 9. Isothermal titration calorimetry (ITC) and affinity electrophoresis analysis of *Ct*CBM6 mutants with xylan-based polysaccharides. The ITC profiles of *Ct*CBM6 mutants Y28A (A), F84A (C) andY28A/F84A (E) with ligand 4-O-methyl glucuronoxylan and of Y28A (B), F84A (D) andY28A/F84A (F) with wheat arabinoxylan, (G) qualitative binding analysis of *Ct*CBM6 by affinity electrophoresis using native–PAGE (6.5%, w/v) gels run at 4 °C with and without substrate, (1) Native–PAGE gel showing the migration of Bovine serum albumin (BSA), used as a reference protein (lane 1), wild-type (WT) *Ct*CBM6 (lane 2), (2) Native–PAGE gel containing 4-O-methyl glucuronoxylan (0.1%, w/v) showing the retardation and hence binding of WT *Ct*CBM6 (lane 2), reduced binding of mutant Y28A (lane3), no binding of mutants F84A (lane 4) and Y28A/F84A of *Ct*CBM6 (lane 5), (3) Native–PAGE gel containing wheat arabinoxylan (0.1%, w/v) showing binding of WT *Ct*CBM6 (lane 2), reduced binding of mutant Y28A (lane 3) and no binding of F84A (lane4) and Y28A/F84A (lane 5).

Table 5
Qualitative binding analysis of *Ct*CBM6 mutants by ITC and affinity electrophoresis (AE).

Protein	Ligand	ITC	AE
Y28A	Wheat arabinoxylan	++[a]	++[a]
Y28A	4-O-methyl glucuronoxylan	+[b]	+[b]
F84A	Wheat arabinoxylan	–[c]	–[c]
F84A	4-O-methyl glucuronoxylan	–	–
Y28A/F84A	Wheat arabinoxylan	–	–
Y28A/F84A	4-O-methyl glucuronoxylan	–	–

[a] (++) indicates weak binding.
[b] (+) indicates very weak binding. Although the affinity was too low to accurately convolute by isotherm but the binding was clearly detectable by AE.
[c] (–) indicates no binding.

Assessment of the contribution of binding residues by site-directed mutagenesis

The structure of *Ct*CBM6 suggests that the protein contain two putative clefts, termed clefts A and B. Structure analysis suggest that only cleft A is functional as cleft is not available for carbohydrate recognition due to steric hindrance. When *Ct*CBM6 was superposed with other CBM6 structures (Fig. 7B) as well as docking studies with xylo-oligosacharides (Table 1S, Fig. 8A–D) suggest that in cleft A Tyr28 and Phe84 might constitute the two key residues involved in ligand recognition. Thus, *Ct*CBM6 mutant derivatives Y28A, F84A and the double mutant Y28A/F84A were produced. ITC analyses of the mutants could provide only qualitative information on their binding to ligands because the affinities were too low to obtain proper titration curves (Fig. 9A–F). The results revealed that F84A and Y28A/F84A completely lost their binding affinity towards both glucuronic acid as well as arabinose substituted xylans. However, the binding affinity of Y28A was significantly reduced but not entirely as could be observed by affinity electrophoresis using 4-O-methyl glucuronoxylan and wheat arabinoxylan (Figs 9G, 2 and 3) and also by ITC analysis (Fig. 9A and B) as summarized in Table 5. The results confirmed that Tyr 28 and Phe84 are the key amino acid residues playing a critical role in the ligand binding. Therefore, only the cleft A is the carbohydrate binding platform of *Ct*CBM6.

Conclusions

Data presented here revealed that *Ct*CBM6 display β-jelly roll β-sandwich fold with two potential binding sites; clefts A and B. Cleft A, located on the loops connecting the two β-sheets is decorated by the side chains of Phe-84, Tyr-28 and Asn-112 whereas cleft B is located in the concave surface of the protein where residues Tyr-34 and Tyr-104 might contribute to ligand recognition. *Ct*CBM6 demonstrated affinity towards α-L-arabinosyl and glucuronic acid substituted xylans. The equilibrium association constant (K_a) determined by affinity electrophoresis and isothermal titration calorimetry suggests that *Ct*CBM6 displays similar affinities of against highly substituted xylans (wheat- and rye-arabinoxylans) and poorly substituted xylans (birch-, beech- and oat spelt-xylan). *Ct*CBM6 is unable to interact with other plant cell wall carbohydrates. The functional importance of the two putative clefts for carbohydrate recognition was investigated by mutagenesis analysis. The data revealed that two cleft A protein mutant derivatives, the single mutant F84A and the double mutant Y28A/F84A, completely lost the ability to recognize xylans, whereas, the single mutant Y28A was able to recognize glucuronic acid and arabinose decorated xylan polysaccharides but with significantly reduced affinity. These results confirm that cleft A, through the action of Tyr28 and Phe84, constitutes the functional carbohydrate binding platform. The protein melting of *Ct*CBM6 at 83 °C conveyed the thermostable nature of the protein. The presence of Ca^{2+} ions increased only marginally the structure stability of *Ct*CBM6 by shifting the melting peak by 2 °C to 85 °C. Overall the data confirms that within the CBM6 family a subset of proteins have evolved the capacity to recognize xylans through cleft A, while in this cases cleft B remain non-functional.

Acknowledgments

The authors thank Arun Dhillon for his help with CD spectroscopy experiment. We thankfully acknowledged Dr. P. Saravanan for his help with the MD simulation study. The funder has no role in designing the experiments, preparation of this manuscript and bearing any further publication charges. The research work in part, was supported by Cutting-edge Research Enhancement and Scientific Training (CREST) Fellowship from Department of Biotechnology, Ministry of Science and Technology to Arun Goyal. The scholarships from IIT Guwahati to AKV are gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.abb.2015.03.026.

References

[1] R.H. Doi, A. Kosugi, Nat. Rev. Microbiol. 2 (2004) 541–551.
[2] Y. Shoham, R. Lamed, E.A. Bayer, Trends Microbiol. 7 (1999) 275–281.
[3] H.J. Gilbert, Plant Physiol. 153 (2010) 444–455.
[4] D. Guillen, S. Sanchez, R. Rodriguez-Sanoja, Appl. Microbiol. Biotechnol. 85 (2010) 1241–1249.
[5] S. Cuyvers, E. Dornez, J.A. Delcour, C.M. Courtin, Crit. Rev. Biotechnol. 32 (2012) 93–107.
[6] R. Novy, K. Yaeger, S. Monsma, M. McCormick, J. Berg, O. Shoseyov, E. Shpigel, D. Seigel, A. Goldlust, G. Efroni, Y. Singer, D. Kilburn, P. Tomme, N. Gilkes, FASEB J. 11 (1997) A1151.
[7] A.B. Boraston, B.W. McLean, M.M. Guarna, E. Amandaron-Akow, D.G. Kilburn, Protein Expr. Purif. 21 (2001) 417–423.
[8] B. Rodriguez, M. Kavoosi, J. Koska, A.L. Creagh, D.G. Kilburn, C.A. Haynes, Biotechnol. Prog. 20 (2004) 1479–1489.
[9] R. Johansson, L.C. Gunnarsson, M. Ohlin, S. Ohlson, J. Mol. Recognit. 19 (2006) 275–281.
[10] I. Levy, O. Shoseyov, Biotechnol. Adv. 20 (2002) 191–213.
[11] M.R. Phelps, J.B. Hobbs, D.G. Kilburn, R.F. Turner, Biotechnol. Bioeng. 46 (1995) 514–524.
[12] J.A. Francisco, C. Stathopoulos, R.A. Warren, D.G. Kilburn, G. Georgiou, Biotechnology (N Y) 11 (1993) 491–495.
[13] A.A. Wang, A. Mulchandani, W. Chen, Biotechnol. Prog. 17 (2001) 407–411.
[14] A.A. Wang, A. Mulchandani, W. Chen, Appl. Environ. Microbiol. 68 (2002) 1684–1689.
[15] O. Shoseyov, Z. Shani, I. Levy, Microbiol. Mol. Biol. R. 70 (2006) 283–295.
[16] J. Henshaw, A. Horne-Bitschy, A.L. van Bueren, V.A. Money, D.N. Bolam, M. Czjzek, N.A. Ekborg, R.M. Weiner, S.W. Hutcheson, G.J. Davies, A.B. Boraston, H.J. Gilbert, J. Biol. Chem. 281 (2006) 17099–17107.
[17] V.M.R. Pires, J.L. Henshaw, J.A.M. Prates, D.N. Bolam, L.M.A. Ferreira, C.M.G.A. Fontes, B. Henrissat, A. Planas, H.J. Gilbert, M. Czjzek, J. Biol. Chem. 279 (2004) 21560–21568.
[18] J.L. Henshaw, D.N. Bolam, V.M.R. Pires, M. Czjzek, B. Henrissat, L.M.A. Ferreira, C.M.G.A. Fontes, H.J. Gilbert, J. Biol. Chem. 279 (2004) 21552–21559.
[19] M.A. Correia, V.M. Pires, H.J. Gilbert, D.N. Bolam, V.O. Fernandes, V.D. Alves, J.A. Prates, L.M. Ferreira, C.M. Fontes, FEMS Microbiol. Lett. 300 (2009) 48–57.
[20] A.L. van Bueren, C. Morland, H.J. Gilbert, A.B. Boraston, J. Biol. Chem. 280 (2005) 530–537.
[21] A.B. Boraston, V. Notenboom, R.A. Warren, D.G. Kilburn, D.R. Rose, G. Davies, J. Mol. Biol. 327 (2003) 659–669.
[22] S. Ahmed, A.S. Luis, J.L.A. Bras, C.M.G.A. Fontes, A. Goyal, Biocatal. Biotransfor. 31 (2013) 217–225.
[23] S. Ahmed, A.S. Luis, J.L. Bras, C.M. Fontes, A. Goyal, Biochemistry (Mosc.) 78 (2013) 1272–1279.
[24] A.C. Fernandes, C.M. Fontes, H.J. Gilbert, G.P. Hazlewood, T.H. Fernandes, L.M. Ferreira, Biochem. J. 342 (Pt 1) (1999) 105–110.
[25] H.J. Bae, G. Turcotte, H. Chamberland, S. Karita, L.P. Vezina, FEMS Microbiol. Lett. 227 (2003) 175–181.
[26] J.L. Henshaw, D.N. Bolam, V.M. Pires, M. Czjzek, B. Henrissat, L.M. Ferreira, C.M. Fontes, H.J. Gilbert, J. Biol. Chem. 279 (2004) 21552–21559.
[27] A.B. Boraston, D.N. Bolam, H.J. Gilbert, G.J. Davies, Biochem. J. 382 (2004) 769–781.
[28] H. Hashimoto, Cell. Mol. Life Sci. 63 (2006) 2954–2967.

[29] E.A. Baydoun, M.C. Hobbs, M.H. Delarge, M.J. Farmer, K.W. Waldron, C.T. Brett, Biochem. Soc. Trans. 19 (1991) 250S.
[30] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, Nucleic Acids Res. 25 (1997) 4876–4882.
[31] M.Y. Zhou, S.E. Clark, C.E. Gomez-Sanchez, Biotechniques 19 (1995) 34–35.
[32] U.K. Laemmli, Nature 227 (1970) 680–685.
[33] H. Mach, C.R. Middaugh, R.V. Lewis, Anal. Biochem. 200 (1992) 74–80.
[34] P. Tomme, A. Boraston, J.M. Kormos, R.A. Warren, D.G. Kilburn, Enzyme Microb. Technol. 27 (2000) 453–458.
[35] K. Takeo, J. Chromatogr. A 698 (1995) 89–105.
[36] A.L. Carvalho, A. Goyal, J.A. Prates, D.N. Bolam, H.J. Gilbert, V.M. Pires, L.M. Ferreira, A. Planas, M.J. Romao, C.M. Fontes, J. Biol. Chem. 279 (2004) 34785–34793.
[37] A.L. Carvalho, F.M. Dias, T. Nagy, J.A. Prates, M.R. Proctor, N. Smith, E.A. Bayer, G.J. Davies, L.M. Ferreira, M.J. Romao, C.M. Fontes, H.J. Gilbert, Proc. Natl. Acad. Sci. U, S, A. 104 (2007) 3089–3094.
[38] A.B. Boraston, P. Tomme, E.A. Amandoron, D.G. Kilburn, Biochem. J. 350 (Pt 3) (2000) 933–941.
[39] N.R. Gilkes, E. Jervis, B. Henrissat, B. Tekant, R.C. Miller Jr., R.A. Warren, D.G. Kilburn, J. Biol. Chem. 267 (1992) 6743–6749.
[40] I.A. Dvortsov, N.A. Lunina, L.A. Chekanovskaya, W.H. Schwarz, V.V. Zverlov, G.A. Velikodvorskaya, Microbiology 155 (2009) 2442–2449.
[41] M. Abou-Hachem, E.N. Karlsson, P.J. Simpson, S. Linse, P. Sellers, M.P. Williamson, S.J. Jamieson, H.J. Gilbert, D.N. Bolam, O. Holst, Biochemistry 41 (2002) 5720–5729.
[42] A. Sali, L. Potterton, F. Yuan, H. van Vlijmen, M. Karplus, Proteins 23 (1995) 318–326.
[43] A. Fiser, R.K. Do, A. Sali, Protein Sci. 9 (2000) 1753–1773.
[44] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, J. Mol. Biol. 7 (1963) 95–99.
[45] M.A. Sainz-Polo, S.V. Valenzuela, B. Gonzalez, F.I. Pastor, J. Sanz-Aparicio, J. Biol. Chem. 289 (2014) 31088–31101.
[46] M.J. Sippl, Proteins 17 (1993) 355–362.
[47] M. Wiederstein, M.J. Sippl, Nucleic Acids Res. 35 (2007) W407–W410.
[48] D.S. Goodsell, A.J. Olson, Proteins 8 (1990) 195–202.
[49] C.D.S., W.L. DeLano, The PyMOL Molecular Graphics System, San Carlos, 2002.
[50] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1998) 1639–1662.
[51] D.T. Jones, J. Mol. Biol. 292 (1999) 195–202.
[52] S.M. Kelly, T.J. Jess, N.C. Price, Biochim. Biophys. Acta 1751 (2005) 119–139.
[53] C. Louis-Jeune, M.A. Andrade-Navarro, C. Perez-Iratxeta, Proteins (2011).
[54] A. Lammerts van Bueren, A.B. Boraston, J. Mol. Biol. 340 (2004) 869–879.
[55] L. Holm, P. Rosenstrom, Nucleic Acids Res. 38 (2010) W545–W549.
[56] D.W. Abbott, A.L. van Bueren, Curr. Opin. Struct. Biol. 28 (2014) 32–40.
[57] D.W. Abbott, E. Ficko-Blean, A.L. van Bueren, A. Rogowski, A. Cartmell, P.M. Coutinho, B. Henrissat, H.J. Gilbert, A.B. Boraston, Biochemistry 48 (2009) 10395–10404.
[58] A.B. Boraston, D. Nurizzo, V. Notenboom, V. Ducros, D.R. Rose, D.G. Kilburn, G.J. Davies, J. Mol. Biol. 319 (2002) 1143–1156.
[59] V. Notenboom, A.B. Boraston, P. Chiu, A.C. Freelove, D.G. Kilburn, D.R. Rose, J. Mol. Biol. 314 (2001) 797–806.
[60] L. Szabo, S. Jamal, H. Xie, S.J. Charnock, D.N. Bolam, H.J. Gilbert, G.J. Davies, J. Biol. Chem. 276 (2001) 49061–49065.