Cloning, expression, purification and biochemical characterization of a novel α-L-arabinofuranosidase of family 43 glycoside hydrolase (GH43) from *Clostridium thermocellum*

PhD Thesis

by

SHADAB AHMED

September 2012

Department of Biotechnology
Indian Institute of Technology Guwahati
ASSAM, INDIA

Cloning, expression, purification and biochemical characterization of a novel α-L-arabinofuranosidase of family 43 glycoside hydrolase (GH43) from *Clostridium thermocellum*

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

SHADAB AHMED

Under supervision of

Professor Arun Goyal

September 2012

Department of Biotechnology
Indian Institute of Technology Guwahati
Guwahati 781 039, Assam, India

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOTECHNOLOGY

STATEMENT

I do hereby declare that the content embodied in this thesis is the result of investigations carried out by me in the Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati, India under the guidance of Professor Arun Goyal.

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

September, 2012 *Shadab Ahmed*

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOTECHNOLOGY

CERTIFICATE

It is certified that the work described in this thesis entitled **"Cloning, expression, purification and biochemical characterization of a novel α-L-arabinofuranosidase of family 43 glycoside hydrolase (GH43) from *Clostridium thermocellum*"** by Mr. Shadab Ahmed 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 in the Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati, India and this work has not been submitted elsewhere for a degree.

Arun Goyal
Professor and Head
(Thesis Supervisor)
Department of Biotechnology
IIT Guwahati
Guwahati-781039

ACKNOWLEDGEMENTS

This thesis has been kept on track and been seen through to completion with the support and encouragement of many people including my supervisor, doctoral committee members, well wishers, my friends and colleagues. I would like to thank all those people who made this thesis possible and an unforgettable experience for me. At the end of my thesis, it is a pleasant task to express my thanks to all those who contributed in many ways and had profound impact in the success of this study deserves special acknowledgment.

At this moment of accomplishment, first of all I am extremely indebted to my thesis supervisor, Professor Arun Goyal, Department of Biotechnology, IIT Guwahati. This work would not have been possible without his guidance, support and encouragement. Under his guidance I successfully overcame many difficulties and learned a lot. I can't forget his how patiently he listened to my problems and provided the necessary instructions. Despite of his ill health, he used to review my thesis progress, give me valuable suggestions and made corrections numerous times. His unflinching convictions will always inspire me, and I hope to continue to work with his noble thoughts. I earnestly thank him for inculcating in me scientific temperament and appreciable work ethics which helped me to achieve this goal.

I would also like to express my sincere gratitude to all my doctoral committee members Dr. Rakhi Chaturvedi, Dr.(Mrs.) Gurvinder Kaur Saini, and Prof. Mohammed Jawed for their valuable suggestions and constructive criticism

during my progress report presentations that has led to the successful completion of my thesis.

I sincerely thank Prof. Carlos Fontes, Faculty of Veterinary Medicine, Technical University, Lisbon, for providing lab facilities and his student Ana for helping me to carry out molecular biology experiments.

I would also like to express my sincere gratitude to Prof. M.N.Gupta, Department of Chemistry, IIT Delhi and his PhD student Saurabh Gautam for circular dichroism (CD) analysis of my recombinant proteins.

I would also like to thank the present and previous heads of the Department of Biotechnology, IIT Guwahati, Prof. P. Goswami and Prof. Arun Goyal for providing me with the necessary facilities.

I am also thankful to my seniors Dr. Ravi Kiran Purama, Dr. Avishek Majumder, Dr. Seema Patel and Dr. Deepmoni Deka for their help and suggestions.

I am immensely thankful to my research group members Arabinda, Saprativ, Rishikesh, Soumyadeep, Anil, Shraddha, Damini, Deeplina, Jagan, Vivek, Shuchi, Rviwoo, Arun and Aruna. I am grateful to all my lab mates of Department of Biotechnology for their cooperation and support.

Hearty thanks to my undergraduate trainees Tushar, Vikas, Rahul and Puneet for their wonderful ideas during various stages of work.

I am thankful to my friends Vijay, Vijay Ravi, Dada, Rachna, Asim, Yata, Chinti, Ravi, Rajesh, Ajju, KK,, Kannadi, Talli and Genda for their moral support during the tenure of my PhD.

I wish to acknowledge the supports received from other teaching and non teaching staffs of the Department of Biotechnology, IIT Guwahati.

I wish to acknowledge IIT Guwahati for providing financial assistance and experimental facilities and also University Grants Commission, Govt. of India, New Delhi for Maulana Azad National Fellowship (MANF).

My PhD. endeavor would not have been successful without the love, trust, support and blessings of my parents. I owe my achievements to my family. I also thank my brother Niaz, sister Sheerin and cousins Arif, Aziz and Benazir for their constant encouragement and moral support during my PhD period.

From the depth of my heart I express my deep sincere gratitude to the Almighty for the Blessings and strength bestowed upon me to complete this work.

Shadab Ahmed
September 2012

SYNOPSIS

Introduction

Plant cell wall is mainly composed of complex structural polysaccharides like cellulose and hemicellulose. The hetero-polymers of pentoses like D-xylose, L-arabinose and hexoses *viz*. D-mannose, D-glucose and D-galactose constitutes the hemicellulose. Often, xylans are hetero-polysaccharides with 1,4-linked-β-D-xylopyranose backbone chains comprising of arabinose, glucuronic acid or its 4-O-methyl ether, acetic, ferulic and *p*-coumaric acids side chains depending mainly on the source of xylans. The linear xylan backbone is generally substituted with a variety of sugars and other moieties.

Rye arabinoxylans contain arabinose and xylose in the A/X ratio of 0.49–0.82 and also ferulate residues attached to arabinose as esters at its O-5 position but in wheat arabinoxylans the arabinose to xylose ratios (A/X) varies from 0.47 to 0.58. The cereal arabinoxylans are composed mostly of a backbone of 1,4-linked-β-D-xylopyranosyl residues substituted with single *α*-arabinofuranosyl substituents attached to the O-2, O-3 or to both O-2,3 of the xylose residues. It has been documented that *α*-L-arabinofuranosyl and to a lesser extent *α*-L-arabinopyranosyl side chains are attached to the *β*-D-galactopyranose main chain by 1,3- and 1,6-linkages in type II arabinogalactans. The *α*-L-arabino-furanosidase (*α*-L-arabinofuranosidase, EC 3.2.1.55) are enzymes known to release terminal *α*-1,2-, *α*-1,3- and *α*-1,5 *α*-L-arabinofuranoside residues from hemicellulose such as arabinoxylan and other L-arabinose containing polysaccharides. Arabinofuranosidases have been reported from a few glycoside hydrolase families (GHs) *viz*., GH30, GH43, GH51, GH54 and GH62. The *α*-L-arabinofuranosidases have been used synergistically

with other cellulose degrading enzymes in agro-industrial processes. Lignocelluloses are known to have cellulose as major content but they also contain 20% hemicellulose which mainly consists of pentoses such as xylose and arabinose. In the past α-L-arabinofuranosidases received less importance in the production of bio-ethanol because pentoses are less efficiently converted to ethanol than hexose sugars. But recently they have been used along with *Candida shehatae* that utilizes the pentose sugars for bio-ethanol production from cellulosic waste like mango and poplar leaves. Also L-arabinose has been shown to inhibit intestinal sucrase and thereby reducing the glycaemic response after sucrose ingestion in animals. The carbohydrate binding modules (CBMs) are the non-catalytic modules known to help or bring the catalytic modules in close proximity to its substrates and also some CBMs are known to stabilize the enzyme (catalytic module) structure and increase its temperature resistance. They have been used recently as molecular probe for the analysis of plant cell wall polymers. The CBMs may be found to contain up to 200 amino acids and can be found attached as single, double or triple domain in one protein, located at both C- or N-terminal within the parental protein. All α-L-arabinofuranosidases have the same inverting mechanism of catalysis, but the enzyme activities are different.

Present work

The present investigations are carried out on the "**Cloning, expression, purification and biochemical characterization of a novel α-L-arabinofuranosidase of family 43 glycoside hydrolase (GH43) from *Clostridium thermocellum***". In the present study molecular characterization of α-L-arabinofuranosidase (*Ct*Ara*f*) of family 43 glycoside hydrolase GH43 was carried out. The genes encoding for *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A

and *Ct*CBM6B were cloned, expressed, purified and functionally characterized. The influence of CBMs on catalytic module and their binding affinities towards soluble and insoluble polysaccharides were explored. To our knowledge this is the first report of any α-L-arabino-furanosidase having the ability to hydrolyze both 4-nitrophenyl-α-arabinofuranoside (*p*NPA*f*) and 4-nitrophenyl-α-arabinopyranoside (*p*NPA*p*). The thesis work is comprised of 5 Chapters.

Chapter 1 is the General Introduction which embodies the brief review of literature dedicated to the importance of cellulose and hemicelluloses and their structural features. It mainly focuses on the enzymes having the capability to hydrolyze plant cell-wall cellulose and hemicellulose. It illustrates sequence-similarity based classification of cellulases and hemicellulases belonging to different glycoside hydrolase (GH) families. It also describes different GHs categorized in to various 'clans' based on fold of proteins or core structural features. This chapter reviewed family 43 glycoside hydrolases and associated non-catalytic carbohydrate binding module (CBM), especially, CBM6. The chapter reviews about the type of core architecture seen in various family 43 GHs, their enzyme activities, substrate specificity, active site and active site residues. The various type of α-L-arabinofuranosidases belonging to family 43 GH, have been discussed and their biochemical properties have been highlighted. The various types of binding clefts or binding sites possible in CBM6 have also been elaborated. Previous literatures highlighting their ligand or polysaccharide binding capabilities have also been demonstrated. Previously characterized recombinant family 43 GHs and CBM6 from different bacteria have been illustrated. The chapter also reviews the potential applications of α-L-arabinofuranosidase.

Chapter 2 describes the cloning, expression and purification of genes encoding for *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B, extracted from the genomic DNA of *Clostridium thermocellum* ATCC 27405 (GenBank Accession No: ABN52503.1). The molecular architecture revealed an N-terminal catalytic family 43 glycoside hydrolase module (*Ct*GH43, 903bp) followed by two carbohydrate binding modules (CBMs) *viz.* *Ct*CBM6A (405 bp) and *Ct*CBM6B (402 bp) at the C-terminus. PCR amplified fragment of gene encoding *Ct*43Ara*f* showed a band of around 1710 bp, whereas, the truncated catalytic module *Ct*GH43 displayed a PCR amplified fragment of 903 bp approx. *Ct*CBM6A and *Ct*CBM6B showed PCR amplified fragments of 405 bp and 402 bp, respectively. *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were digested with *Nhe*I-*Xho*I restriction enzymes and the digested fragments extracted from gel and purified using gel extraction kit. Similarly, the pET vectors (pET-21a and pET-28a) were also digested with *Nhe*I-*Xho*I restriction enzymes and the digested fragment of each was purified by gel extraction kit.

The gene encoding for catalytic module *Ct*43Ara*f* and its truncated derivatives *Ct*CBM6A and *Ct*CBM6B were ligated with pET-21a vector and *Ct*GH43 was ligated with pET-28a vector. *E. coli* (DH5α) competent cells were transformed with ligated samples of each derivative mentioned above. The transformed cells were spread on LB agar plates with appropriate antibiotics and grown at 37°C, which resulted in many colonies after 12h incubation. The plasmid DNA of each recombinant derivative was isolated from the colonies after growing in 5 ml LB medium supplemented with appropriate antibiotics at 37°C and at 180 rpm. The cloning of *Ct*43Ara*f* and its truncated derivatives in pET-21a and pET-28a vectors was

confirmed by restriction digestion analysis of the recombinant plasmid DNA as well as by DNA sequencing of the recombinant plasmids. The *E. coli* BL-21(DE3) competent cells were transformed with recombinant plasmid DNA of full length *Ct*43Ara*f* as well as its truncated derivatives *Ct*CBM6A and *Ct*CBM6B, whereas, *E. coli* BL-21(DE3) pLysS competent cells were transformed with recombinant plasmid DNA of *Ct*GH43, for expression of recombinant proteins. Over expression of recombinant proteins was achieved by using IPTG as inducer at a final concentration of 1 mM. The hyper-expression of the recombinant proteins were checked and confirmed by SDS-PAGE gel analysis. After confirming the successful expression, the recombinant proteins were purified from their respective cell free extracts by immobilized metal ion chromatography (IMAC) using HiTrap chelating columns. The degree of purification obtained after this purification step was higher than 95%. The catalytic enzymes purified by IMAC displayed approx. molecular sizes of 63 kDa (*Ct*43Ara*f*), 34 kDa (*Ct*GH43) on SDS-PAGE. The non-catalytic carbohydrate binding modules (*Ct*CBM6A and *Ct*CBM6B) showed molecular sizes of 14.85 kDa and 14.74 kDa, respectively, on SDS-PAGE. The amount of protein obtained from 100 ml of cultures *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B after IMAC purification were 0.9 mg (*Ct*43Ara*f*), 1.08 mg (*Ct*GH43), 0.82 mg (*Ct*CBM6A) and 0.72 mg (*Ct*CBM6B).

Chapter 3 elaborates the substrate specificity and kinetic properties of catalytic modules *viz.* *Ct*43Ara*f* and *Ct*GH43 from *C. thermocellum*. *Ct*43Ara*f* and *Ct*GH43 showed optimal enzyme activities at pH 5.7 and 5.4, respectively, and the optimal temperature for both was 50ºC. *Ct*43Ara*f* and *Ct*GH43 displayed more or less similar pH and thermal stability profiles. *Ct*43Ara*f* was stable over a pH range of 5-

6.8 while *Ct*GH43 was stable in the pH range 5-6.3. Both the enzymes displayed significant loss of activity when enzyme substrate reaction was carried out at 60°C or higher temperature. *Ct*43Ara*f* and *Ct*GH43 showed maximum activity with rye arabinoxylan 4.7 U/mg and 5.0 U/mg, respectively. The catalytic enzymes *Ct*43Ara*f* and *Ct*GH43 also displayed noticeable activities with soluble wheat arabinoxylan (2.5 U/mg and 2.7 U/mg), oat spelt xylan (1.7 U/mg and 1.8 U/mg) and insoluble wheat arabinoxylan (1.2 U/mg and 1.1 U/mg). They also displayed considerable activities with beechwood xylan (1.0 U/mg and 0.9 U/mg) and birchwood xylan (0.7 U/mg and 0.8 U/mg). Both *Ct*43Ara*f* and *Ct*GH43 primarily displayed activity against arabinoxylans with the ability to act on both O-2 and O-3 substituted xyloses to release L-arabinose, implying that they possess both α-1,2- and α-1,3- type of α-arabinofuranosidase activities. The catalytic modules were able to act and degrade synthetic substrates *p*-nitrophenyl-α-L-arabinofuranoside (*p*NP-Ara*f*) as well as *p*-nitrophenyl-*α*-L-arabinopyranoside (*p*NP-Ara*p*). The catalytic efficiencies (k_{cat}/K_M) of *Ct*43Ara*f* and *Ct*GH43 were approximately, 3-fold higher with *p*NP-Ara*f* as compared to *p*NP-Ara*p*. Based on the catalytic efficiencies of *Ct*43Ara*f* and *Ct*GH43 with natural as well as synthetic substrates, it was evident that both these catalytic modules are predominantly *α*-L-arabinofuranosidase. *Ct*43Ara*f* and *Ct*GH43 belong to subclass AXHB-md2,3 exo *α*-L-arabinofuranosidase that release terminal non-reducing arabinose from singly or doubly substituted xylose. The catalytic activity of *Ct*43Ara*f* and *Ct*GH43 increased significantly to more than two-fold in presence of Ca^{2+} and Mg^{2+} salts, implying that these ions may be required as cofactors. A marginal increase in activity was also seen with salts of nickel, manganese and zinc. However, the activity was reduced when concentration of salts of copper, mercury, cobalt and silver increased to more than 10 mM in the reaction mixture. This suggests that these ions

play no role at the catalytic site of enzyme and their presence at active site might be disrupting the enzyme substrate complex formation.

The TLC analysis revealed that L-arabinose was the main sugar that was released as a result of reaction of catalytic modules with rye arabinoxylan, wheat arabinoxylan and oat spelt xylan. TLC showed only single spot corresponding to L-arabinose. HPAEC analysis of the enzyme-substrate reaction products gave only the peak for L-arabinose, which confirmed that L-arabinose was the only monosaccharide that was released after the degradation of substrates (rye arabinoxylan, wheat arabinoxylan and oat spelt xylan) by catalytic enzyme *Ct*43Ara*f*. The TLC and HPAEC analysis confirmed that *Ct*43Ara*f* and *Ct*GH43 released only *α*-L-arabinosyl side chain sugars from arabinoxylans in an exo-acting manner. Protein-melting curves of full length protein *Ct*43Ara*f* and its truncated derivative *Ct*GH43 showed that in presence of Ca^{2+} ions, protein-melting peaks of *Ct*GH43 and CBMs (CBM6A-CBM6B) shifted to higher temperature implying that they impart thermal stability. This was further confirmed by the reversal of melting peaks to original positions in presence of EGTA. In addition to this the catalytic efficiencies of both *Ct*43Ara*f* and *Ct*GH43 were markedly decreased in presence of EGTA implying that calcium may be a crucial co-factor for enzyme catalysis. The comparison of protein-melting curves of *Ct*43Ara*f* and *Ct*GH43, suggested that the CBMs (CBM6A-CBM6B) and the catalytic module are melting independently.

Chapter 4 illustrates the binding analysis of *Ct*CBM6A and *Ct*CBM6B of *Ct*Araf43 from *Clostridium thermocellum*. The binding analysis of CBMs (*Ct*CBM6A and *Ct*CBM6B) to soluble polysaccharides was evaluated using affinity gel electrophoresis. The equilibrium association constant (K_a) of the CBMs was

determined for rye arabinoxyalan and oat spelt xylan by measuring the relative migration distance of proteins on native PAGE gels in the presence of above substrates. Both *Ct*CBM6A and *Ct*CBM6B showed less affinity for rye arabinoxylan (for which the catalytic modules showed maximum activity) but with oat spelt xylan *Ct*CBM6B showed higher binding affinity which was evident from the greater retardation during affinity gel electrophoresis. The normalized curve showed the relative mobility (extent of retardation) of CBMs (*Ct*CBM6A and *Ct*CBM6B) with different concentrations of rye arabinoxylan and oat spelt xylan. The normalized curve clearly demonstrated that *Ct*CBM6B is more retarded as compared to *Ct*CBM6A with oat spelt xylan. It was also visible from the normalized curve by comparing the concentration of oat spelt xylan with extent of retardation reflected the 1:1 *Ct*CBM6B to polysaccharide stoichiometry. Both CBMs displayed very less affinity with rye arabinoxylan. The equilibrium association constant (K_a) values for *Ct*CBM6A and *Ct*CBM6B were determined from a plot of 1/r versus substrate concentrations (rye arabinoxylan and oat spelt xylan). The K_a of both CBMs with rye arabinoxylan was similar 5.6 (%, w/v) indicating very less affinity for this substrate. *Ct*CBM6B had much higher affinity 100 (% w/v) as compared to *Ct*CBM6A 3.0 (%, w/v) with oat spelt xylan. *Ct*CBM6A displayed very low affinity for both rye arabinoxylan as well as oat spelt xylan, indicating that, it is different from the xylan binding CBM6 reported by Czjzek *et al*. (2004).

Chapter 5 describes the *in silico* structure and substrate or ligand binding properties of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B from *Clostridium thermocellum.* The phylogenetic analysis of *Ct*GH43 showed close evolutionary relationship with family 6 Carbohydrate Binding Modules (CBM6) from *Clostridium cellulolyticum,*

Clostridium papyrosolvens and *Acetovibrio cellulyticus*. The comparative modeling of family 43 glycoside hydrolase (*Ct*GH43) from *Clostridium thermocellum* was performed based on crystal structures of arabinoxylan arabinofuranohydrolases (*Bacillus subtillis*, 3C7F), β-1,4-xylosidase (*Bacillus subtillis*, 1YIF), xylan β-1,4-xylosidase (*Bacillus halodurans*, 1YRZ), β-D-xylosidase (*Geobacillus stearothermophilus*, 2EXH) and endo-1,5-α-L-arabinanase (*Bacillus thermodenitrificans* TS-3, 1WL7) from PDB. The secondary structure analysis of *Ct*GH43 established that in most of the sequences, β-sheets and coils dominated followed by alpha helix, beta strands and beta turns. The structure having lowest MODELLER objective function was selected. The 3-D structure revealed typical 5-fold β–propeller. Energy minimization and validation with VERIFY 3D indicated acceptability of the predicted 3-D structure. The Ramachandran plot analysis by RAMPAGE confirmed that family 43 glycoside hydrolase (*Ct*GH43) contains less or negligible segments of helices. It also showed that out of 301 residues, 267 (89.3%) were in most favoured region, 23 (7.7%) were in allowed region and 9 (3.0%) were in outlier region. Molprobity analysis of Ramachandran also confirmed that the predicted 3-D structure of *Ct*GH43 is acceptable. The ligand binding analysis by QSiteFinder and THEMATICS gave the active site residues and the regions. The most probable active site residues were identified by Molecular Virtual docker (MVD), which showed the catalytic triad formation between two Asp and a Glu residues. One Asp acts as the catalytic general base, the Glu acts as the catalytic general acid and the other Asp is responsible for pKa modulation. The best docking scores were obtained with xylotetraose as the ligand which gave a clue that the enzyme attacks xylans and related structures. IUPred analysis of *Ct*GH43 showed no disordered region. The predicted 3-D structure of *Ct*GH43 affirms the fact that it has the similar core 5-fold

β–propeller architecture common to family 43 GHs, so it has the same inverting mechanism of action with the formation of catalytic triad for catalysis of polysaccharides.

The sequence analyses and structure prediction of *Ct*CBM6A and *Ct*CBM6B were carried out. The secondary structures of *Ct*CBM6A and *Ct*CBM6B showed that they are mainly composed of β-strands. The 3-dimensional structures of *Ct*CBM6A and *Ct*CBM6B were generated using *Clostridium thermocellum* (1GMM), *Celvibrio mixtus* (1UXX), *Clostridium cellulolyticum* (2v4v) and *Clostridium stercorarium* (1UY3) as template crystal structures from PDB database. The structure prediction of *Ct*CBM6A and *Ct*CBM6B were carried out, which revealed typical β-sandwich structure, common to family 6 CBMs. Structure validation was done using an array of validation programs, which involved mainly the plotting and analysis of the Ramachandran Plot and other parameters. The final outcome of the above analysis confirmed that the 3-D models of *Ct*CBM6A and *Ct*CBM6B were acceptable and had stable conformation. Docking energies analysis indicated that the binding of both CBMs is stronger with xylose-based ligands such as xylopyranose, arabinopyranose and arabinofuranose. However, *Ct*CBM6B showed better docking scores with arabinopyranose and arabinofuranose sugars than *Ct*CBM6A. The docking results of CBM6B were in agreement with experimental results of affinity gel electrophoresis result with oat spelt xylan which contains xylopyranose and arabinofuranose sugar residues.

The Circular Dichroism (CD) analysis of *Ct*GH43 showed that β-sheets and random coils were the main secondary structural elements and very few α-helices were present. The CD analysis of *Ct*CBM6A showed appreciable α-helical content although two third of the secondary structure was made up of random coils followed

by β-sheets. Similar to *Ct*GH43, *Ct*CBM6B also showed very few α-helices while most of the secondary structure consisted of random coils and β-sheets. The results of CD analysis of both *Ct*GH43 and *Ct*CBM6B were in agreement with the secondary structure prediction of the same proteins by PSIPRED VIEW software with slight difference in number of α-helices. These observations confirmed that the secondary structure of *Ct*GH43 and *Ct*CBM6B were mostly made up of β-sheets and random coils with very few α-helices. However, *Ct*CBM6A showed greater α-helical content as compared with *Ct*GH43 and *Ct*CBM6B.

CONTENTS

Chapter 3. Biochemical, functional and structural characterization of catalytic modules *Ct*43Ara*f* and *Ct*GH43 from *Clostridium thermocellum*

Chapter 1

General Introduction

1. Carbohydrates

Carbohydrates are unarguably, the most abundant biomolecules found on the surface of earth (Bayer *et al*., 1998; Bayer *et al*., 2000). They are present everywhere in nature as they are found in plant biomass, exoskeleton of insects, bacterial cell surfaces, bio-films, and also on mammalian cell surfaces. The utility of carbohydrates is dependent on their overall structural arrangement *viz.* length of the carbohydrate, its sugar composition, the position of the anomeric carbon and the type of glycosidic linkages formed between sugar monomers (Fontes and Gilbert, 2010). These structural features provide for millions of different possible combinations of carbohydrate structures, and each structure is suited for serving its function in nature.

1.1 Structural polysaccharides in plants

Structural polysaccharides act as protective walls of plant cells (Gilbert, 2011). Cellulose, chondroitin sulphates, hyaluronic acids and chitin are examples of structural polysaccharides. The terrestrial biomass contributed by plants is mainly composed of plant cell-wall materials *viz.* cellulose and hemicellulose (Bayer *et al.,* 2007). These are the most abundant naturally occurring substances found on the earth.

They account for more than 60% of all global carbon bound to organic matter according to Ramamurthy *et al*. (1992). Polysaccharides are macromolecular carbohydrates comprising chain of monosaccharides linked together by glycosidic bonds, formed as a result of condensation reaction (Berg, 2007). The bulk of plant cell wall material is cellulose, a homopolymer of β-1,4-glucose which takes on an overall linear shape. Cellulose is hypothesized to exist in two forms, crystalline and amorphous. In the crystalline form cellulose chains self associate via intra- and intermolecular hydrogen bonds and van der Waals forces to form cellulose fibrils and micro-fibrils, which are insoluble and provide the majority of tensile strength to the plant cell wall. Amorphous cellulose lacks this higher order structure and are more susceptible to increased degradation. Plant cell walls also contain a number of other sugar polymers termed hemicellulose which includes xylan (β-1,4-linked xylose), laminarin (β-1,3-linked glucose), mannan (β-1,4-linked mannose) and lichenan (mixed 1,3-1,4-β-D-glucan) (Scheller and Ulvskov, 2010). Structural polysaccharides are either fibrous-polysaccharides, for example, cellulose in higher plants and some algae or matrix-polysaccharides, e.g. arabinoxylans, galactomannans or pectins in plants (Cosgrove, 1999; Shallom and Shoham, 2003). The other main structural polysaccharide found within the plant cell wall are pectins and substituted heteropolysaccharides composed of a α-1,4-D-galacturonic acid backbone with rhamnose, galactose and arabinose substituents (Bayer *et al.,* 1998).

1.1.1 Cellulose

Cellulose is the main polysaccharide that provides structural integrity to plants and is one of the most important substances occurring in nature. About 40% of the plant carbon is contained by cellulose (Demain *et al*., 2005). However, the percentage of cellulose in plant varies depending on the origin. Its occurrence is widespread, from higher plants to primitive organism such as sea-weads, flagellets and bacteria. Other cellulose-containing materials include agriculture residues, water plants, grasses etc. The cellulose are linear homopolymer of consisting of regio and enantiopure D-glucopyranose (also known as anhydroglucose units) chains linked together by β-(1→4)-glycosidic linkages as shown in Fig. 1.1. The chemical compositions of cellulose containing materials from different sources are listed in Table 1.1. The Table 1.1 illustrates the percentage of cellulose, hemicellulose and lignin present in different sources. Cotton, henequen plant and sunn hemp have shown very high amount of cellulose content (Table 1.1).

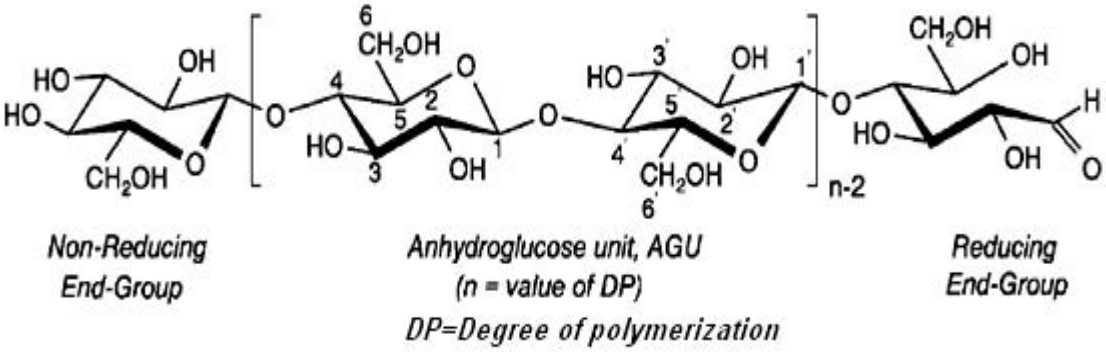


Fig. 1.1 Chemical structure of cellulose showing β-(1→4) glycosidic linkages, reducing end and non-reducing end (Somerville, 2006).

Table 1.1 Chemical composition of cellulose containing substances from different sources (Adapted from Hon (1996) and Klemm *et al*., 1998).

Sources	Composition (%)		
	Cellulose	Hemicellulose	Lignin
Hardwood	42-47	25-35	16-24
Softwood	40-44	25-29	25-31
Bagasse (sugarcane or sorghum)	40	30	20
Coir (from husk of coconut)	32-43	10-20	43-49
Corn cobs	45	35	15
Corn stalks	35	25	35
Cotton	95	2	1
Flax, *Linum usitatissimum* (retted)	71	21	2
Flax, *Linum usitatissimum* (unretted)	63	12	3
Hemp (*Cannabis sativa*)	70	22	6
Henequen (*Agave fourcryodes*)	78	4-8	13
Istle leaf fibres	73	4-8	17
Jute (*Corchorus olitorius*)	71	14	13
Kenaf (*Hibiscus cannabinus*)	36	21	18
Ramie (*Boehmeria nivea*)	76	17	1
Sisal (*Agave sisalana*)	73	14	11
Sunn (*Crotalaria juncea* L)	80	10	6
Wheat straw	30	50	15

1.1.2 Hemicellulose

Hemicelluloses are usually hetero-polysaccharides present in plant cell walls containing β-(1→4)-linked backbones of glucose, mannose or xylose (Scheller and Ulvskov, 2010). They commonly occur in nature as xyloglucans, xylans, arabinoxylans, arabinogalactans, mannans, glucomannans, β-(1→3,1→4)-glucans and pectins (Schadel *et al.*, 2009). Hemicelluloses are known to be synthesized in the Golgi apparatus (membranes) by various glycosyltransferases. Many glycosyltransferases involved in the process of biosynthesis of xyloglucans and

mannans are known (Schadel *et al.*, 2009). In contrast, the biosynthesis of xylans and β-(1→3,1→4)-glucans remains very elusive, and recent studies have led to more questions than answers (Scheller and Ulvskov, 2010). The predominant hemicellulose in many primary cell walls is xyloglucan. Other hemicelluloses found in primary and secondary walls include glucuronoxylan, arabinoxylan, glucomannan, and galactomannan (Schadel *et al.*, 2009). Hemicelluloses can be extracted from cell walls by alkaline treatment (Cheng *et al*., 2010).

1.1.2.1 Xylans

Xylans are the most abundantly found hemicellulose in plants, which has a backbone of β-(1→4)-linked xylose residues. The linear xylan backbone is generally substituted with a variety of sugars and other moieties. Xylans are substituted with α-(1→2)-linked glucuronosyl and 4-O-methyl glucuronosyl residues in the secondary wall of dicots (Scheller and Ulvskov, 2010). Often, xylans are hetero-polysaccharides with 1,4-linked-β-D-xylopyranose backbone chains substituted with arabinose, glucuronic acid, or its 4-O-methyl ether, acetic, ferulic, and *p*-coumaric acids side chains depending mainly on the source of xylans (Numan and Bhosle, 2006). Most hemicellulases are anchored in a β-1,4-linkage and the main backbone is branched, whereas, the individual sugars may be acetylated or methylated (Numan and Bhosle, 2006; Saha, 2003). The L-arabinosyl residues are often found in hemicelluloses, such as arabinan, arabinoxylan, gum arabic and arabinogalactan.

1.1.2.2 Glucouronoxylans

They have β-(1-4)-xylose backbone with 4-O-methylglucuronic acid side-chains. Sometimes arabinose and O-acetyl side-chains may also be found attached to

the main chain (Fig. 1.2). Glucouronoxylans are the major polysaccharide of secondary walls of dicot plants (Urbanowicz *et al*., 2012).

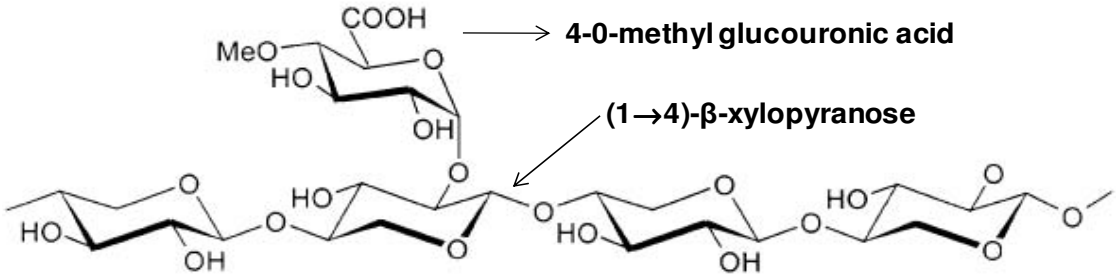


Fig. 1.2 Chemical structure of glucouronoxylan (4-O-methyl-D-glucourono-D-xylan) adapted from Nishitani and Nevins (1991).

1.1.2.3 Arabinoxylans

Arabinoxylans are composed of a β-(1-4)-D-xylopyranosyl backbone, with one or more- L-arabinofuranosyl residues substituted at position 2 or 3 (Fig. 1.3). On the other hand, in many cases, hexoses and hexuronic acids are also present, but as minor constituents. The hemicellulose fraction is composed mainly of xylans that are linked to cellulose fibrils by hydrogen bonds. The xylose residues tend to be acetylated at positions O–2 and O–3, whereas other groups, such as ferulic acid, tend to be bound to positions O–2 or O–3 and lignin, forming cross-links that entrap the polymer (Cosgrove, 1999). A combination of enzymes is therefore required for the release of arabinoxylans from the cell matrix and for its hydrolysis (Saha, 2001; Saha, 2003). Arabinoxylans are polysaccharides found in the bran of grasses and cereals (wheat, rye, barley, oat, rice, corn and sorghum). They are the major hemicellulose component of flour and bran (dietary fibres). Since xylose and arabinose are both pentoses, arabinoxylans are usually classified as pentosans. 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). The structural features of hetero xylans are depicted below which shows a non-substituted xylopyranose (Fig. 1.3A) and substituted xylopyranose (Fig. 1.3B, Fig. 1.3C and Fig. 1.3D). D-xylopyranose can be substituted at O-2, O-3 and O-5 position with L-arabinose as shown in Fig. 1.3B, Fig. 1.3C and Fig. 1.3D, respectively. The schematic representation of xylan substituted with L-arabinose shows the structure of these xylans and enzymes acting on them (Fig. 1.4). The (1→4)-β-D-xylan endohydrolase acts and degrades only the main chain xylan (1→4)-β-D-xylopyranose) as shown in Fig. 1.4. The main chain D-xylopyranose ring can be singly or doubly substituted with L-arabinose and the α-L-arabinofuranosidases are capable of attacking both singly and doubly substituted xylans (Fig. 1.4). The non-reducing (1→4)-β-D-xylopyranose are acted upon by β-D-xylosidases (Fig. 1.4).

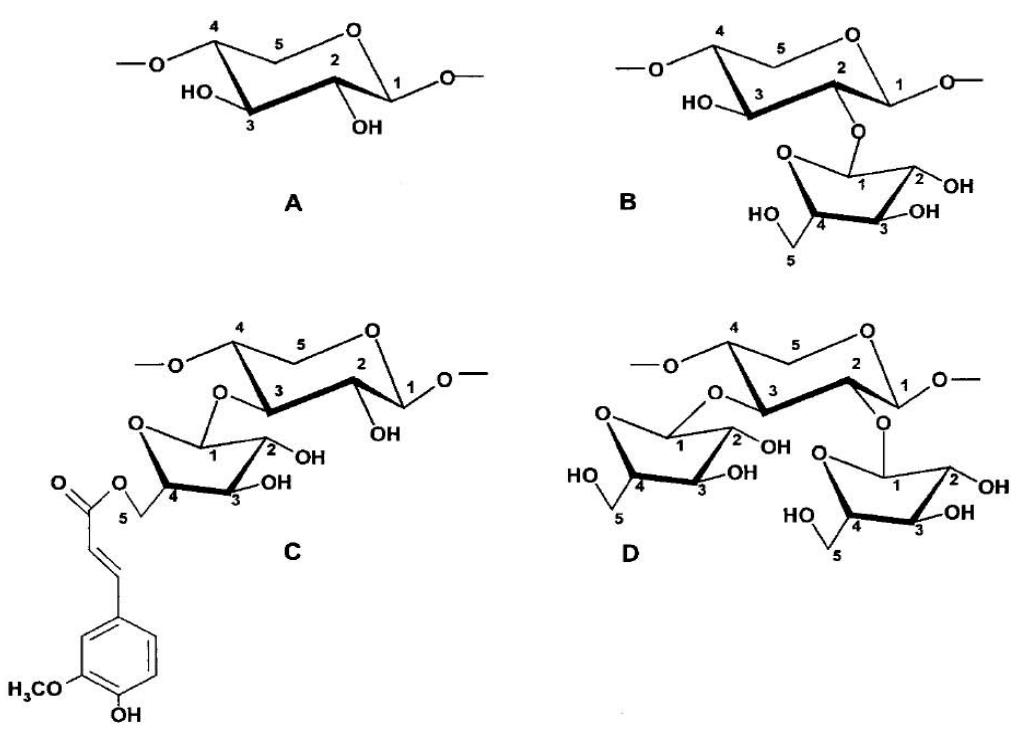


Fig. 1.3 Chemical structure of arabinoxylan. **A)** non-substituted D-xylopyranose, **B)** O-2 substituted D-xylopyranose, **C)** O-3 substituted D-xylopyranose and also the attachment point of ferulic acid to O-5 of L-arabinfuranose, **D)** O-2 and O-3 substituted D-xylopyranose with L-arabinofuranosyl residues (Cyran *et al*., 2003).

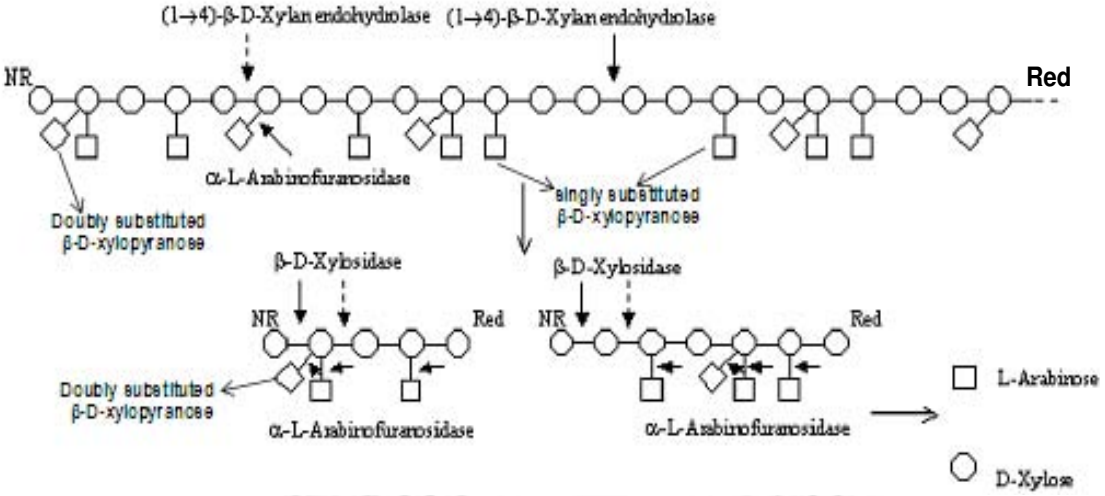


Fig. 1.4 Schematic representation of a typical xylan component of hemicellulose and the respective enzymes acting on them (Numan and Bhosle, 2006; Saha, 2003). NR represent the non-reducing end; Red represents the reducing end and the arrows depict the enzymes acting at particular site *viz.* xylan endohydrolase, α-L-arabinofuranosidase and β-D-xylosidase.

1.1.2.4 Xyloglucans

Xyloglucan are composed of β-(1→4)-glucose units, substituted with (1→6)-D-xylose side chains (Fig. 1.5) (Fry, 1989; Brennan and Harris, 2010). Xyloglucans are heavily branched β(1→4)-glucans (Fig. 1.5). Xyloglucan makes up 20-25% (dry weight) of the type I primary cell walls (all flowers except the grass family) of all dicots and monocot species except grasses (type II primary cell wall) and cereals, where they are found in very less extent (2-5%) (Fry, 1989).

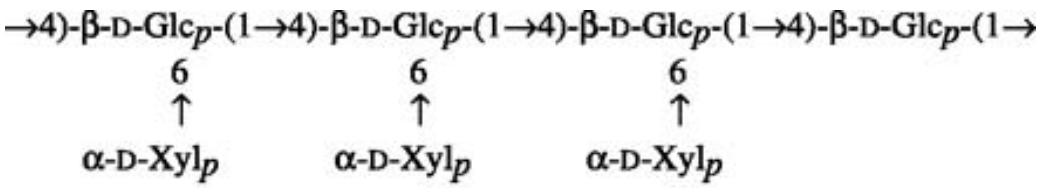


Fig. 1.5 Schematic representation of xyloglcan structure showing β-(1→4)-glucose units substituted at (1→6) position with D-xylopyranose sugars (Brennan and Harris, 2010).

1.1.2.5 Arabinans

Arabinans, of cell-wall pectic-substances, comprise 1,5-α-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. The branched arabinan display a main chain of 1,5-α-linked-L-arabinofuranose substituted with L-arabinofuranosyl residues at 1,3 or 1,2 position as depicted in Fig. 1.6. The linear arabinan show a linear main chain of 1,5-α-linked-L-arabinofuranose (Fig. 1.7).

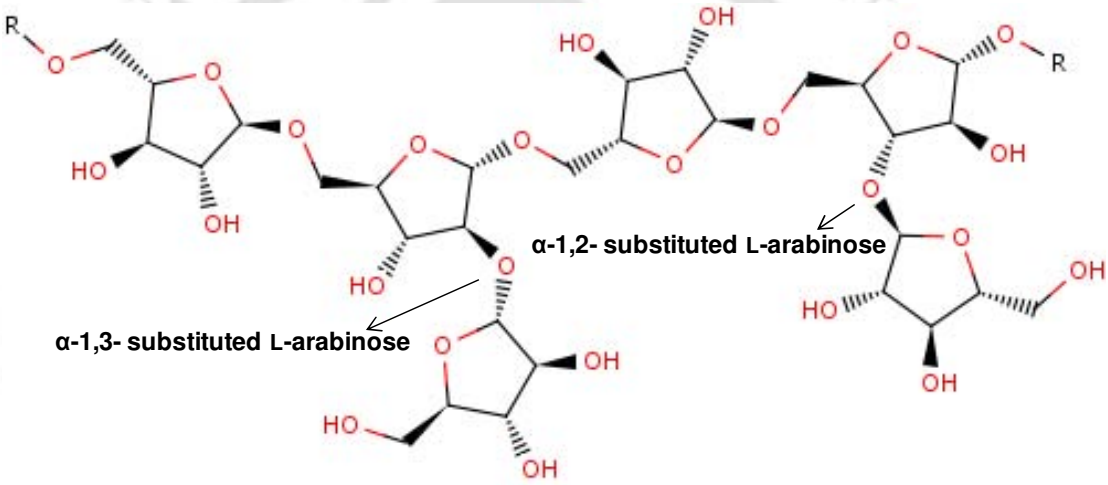


Fig. 1.6 Chemical structure of branched arabinan (http://www.brenda-enzymes.org, Technische Universitat Braunschweieg).

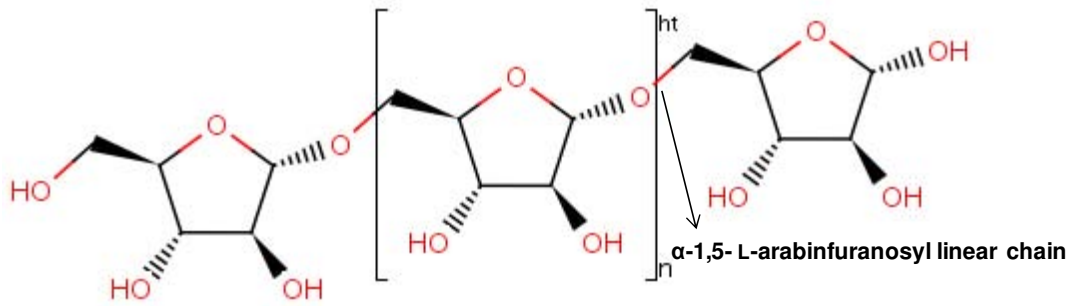


Fig. 1.7 Chemical structure of linear arabinan (http://www.brenda-enzymes.org, Technische Universitat Braunschweieg).

1.1.2.6 Galactans

Galactans are polymers of β-D-galactopyranose. They are sometimes included in the hemicellulose group, but a recent review by Scheller and Ulvskov (2010) suggested that they should not be included in the heterogeneous group of hemicelluloses. However, at present they are classified as a hemicellulose. Scheller and Ulvskov gave reason that since galactans appear to be part of pectin molecules, at least in the initial synthesis stage. The galactans are also found to be substituted with arabinosyl resides in arabinogalactans. These are of two types, Type I which is rich in 1,4 linkages and type II, contains terminal 1,3,5-linked arabinose or 3,5-linked arabinose substituted to main chain of galatan (e.g larchwood arabinogalatan).

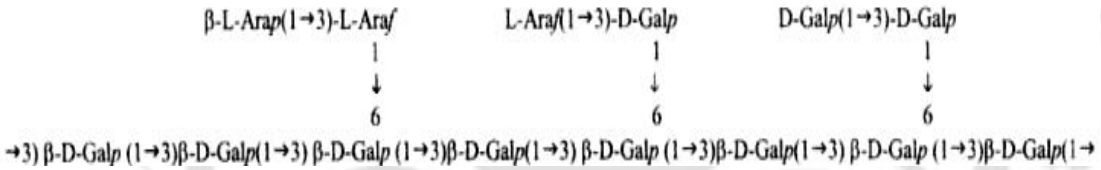


Fig. 1.8 Arabinogalactan type II Structure from Larchwood (Heinze *et al*., 2005)**.**

1.1.2.7 Mannans

Mannans are heteropolymers and contains β-(1→4)-D-mannopyranose backbone. The β-(1→4)-D-mannopyranose backbone is usually substituted with galactose or glucose.

1.1.2.7.1 Galactomannans

Galactomannan is composed of a β-(1→4)-D-mannopyranose backbone to which single D-galactosyl units are attached to C-6 of some of the D-mannosyl residues as shown in Fig. 1.9 (McCleary and Matheson, 1983; Cui *et al.,* 2009). The extent of galactopyranose substitution invariably affects the solubility, viscosity and interaction with other polysaccharides (Cui *et al.,* 2009).

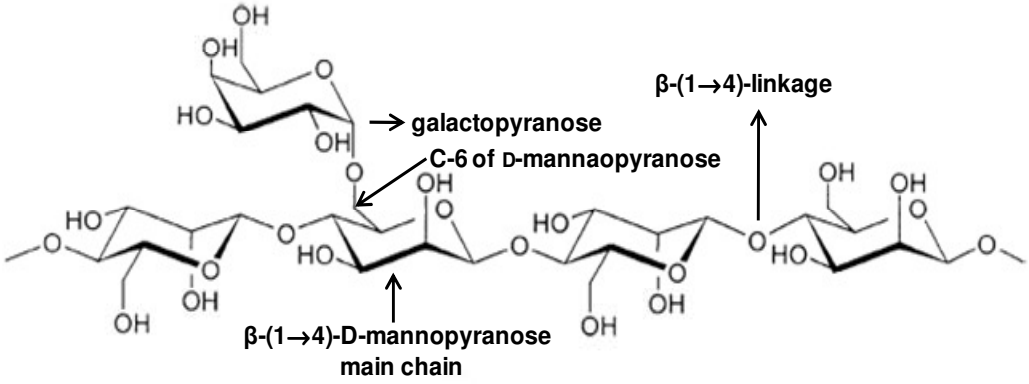


Fig. 1.9 Chemical structure of galactomannan (Cui *et al.,* 2009).

1.1.2.7.2 Glucomannan

Glucomannans are the hetero-polymers of β-(1→4)-D-mannopyranose backbone and β-1,3 substituted β-D-glucopyranose (Fig. 1.10). They are generally found in the secondary walls of softwood (Maeda *et al*., 1980).

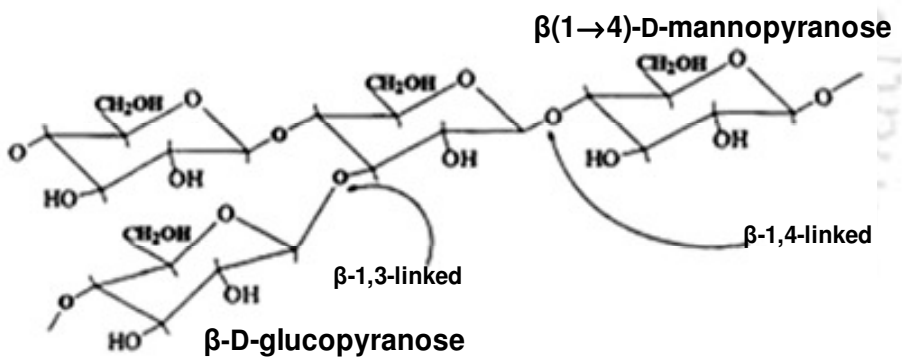


Fig. 1.10 Chemical structure of glucomannan (Maeda *et al*., 1980).

1.1.3 Lignin

The main building blocks of lignin are the hydroxycinnamyl alcohols (or monolignols) coniferyl alcohol and sinapyl alcohol, with typically minor amounts of *p*-coumaryl alcohol (Fig. 1.11). Lignins are large group of aromatic polymers that result from the oxidative combinatorial coupling of 4-hydroxyphenylpropanoids as shown in Fig. 1.11 (Boerjan *et al*., 2003; Ralph *et al*., 2004). Lignin is found in all

vascular plants, mostly between the cells and also within the cells and in the cell walls. It usually occurs as complex structure bound to the hemicelluloses in wood.

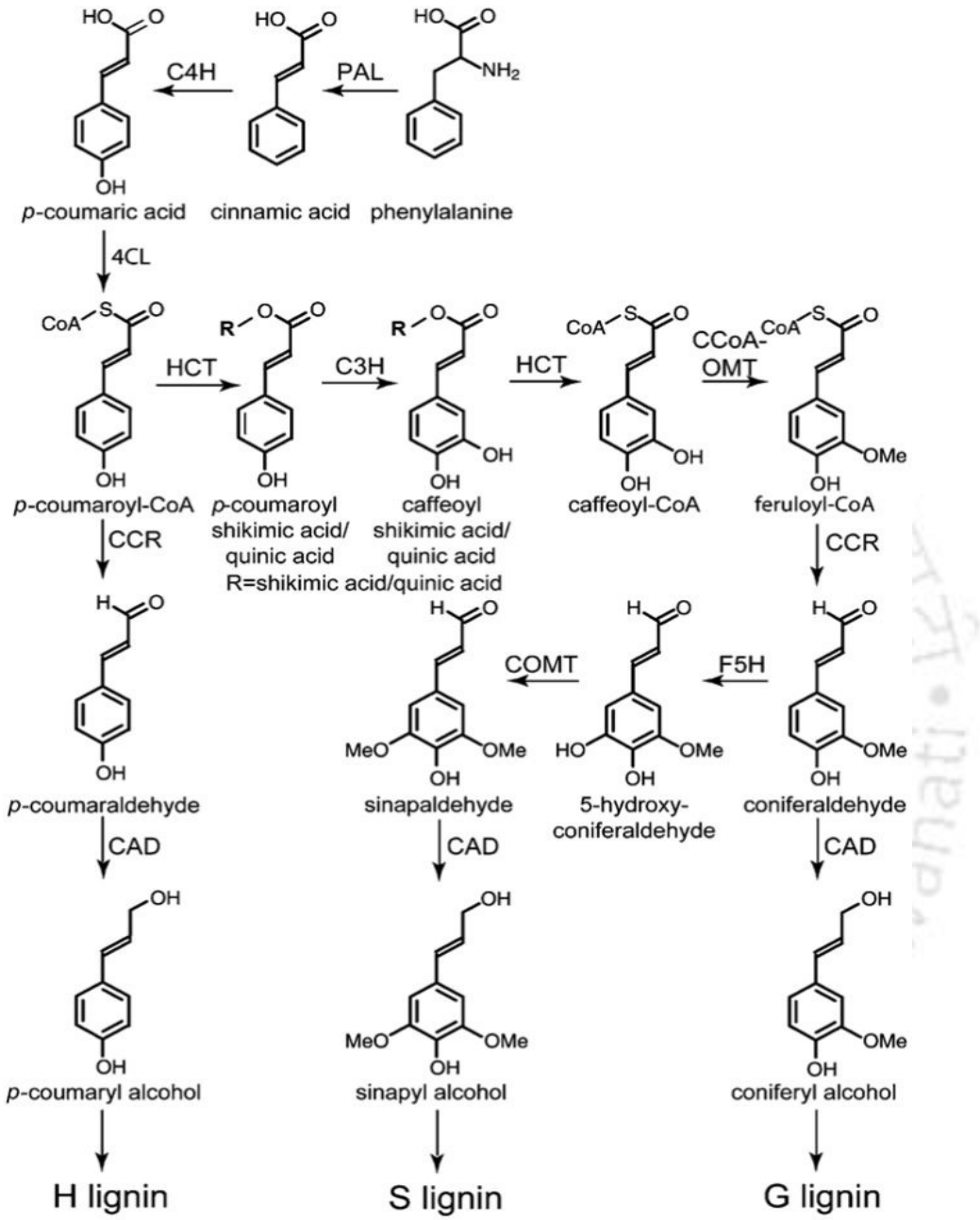


Fig. 1.11 The main biosynthetic pathway for synthesis of the monolignols p-coumaryl, coniferyl, and sinapyl alcohol (Boerjan *et al*., 2003). Phenylalanine ammonia-lyase (PAL); Cinnamate-4-hydrolyase (C4H); 4-coumarate:CoA ligase (4CL); p-coumarate-3-hydrolyase (HCT), p-hydroxycinnamoyl-CoA:Quinate/Shikmate-p-hydroxycinnamoyltranferase (CCoAOMT), Caffeoyl-CoA-O-methyltransfersae (CCR), Cinnamoyl CoA reductase ferrulate-5-hydroxylyase (F5H), Caffeic acid-O-methyltransferase (COMT); Cinnamoyl alcohol dehydrogenase (CAD) (Adapted from Borjan *et al*., 2003 and Ralph *et al*., 2004).

The units resulting from the hydroxycinnamyl alcohols, when incorporated into the lignin polymer, are called guaiacyl (G), syringyl (S), and *p*-hydroxyphenyl (H) units (Figs. 1.11 and Fig 1.12). The β-O-4 linked aryl ethers are found most abundantly in the syringyl lignin of poplar (Fig.1.12). Apart from this, guaicyl (G), sinapyl p-hydroxybenzoate-derived, phenyl coumaran, biphenyl ether, resinol units were also present as components of the poplar lignin (Fig. 1.12).

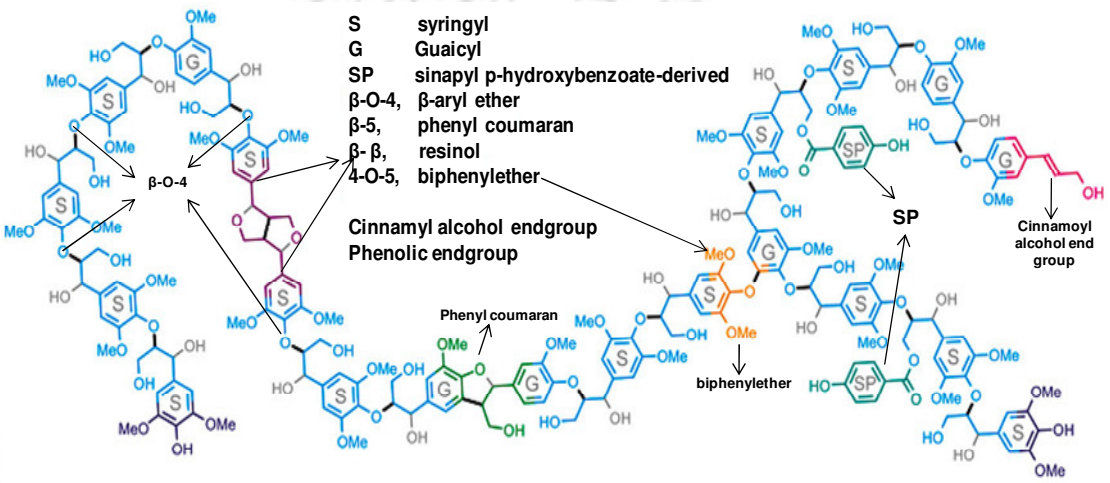


Fig. 1.12 A schematic representation of a lignin polymer from poplar, as predicted from NMR-based lignin analysis. The poplar lignin polymer of around 22 units shows high S (syringyl) lignin with a prevalence of β-aryl ether linkages. Sinapyl p-hydroxybenzoate derived, phenyl coumaran, resinol, biphenyl units were also present in the poplar lignin polymer (Adapted from Boerjan *et al*., 2003 and Stewart *et al*., 2009).

1.1.4 Pectin

Pectins are a family of complex polysaccharides that contain 1,4-linked α-D-galactosyluronic residues. Three pectic polysaccharides, homogalacturonan, rhamnogalacturonan-I and substituted galacturonans have been isolated from primary plant cell walls (Ridley *et al.*, 2001). Homogalacturonan is a linear chain of 1,4-linked α-D-galactosyluronic residues, in which some of the carboxyl groups are methyl esterified (Fig. 1.13). Homogalacturonan may also be O-acetylated at the C-2 and C-3 positions. Rhamnogalacturonan-I is a family of pectic polysaccharides containing a

backbone of the repeating disaccharide [→4)-α-D-GalpA-(1→2)-α-L-Rhap-(1→] as shown in Fig. 1.14 (Ridley *et al.*, 2001). The minor component of the pectin backbone is rhamnogalacturonan-II. This is not structurally related to rhamnogalacturonan-I since its backbone is composed of 1,4-linked α-D-galactosyluronic units (dimers) like homogalacturonan (Fig. 1.15). Rhamnogalacturonan-II is present in primary walls predominantly as a dimer of 1,4-linked α-D-galactosyluronate that is cross-linked by a 1:2 borate-diol ester as shown in Fig. 1.15 (Kobayashi *et al.*, 1996; Ishii *et al.*, 1999). This cross-link is formed between OH-2 and OH-3 of the apifuranosyl residues (βApi*f*) in each monomeric rhamnogalactoronan-II subunit (Fig. 1.15). Rhamnogalactoronan-II is of interest as it occurs in relatively high amounts in wine and other fruit juices and it has been demonstrated that it binds heavy metals and has immunomodulating activities (Ridley *et al*., 2001).

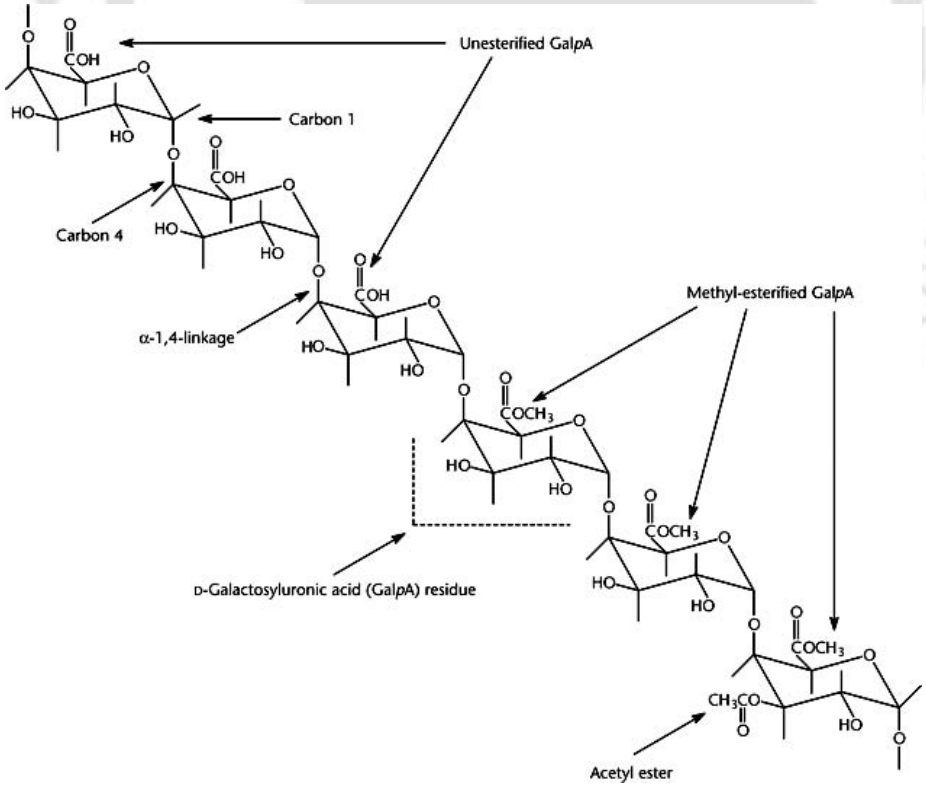


Fig. 1.13 Chemical structure of homogalacturonan showing 1,4-linked α-D-galactopyranosyluronic acid residues (GalpA). The carboxyl groups of the GalpA residues are often methyl-esterified. Some of the hydroxyl groups may be O-acetylated (Ridley *et al*., 2001).

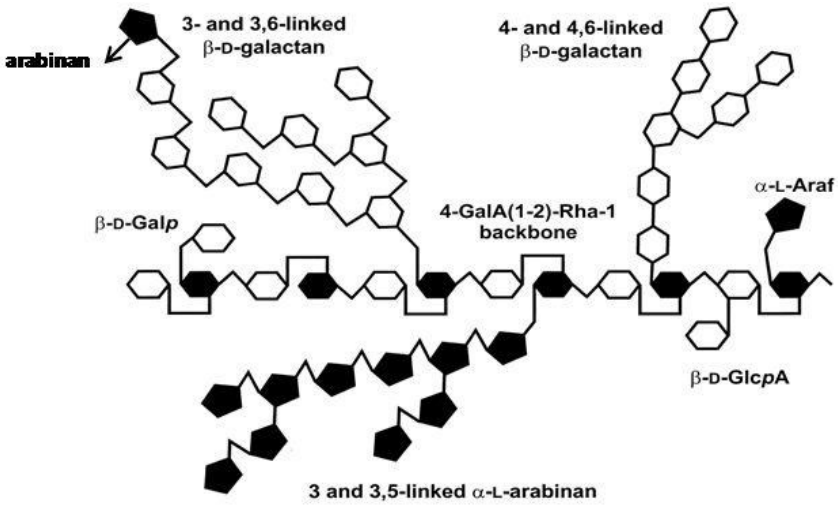


Fig. 1.14 A schematic presentation showing the major structural features of rhamnogalacturonan I. The backbone is composed of the disaccharide repeating unit [1→4- α-D-galactopyranoseA-(→2)-α-L-rhamnopyranose-(1→]. Branched and linear oligosaccharides composed predominantly of α-L-arabinofuranose and β- D -galactopyranose residues are linked to C4 of some of the rhamnopyranose residues. Some of the rhamnopyranose residue may also be O-acetylated at C2 and/or C3 (Ridley *et al*., 2001).

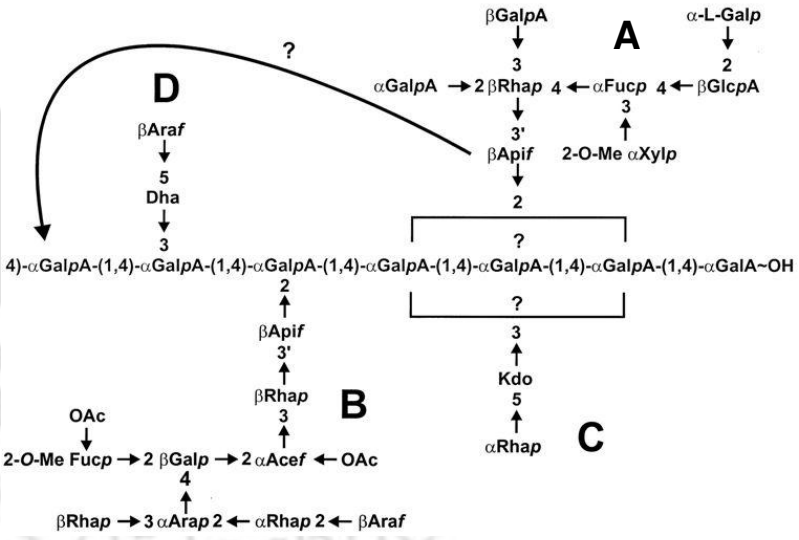


Fig. 1.15 The primary structure of rhamnogalacturonan II showing backbone of 1→4-linked α-D-GalpA dimers. Four structurally different oligosaccharide side chains (A, B, C and D) are linked to the rhamnogalacturonan II backbone. The two disaccharides, α-L-Rha*p*-(1→5)-Kdo (3-deoxy-D-manno-octulopyranosylonic acid) and β-L- Ara*f*-(1→5)-Dha, which are attached to position 3 of two of the 4-linked Gal*p*A residues, have been omitted. 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*) (Iishi *et al*., 1999).

1.2 Carbohydrate-active enzymes

Carbohydrates are dynamic molecules that are constantly synthesized and broken down. There are varieties of enzymes involved in the synthesis as well as breakdown of carbohydrates. The glycosyltransferases (GTs) are mainly involved in the formation of the glycosidic bond or biosynthesis of carbohydrates. The polysaccharide lyases (PLs), carbohydrate esterase (CEs) and glycoside hydrolases (GHs) are concerned with the breakdown of polysaccharides. In summary, the carbohydrate-active enzymes are grouped into 262 families based on amino acid sequence similarity and are listed in the continually updated carbohydrate-active enzyme (CAZy) database (www.cazy.org). Out of 262 carbohydrate-active enzymes, about 130 families belong to GHs (Cantarel *et al*., 2009). These enzymes belonging to 130 GH families have been reported from nearly 1884 species of bacteria, 130 species of archaea and 65 species of eukaryote (www.cazy.org). A close inspection of the genomes listed within the database reveals the fact that 1-3% of the genome of most organisms is devoted to encoding glycosyltransferases (GTs) and glycoside hydrolases (GHs) (www.cazy.org). The information available at CAZy database provide a wealth of gene sequences (many yet to be characterized) to study the structure and function of carbohydrate-active enzymes (Cantarel *et al.*, 2009).

1.2.1 Glycosyltransferases

Glycosyltransferases (GTs) are enzymes that catalyse the transfer of any sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds (Sinnot, 1990). These enzymes utilize 'activated' sugar phosphates as glycosyl donors and catalyze glycosyl group transfer to a nucleophilic group, usually an alcohol (Campbell *et al*., 1997). The product of glycosyl transfer may be an O-, N-, S-, or C-glycoside; the glycoside may be part of a monosaccharide,

oligosaccharide or polysaccharide (Lairson *et al*., 2008). As of now, nearly 94 families of GTs have been recognized and are listed in the CAZy database (http://www.cazy.org/GlycosylTransferases.html). There are almost over 8000 gene sequences in GenBank and the crystal structure of 36 GTs has been solved to date (http://www.cazy.org/GlycosylTransferases.html). Many GTs have been reported from a wide range of bacterial population *viz. Acidophilium, Actinoplanes, Bacillus, Clostridium, Gloeobacter, Lactobacillus,* etc (Coutinho *et al*., 2008).

1.2.2 Polysaccharide lyases

Polysaccharide lyases (PLs) are group of enzymes that break the uronic acid-containing polysaccharide chains following a β-elimination mechanism to liberate an unsaturated hexenuronic acid residue and a new reducing end (http://www.cazy.org/Polysaccharide-Lyases.html). The CAZy database contains 22 families of these enzymes classified, based on amino acid sequence similarities, which also reflect their structural features (Lombard *et al*., 2010). As of now, more than a thousand gene sequences of PLs have been deposited in the GenBank and out of them 546 sequences are from bacteria. Nearly 98 PLs have been characterized to date and the crystal structures of only 12 PLs have been solved (http://www.cazy.org/Polysaccharide-Lyases.html).

1.2.3 Carbohydrate esterase

The carbohydrate esterase (CEs) catalyzes the de-O or de-N-acylation of substituted saccharides. Two types of substrates have been considered for carbohydrate esterases: i) those in which the 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 ii) those 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 hydrolases (GHs) are enzymes that catalyze the hydrolysis of the glycosidic linkage of glycosides, leading to the formation of a sugar hemiacetal or hemiketal and the corresponding free aglycone. Glycoside hydrolases are also referred to as glycosidases, and sometimes also as glycosyl hydrolases. Glycoside hydrolases can catalyze the hydrolysis of O-, N- and S-linked glycosides. The glycoside hydrolases are a group of enzymes that exists in most living organism (Cantarel *et al*., 2009). They hydrolyze the glycosidic linkages between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (http://www.cazy.org/Glycoside-Hydrolases.html). The carbohydrate-active enzymes database (CAZy) provides a continuous updated list of the glycoside hydrolase families. They are classified into different families based on homology of their primary sequence (Henrissat, 1991). As of now, the GHs are grouped into 130 families based on amino acid sequence with more than 30,000 entries in the CAZy database (www.cazy.org). GHs are also grouped into 14 clans based on the fold of proteins as it was found to be better conserved than their amino acid sequence (Cantarel *et al.*, 2009). Even the closely related catalytic GHs are very often found to differ in substrate specificity (http://www.cazy.org/), while enzymes with very different enzyme activities are found among glycoside hydrolase homologs (Coutinho *et al.,* 2003). Small changes in the primary structure of glycoside hydrolases are able to change their substrate specificity (Andrews *et al.*, 2000). Unlike other CAZymes (GTs, CLs, and CEs), the structural data on GHs has clearly displayed several different folds, such as $(\alpha/\alpha)_6$, β-helix, β-propeller, β-jelly roll and the $(\alpha/\beta)_8$ TIM

barrel motif. The $(\alpha/\beta)_8$ TIM barrel is found in the majority of GHs listed in CAZy database (Cantarel *et al*., 2009).

GHs are important plant cell wall degrading enzymes and thus cell wall degradation by microbial enzymes is pivotal to many biological and industrial processes (Gilbert *et al*., 2010). The polysaccharides of plant cell walls are relatively recalcitrant to enzymatic degradation and due to this the microbes, with time, have evolved and developed a complex enzymatic systems in order to counter this problem. For example, *Clostridium thermocellum* and *Clostridium cellulolyticum* secrete a mega-dalton multi-modular enzyme complex called the "cellulosome" (Bayer *et al*., 2006; Fontes and Gilbert, 2010). The cellulosome is a macromolecular complex, whose components interact in a synergistic manner to catalyze the efficient degradation of cellulose. The cellulosome complex is composed of numerous kinds of cellulases and related enzyme subunits, which are assembled into the complex by virtue of a unique type of scaffolding subunit (Bayer *et al.,* 1998; Bayer *et al.,* 2004).

1.2.4.1 Glycoside hydrolases and their modular nature

A module is defined as a contiguous amino acid sequence within a larger sequence that folds independently and has an individual function but together increase the overall efficiency of the enzyme (Doi and Kosugi, 2004). The glycoside hydrolases are often found to exhibit a modular architecture comprising a catalytic module fused or attached with one or more ancillary modules via linker peptides. The ancillary modules are usually the carbohydrate binding modules (CBMs). The first indication that these enzymes contained distinct independent functioning modules was reported by Davies and Henrissat (1995, 2002) and later many researchers reported the same (Bayer *et al*., 2004). Fig. 1.16 shows typical example of a modular family 33

glycoside hydrolase (GH33) from *Micromonospora viridifaciens* displaying a catalytic GH33 module which is attached to a non catalytic family 22 carbohydrate binding module (CBM22) via a linker. The catalytic GH33 from *Micromonospora viridifaciens* was found to exhibit sialidase activity (Gaskell *et al*., 1995).

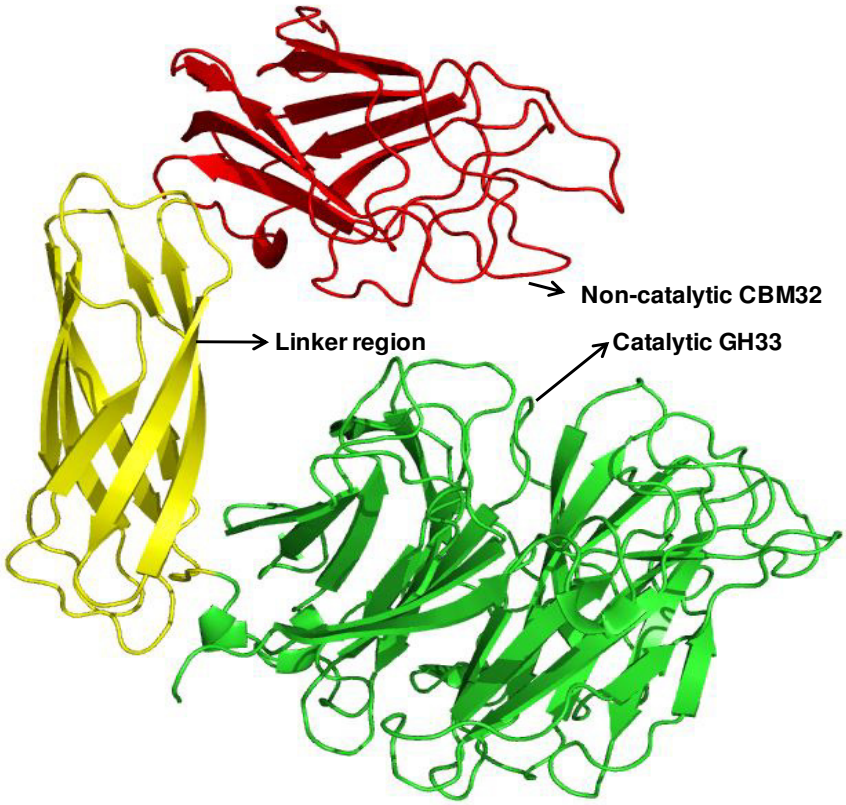


Fig. 1.16 The 3-dimensional structure (PDB Code 1EUT) of full length family 33 glycoside hydrolase (*Mv*GH33, sialidase) from *Micromonospora viridifaciens* as reported by Gaskell *et al.* (1995). The catalytic GH33 is shown in green; linker in yellow and carbohydrate-binding module (CBM32) is shown in red.

1.2.4.2 Cellulosome structure

The cellulosome is a macromolecular complex, whose components interact in a synergistic manner to catalyze the efficient degradation of cellulose (Bayer *et al.*, 2007). The cellulosome complex comprises numerous kinds of cellulases and related enzyme subunits, which are assembled into the complex by virtue of a unique type of scaffolding subunit known as scaffoldin (Bayer *et al*., 2004). The cellulosomal

enzymes from *Clostridium thermocellum* range in molecular size from about 40 to 180 kDa (Fontes and Gilbert, 2010). The realization about the multi-enzyme complex (cellulosome) occurred gradually with discoveries of different catalytic and binding domains, dockerins and scaffoldins, first in clostridial species and then subsequently in other bacteria (Doi *et al*., 1994; Belaich *et al.,* 1997; Doi and Kosugi, 2004). Each of the cellulosomal subunits consists of a multiple set of modules, two classes of which (dockerin domains on the enzymes and cohesin domains on scaffoldin) govern the incorporation of the enzymatic subunits into the cellulosome complex (Bayer *et al*., 1998). The cellulosomal enzymes are usually members of the glycosyl hydrolase families of enzymes, which hydrolyze oligosaccharides and polysaccharides (Henrissat, 1991; Fontes and Gilbert, 2010).

1.2.4.3 Glycoside hydrolases and their activity

The glycosidic bond linking two carbohydrates together or a carbohydrate to another non-sugar (aglycone) compound is generally a very stable bond with an estimated half life of 5 million years (Wolfenden *et al*. 1998). The GHs act to hydrolyze these glycosidic bonds and thereby increase bond cleavage efficiency up to ~ 10^{17} fold (Wolfenden *et al*. 1998). The substrates degraded by GHs include plant cell wall polysaccharides such as cellulose, hemicellulose and lignin.

1.2.4.4 Mechanism of action of Glycoside hydrolase

Glycoside hydrolases usually hydrolyze glycosidic bonds by adding water across the glycosidic bond. Asp or Glu are the most commonly found catalytically active residues, although a few exceptions do exist (Davies *et al*., 2002; Rajan *et al*., 2003). There are two fundamental mechanisms of action by which these enzymes

hydrolyze the substrates (www.cazy.org). These include i) Retaining mechanism and ii) Inverting mechanism and are explained in the sub sections 1.2.4.4.1 and 1.2.4.4.2.

1.2.4.4.1 Retaining mechanism

Glycoside hydrolase acting by retaining mechanism generally have a nucleophile and catalytic acid/base residues approximately 6Å apart (Zechel and Withers, 2000). This reaction proceeds *via* a double displacement mechanism (Fig 1.17A). In the first step, the nucleophilic attack occurs at the anomeric centre with immediate protonation of the glycosidic oxygen by the general acid (Fig 1.17A). This occurs through a transition state that has oxocarbenium ion character and results in the formation of a glycosyl-enzyme intermediate and results in retention of stereochemistry at the anomeric carbon as shown in Fig. 1.17A (Zechel and Withers, 2000). Hydrolysis of the intermediate occurs and simultaneously, the general base deprotonates an incoming water molecule, which further attacks the anomeric centre causing hydrolysis of the glycosyl-enzyme intermediate (Fig 1.17A).

1.2.4.4.2 Inverting mechanism

Inverting glycoside hydrolases require a larger distance between general base and general acid residues in order to accommodate the direct attack of a water molecule and the substrate (Zechel and Withers, 2000). This reaction proceeds *via* a single displacement mechanism (Fig. 1.17B). Water is deprotonated by the general base and attacks the anomeric centre while the general acid concomitantly protonates the leaving group (Fig. 1.17B). The inverting mechanism proceeds *via* an oxacarbenium ion-like transition state and results in overall inversion of stereochemistry at the anomeric carbon (Zechel and Withers, 2000).

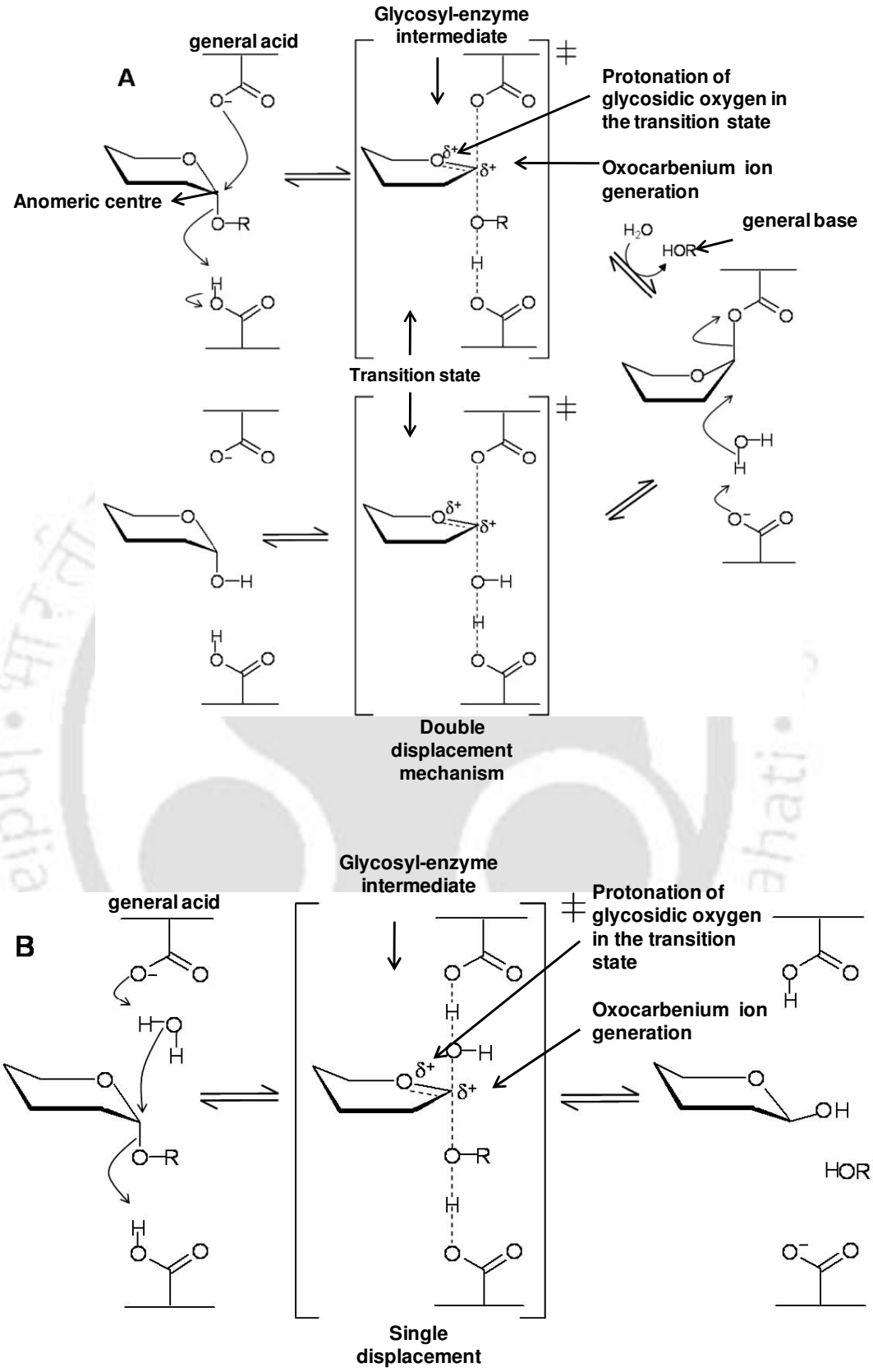


Fig. 1.17 Mechanisms of action of hydrolysis by glycoside hydrolases. **A)** Retaining mechanism, **B)** Inverting mechanism.

1.3 Family 43 glycoside hydrolases

The family 43 glycoside hydrolase (GH43) includes enzymes with the different activities like β-xylosidase (EC 3.2.1.37), β-1,3-xylosidase (EC 3.2.1.-), xylanase (EC 3.2.1.8), galactan 1,3- β-galactosidase (EC 3.2.1.145), α-L-arabinofuranosidase (EC 3.2.1.55) and arabinanase (EC 3.2.1.99). Family 43 GHs belong to GH-F clan of related families showing fold-5-β-propeller structure at its core (http://www.cazy.org/GH43.html). Its mechanism of action is by hydrolysis via overall inversion of the anomeric configuration. The presence of a catalytic triad of Asp, Glu and Asp has been reported at the active site as shown in Fig. 1.18 (http://www.cazy.org/GH43.html; Vandermarliere *et al*., 2009). The family 43 GHs act on both celluloses and hemicelluloses.

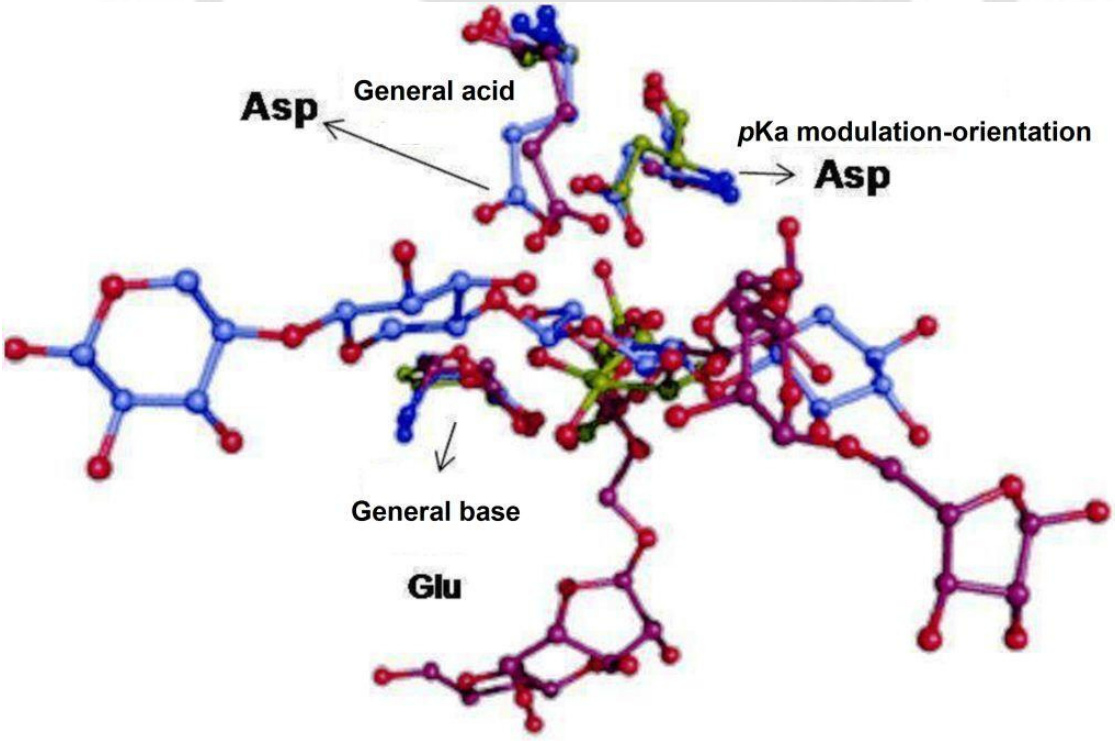


Fig. 1.18 General configuration of catalytic residues displaying orientation of a catalytic triad of Asp, Glu and Asp.

1.3.1 α-L-Arabinofuranosidases

The most recent classification scheme based on amino acid sequences, primary structure similarities and hydrophobic cluster analysis has classified α-L-arabinofuranosidases (α-L-Ara*f*) into five glycosyl hydrolase (GHs) families i.e. GH30, GH43, GH51, GH54 and GH62 (Zhou *et al*., 2012; Cartmell *et al*., 2011; Sørensen *et al.,* 2006; Guais *et al.,* 2010; Hashimoto *et al*., 2010). The α-L-arabinofuranoside arabinofuranohydrolases (α-L-Ara*f*, EC 3.2.1.55) are the enzymes involved in the hydrolysis of L-arabinosyl linkages. These enzymes have been purified from several bacteria, fungi and plants (Numan and Bhosle, 2005). They form an array of GHs required for the complete degradation of arabinose containing polysaccharides (Saha, 2000; Saha, 2003). The action of these enzymes accelerates the hydrolysis of the glycosidic bonds by more than 1000 fold, making them one of the most efficient catalysts known (Saha, 2000). Such enzymatic hydrolysis release soluble substrates which are utilized by both prokaryotic and eukaryotic microorganisms (Bayer *et al*., 2000). The α-L-Ara*f* specifically catalyze the hydrolysis of terminal non-reducing- α-L-1,2-, α-L-1,3-, and α-L-1,5-arabinofuranosyl residues from different oligosaccharides and polysaccharides (Numan and Bhosle, 2005). The α-L-Ara*fs* do not distinguish between the saccharides linked to the arabinofuranosyl moiety, which enables them to exhibit wide substrate specificity (Saha, 2003). Effective hydrolysis of α-L-arabinofuranosyl residues from various homohemicellulosic polysaccharides (branched arabinans, debranched arabinans), heteropolysaccharides (arabinogalactans, arabinoxylans, arabinoxyloglucans, glucuronoarabinoxylans, etc.), pectic and different glycoconjugates is carried out by α-L-Ara*f* (Cartmell *et al*., 2011; Mckee *et al*. 2012).

1.3.1.1 Biochemical properties of α-L-arabinofuranosidases

The biochemical properties of α-L-arabinofuranosidases (α-L-Araf) from various sources have been characterized (Numan and Bhosle, 2005). The molecular size of the α-L-Ara*f* varies and it can be as low as 30 kDa to as high as 495 kDa, although the usual size ranges between 30-80 kDa (Saha, 2003; Numan and Bhosle, 2005). The α-L-Ara*f* shows optimum pH range of 5-7 but there are few exceptions (few fungal α-L-Ara*f*). Similarly, the optimum temperature range is very wide (50-90°C) (Cartmell *et al*., 2011; McKee *et al*., 2012). The activities of α-L-Ara*f* are affected by metal ions, ionic and nonionic detergents, chelating and reducing agents depending on the enzyme and concentration of the agent used (Numan and Bhosle, 2005). In some cases, metal ions such as Ag^{+}, Cu^{2+}, Hg^{2+}, Zn^{2+}, Cd^{2+} and Co^{2+} displayed inhibitory effect on α-L-Ara*f* (Shin *et al*., 2003; Khandeparkar *et al*., 2008).

1.3.1.2 Molecular cloning of α-L-arabinofuranosidases

Some α-L-Ara*f* has been studied at the molecular level. The genes coding for these enzymes (α-L-Ara*f*) were identified, cloned and expressed in different bacterial and fungal systems. A few genes of α-L arabinofuranosidase have been sequenced and the evolutionary relationship among some of the sequenced proteins has been reported using the phylogenetic analysis (Degrassi *et al*., 2003). The gene sequencing results as well as the crystal structure studies reported previously indicated the presence of CBMs in some of the reported enzymes. The CBMs may take part in increasing the efficiency of the enzyme function (Numan and Bhosle, 2005). However, the possible roles of α-L-Ara*f* in the release of arabinofuranosyl residues is not yet clear (Numan and Bhosle, 2005).The α-L-Ara*f* (AkAbfB) from *Aspergillus kawachii* was reported to contain an arabinose-binding domain or module (ABD) which showed a number of

distinct characteristics that are different from those of carbohydrate-binding module (Shallom and Shoham, 2003).

1.3.1.3 Applications α-L-arabinofuranosidases of family 43 glycoside hydrolase

α-L-Arabinofuranosidases (α-L-Ara*f*) have been used synergistically along with xylanase for complete degradation of hemicellulose to pentose and hexose sugars for ethanol production and in the paper and pulp industry (Numan and Bhosle, 2005). α-L-Arabinofuranosidases have several applications which are described in sub sections 1.3.1.3.1, 1.3.1.3.2 and 1.3.1.3.3.

1.3.1.3.1 α-L-Arabinofuranosidase in acetic acid production and quality of bread

Pentose sugars (soluble pentosans) such as arabinose and xylose is important functional ingredient in bread and their positive role in bread texture and staling is well known (Gobbetti *et al*., 2000; Devesa *et al*., 2003). Wheat flour added to the dough may be moderately hydrolysed by wheat flour degrading enzymes and especially, by exogenous enzymes, such as xylan degrading system including α-L-arabinofuranosidase (Saha, 2000). The pentose sugars released by α-L-Ara*f* can be easily converted to acetic acid to be used as raw material for textile finishing agents (Saha, 2003).

1.3.1.3.2 α-L-Arabinofuranosidase in pulp and paper industry

Several commercial xylanase preparations are available for the treatment of pulp (Gubitz *et al*., 1997). Application of α-L-arabinofuranosidase further enhances the delignification of pulp as the enzyme acts to release arabinose side chain that retard the action of other bleaching enzymes (Numan and Bhosle, 2006). The removal

of lignin from semi-bleached kraft pulp was improved when the pulp was treated with α-L-arabinofuranosidase from *Bacills stearothermophilus L1* together with xylanase (Dhiman *et al*., 2000; Numan and Bhosle, 2006). The enzyme acted synergistically with a thermophilic xylanase in the delignification process, releasing 19.2% of lignin (Dhiman *et al*., 2000; Numan and Bhosle, 2006). The delignification obtained using the combined enzyme treatment is more than the sum of the amounts obtained using the enzymes separately (Gubitz *et al*., 1997).

1.3.1.3.3 L-Arabinose as antiglycemic agent

L-Arabinose released from arabinose containing polysaccharides by α-L-Ara*f* can be used as food additive because of its sweet taste and low uptake due to its poor absorption by the human body (Numan and Bhosle, 2006). It has been reported that L-arabinose selectively inhibits intestinal sucrase in a competitive manner and thus reduces the glycemic response after sucrose ingestion in animals so α-L-Ara*f* can be used as prebiotics (Saha, 2000; Numan and Bhosle, 2006).

1.4 Carbohydrate binding modules

The carbohydrate binding modules (CBMs) were described earlier as cellulose-binding domains (CBDs) as at that time only cellulose binding domains were discovered. As more diverse carbohydrate binding ligands were identified for new CBMs, the term CBDs was changed to CBMs. The first identification of a cellulose binding domain was illustrated in 1986 by the proteolytic degradation of a cellulase produced from the fungus *Trichoderma reesei* (Saha, 2003). It was observed that one domain retained the cellulase activity and the other domain had cellulose-binding capacity (Saha, 2003). Carbohydrate binding modules (CBMs) may be

defined as independently folding modules, occurring alongside the carbohydrate-active enzymes. The CBMs are the non-catalytic modules known to help or bring the catalytic modules in close proximity to its substrates and also some CBMs are known to stabilize the structure of catalytic modules and increase its thermal stability (Boraston *et al*., 2004; Henshaw *et al*., 2006; Dvortsov *et al*., 2009). The CBMs may be found to contain up to 200 amino acids and can be found attached as single, double or triple domain in one protein, located at both C- or N-terminal within the parental protein (Shoseyov *et al*., 2006). Similar to the catalytic modules of glycoside hydrolases, CBMs have been grouped into 64 families according to the data in Carbohydrate Active enZYmes (CaZy) database, based upon amino acid sequence similarity (http://www.cazy.org/Carbohydrate-Binding-Modules.html).

1.4.1 Binding site architecture of carbohydrate binding modules

The sequence-based carbohydrate binding module (CBM) families have been further subdivided into three types based on mechanism of ligand binding and binding site topology (Fig. 1.19). The type A, CBMs bind to insoluble crystalline cellulose or chitin and have flat platform-like binding sites (Boraston *et al.,* 2004; Abbott *et al*., 2009; Abbott *et al*., 2012,). The planar arrangement of aromatic residues of CBMs within the binding site, allows it to rest on the flat hydrophobic surface of a crystalline ligand (Fig. 1.19A). The type B, CBMs have a binding site that exists as an extended groove, within the groove there are multiple sugar sub sites (Fig. 1.19B). These CBMs bind chains of soluble polysaccharides and both polar and nonpolar interactions drive specificity of the binding. The type C, CBMs, considered lectin-like, have a shallow binding pocket ideal for binding mono-, di- or tri-saccharides (Fig. 1.19C). As with the Type B CBMs both polar and apolar interactions drive

specificity; however, the hydrogen bonding interactions are more dominant in type C CBMs (Boraston *et al*. 2004).

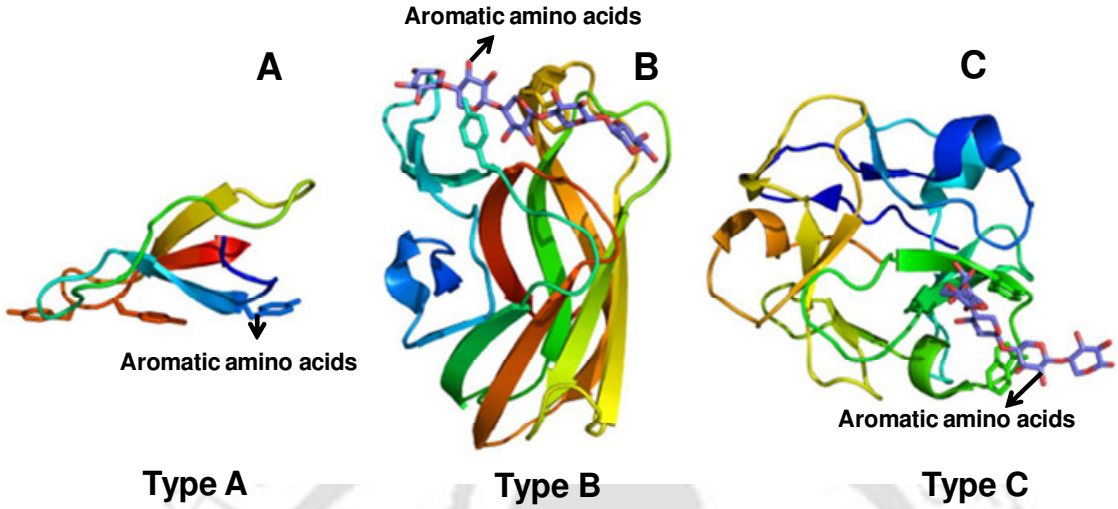


Fig. 1.19 Three types of binding site architecture and ligand binding in CBMs. Aromatic amino acids involved in binding are shown in stick **A)** Type A, cellulose binding *Hj*CBM1 from *Hypocrea jecorina* (PDB code 1CBH), **B)** Type B, *Ct*CBM6 from *Clostridium thermocellum* in complex with xylopentaose (PDB code 1UXX), **C)** Type C, *Sl*CBM13 from *Streptococcus lividans* in complex with xylopentaose (PDB code 1MC9).

1.4.2 Carbohydrate binding module clans based on fold of their 3-dimensional structure

Currently seven families of folds are reported for CBMs in the CAZy database (Cantarel *et al*. 2008). The seven different folds include the β-sandwich fold, the β-trefoil fold, the oligonucleotide-carbohydrate binding fold (OB), the knottin fold, the hevein fold, the hevein-like fold and a unique fold (http://www.cazy.org/CBM6.html). The most frequent is the β-sandwich fold which is common to plant legume lectins and animal galectins such as CBM6s reported by van Beuren *et al*. (2005) and Vandermarliere *et al*. (2009). The second most common fold is the β-trefoil fold reported in CBM6s from *Clostridium cellulolyticum* and *Clostridium thermocellum* by Abbott *et al.* (2009) and Czjzek *et al*. (2001), respectively.

1.4.3 Functions of carbohydrate binding modules

Carbohydrate binding modules (CBMs) increase the ability of glycoside hydrolases to efficiently degrade the polysaccharide substrates. This occurs through two different roles that CBMs play in polysaccharide breakdown, i) they localize the soluble enzyme to its target substrate (Shallom and Shoham, 2003), ii) they are associated with a catalytic domain and contain multiple clefts to attach with substrate, thereby enhancing the catalytic activity (Henshaw *et al.,* 2004; Hashimoto *et al*., 2006) iii) they can act as fusion proteins capable of binding to a cellulose matrix and consequently used in a protein purification (Shpigel *et al*., 1998).

The proximity effect describes the binding of the CBM that brings the catalytic module in close proximity (or association) to the substrate and maintains it for a prolonged period. This effect is seen primarily on insoluble substrates such as cellulose and xylan (Tomme *et al*. 1988; Boraston *et al*. 2004). The CBMs bring about the close proximity of catalytic enzyme to its target substrates by two effects as described in sub sections 1.4.3.1 and 1.4.3.2.

1.4.3.1 Targeting effect of carbohydrate binding modules

The targeting effect is observed where CBMs (associated with catalytic module) specifically bind to a substrate thereby bringing the substrate in close proximity of the enzyme (Shallom and Shoham, 2003). For example, there are many cellulose specific CBMs, however, type A cellulose binding CBMs bind crystalline cellulose, whereas, type B cellulose binding CBMs bind non-crystalline components of cellulose. Thus, two different binding mechanisms are driving the recognition of different components of cellulose substructure. This targeting effect would drive

hydrolysis in specific regions of cellulose, rather than just bringing the catalytic modules into proximity as described above (Carrard *et al*. 2000; Boraston *et al*. 2004).

1.4.3.2 Disruptive effect of carbohydrate binding modules

The disruptive effect may arise due to disruption of the polysaccharide fibres due to CBM permeation within the fibres. It was suggested that the CBMs bind and disrupt the crystalline cellulose (Din *et al*. 1994) or chitin (Vaaje-Kolstad *et al*. 2005) allowing the release of any non-covalently attached fibres thereby exposing the sites for polysaccharide hydrolysis.

1.4.4 Carbohydrate binding modules and multi-valency

Sometimes, more than one CBM may be found within a glycoside hydrolase. CBMs can occur side by side with one another within the enzyme though this is not always the case as the CBMs may also be separated by other modules. More than one CBM from the same family may occur in an enzyme, however, CBMs from different families may also occur within the same enzyme. CBM present side by side may show increased affinity for ligand over that of the individual CBMs though this is not always true (Tomme *et al*. 1998; Boraston *et al*. 2004; Abbott *et al*. 2009). Multivalent binding (two side by side CBMs binding to substrate) can help maintain the CBMs proximity to the carbohydrate surface or fine tune targeting.

1.4.5 Family 6 carbohydrate binding module

Family 6 carbohydrate binding module (CBM6s) are different from other CBM families in that, these modules are known to contain multiple distinct ligand binding sites. The CBMs of family 6 are known to contain modules of diverse

specificity and variation in the location of substrate binding site with respect to their 3-dimensional structure (Henshaw *et al*., 2004). This variation in ligand recognition is exemplified in CBM family 6 (CBM6), which contains proteins that recognize xylan (Cezjek *et al*., 2004), cellulose (β-1,4-linked glucose homopolymer) (Henshaw *et al*., 2004) , laminarin (β -1,3-linked glucose homopolymer), and β-1,4- and β -1,3-mixed linked β -glucans such as lichenan (Pires *et a*l., 2004; Boraston *et al*., 2004). In all crystal structures of CBM6, characterized to date, the conserved amino acid residues *viz.* Tyr-33, Trp-92 and Asn-120 have been reported to play a critical role in ligand binding (Pires *et al.*, 2004; Fontes *et al.,* 2004; Michel *et al*., 2009). The studies on CBMs that recognize the branched hemicelluloses xylan and galactomannan indicated that the side chains of decorated polysaccharides are usually solvent exposed and do not restrict ligand binding or represent the determinants of specificity (Hashimoto *et al*., 2006; Abbott *et al*., 2009; Montanier *et al*., 2011).

1.4.5.1 Type A family 6 carbohydrate binding module

Type A family 6 carbohydrate binding module (CBM6A) generally binds to crystalline cellulose and contains a planar hydrophobic ligand binding surface (Boraston *et al.,* 2004). CBM families, such as 1, 2A, 3A, 5 and 10 are known to contain "Type A" modules which bind crystalline polysaccharides. The ligand specificity of different members of this family is highly conserved even though they are present along with different glycoside hydrolases such as cellulases (Pires *et al*., 2004; Henshaw *et al.*, 2004), xylanases (Michel *et al*., 2009), mannanase (Boraston *et al.*, 2003), acetyl xylan esterase and arabinofuranosidases (Fontes and Gilbert, 2010).

1.4.5.2 Type B family 6 carbohydrate binding module

Type B family 6 carbohydrate binding modules (CBM6B) are believed to be evolved from 11 different families. They bind to individual polysaccharide chains and accommodate their target ligands in a cleft of varying depth and can also have multiple clefts for binding (Czjzek *et al.,* 2001; vanBueren *et al*., 2005). One of the features that distinguish Type B CBMs from lectins is the mechanism of ligand recognition. Each binding site in lectins recognizes one or two sugars through an extensive network of hydrogen bonds, while Type B CBMs generally accommodate four to six sugars, with specificity conferred primarily by the conformation of the ligand, which reflects the topology of the binding site.

1.4.5.3 Type C family 6 carbohydrate binding module

Type C family 6 carbohydrate binding module (CBM6C) are commonly referred to as small sugar-binding CBMs and known to exist in CBM families 9, 13, 14, 18 and 32 (Boraston *et al*., 2004). They have lectin-like properties and bind to mono-, di- as well as tri-saccharides. There are CBM6Cs that recognize crystalline cellulose, non-crystalline cellulose, chitin, β-1,3-glucans and β-1,3-1,4-mixed linkage glucans, xylan, mannan, galactan and starch, while some CBMs display 'lectin-like' specificity and bind to a variety of cell-surface glycans (Abbott *et al.,* 2004; Abbott *et al.*, 2009).

1.4.6 Applications of carbohydrate binding modules

Three basic features have led to CBMs being perfect candidates for many 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 can be controlled by modifying the structure using tools such as site directed mutagenesis and therefore the right solution can be adopted for a particular problem (Abbott *et al*., 2009).

1.4.6.1 Bioprocessing

Bio-specific affinity purification (affinity chromatography) has become one of the most rapidly developing divisions of immobilized affinity ligand technology. Many protein entities have been expressed when fused to CBMs, establishing CBMs as high-capacity purification tags for the isolation of biologically active target peptides at relatively low cost (Shosheyov *et al.,* 2006; Bolam *et al*., 2004; Abbott *et al*., 2009). A high-level production of a cellulose binding domain (CBD_{Cex}) of cellulase (Cex) from *Cellulomonas fimi* expressed in *E. coil*, served as affinity tag in a novel secretion-affinity fusion system for purification of recombinant exoglucanase (Hasenwinkle *et al*., 1997). Also, a strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *Escherichia coli* for purification of recombinant proteins was developed (Kavoosi *et al*., 2007).

1.4.6.2 Cell immobilization using carbohydrate binding modules

Surface-exposed CBMs can be an efficient means of whole-cell immobilization. Whole-cell immobilization by cellulosic material was first demonstrated when an *E. coli* surface anchored CBM, derived from *Cellulomonas fimi*, was attached to cellulose (Saha, 2000). The cells bound tightly to cellulose at a wide range of pH and the extent of immobilization was dependent on the surface of exposed CBM (Saha, 2000; Numan and Bhosle, 2006).

1.4.6.3 Bio-engineering of carbohydrate binding modules for different applications

The potential of carbohydrate binding modules (CBMs) for modifying the characteristics of several enzymes has been reported. The basic approach in CBM engineering was to replace or add a CBM in order to improve hydrolytic activity. Addition of a CBM derived from cellobiohydrolase II of *Trichoderma reesei* to *Trichoderma harzianum* chitinase resulted in increased hydrolytic activity of insoluble substrates (Shoseyov *et al*., 2006). The replacement of the CBM of endo-1,4-β-glucanase from *Bacillus subtilis* with the CBM of exoglucanase I (Texl) from *Trichoderma viride* conferred higher binding with enhanced hydrolytic activity on the microcrystalline cellulose (Shoseyov *et al*., 2006). In addition, the hybrid enzyme was more resistant to the tryptic digestion (Shoseyov *et al*., 2006).

1.5 The microorganism

Clostridium thermocellum is an anaerobic, thermophilic, cellulolytic and ethanologenic, Gram-positive bacterium capable of directly converting cellulose biomass into ethanol (Johnson *et al*., 1982; Bayer *et al*., 2000). The general cellular structure of *C. thermocellum* is similar to most rod-shaped bacteria as shown in Scanning Electron Microscopic images in Fig. 1.20 (Lamed *et al*., 1987). The unstained *C. thermocellum* YS cells show typical rod-shaped cells (Fig. 1.20A). However, the fibrous and protuberant structures on the cell surface are visible only when the cell is stained with cationized ferritin as shown for *C. thermocellum* ATCC 27405 (Fig. 1.20B). The cationized ferritin (CF) stained *Clostridium thermocellum* YS structure as seen in Transmission Electron Microscope (TEM) is displayed in Fig. 1.21. The TEM image of *C. thermocellum* clearly showed fibrous and protuberant structures on the cell surface (Fig. 1.21). The three major types of labelling: the

monolayer (m) of CF particles which envelop the entire cell surface, the fibrous structures (f) which sometimes connect two adjacent cells, and the nodulous protuberances (p) which appear in large numbers over the entire cell surface (Fig. 1.21). In tangentially sectioned areas of the cell surface (t), there are indications that the protuberances may be interconnected secondarily by low-lying structures (l) (Fig. 1.21).

The capability of *Clostridium thermocellum* cellulosomal enzymes to directly convert the cellulose biomass into a usable energy source *viz*. bio-fuel or ethanol makes it useful. However, there are some shortfalls in applying the organism to practical applications due to it having low ethanol yield, at least partially due to branched fermentation pathways that produce acetate, formate and lactate along with ethanol. Recent research has been directed to optimizing the ethanol-producing metabolic pathway in the hopes of creating more efficient biomass conversion (Zhang and Lynd, 2005). Biotechnological research has shown that the cellulose degrading bacteria produces a large, complex cellulase system known as the cellulosome which consists about 20 catalytic proteins that are involved in the bacteria's adherence to cellulose, breakdown and regulation of cellulose degradation and the transport of sugar monomers (Bayer *et al*., 2000; Fontes and Gilbert, 2010).

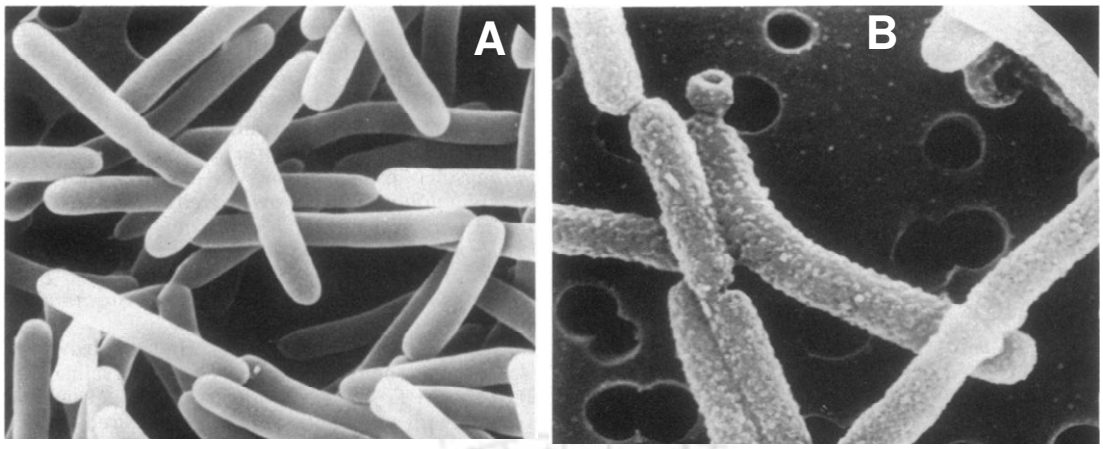


Fig. 1.20 Scanning electron microscope (SEM) images of *Clostridium thermocellum,* **A)** showing normal rod shaped cells of *Clostridium thermocellum* YS and **B**) shows cationized ferritin (CF) stained *Clostridium thermocellum* ATCC 27405 rod shaped cells with protuberant structures on cell surface (adapted from Lamed *et al*., 1987).

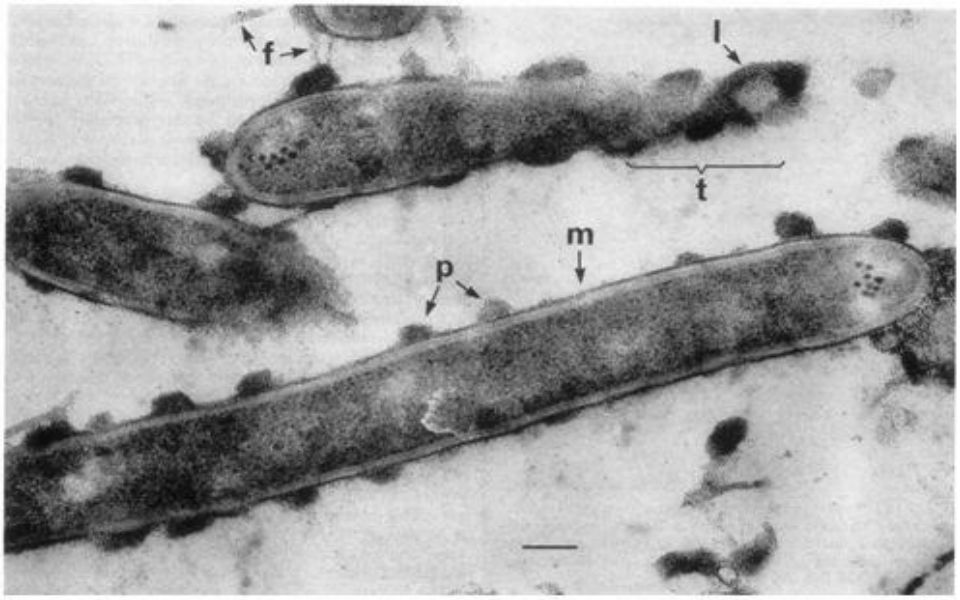


Fig. 1.21 Transmission electron microscope (TEM) image of cationized ferritin (CF) stained *Clostridium thermocellum* YS grown on cellobiose. The three major types of labelling: the monolayer (m) of CF particles enveloping the entire cell surface, the fibrous structures (f) which connecting two adjacent 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).

1.6 Objectives of the present study

1.6.1 Why study family 43 GHs and associated family 6 CBMs from Clostridium thermocellum

The reason for selecting the family 43 glycoside hydrolase (GH43) and associated family 6 carbohydrate binding modules (CBM6s) from *Clostridium thermocellum* are summarized as given below:

1. *Clostridium thermocellum* genome contains a complex cellulosome that contains a large number of enzymes, capable of degrading cellulose and hemicellulose.

2. *Clostridium thermocellum* cellulosomal enzyme complex is 50 times more efficient in degrading polysaccharides as reported by Fontes and Gilbert (2010).

3. Functional characterization of family 43 glycoside hydrolase (GH43) is important as all the enzymes belonging to family 43 GH may have the same inverting mechanism of catalysis though having different activites (Cantarel *et al.*, 2009).

4. Family 43 GHs are generally hemicellulases which are essential for complete degradation of xylans, arabinoxylans, galactans, arabinans and arabinogalactans into respective pentose sugars that can be utilized for bioethanol production.

5. It is important to determine the role of CBMs in substrate binding if they are altering the activity of the catalytic modules.

In the present study the full length family 43 glycoside hydrolase designated as *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were investigated and functionally characterized. The full length family 43 glycoside hydrolase (*Ct*43Ara*f*) molecular architecture of the primary sequence showed that it consisted of an N-terminal glycoside hydrolase catalytic module (903bp, *Ct*GH43) and two carbohydrate binding modules (CBMs) named as *Ct*CBM6A (405 bp) and *Ct*CBM6B (402 bp) at the C-terminus. The influence of CBMs on catalytic modules and its binding affinities towards soluble and insoluble polysaccharides were also analyzed. This is necessary as CBM6 had been found to bind substrate in two different places, known as cleft A and cleft B (Abbott *et al*., 2009). Biochemical and functional characterization of *Ct*43Ara*f* and *Ct*GH43 becomes essential as all the family 43 glycoside hydrolases may have the same inverting mechanism of catalysis, but their enzyme activity or substrate specificity are different (http://www.cazy.org/GH43.html; 30). The following objectives were defined for the thesis work:

1.6.2 Specific Objectives

1. PCR amplification, cloning of full length family 43 glycoside hydrolase (*Ct*43Ara*f*) and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B from *Clostridium thermocellum.*

2. Cloning, expression and purification of family 43 glycoside hydrolase (Ct43Araf) and its truncated derivatives CtGH43, CtCBM6A and CtCBM6B from *Clostridium thermocellum* ATCC 27405.

3. Biochemical, functional and structural characterization of catalytic modules Ct43Araf and CtGH43 from *Clostridium thermocellum.*

4. Binding analysis of *Ct*CBM6A and *Ct*CBM6B of *Ct*43Ara*f* from *Clostridium thermocellum*.

5. *In silico* analyses of substrate-function and ligand binding sites of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B from *Clostridium thermocellum*.

References

Abbott, D.W. and Boraston, A.B. (2012) Quantitative approaches to the analysis of carbohydrate-binding module function. Methods Enzymol. 510, 211-231.

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

Andrews, S.R., Charnock, S.J., Lakey, J.H., Davies, G.J., Claeyssens, M., Nerinckx, W., Underwood, M., Sinnott, M.L., Warren, R.A. and Gilbert, H.J. (2000) Substrate specificity in glycoside hydrolase family 10. Tyrosine 87 and leucine 314 play a pivotal role in discriminating between glucose and xylose binding in the proximal active site of *Pseudomonas cellulosa* xylanase 10A. J. Biol. Chem. 275, 23027-23033.

Bayer, E.A., Belaich, J.P., Shoham, Y. and Lamed, R. (2004) The cellulosomes: multi-enzyme machines for degradation of plant cell wall polysaccharides. Annu. Rev. Microbiol. 58, 521-554.

Bayer, E.A., Chanzy, H., Lamed, R. and Shoham, Y. (1998) Cellulose, cellulases and cellulosomes. Curr. Opin. Struct. Biol. 8, 548-557.

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

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

Bayer, E. A., Morag, E., Lamed, R., Yaron, S. and Shoham, Y.(1998b) Cellulosome structure: Four-pronged attack using biochemistry, molecular biology, crystallography and bioinformatics, In *Carbohydrases from Trichoderma reesei and Other Microorganisms*, Claeyssens, M., Nerinckx, W., and Piens, K. (Eds.), The Royal Society of Chemistry, London, 39-67.

Bayer, E.A., Shoham, Y. and Lamed, R., (2000) Cellulose-decomposing prokaryotes and their enzyme systems. (3rd ed.) Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., and Stackebrandt, E. (ed.), In *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community*, 2, 578-617.

Bayer, E. A., Shoham, Y. and Lamed R. (2000) The cellulosome-an exocellular organelle for degrading plant cell wall polysaccharides. In Doyle, R.J. (ed.). *Glycomicrobiology*. Kluwer Academic/Plenum Publishers, New York. 387-439.

Belaich, J.P., Tardif, C., Belaich, A. and Gaudin, C. (1997) The cellulolytic system of *Clostridium cellulolyticum,* J. Biotechnol. 57, 3–14.

Berg, J.M. (2007) In *Biochemistry,* (6th ed). New York: W.H. Freeman. 310-323.

Bolam, D.N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B. and Gilbert, H.J. (2004) X4 modules represent a new family of carbohydrate-binding modules that display novel properties. J. Biol. Chem. 279, 22953-22963.

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

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

Boerjan, W., Ralph, J. and Baucher, M. (2003) Lignin biosynthesis. Annu. Rev. Plant Bio. 54, 519-546.

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.

Campbell, J.A., Davies, G.J., Bulone, V. and Henrissat, B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J. 326:929-939.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37, 233-238.

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

Chen, H., Li, X., Blum, D. and Ljungdahl, L. (1998) Two genes of the anaerobic fungus *Orpinomyces* sp. strain PC-2 encoding cellulases with endoglucanase activities may have arisen by gene duplication, FEMS Microbiol. Lett. 159, 63–68.

Czjzek, M., Bolam, D.N., Mosbah, A., Allouch, J., Fontes, C.M.G.A., Ferreira, L.M., Bornet, O., Zamboni, V., Darbon, H., Smith, N.L., Black, G.W., Henrissat,

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

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

Cosgrove, D.J. (1999) Enzymes and other agents that enhance cell wall extensibility. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50, 391–417.

Coutinho, P.M., Deleury, E., Davies, G.J. and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. J. Mol. Biol. 328, 307-317.

Coutinho, P.M., Stam, M., Blanc, E. and Henrissat, B. (2003) Why are there so many carbohydrate-active enzyme-related genes in plants? Trends Plant Sci. 8, 563-565.

Cui, S.W. and Wang, Q. (2009) Cell wall polysaccharides in cereals: chemical structures and functional properties. Struct. Chem. 20, 291-297.

Cyran, M., Courtin, C.M. and Delcour, J.A. (2003) Structural features of arabinoxylans extracted with water at different temperatures from two rye flours of diverse bread making quality. J. Agric. Food. Chem. 51, 4404-4416.

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

Davies, G.J. and Henrissat, B. (2002) Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era. Biochem. Soc. Trans. 30, 291-297.

Degrassi, G., Vindigni, A. and Venturi, V. (2003) A thermostable alpha-arabinofuranosidase from xylanolytic *Bacillus pumilus*: purification and characterisation. J. Biotechnol. 101, 69-79.

Demain, A.L., Newcomb, M. and Wu, J.H. (2005) Cellulose, clostridia and ethanol. Microbiol. Mol. Biol. Rev. 69, 124-154.

Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. Curr. Opin. Microbiol. 11, 87–93.

Devesa, A., Antonia, M. and Anaya, M. (2003) Influence of pentosans on texture of starch gels during storage, and effects after enzyme treatment. Eur. Food. Res. Technol. 216, 323-330.

Dhiman, S.S., Sharma, J. And Battan, B. (2000) Industrial applications and future prospects of microbial xylanases: a review. Bioresources, 3, 1-26.

Din, N., Forsythe, I.J., Burtnick, L.D., Gilkes, N.R., Miller, R.C.Jr., Warren, R.A. and Kilburn, D.G. (1994) The cellulose-binding domain of endoglucanase A (CenA) from *Cellulomonas fimi*: evidence for the involvement of tryptophan residues in binding. Mol. Microbiol. 11, 747-55.

Doi, R.H., Goldstein, M., Hashida, S., Park, J.S. and Takagi, M. (1994) The *Clostridium cellulovorans* cellulosome. Crit. Rev. Microbiol. 20, 87-93.

Doi, R.H. and Kosugi, A. (2004) Cellulosome: plant-cell-wall degrading enzyme complexes. Nat. Rev. Microbiol. 2, 541-551.

Dvortsov, I.A., Lunina, N.A., Chekanovskaya, L.A., Schwarz, W.H., Zverlov, V.V. and 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.

Fanutti, C., Ponyi, T., Black, G.W., Hazlewood, G. P. and Gilbert, H. J. (1995) The conserved non catalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. J. Biol. Chem. 270, 29314-29322.

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

Fontes, C.M., Ponte, P.I., Reis, T.C., Soares, M.C., Gama, L.T., Dias, F.M. and Ferreira, L.M. (2004) A family 6 carbohydrate-binding module potentiates the efficiency of a recombinant xylanase used to supplement cereal-based diets for poultry. Br. Poult. Sci. 45, 648-656.

Fry, S.C. (1989) The structure and functions of xyloglucan. J. Exp. Bot. 40, 1-11.

Gaskell, A., Crennell, S. and Taylor, G (1995) The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure, 3, 1197-1205.

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

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

Guais, O., Tourrasse, O., Dourdoigne, M., Parrou, J.L. and Francois, J.M. (2010) Characterization of the family GH54 α-l-arabinofuranosidases in *Penicillium funiculosum*, including a novel protein bearing a cellulose-binding domain. Appl. Microbiol. Biotechnol. 87, 1007-1021.

Gu¨bitz, G.M., Haltrich, D., Latal, B. and Steiner, W. (1997) Mode of depolymerisation of hemicellulose by various mannanases and xylanases in relation to their ability to bleach softwood pulp. Appl. Microbiol. Biotechnol. 47, 658-621.

Hasenwinkle, D., Jervis, E., Kops, O., Liu, C., Lesnicki, G., Haynes, C.A. and Kilburn, D.G. (1997) Very high-level production and export in *Escherichia coli* of a cellulose binding domain for use in a generic secretion-affinity fusion system. Biotechnol. Bioeng. 55, 854-863.

Hashimoto, K., Yoshida, M. and Hasumi, K. (2011) Isolation and characterization of CcAbf62A, a GH62 α-L-arabinofuranosidase, from the basidiomycete *Coprinopsis cinerea*. Biosci. Biotechnol. Biochem. 75, 342-345.

Heinze, T., Schwikal, K. and Barthel, S. (2005) Ionic liquids as reaction medium in cellulose functionalization. Macromol. Biosci. 5, 520-525.

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

Henshaw, J.L., Bolam, D.N., Pires, V.M., Czjzek, M., Henrissat, B., Ferreira, L.M., Fontes, C.M.G.A. and Gilbert, H.J. (2004) The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. J. Biol. Chem. 279, 21552-21559.

Henshaw, J., Horne-Bitschy A., van Bueren, A.L., Money, V.A., Bolam, D.N., Czjzek, M., Ekborg, N.A., Weiner, R.M., Hutcheson, S.W., Davies, G.J.,

Boraston, A.B. and Gilbert, H.J. (2006) Family 6 carbohydrate binding modules in β-Agarases display exquisite selectivity for the non-reducing termini of agarose chains. J. Biol. Chem. 281, 17099-17107.

Hon, D.N.Sr. (1996) Functional polymers: a new dimensional creativity in lignocellulosic chemistry. In *Chemical Modification of Lignocellulosic Materials*. Dekker, New York, 1-10.

Ishii, T., Matsunaga, T., Pellerin, P., O'Neill, M.A., Darvill, A. and Albersheim, P. (1999) The plant cell wall polysaccharide rhamnogalacturonan II self-assembles into a covalent cross-linked dimer. J. Biol. Chem. 274, 13098-13109.

Ishii, T. and Ono, H. (1999) NMR spectroscopic analysis of the borate diol esters of methyl apiofuranoside. Carbohydr. Res. 321, 257-260.

Khandeparkar, K. and Bhosle, N.B. (2007) Application of thermoalkalophilic xylanase from *Arthrobacter* sp. MTCC 5214 in biobleaching of kraft pulp. Bioresour. Technol. 98, 897-903.

Kavoosi, M., Creagh, A.L., Kilburn, D.G. and Haynes, C.A. (2007) Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *Escherichia coli*. Biotechnol. Bioeng. 98, 599-610.

Klemm, D., Philipp, B., Heinze, T., Heinze, U. and Wagenknecht, W. (1998) In *Comprehensive Cellulose Chemistry*, (ed 1), (Vol 1 and 2), Wiley-VCH, Weinheim.

Kobayashi, M., Matoh, T. and Azuma, J. (1996) Two chains of rhamnogalacturonan II are cross linked by borate-diol ester bonds in higher plant cell walls. Plant Physiol. 110, 1017–1020.

Lairson, L.L., Henrissat, B., Davies, G.J. and Withers, S.G. (2008) Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. 77, 521-555.

Lamed, R., Naimark, J., Morgenstern, E. and Bayer, E.A. (1987) Specialized cell surface structures in cellulolytic bacteria. J. Bacteriol. 169, 3792-3800.

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

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

McKee, L.S., Peña, M.J., Rogowski, A., Jackson, A., Lewis, R.J., York, W.S., Krogh, K.B., Viksø-Nielsen, A., Skjøt, M., Gilbert, H.J. and Wright, J.M (2012) Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. Proc. Natl. Acad. Sci. (USA). 109, 6537-6542.

Michel, G., Barbeyron, T., Kloareg, B. and Czjzek M. (2009) The family 6 carbohydrate-binding modules have coevolved with their appended catalytic modules toward similar substrate specificity. Glycobiology, 19, 615-623.

Montanier, C., Money, V.A., Pires, V.M.R., Flint, J.E., Pinheiro, B.A., Goyal, A., Prates, J.A.M., Izumi, A., Stalbrand, H., Morland, C., Cartmell, A., Kolenova, K., Topakas, E., Dodson, E.J., Bolam, D.N., Davies, G.J., Fontes, C.M.G.A. and Gilbert, H.J. (2009) The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. PLoS Biol. 31, e71.

Montanier, C.Y., Correia, M.A., Flint, J.E., Zhu, Y., Baslé, A., McKee, L.S., Prates, J.A., Polizzi, S.J., Coutinho, P.M., Lewis, R.J., Henrissat, B., Fontes, C.M.

and Gilbert, H.J. (2011) A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. J. Biol. Chem. 286, 22499-22509.

Nishitani, K. and Nevins, D.J. (1991) Glucuronoxylan xylanohydrolase. A unique xylanase with the requirement for appendant glucuronosyl units. J. Biol. Chem. 266, 6539-6543.

Numan, M.T. and Bhosle, N.B. (2006) Alpha-L-arabinofuranosidases: the potential applications in biotechnology. J. Ind. Microbiol. Biotechnol. 33, 247-260.

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

Rajan, S.S., Yang, X., Collart, F., Yip, V.L., Withers, S.G., Varrot, A., Thompson, J., Davies, G.J. and Anderson, W.F. (2004) Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD^+/Mn^{2+} -dependent phospho-alpha-glucosidase from *Bacillus subtilis*. Structure, 12, 1619-1629.

Ralph, J., Lapierre, C., Marita, J.M., Kim, H., Lu, F., Hatfield, R.D., Ralph, S., Chapple, C., Franke, R., Hemm, M.R., Van Doorsselaere, J., Sederoff, R.R., O'Malley, D.M., Scott, J.T., MacKay, J.J., Yahiaoui, N., Boudet, A., Pean, M., Pilate, G., Jouanin, L. and Boerjan, W. (2004) Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. Phytochemistry, 57, 993-1003.

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

Ramamurthy, V., Thacker, S.P. and Kothari, R.M. (1992) Optimised protocol for the pilot-scale preparation of fungal cellulase. J. Ind. Microbiol. 9, 121-125.

Saha, B.C. (2000) Alpha-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. Biotechnol. Adv. 18, 403-423.

Saha, B.C. (2003) Hemicellulose bioconversion. J. Ind. Microbiol. Biotechnol. 30, 279-291.

Scheller, H.V. and Ulvskov, P. (2010) Hemicellulose. Annu. Rev. Plant. Biol. 61, 263-289.

Schladel, C., Blochl, A., Richter, A. and Hoch, G. (2009) Short-term dynamics of nonstructural carbohydrates and hemicelluloses in young branches of temperate forest trees during bud break. Tree Physiol. 29, 901–911.

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

Shpigel, E., Elias, D., Cohen, I. R. and Shoseyov, O. (1998) Production and purification of a recombinant human hsp60 epitope using the cellulose-binding domain in *Escherichia coli*. Protein Expr. Purif. 14, 185-191.

Sørensen, H.R., Jørgensen, C.T., Hansen, C.H., Jørgensen, C.I., Pedersen, S. and Meyer, A.S. (2006) A novel GH43 alpha-L-arabinofuranosidase from *Humicola insolens*: mode of action and synergy with GH51 alpha-L-arabinofuranosidases on wheat arabinoxylan. Appl. Microbiol. Biotechnol. 73, 850-861.

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

Sinnott, M.L. (1990) Catalytic mechanisms of enzymatic glycosyl transfer. Chem. Rev. 90, 1171-1202.

Stewart, J.J., Akiyama, T., Chapple, C., Ralph, J. and Mansfield, S.D. (2009) The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. Plant Physiol. 150, 621-635.

Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A.L., Sturch, K., Gilkes, N.R., Haynes, C.A., Warren, R.A. and Kilburn, D.G. (1998) Characterization and affinity applications of cellulose-binding domains. J. Chromatogr. B. 715, 283–96.

Urbanowicz, B.R., Peña, M.J., Ratnaparkhe, S., Avci, U., Backe, J., Steet, H.F., Foston, M., Li, H., O'Neill, M.A., Ragauskas, A.J., Darvill, A.G., Wyman, C., Gilbert, H.J. and York, W.S. (2012) 4-O-methylation of glucuronic acid in Arabidopsis glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. Proc. Natl. Acad. Sci. (USA). 109, 14253-14258.

van Bueren, A.L., Morland, C., Gilbert, H.J. and Boraston, A.B. (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of β-1,3-linked glcans by presenting a unique ligand binding surface. J. Biol. Chem. 280, 530-537.

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

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

Wolfenden, R., Lu, X. and Young, G. (1998). Spontaneous hydrolysis of glycosides. J. Am. Chem. Soc. 120, 6814-6815.

Zechel, D.L., Withers, S.G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. Acc. Chem. Res. 33, 8-11.

Zhang, Y.H. and Lynd, L.R. (2005) Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. Proc. Natl. Acad. Sci. (USA). 102, 7321-7325.

Zhou, J., Bao, L., Chang, L., Zhou, Y. and Lu, H. (2012) Biochemical and kinetic characterization of GH43 β-d-xylosidase/α-L-arabinofuranosidase and GH30 α-l-arabinofuranosidase/β-d-xylosidase from rumen metagenome. J. Ind. Microbiol. Biotechnol. 39, 143-152.

Chapter 2

Cloning, expression and purification of family 43 glycoside hydrolase (*Ct*43Ara*f*) and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B from *Clostridium thermocellum* ATCC 27405

2.1 Introduction

The most easily available and useful biopolymers on the earth are derived from plant cell walls. The major structural components of primary plant cell wall are mainly celluloses and hemicelluloses. These are hydrolysed by a large number of glycoside hydrolases, such as cellulases, xylanases, xylosidases, arabinofuranosidases, galactosidases, glucouronidases, glucanases etc. (Bayer *et al*., 2000a; Bayer *et al*., 2000b). The cellulolytic and hemicellulolytic complex commonly referred to as cellulosome, have been reported to have 20 or more polypeptides and many of which have been sequenced (Shoham *et al.*, 1999; Bayer *et al.*, 2004; Fontes and Gilbert, 2010). The glycoside hydrolases (GHs) have been grouped into more than 127 families based on the nomenclature of the CAZy (**C**arbohydrate-**A**ctive en**Zy**mes or CAZymes) database (http://www.cazy.org/fam/acc_GH.html; Cantarel *et al*., 2009), which describes families of structurally similar catalytic and carbohydrate binding modules (Davies and Henrissat 2002; Cantarel *et al.*, 2009). Now, structurally related

enzymes from same GH family may have different substrate specificities, that is to say a given enzyme may hydrolyze a set of diverse substrates. These diversities reflect their significance in specific biological processes at different stages of a plant owing to the fact that the GHs in plant genomes often belong to multi-gene family (Gilbert, 2010). The family 43 glycoside hydrolases (GH43) currently contain nearly 80 open reading frames and the well characterized enzymes so far have shown activities like β-xylosidase (EC 3.2.1.37), β-1,3-xylosidase (EC 3.2.1.72), α-L-arabinofuranosidase (EC 3.2.1.55), bi-functional β-1,3-xylosidase/α-L-arabinofuranosidase, endo-arabinase (EC 3.2.1.99), β-xylanase (EC 3.2.1.8) and galactan 1,3-β-galactosidase (EC 3.2.1.145) (http://www.cazy.org/GH43.html). All the members of GH43 have been reported to have the same holding and catalysis in the same inverting mechanism although the activities are different (Cartmell *et al.*, 2011).

The non-catalytic polysaccharides recognizing modules were initially named cellulose binding domain (Bayer, 2000a). Later on the term CBM (carbohydrate binding module) evolved reflecting the diversity in ligand specificity of these proteins (Boraston *et al.*, 2004; Shoseyov *et al*., 2006). Based on the amino acid sequence similarity, currently there are 64 defined families of CBMs (http://www.cazy.org/Carbohydrate-Binding-Modules.html), and they display substantial diversity in terms of ligand specificity. The CBMs append to the glycoside hydrolases that degrade the soluble or insoluble polysaccharides (Bolan *et al*., 2004). The CBMs are categorized into three types based on their structure and functional similarities: (i) surface binding (type A), (ii) binding with single polysaccharide chain (type B) and (iii) small sugar binding (type C) (van Bueren *et al.,* 2005; Shoseyov *et al*., 2006). In previous studies the family 6 carbohydrate binding modules (CBM6) have been shown to bind the insoluble cellulose involves synergistic interactions

between cleft A and cleft B (Henshaw *et al.,* 2004; Pires *et al.,* 2004). This also suggested that CBM6 contains two binding sites that display differences in ligand specificity, supporting the view that distinct binding clefts with different specificities can contribute to the variation in ligand recognition displayed by family 6 CBMs (Henshaw et al., 2006). In another study it was found out that affinity for extended β-1,3-glucooligosaccharides is conferred by interactions with the surface of the protein located between the two binding sites common to CBM6s and thus reveals a third ligand binding site in family 6 CBMs (van Bueren *et al*., 2005). It was reported that, how the multiple binding clefts and highly unusual protein surface of family 6 CBMs enables them to display an extensive range of specificities or binding affinities (van Bueren *et al*., 2005). In the present study the gene encoding *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were cloned from *Clostridium thermocellum* ATCC 27405 expressed in *Escherichia coli* and purified by immobilized metal ion affinity chromatography (IMAC) in a single step.

2.2 Materials and Methods

2.2.1 Chemicals, reagents and kits

The oligonucleotide primers for PCR amplification of *Ct*43Ara*f*, *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were procured from Stabvida, Portugal. *Pfu* DNA polymerase was supplied by Stratagene, USA. The dNTPs and $MgCl_2$ were acquired from Bioline, USA. PCR tubes (0.2 ml) used were from Axygen, Germany. Restriction enzymes *Nhe*I, *Bgl*II and *Xho*I were purchased from Fermentas, Germany. The expression vectors, pET-21a and pET-28a, were purchased from Novagen, EMD4 BioScience, Germany. T_4 DNA ligase, 10x ligase buffer were purchased from Promega, USA. RNAse solution (20 mg/ml), 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 obtained from Sigma-Aldrich Pvt. Ltd. USA. The protein staining dye, Coomassie Brilliant Blue R250 was from Himedia, 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 salts (EDTA), glucose, sodium hydroxide, sodium dodecyl sulphates (SDS), LB medium and SOC medium components were supplied by Himedia, India. The antibiotics, ampicillin was supplied by Himedia, 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 was purchased from Amresco LLC, USA. The genomic DNA of

Clostridium thermocellum and DNA marker III (NZYTech, Lda.) were gifts from Professor Carlos M.G.A Fontes, Faculdade de Medicina Veterinaria, Avenida da Universidade Técnica, Lisbon, Portugal.

2.2.2 Microorganisms

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

2.2.3 PCR amplification of gene encoding for *Ct*43Ara*f* and truncated derivative *Ct*GH43

The oligonucleotide primers were designed and the DNA encoding the full length gene *Ct*43Ara*f* (GenBank Accession No: ABN52503.1) and the truncated derivative *Ct*GH43 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 2.2.5.

Table 2.2.1 Oligonucleotide primers used for cloning full length *Ct*43Ara*f* and truncated derivative *Ct*GH43 from *Clostridium thermocellum*. The bold and underlined sequence specifies the restriction enzyme cut site.

Construct	Primer sequence
Ct*43Ara*f	Forward: 5'-CTC**GCTAGC**GCTGCCGATTATCCG-3' Reverse: 5'-CAC**CTCGAG**AATTATGCCACTACTGC-3'
***Ct*GH43**	Forward: 5'-CTC**GCTAGC**GCTGCCGATTATCCG-3' Reverse: 5'-CAC**CTCGAG**TTACTGTTTCAAGCCATC-3'

Table 2.2.2 PCR mixture for amplification of full length C*t*43Ara*f* and truncated derivative *Ct*GH43 from *Clostridium thermocellum.*

PCR components	Volume (µl)	Final concentration
10x reaction buffer	5.0	1x
100 mM dNTP mix	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, pH 8.0	39.7	--
Genomic DNA (15.4 µg/ml)	0.5	7.7.ng
Pfu DNA polymerase (2.5 U/µl)	1.0	2.5 U
Total	50.0	--

Table 2.2.3 PCR cycles for amplification of full length *Ct*43Ara*f* and truncated derivative *Ct*GH43 from *Clostridium thermocellum*.

Steps		Time (min)
I. Denaturation at 94℃		4
II. 30 cycles of	i) Denaturation at 94℃	1
	ii) Annealing at 55℃	1
	iii) Extension at 72℃	2
III. Final extension at 72℃		10

2.2.4 PCR amplification of *Ct*CBM6A and *Ct*CBM6B

Similar to above primers were also designed for the non catalytic CBMs *viz.* *Ct*CBM6A and *Ct*CBM6B where N-terminal primer included the *Nhe*I restriction site (5'-GCTAGC-3') and a C-terminal primer contained an *Xho*I restriction site (5'-CTCGAG-3') at the end of the gene. The primer details are given below (Table 2.2.4). The PCR amplification was carried out using similar 50 µl reaction mixture, as mentioned above (Table 2.2.2). The PCR cycles used for amplification of CBMs are mentioned in Table 2.2.5. The amplified products were visualized on a 0.8% agarose gel and later purified from the agarose gel as described in Section 2.2.5.

Table 2.2.4 Oligonucleotide primers used for cloning full length *Ct*43Ara*f* and truncated derivative *Ct*GH43 from *Clostridium thermocellum*.

Construct	Primer sequence
***Ct*CBM6A**	Forward: 5'-CTC**GCTAGC**CTCAAATATGTAGATC-3' Reverse: 5'-CAC**CTCGAG**TGCGTCTGAGTTGAATTTC-3'
***Ct*CBM6B**	Forward: 5'-CTC**GCTAGC**GTAACTGAAAGAAGTG-3' Reverse: 5'-CAC**CTCGAG**AATTATGCCACTACTGC-3'

Table 2.2.5 PCR cycles for amplification of full length *Ct*43Ara*f* and truncated derivative *Ct*GH43.

Steps		Time (min)
I. Denaturation at 94°C		4.0
II. 30 cycles of	i) Denaturation at 94°C	0.5
	ii) Annealing at 55°C	1.0
	iii) Extension at 72°C	2.0
III. Final extension at 72°C		10

2.2.5 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 an 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).

2.2.5.1 DNA loading dye

The DNA or sample loading dye was prepared by mixing the components mentioned below in Table 2.2.6. 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 2.2.6 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)

2.2.6 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.).

2.2.6.1 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 coloured solution in case of PCR amplified DNAs of *Ct*CBM6A and *Ct*CBM6B and to recombinant plasmids (For higher yield of DNA fragments <500 bp and >4 kb).
6. QIAquick spin column (DNA binding column) was placed in 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 room temperature, 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 17,900***g*** for 1 min at room temperature, 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 room temperature. The flow through was discarded and the column was given an additional spin of 1 min at 17,900***g***, to completely remove the residual ethanol.
9. Now the column containing bound DNA was placed in 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 room temperature and centrifuged at 17,900***g*** for 1 min.
10. The PCR-amplified or plasmids DNAs were eluted from QIAquick spin columns got collected in 1.5 ml sterile microcentrifuge tube. The PCR-amplified or other DNAs were stored at -20°C for further use.

2.2.7 Restriction enzyme digestion of the PCR amplified DNAs

The amplified products were digested with restriction enzymes (RE) *Nhe*I and *Xho*I as per the following set up given below (Table 2.2.7). The reaction mixtures were incubated at 37°C in a water bath (Grant, SUB36) for 90 min. The digested fragments 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 (Sigma-Aldrich).

Table 2.2.7 Restriction enzyme digestion set up of PCR amplified DNAs of *Ct*43Ara*f* and truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B.

RE digestion set up	1x (µl)
10x buffer	6.0
PCR DNAs	30
DNAse free water	18
*Nhe*I (10 U/µl)	3.0
*Xho*I (10 U/µl)	3.0
Total	60

2.2.8 Generation of pET expression vectors for cloning of *Ct*43Ara*f*, *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B amplified fragments

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). Although this system is extremely powerful, it is also possible to control expression levels simply by manipulating the concentration of inducer. The desired genes to be cloned in pET plasmids are under the influence of strong 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 are easy to purify in a single step by making use of affinity chromatography. The pET-21a vectors, employed for cloning of *Ct*43Ara*f*, *Ct*CBM6A and *Ct*CBM6B fragments, had an N-terminal T7-Tag® sequence and a C-terminal His-Tag® sequence (Fig. 2.2.1). The cloning and expression region of the coding strand transcribed by T7 RNA polymerase is shown in pET-21 vector map (Fig. 2.2.1). The pET-28a (+) vectors is incorporated with an N-terminal His-Tag®/thrombin/T7-Tag® configuration in addition to an optional C-terminal His-Tag sequence (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 Fig. 2.2.2.

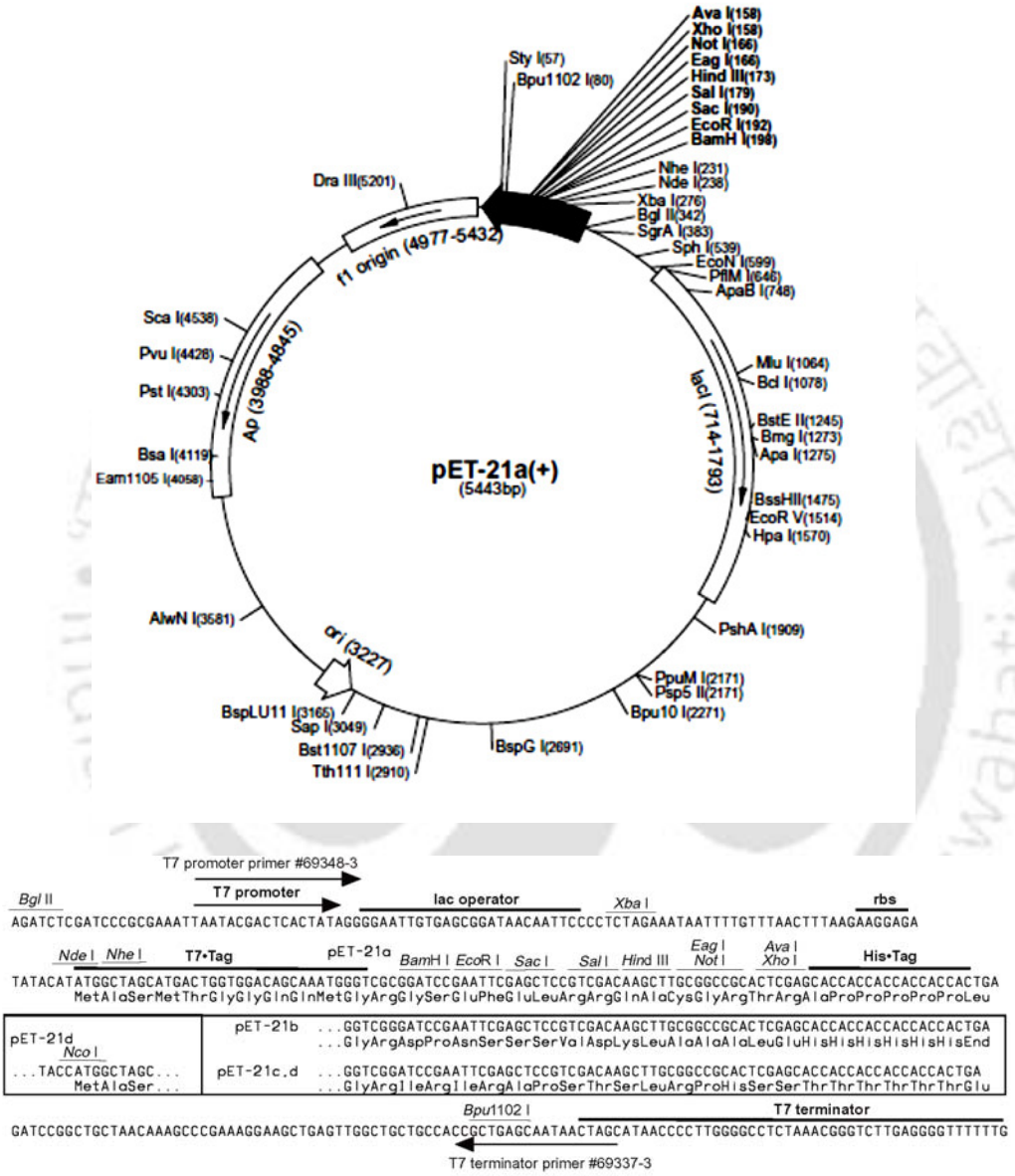


Fig. 2.2.1 Restriction map of the pET-21a(+) expression vector showing multiple cloning site (153-203 bp), restriction enzyme sites, C-terminal His•Tag coding sequence (140-157 bp), T7 promoter (311-327 bp), T7 terminator (26-72 bp), pBR322 origin (3227 bp), ampiciilin marker (3988-4845 bp) and a f1 origin (4977-5432). *Nhe*I cuts at 229, *Bgl*II at 342 and *Xho*I at 158.

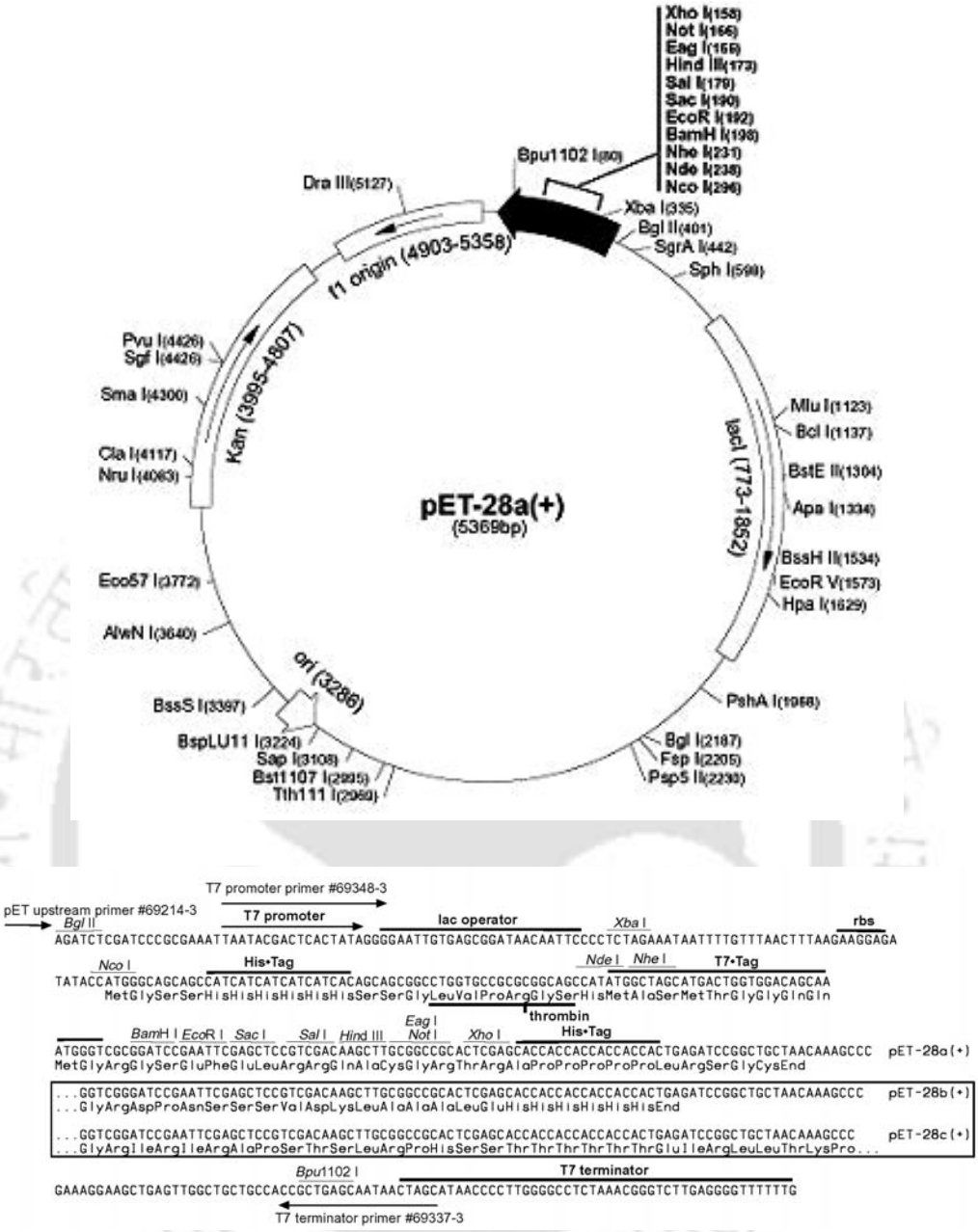


Fig. 2.2.2 Restriction map of the pET-28a(+) expression vector showing multiple cloning site (158-203 bp), restriction enzyme sites, N-terminal His•Tag coding sequence (270-287 bp), C-terminal His•Tag coding sequence (140-157 bp), T7 promoter (370-386), T7 terminator (26-72 bp), pBR322 origin (3286 bp), 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 vectors (pET-21a and pET-28a) were digested with *Nhe*I-*Xho*I to prepare them for cloning of PCR amplified fragments of *Ct*43Ara*f*, *Ct*GH43,

*Ct*CBM6A and *Ct*CBM6B. A stock solution (100 ng/µl) was prepared from the supplied stock of 10 µg for each of the above vectors using DNAse free water (pH 8.0). The restriction enzyme digestions of pET vectors were then carried out as described below (Table 2.2.8). The digestion mixtures were incubated at 37°C in a water bath for 90 min. The *Nhe*I-*Xho*I digested pET vectors were purified from agarose gel as described in Section 2.2.6.

Table 2.2.8 Restriction enzyme digestion set up of pET-21a and pET-28a.

RE digestion set up	**1x (µl)**
10x buffer	2.0
vector (pET-21a or pET-28a) 100 ng/µl	10.0
DNAse free water (Sigma-Aldrich)	6.0
*Nhe*I (10 U/µl)	1.0
*Xho*I (10 U/µl)	1.0
Total	20.0

2.2.9 Ligation of *Nhe*I-*Xho*I digested PCR fragments into pET expression vectors

The *Nhe*I-*Xho*I digested PCR fragments mentioned in Section 2.2.7 were then cloned into pET vectors, which were also digested with same restriction enzymes as described in Section 2.2.8. The concentration of *Nhe*I-*Xho*I restricted pET vectors, after gel extraction was adjusted to 100 ng/µl using DNAse free Sigma water. The full length gene, *Ct*43Ara*f* and carbohydrate binding modules, *Ct*CBM6A and *Ct*CBM6B were cloned into pET-21a vector having ampicillin as selective marker. The truncated catalytic derivative *Ct*GH43 was cloned in pET-28a vector having kanamycin as selective marker. 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*43Ara*f*, *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B required for cloning were calculated as using the following formula (Engler and Richardson, 1982),

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

$$\frac{100 \text{ ng} \times 1{,}71 \text{ kb} \times 3}{5.44 \text{ kb} \quad 1} = 95 \text{ ng of insert (}Ct\text{43Ara}f\text{)}$$

$$\frac{100 \text{ ng} \times 0.9 \text{ kb} \times 3}{5.36 \text{ kb} \quad 1} = 51 \text{ ng of insert (}Ct\text{GH43)}$$

$$\frac{100 \text{ ng} \times 0.4 \text{ kb} \times 3}{5.44 \text{ kb} \quad 1} = 22.22 \text{ ng of insert (}Ct\text{CBM6B)}$$

$$\frac{100 \text{ ng} \times 0.4 \text{ kb} \times 3}{5.44 \text{ kb} \quad 1} = 22.22 \text{ ng of insert (}Ct\text{CBM6B)}$$

The following set up was used for ligation of digested fragments into respective pET vectors as shown below in Table 2.2.9.

Table 2.2.9 Ligation set up for restriction enzyme digestion of pET vectors.

Ligation set up	**1x (µl)**
10x buffer	3.0
pET-21a or pET-28a vector (100 ng)	1.0
RE digested DNA insert (ng)	24.5
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 by calcium chloride method

Day 1

1. 50 µl of culture of *E. coli* (DH5α) from glycerol stocks were 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 grown culture from above was taken and inoculated into 100 ml LB medium in 250 ml conical flask. The flask was incubated at 37ºC with 180 rpm till cell OD reached 0.5-0.6 at 550 nm.
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 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 twice 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 resuspended 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 or -80ºC. (Freeze thawing should not be for longer time).

2.2.11 Preparation of Luria-Bertani (LB) medium

The LB medium was prepared using the components mentioned in Table 2.2.10 as shown below. The contents were dissolved in 800 ml of deionized water, the pH was adjusted to 7.0 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 and kanamycin) were added to autoclaved and cooled LB medium prior to inoculation.

Table 2.2.10 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 2.2.10 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 antibiotics were added in appropriate amount under laminar air flow. Immediately the around 25 ml of medium supplemented with antibiotics was poured in sterile petriplates and the medium was allowed to solidify for 15- 20 min.

2.2.12 Preparation of SOC (super optimal medium with catabolic repression) medium

The SOC (super optimal medium with catabolic repression) was prepared using components as described in Table 2.2.11. It is a modified SOB medium with addition of glucose (Hanahan, 1983; Deutscher, 2008; Park *et al.,* 2012). 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 amount was added to above solution in the laminar hood to finally make SOC medium having above mentioned ingredients in desired concentrations.

Table 2.2.11 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 *E. coli* (DH5α) competent cells by ligated DNA

The *E. coli* (DH5α) competent cells were transformed with recombinant plasmid DNAs of *Ct*43Ara*f*, *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B. 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 to ice for 2-3 min.
5. 1.0 ml of super optimal medium with catabolite repression (SOC) (Hanahan, 1983; Sambrook *et al*., 1989; shown in Section 2.2.12) was added to the transformed cells (previously incubated at 37°C).
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 room temperature 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 antibiotics *viz.* ampicillin (*Ct*42Ara*f*, *Ct*CBM6A and *Ct*CBM6B) and kanamycin (*Ct*GH43) at a final concentration of 100 µg/ml and 50 µg/ml, respectively.
10. The LB agar plates were incubated at 37°C and the cells were grown overnight.

11. The transformation efficiency was calculated using the following formula,

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

From the 30 µl ligation mixture, 10 µl was used for transformation of 200 µl *E. coli* (DH5α) competent cells. The transformed cells were spread on LB agar plates having 100 µg/ml ampicillin (*Ct*43Ara*f*, *Ct*CBM6A and *Ct*CBM6B) and 50 µg/ml kanamycin (*Ct*GH43), respectively and incubated at 37ºC overnight. 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 ampicillin or kanamycin, for all the recombinant derivatives and were grown at 37ºC and at 180 rpm for 12h for isolation of plasmid DNA to check for positive clones.

2.2.14 Isolation of recombinant plasmid DNA

Recombinant plasmid DNAs of *Ct*43Ara*f*, *Ct*CBM6A, *Ct*CBM6B and truncated *Ct*GH43 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.) with RNAse (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 picked randomly from the LB agar plates with appropriate antibiotics and each colony was inoculated into 5.0 ml of LB medium supplemented with appropriate antibiotics. Each culture was 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 twice 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 inverted gently 5-6 times.
5. 300 µl of ice chilled 5 M potassium acetate solution (pH 4.8) was added immediately and the tubes were inverted gently again 5-6 times. This solution neutralizes NaOH added in the previous lysis step, while precipitating the genomic DNA and SDS in an insoluble white, rubbery precipitate. The tubes were kept on ice for 3-5 min.
6. The mixture was then spun 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/isoamyl alcohol; 25:24:1) removing the protein contaminations. The mixture was vortexed 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 room temperature. The solution was mixed well and kept again at room temperature for 2 min. Thereafter, the plasmid DNA was pellet down by centrifuging 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 Tris buffer (10 mM Tris-Cl, pH 8.0) or DNAse free Sigma water (pH 8.0). 10 µl of RNAse A 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 DNAs for identification of positive clones

10 µl of recombinant plasmid DNA of all the derivatives *viz. Ct*43Ara*f,* *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B, isolated by alkaline lysis method as described in Section 2.2.14, was taken in a fresh sterile microcentrifuge tube for restriction enzyme digestion analysis. The recombinant DNA of each of the above mentioned derivatives was digested with restriction enzymes, *Bgl*II and *Xho*I, to check for positive clones following the reaction mixture set up given below in Table 2.2.12. The reaction mixtures for each recombinant derivative were incubated at 37ºC in a water bath (Grant, SUB) for 90 min. The digested products were run on 1% agarose gel as described in Section 2.2.5. The digested fragments *viz.* linearized vectors (5.44 bp, pET-21and 5.36 bp pET-28a) and the respective insert DNA of above mentioned recombinant derivatives were visualized by placing the gel under UV

transilluminator. The digested fragments (insert and vector) were analyzed to check whether the size of insert DNA and linearized vector, pET-21a or 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 clone were also confirmed by performing DNA sequencing of the recombinant plasmids of each of the above mentioned derivatives. 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 above mentioned derivatives were prepared by keeping final concentration of glycerol to 20-25% (v/v). The glycerol stocks stored at -80°C. The positive clones for each of the recombinant derivatives *viz.* *Ct*43Ara*f,* *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were identified and were preserved.

Table 2.2.12 Restriction enzyme digestion set up of recombinant DNA.

Digestion set up	1x (µl)
10x buffer	2.0
DNAse free water (Sigma-Aldrich)	6.0
Recombinant plasmid DNA (approx. 100 ng)	10.0
*Bgl*II (10 U/µl)	1.0
*Xho*I (10 U/µl)	1.0
Total	20.0

2.2.16 Preparation of *E. coli* BL-21 and *E. coli* BL-21 pLysS competent cell by calcium chloride method

The *E. coli* BL-21 (DE3) and *E. coli* BL-21 (DE3) pLysS competent cells were 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 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*43Ara*f,* *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B 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 DNAs of each recombinant derivative were isolated on next day using plasmid miniprep kit (Sigma-Aldrich) following manufacturer's protocol as described below in Section 2.2.17.1.

2.2.17.1 Plasmid isolation protocol (Sigma-Aldrich)

1. 1.5 ml from the each of the grown recombinant 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 twice with another 1.5 ml of grown culture.
3. The resulting cell pellets of each recombinant derivative were resuspended in 200 µl resuspension solution by vortex. RNAse 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 inverted gently 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 activated DNA binding column and centrifuged at 13,000***g*** for 1 min and the flow through in the collection tube was discarded again.
9. The plasmid DNAs bound to column were 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 microcentrifuge tube and 30 µl of DNAse free water (Sigma-Aldrich) 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 then got collected in the sterile microcentrifuge tube.
11. The eluted plasmid DNAs in sterile microcentrifuge tube were stored at -20°C.

2.2.18 Transformation of *E. coli* BL-21(DE3) and *E. coli* BL-21(DE3) pLysS cells with recombinant plasmids for protein expression

10 µl of from each of the recombinant plasmid DNAs isolated by above mentioned method, were used for transformation of *E. coli* BL-21(DE3) for expression of *Ct*43Ara*f, Ct*CBM6A and *Ct*CBM6B. 200 µl of *E. coli* BL-21(DE3) pLysS competent cells was transformed with 10 µl plasmid DNA of *Ct*GH43 for protein expression. The transformation protocol was similar to as described in Section 2.2.13. The transformed cells were spread on LB agar plates having appropriate antibiotics. The recombinant cells harbouring recombinant plasmids were grown at 37ºC under static condition for 12h. The colonies were picked randomly from LB agar

plates containing transformed recombinant cells and inoculated further in 5.0 ml LB medium supplemented with appropriate antibiotics as mentioned in Section 2.2.13, for all the recombinant derivatives. The recombinant cells containing recombinant plasmids were grown at 37ºC in shaking incubator with 180 rpm for analysis of expression of protein.

2.2.19 Overexpression of recombinant proteins

E. coli BL-21(DE3) cells were transformed with recombinant plasmids (10 µl) of *Ct*43Ara*f*, *Ct*CBM6A and *Ct*CBM6B as described in Section 2.2.18. *E. coli* BL-21(DE3) pLysS cells were transformed with 10 µl plasmid DNA of *Ct*GH43. The colonies grown on LB agar plates supplemented with appropriate antibiotics were picked and inoculated in 5 ml LB medium as stated in Section 2.2.13 The necessary antibiotics, ampicillin (100 µg/ml) and kanamycin (50 µg/ml), described in Section 2.2.13 were added to each tube and incubated at 37ºC with 180 rpm to grow up to mid-exponential phase till cell O.D. reached $A_{550} \approx 0.6$ (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 (for all recombinant derivatives *viz. Ct*43Ara*f, Ct*CBM6A and *Ct*CBM6B and *Ct*GH43) with 180 rpm for 24h (Ichinose *et al*., 2008; Ahmed *et al*., 2012). 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. These cells of BL-21(DE3) and BL-21(DE3) pLysS, the source of recombinant proteins *viz. Ct*43Ara*f*, *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were preserved at -80°C as glycerol stocks by

keeping the final glycerol concentration to 20-25% (v/v) as described in Section 2.2.15.

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

The recombinant proteins were separated on SDS-PAGE gel on the basis of their respective molecular size. PAGE is an analytical method used to separate components of a protein mixture based on their size (Laemmli, 1970; Sambrook *et al*., 1989). 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 were 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, it migrated towards the anode 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 basically 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 electrophoretic 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 2.2.20.2-2.2.20.5.

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

0.8 g of bis-acrylamide was weighed and transferred into an amber coloured 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 2.2.13 and 2.2.14 given below. The resolving gels were prepared by adding the all the components in the order as mentioned in Table 2.2.13, in a 25 ml beaker, by keeping acrylamide concentration to 13% (w/v) for *Ct*43Ara*f* and *Ct*GH43 and 14% (w/v) for *Ct*CBM6A and *Ct*CBM6B. Similarly, the stacking gel prepared by dissolving all the components mentioned in Table 2.2.14. The acrylamide concentration in the stacking gel was kept at 4% (w/v).

Table 2.2.13 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 2.2.14 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 below in Table 2.2.15. 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 2.2.15 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 2.2.16 and the pH of the buffer was adjusted to 6.8. The components were dissolved in the order as mentioned in Table 2.2.16 to make 5x sample buffer. However, the final concentration while loading to a SDS-PAGE gel was always kept to 1x by mixing 4 volumes of sample (protein) with 1 volume of 5x sample buffer.

Table 2.2.16 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 a disulfonated triphenylmethane (Fig. 2.2.3). The CBB R-250 dye (detection range of 100-1000 ng of protein) formed a non-covalent complex with proteins, based on a combination of 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), 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.

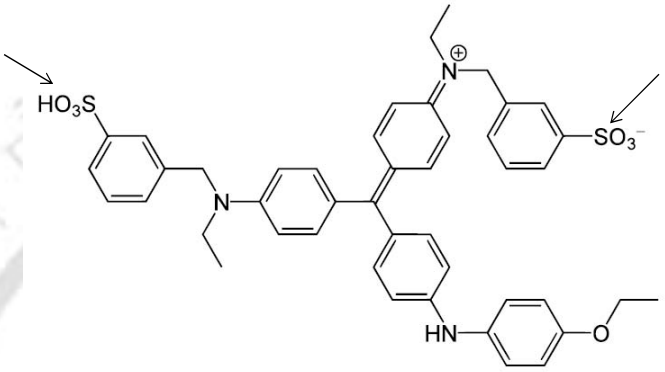


Fig. 2.2.3 Chemical structure of Coomassie Brilliant Blue dye R-250 showing disulfonated triphenylmethane ring.

2.2.21 Purification of recombinant proteins

The *E. coli* BL-21(DE3) and BL-21 (DE3) pLysS cells harbouring recombinant plasmids were grown in 100 ml LB medium in 250 ml flask with appropriate antibiotics as mentioned in Section 2.2.10.2. The recombinant proteins contained a His_6 tag at C-terminal in pET-21a and at N-terminal in pET-28a vectors. 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 recombinant proteins are mentioned Table 2.2.16. All the buffers (binding, elution and cleaning buffer), 0.1M $NiSO_4$ solution and water was filtered through 0.45 µm filters and degassed to avoid back pressure and clogging of column.

Table 2.2.17 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, 4*K*15) 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 passed through a 0.45 µm filter membrane before loading into 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 1.0 ml of 0.1 M $NiSO_4$ solution and the unbound nickel was washed away with 2-5 volumes of water. Then the column was equilibrated with 5-10 volumes of equilibration buffer (Table 2.2.17). The filtered cell free extract (4-4.5) 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 proteins of interest were 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 2.2.17, washed with 2-5 volumes of water and incubated in 1N NaOH at 4°C for 2h. The

column was then washed with 10-20 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's reagent in a microcentrifuge tube. Development of blue colour complex indicated the presence of protein and usually the purified protein was eluted in second fraction as 1 ml HiTrap column was used. The purified recombinant protein *Ct*43Ara*f* was dialyzed against 20 mM sodium phosphate buffer pH 5.7, *Ct*GH43 was dialyzed against 20 mM sodium acetate buffer pH 5.4 and CBMs (*Ct*CBM6A and *Ct*CBM6B) were dialyzed with 20 mM sodium phosphate buffer pH 7.0. The purity and molecular mass of recombinant proteins were verified by SDS-PAGE as described in Section 2.2.14. The purified clostridial recombinant proteins were quantified using Bradford reagent (diluted to 1x from 3x stock) (Bradford, 1976) as described in Section 2.2.22.

2.2.22 Protein estimation by Bradford's method

The concentrations of the recombinant proteins were determined using the Bradford's method of estimating proteins at a wavelength of 595 nm (Bradford, 1976). Bovine serum albumin (BSA) purchased from Sigma-Aldrich Pvt. Ltd. was used as standard protein. A standard plot of OD at 595 nm versus different concentration of BSA in µg/ml was prepared.

2.2.22.1 Preparation of Bradford's 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 was finally diluted to 1 liter with deionized water, filtered (Whatman, No. 1 paper) under dark conditions and stored at 4°C.

Commercial Bradford reagent (Sigma-Aldrich Pvt. Ltd.) was also used for protein content determination. The amount of recombinant 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 enzyme-substrate 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 PCR amplification of full length family 43 glycoside hydrolase (*Ct*43Ara*f*) and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B

A close inspection of 43 glycoside hydrolase (GH43) molecular architecture reveals that it consists of an N-terminal glycoside hydrolase catalytic module (903bp) and two carbohydrate binding modules (CBMs) viz. CBM6A (405 bp) and CBM6B (402 bp) at the C-terminus (Fig. 2.3.1). The genes encoding *Ct*43Ara*f*, *Ct*GH43 *Ct*CBM6A and *Ct*CBM6B were extracted from genomic library of *Clostridium thermocellum* ATCC 27405 (GenBank Accession No: ABN52503.1) and amplified using designed primers incorporated with *Nhe*I and *Xho*I restriction site.

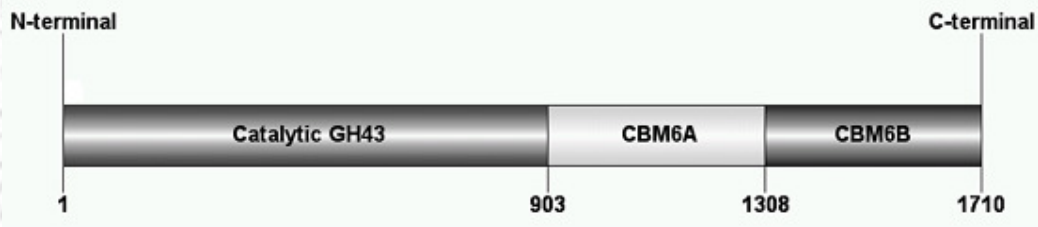


Fig. 2.3.1 Molecular architecture of family 43 glycoside hydrolase (*Ct*43Ara*f*) from *Clostridium thermocellum.*

*Ct*43Ara*f*, *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B (genes) were amplified from *Clostridium thermocellum* using the conditions as mentioned in Section 2.2.3 and were detected on 0.8% agarose gel for *Ct*43Ara*f* (Fig. 2.3.2, lane 3) and for *Ct*GH43 (Fig. 2.3.3, lane 2) as shown below. The CBMs viz. PCR amplified products of 405 bp, *Ct*CBM6A (Fig. 2.3.4, lane 2) and 402 bp, *Ct*CBM6B (Fig. 2.3.4, lane 3) as analyzed on 0.8% agarose gel shown below. 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.

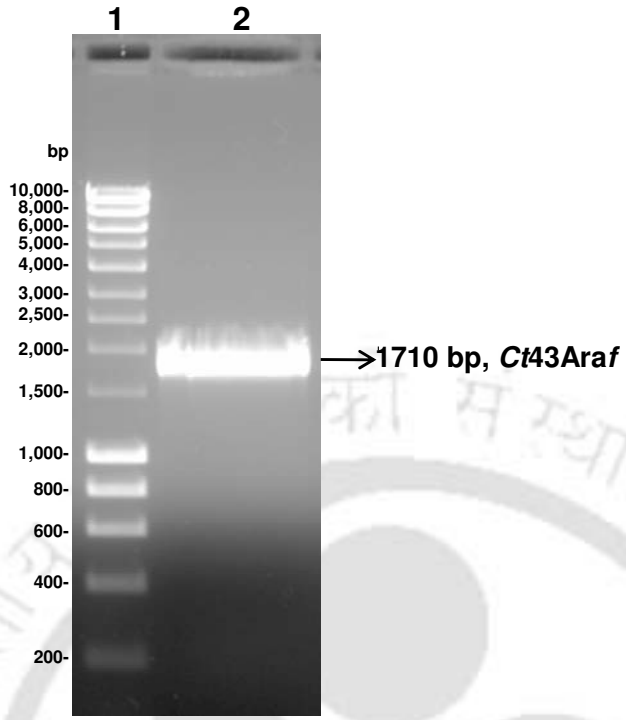


Fig. 2.3.2 Agarose gel (0.8%) showing amplified fragment of *Ct*43Ara*f*. Lane 1: DNA marker (Hyperladder I, Bioline), Lane 2: *Ct*43Ara*f* (1710 bp, approx).

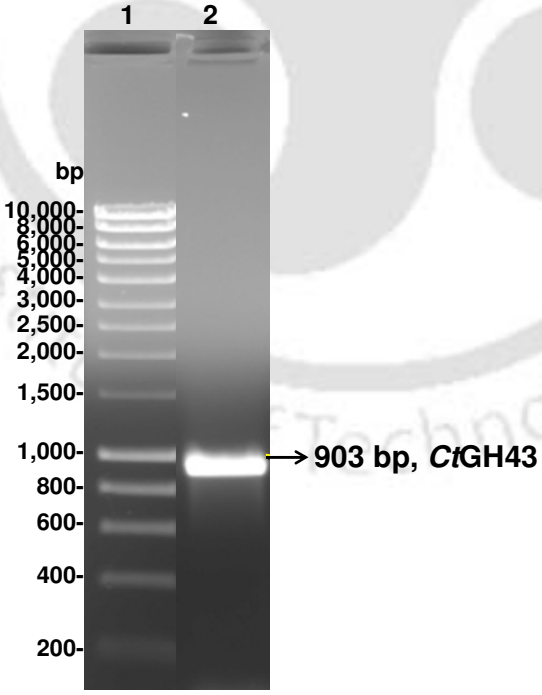


Fig. 2.3.3 Agarose gel (0.8%) showing amplified fragment of *Ct*GH43. Lane 1: DNA marker (Hyperladder I, Bioline), Lane 2: *Ct*GH43 (903 bp, approx.).

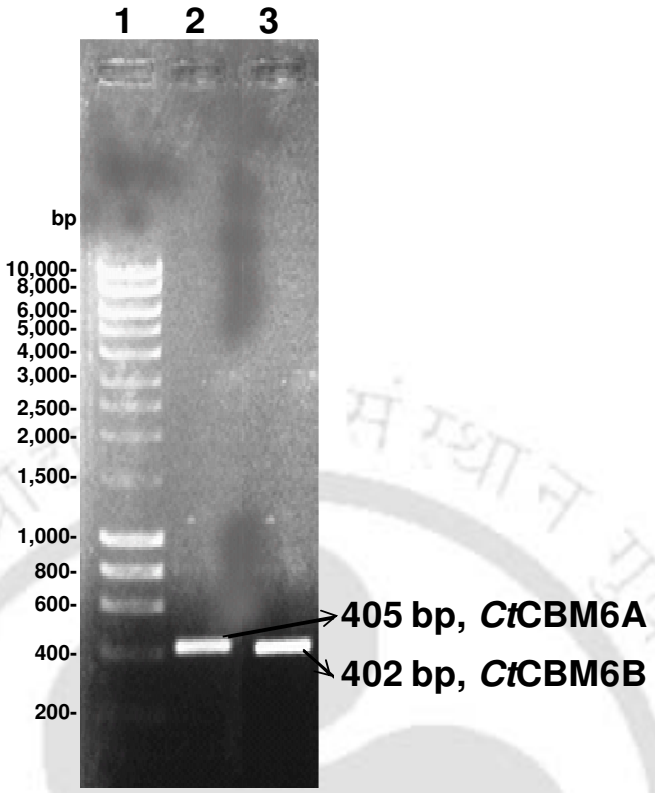


Fig. 2.3.4 Agarose gel (0.8%) gel showing PCR amplification of *Ct*CBM6A and *Ct*CBM6B. Lane 1: DNA marker (Hyper ladder I, Bioline), Lane 2: *Ct*CBM6A (405 bp, approx.), Lane 3: *Ct*CBM6B (402 bp, approx.).

2.3.2 Restriction enzyme digestion of PCR amplified fragments

The above mentioned amplified fragments (*Ct*43Ara*f*, *Ct*GH43 and *Ct*CBM6A and *Ct*CBM6B) were digested with *Nhe*I and *Xho*I restriction enzymes to prepare them for cloning into pET (expression) vectors. Simultaneously, pET-21a and pET-28a vectors were also digested using the same restriction enzymes. *Nhe*I-*Xho*I digested *Ct*43Ara*f* (Fig. 2.3.5, lane 2) and *Ct*GH43 (Fig. 2.3.5, lane 3) fragments were visualized on 1.0 % agarose gel as depicted below. The restriction enzyme digested fragments of *Ct*43Ara*f* and *Ct*GH43 was further used for cloning into pET-21a and pET-28a vectors. The *Nhe*I-*Xho*I digested fragments of *Ct*CBM6A (lane 2) and *Ct*CBM6B (lane 3) are shown in Fig. 2.3.6. The restriction enzyme digestions of pET vectors (pET-21a & pET-28a) are shown in below (Fig. 2.3.7 and 2.3.8) which also depicted their sizes and that the vectors are linearized. The *Nhe*I-*Xho*I digested pET-21a (5.4 kb) is shown in Fig. 2.3.7, lane 2, whereas, lane 2 in Fig. 2.3.8 displayed digested pET-28a (5.36 kb) vector on 1.0 % agarose gel.

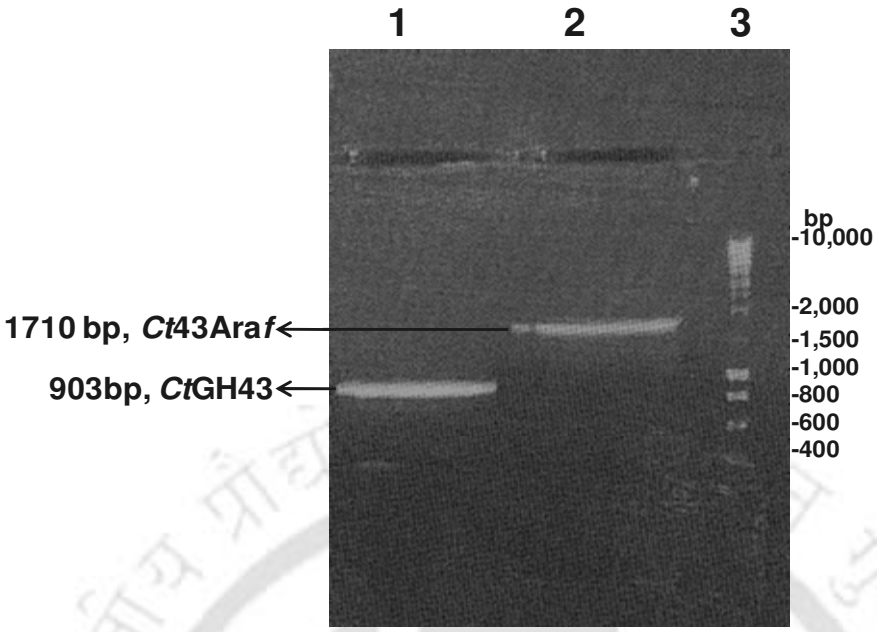


Fig. 2.3.5 Agarose gel (1.0%) showing *Nhe*I-*Xho*I digested PCR fragments of *Ct*43Ara*f* and *Ct*GH43. Lane 1: digested *Ct*GH43, Lane 2: digested *Ct*43Ara*f*, Lane 3: DNA marker III (NZYTech).

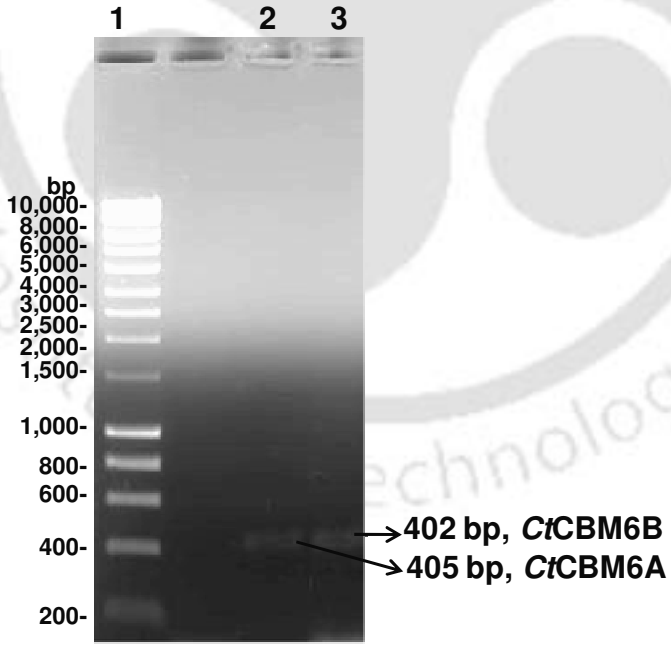


Fig. 2.3.6 Agarose gel (1.0%) showing *Nhe*I-*Xho*I digested PCR fragments of *Ct*CBM6A and *Ct*CBM6B. Lane 1: DNA marker (Hyper ladder I, Bioline), Lane 2: digested *Ct*CBM6A, Lane 3: digested CBM6B**.**

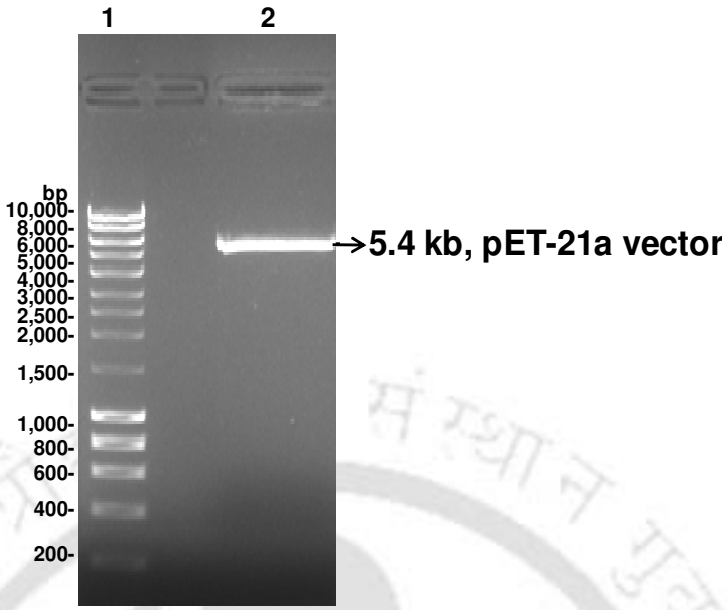


Fig. 2.3.7 Agarose gel (1.0%) showing *Nhe*I-*Xho*I digested pET-21a vector. Lane 1: DNA marker (Hyper ladder I, Bioline), Lane 2: 5.4 kb approx., digested pET-21a vector.

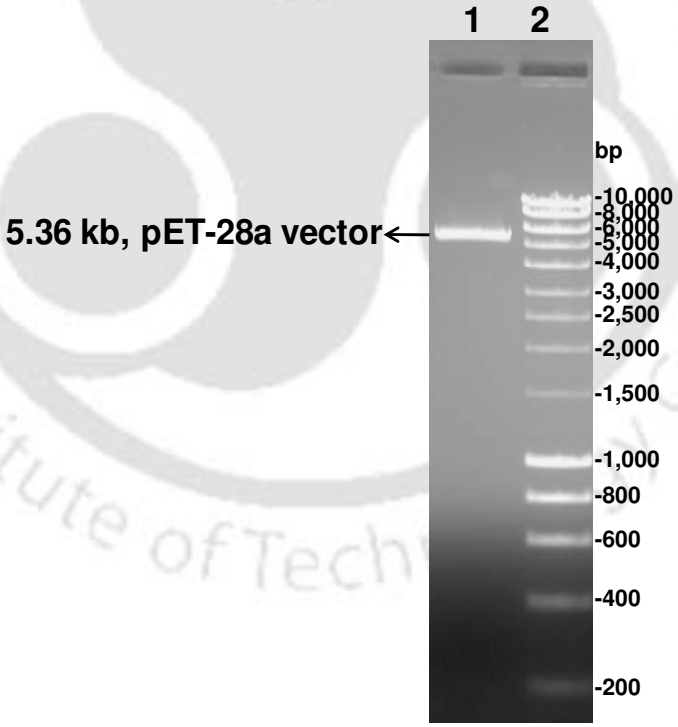


Fig. 2.3.8 Agarose gel (1.0%) showing *Nhe*I-*Xho*I-digested pET-28a vector. Lane 1: DNA marker (Hyper ladder I, Bioline), Lane 2: 5.36 kb approx., digested pET-28a vector.

2.3.3 Cloning, expression and purification of recombinant proteins *Ct*43Ara*f*, *Ct*GH43 and *Ct*CBM6A and *Ct*CBM6B

2.3.3.1 Cloning of NheI-XhoI digested fragments into pET expression vector systems

The ligation of recombinant derivatives into respective pET vectors were carried out as described in Materials and Methods Section 2.2.9. Recombinant derivatives *Ct*43Ara*f,* *Ct*CBM6A and *Ct*CBM6B were constructed in *NheI/XhoI* digested pET-21a and *Ct*GH43 in pET-28a vector.

2.3.3.2 Transformation of *E. coli* (DH5α) competent cells with recombinant plasmid DNA

E. coli (DH5α) competent cells were transformed with recombinant plasmid DNA of each derivative *viz.* *Ct*43Ara*f,* *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B as described in Section 2.2.13. The transformed cells in all the LB agar plates grown at 37°C under static condition for 12-16 h showed many colonies. The grown colonies observed on LB agar plates were counted and the transformation efficiency was calculated using the formula discussed in Section 2.2.13 and is displayed in Table 2.3.1.

Table 3.3.1 Transformation efficiency of *E. coli* (DH5α) cells.

Clone	insert DNA (ng)	No. of colonies	Transformation Efficiency (cfu/µg)
+ve control (pUC18)	1.0	279	2.8×10^5
*Ct*43Ara*f*	2.0	119	6.0×10^4
*Ct*GH43	2.0	209	1.0×10^5
*Ct*CBM6A	2.0	259	1.3×10^5
*Ct*CBM6A	2.0	217	1.1×10^5

2.3.3.3 Isolation of plasmids

The plasmid DNAs were isolated for each recombinant derivative from the 5.0 ml LB medium tubes supplemented with appropriate antibiotics, using the plasmid miniprep kit as described in Section 2.2.17. The figures depicted below show the plasmid DNAs for each of the above mentioned recombinant derivatives *viz.* *Ct*43Ara*f*, *Ct*GH43 and *Ct*CBM6A and *Ct*CBM6B. The figures shown below displayed recombinant plasmids of *Ct*43Ara*f* (Fig. 2.3.9, lane 3-4), *Ct*GH43 (Fig. 2.3.9, lane 2) and *Ct*CBM6A (Fig. 2.3.10, lane 2) and *Ct*CBM6B (Fig. 2.3.10, lane 3) as analyzed on 0.8% agarose gel.

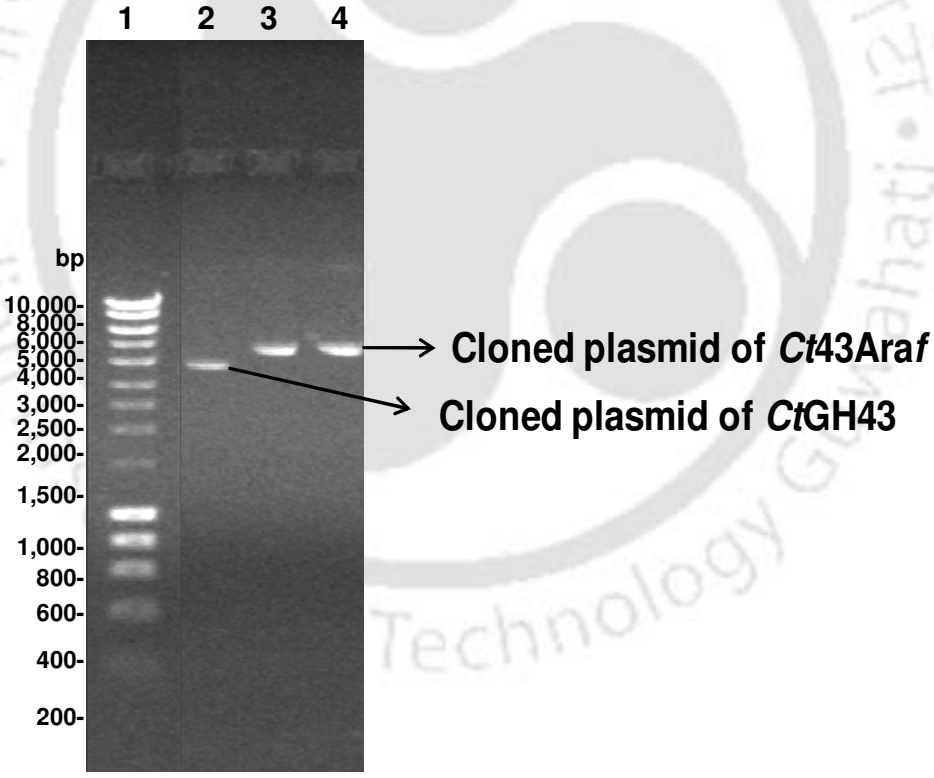


Fig. 2.3.9 Agarose gel (0.8%) showing recombinant plasmids *Ct*43Ara*f* and C*t*GH43. Lane 1: DNA marker (Hyperladder I, Bioline), Lane 2: recombinant plasmid of *Ct*GH43, Lane3-4: recombinant plasmids of *Ct*43Ara*f*.

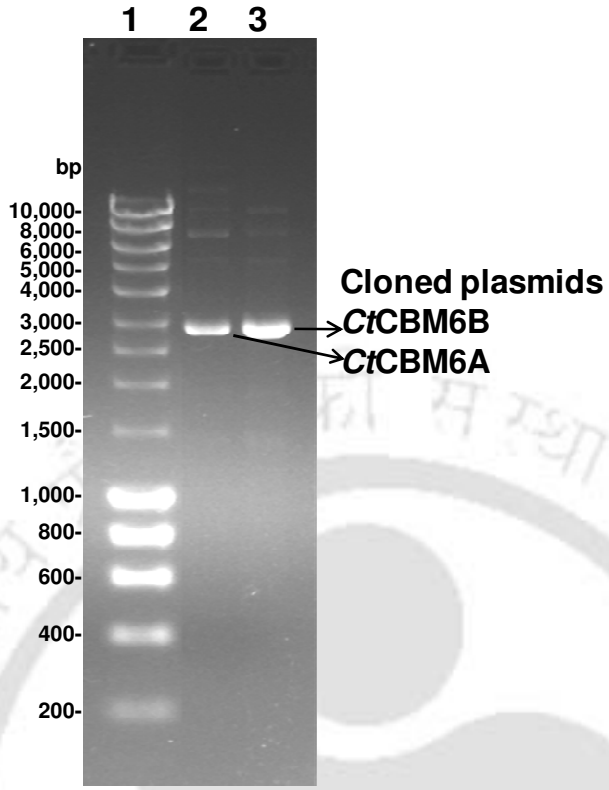


Fig. 2.3.10 Agarose gel (0.8%) showing recombinant plasmids *Ct*CBM6A and *Ct*CBM6B. Lane 1: DNA marker (Hyperladder I, Bioline), Lane 2: recombinant plasmid of *Ct*CBM6A, Lane 3: recombinant plasmid of *Ct*CBM6B.

2.3.3.4 Screening of recombinant plasmid DNAs for identification of positive clones

The positive clones were identified by performing restriction digestion analysis of the recombinant plasmids as shown below in figures (Fig. 2.3.11-2.3.13). The screening for recombinant plasmid *Ct*43Ara*f* by restriction enzyme digestion is depicted below (Fig 2.3.11, lane 2). The *Bgl*II-*Xho*I digested fragment of *Ct*43Ara*f* showed two distinct, the usual 5.4 kb pET-21a and 1.82 kb insert (*Ct*43Ara*f*). Since *Bgl*II is located 111 bases away as compared to *Nhe*I, the *Bgl*II-*Xho*I digested fragment of *Ct*43Ara*f* consequently shows a size of 1.82 kb (Fig. 2.3.11, lane 2). This is the reason for observing the insert band close to 2000 bp band of DNA ladder. So, we can conclude from above results that the clone harbours the desired recombinant

plasmid (*Ct*43Ara*f*) and subsequently we were able to express the desired protein using recombinant plasmid shown in Fig. 2.3.12 shown below. Similar to above cloned plasmid of *Ct*GH43 was also digested using *Nhe*I-*Xho*I, which resulted in two fragments (Fig. 2.3.12, lane 2) detected on 1.0 % agarose gel. The fragments obtained were 5.36 kb pET-21a vector and 903 bp insert and this was the basis for selection of the positive clone for truncated derivative *Ct*GH43. Similarly, *Bgl*II-*Xho*I digestion of CBMs resulted in two fragments in each case; one was 5.4 kb pET-21a vector and the others were insert *viz*. *Ct*CBM6A (Fig. 2.3.13, lane 2) and *Ct*CBM6B (Fig. 2.3.13, lane 3).

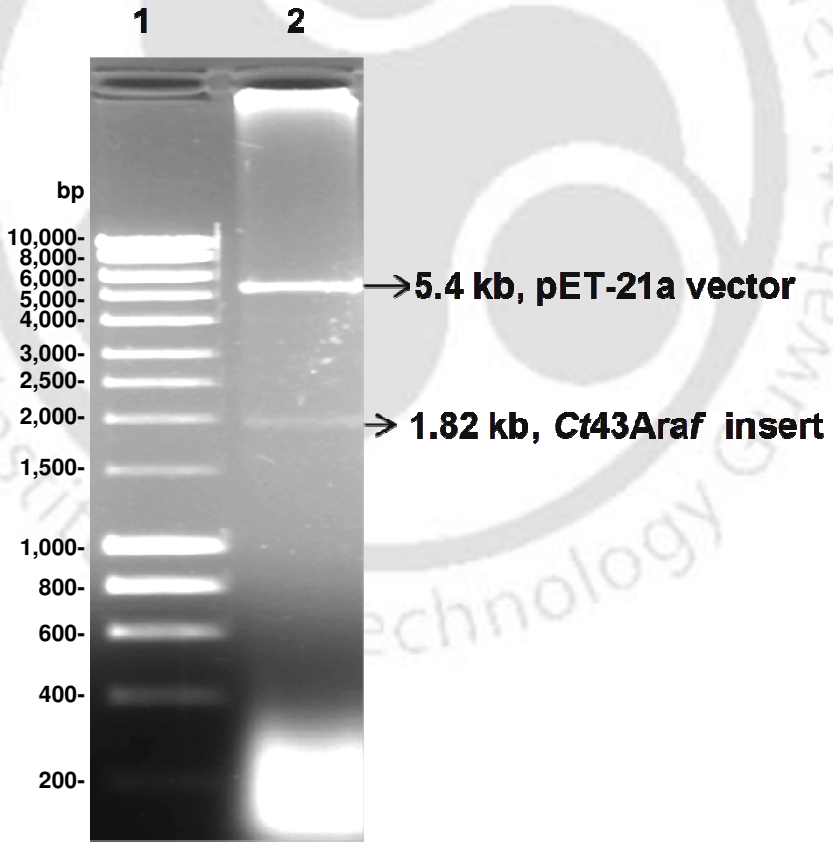


Fig. 2.3.11 Agarose gel (1.0%) showing *Bgl*II-*Xho*I digested recombinant plasmid containing *Ct*43Ara*f* fragment. Lane 1: Hyper ladder I (Bioline), Lane 2: *Bgl*II-*Xho*I digested 1.82 kb fragment of *Ct*43Ara*f* (insert) and vector (5.4 kb, pET-21a).

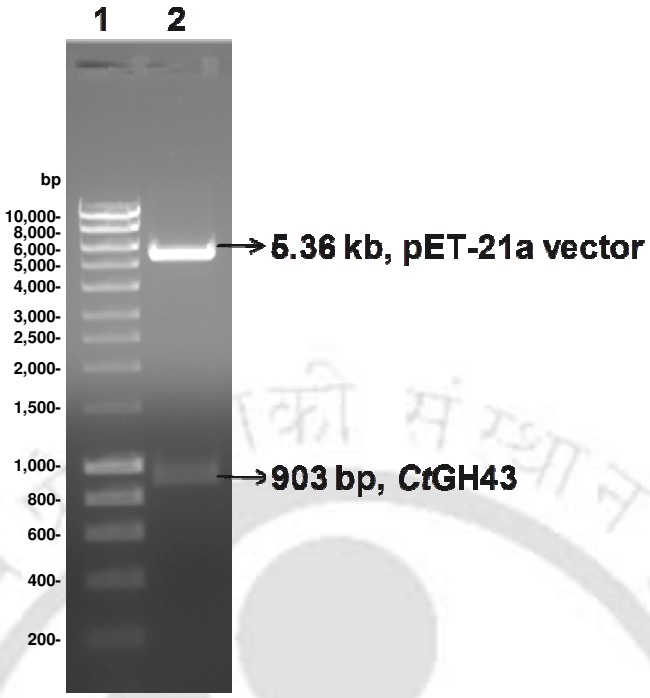


Fig. 2.3.12 Agarose gel (1%) showing *Nhe*I-*Xho*I digested recombinant plasmid containing *Ct*GH43 fragment. Lane 1: DNA marker (Hyper ladder I, Bioline), Lane 2: showing insert (903 bp, *Ct*GH43) and vector (5.46 kb, pET-28a).

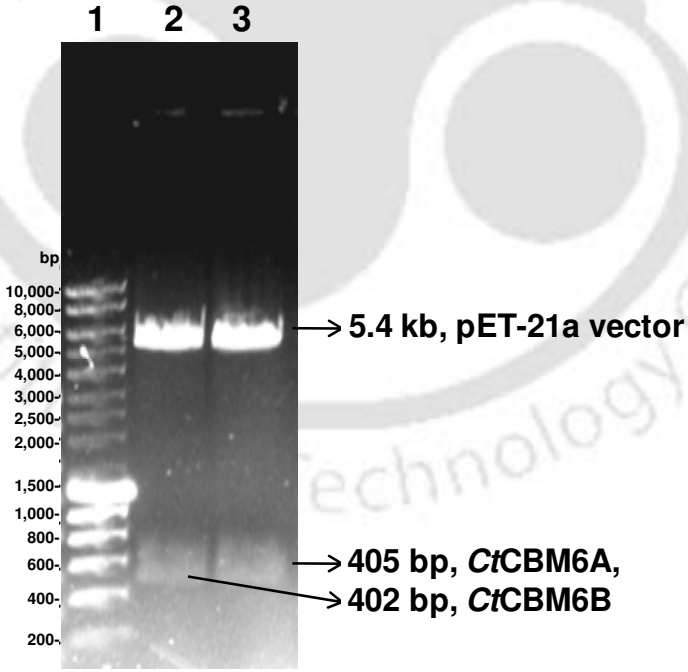


Fig. 2.3.13 Agarose gel (1%) showing *Bgl*II-*Xho*I digested recombinant plasmids containing *Ct*CBM6A and *Ct*CBM6B fragments. Lane 1: Hyper ladder I (Bioline), Lane 2: showing insert (405 bp, *Ct*CBM6A) and vector (5.4 kb, pET-21a), Lane 3: showing insert (402 bp, *Ct*CBM6B) and vector (5.4 kb, pET-21a).

The cloning of truncated catalytic derivative *Ct*GH43 was confirmed by sequencing the plasmid DNA. The sequencing results confirmed that the recombinant plasmid harbours the desired gene. The DNA sequence of *Ct*GH43 is given below in fasta format (Fig. 2.3.14A and 2.3.14B).

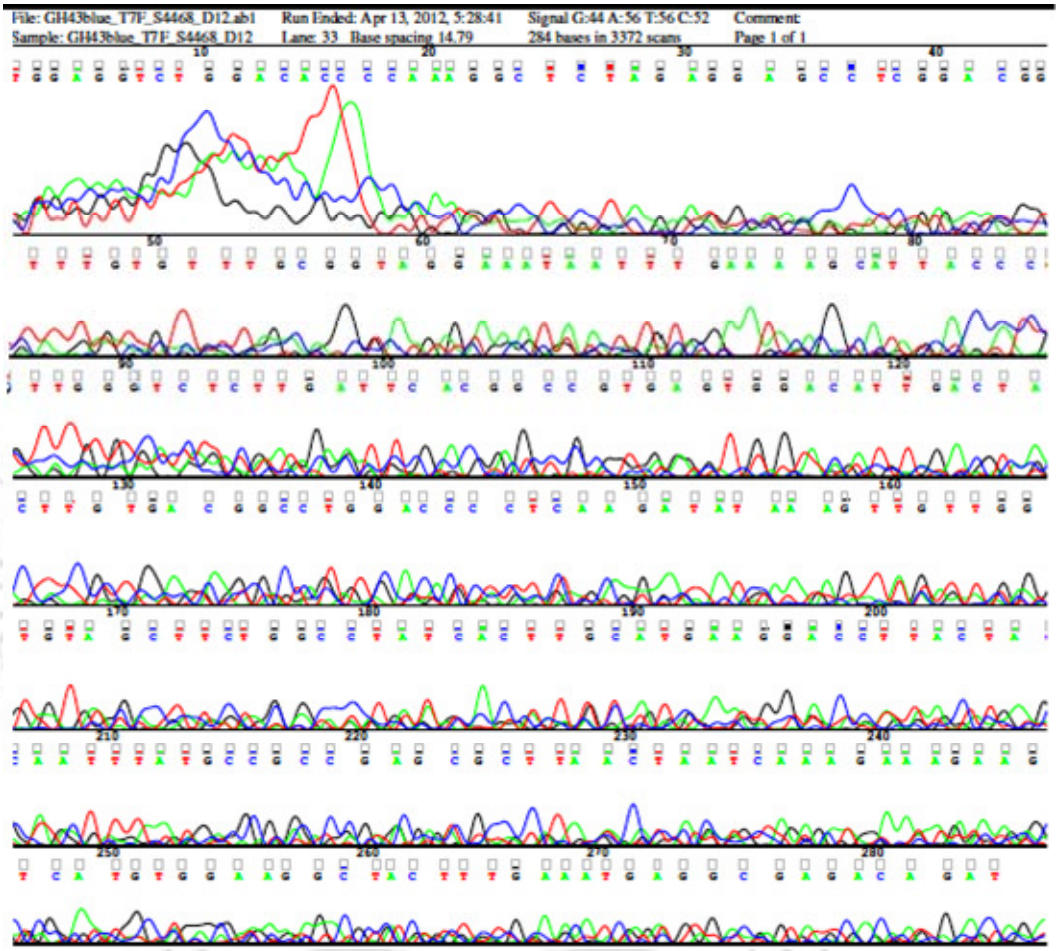


Fig. 2.3.14A DNA sequencing results of cloned plasmid of *Ct*GH43 (no mutation detected after comparison with DNA sequence of gene for *Ct*GH43).

```
>Seqresult_GH43
GLPIIRFSHSVLLQTQPQSCNGRLYIYCSHDSDATPGQSTYNIPDITCIST
DDLKNWTDHGEVFNAKRDSRWASVSWAPSIVYRNNKFYLYYGNGGNGIGVA
VSDSPTGPFKDPLPGPLVSWNTPGVQPAQNMWLFDPGVFVDDDGQAYMYFG
GNGQNNIRVIKLGNDMISTVGSAMTMSAPRFFEAAYMHKYNGKYYFSYASD
FSQGASKIEYMMSDKPTTGFQYKGVILPQPPDNYSNNNHHAIVEYKGNWYV
VYHNRTVAKQRGLDPVYQRNVCIDQMFYNADGTIKQVVPTVDGLKQ
```

Fig. 2.3.14B DNA sequencing results of cloned plasmid of *Ct*GH43 (in FASTA format).

2.3.3.5 Hyper-expression analysis and purification of recombinant proteins

The *E. coli* BL-21 pLysS competent cells were transformed with recombinant plasmids of *Ct*GH43. The *E. coli* BL-21 cells were transformed with recombinant plasmids of full length family 43 glycoside hydrolase (*Ct*43Ara*f*) and its truncated derivatives, *Ct*CBM6A and *Ct*CBM6B. The colonies were picked randomly and grown in 5 ml LB medium supplemented with appropriate antibiotics, as described in Section 2.2.13. The cells were induced for overexpression of protein at mid exponential stage as described in Section 2.2.19. Hyper-expression of protein was analyzed on SDS-PAGE, by loading uninduced as well as the induced cells of recombinant protein on adjacent wells, as depicted in Fig. 2.3.15 to Fig. 2.3.18. The hyper-expression of *Ct*43Ara*f* is shown in Fig. 2.3.15 (lane 3), *Ct*GH43 in Fig. 2.3.16 (lane 3 and lane 4), *Ct*CBM6A in Fig. 2.3.17 (lane 3) and the expression of *Ct*CBM6B is displayed in Fig. 2.3.18 (lane 3).

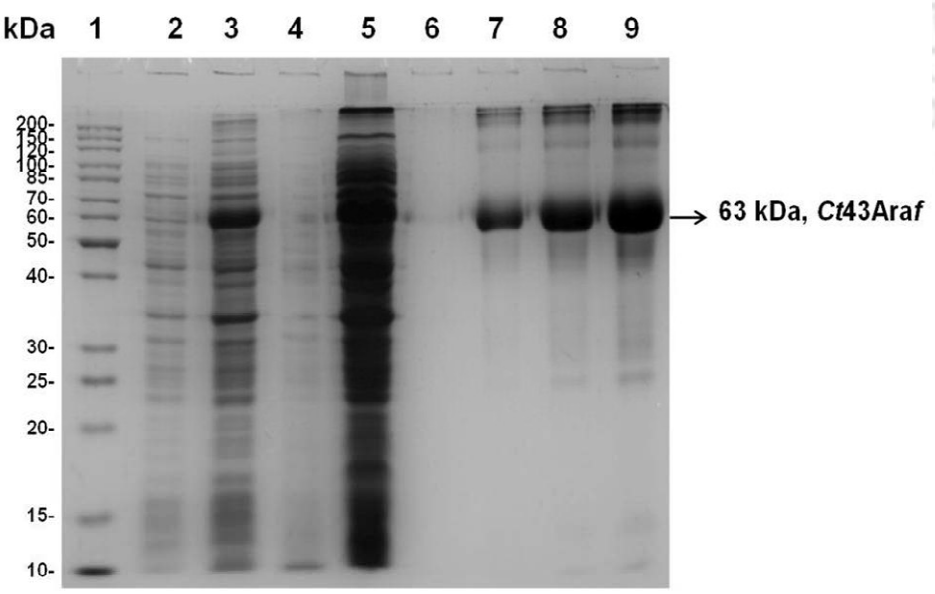


Fig. 2.3.15 SDS-PAGE (13.0%) gel showing expression and purification of recombinant protein *Ct*43Ara*f* in *E. coli* BL-21 cells, Lane1: Unstained Page Ruler marker (Fermentas), Lane 2: Uninduced *Ct*43Ara*f* cells, Lane 3: Induced *Ct*43Ara*f* cells, Lane 4: Cell pellet (cell debris after sonication), Lane 5: Cell free extract, Lane 6: Purified un-dialyzed *Ct*43Ara*f*, Lane 7: Dialyzed *Ct*43Ara*f* (63 kDa approx.).

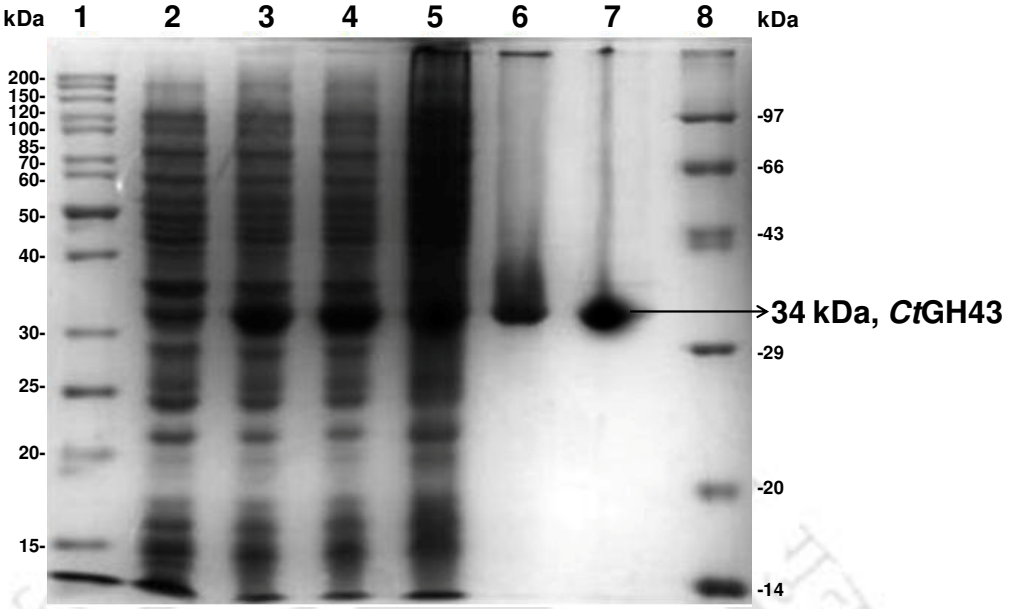


Fig. 2.3.16 SDS-PAGE (13.0%) gel showing expression and purification of recombinant *Ct*GH43 in *E. coli* BL-21 pLysS cells. Lane 1: Page Ruler protein marker (Fermentas), Lane 2: Uninduced *Ct*GH43 cells, Lane 3-4: Induced *Ct*GH43 cells, Lane 5: Cell free extract (*Ct*GH43), Lane 6-7: *Ct*GH43 affinity column purified (34 kDa), Lane 8: Mid range protein marker (Bangalore Genei).

The overexpressed recombinant proteins were purified as described in Section 2.2.21 and then dialyzed to remove salts. Fig. 2.3.15 and Fig. 2.3.16 show SDS-PAGE (13.0%) analysis of expression and purification of *Ct*43Ara*f* and *Ct*GH43, respectively. The above enzymes expressed as soluble proteins, as can be seen in cell free extracts (Fig. 2.3.15, lane 5 and Fig. 2.3.16, lane 5). The purified *Ct*43Ara*f* and *Ct*GH43 showed homogenous band of approx. sizes 63 kDa and 34 kDa, respectively. *Ct*CBM6A and *Ct*CBM6B were also hyper-expressed and checked for homogenous bands of purified proteins on SDS-PAGE (14.0%). Both *Ct*CBM6A and *Ct*CBM6B recombinant proteins were expressed as soluble protein as they can be seen in cell free extract as shown in Fig. 2.3.17, lane 4 and Fig. 2.3.18, lane 4, respectively. The purified *Ct*CBM6A showed molecular size of 15 kDa (Fig. 2.3.17, lane 7) and purified *Ct*CBM6B showed molecular size of 14 kDa (Fig. 2.3.18, lane 6).

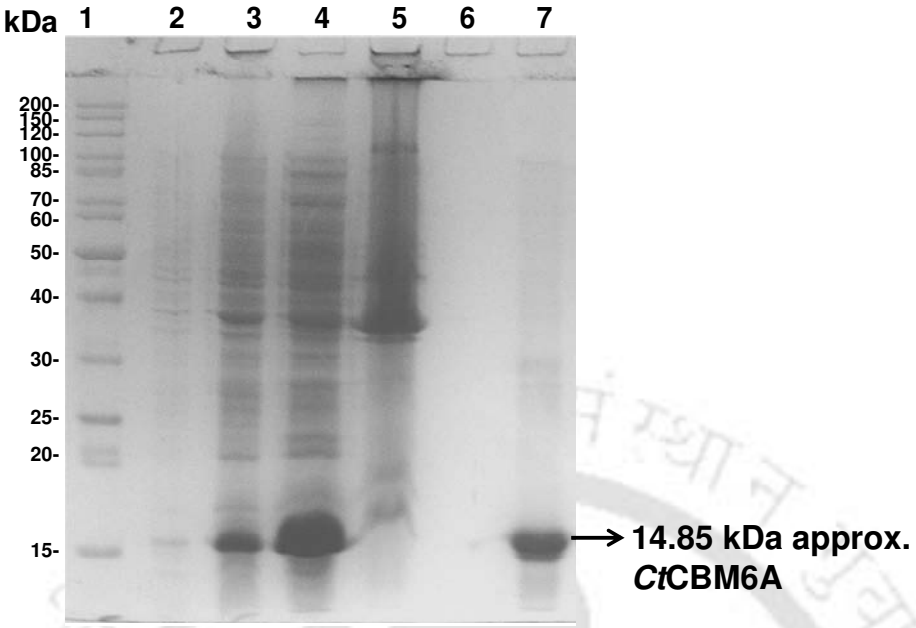


Fig. 2.3.17 SDS-PAGE (14.0%) gel showing expression and purification of *Ct*CBM6A. Lane 1: Page Ruler protein marker (Fermentas), Lane 2: Uninduced *Ct*CBM6A cells, Lane 3: Induced *Ct*CBM6A cells, Lane 4: Cell free extracts (*Ct*CBM6A), Lane 5: Cell pellet (after sonication), Lane 6: Last wash before protein elution, Lane 7: Column purified *Ct*CBM6A of 14.85 kDa approx.

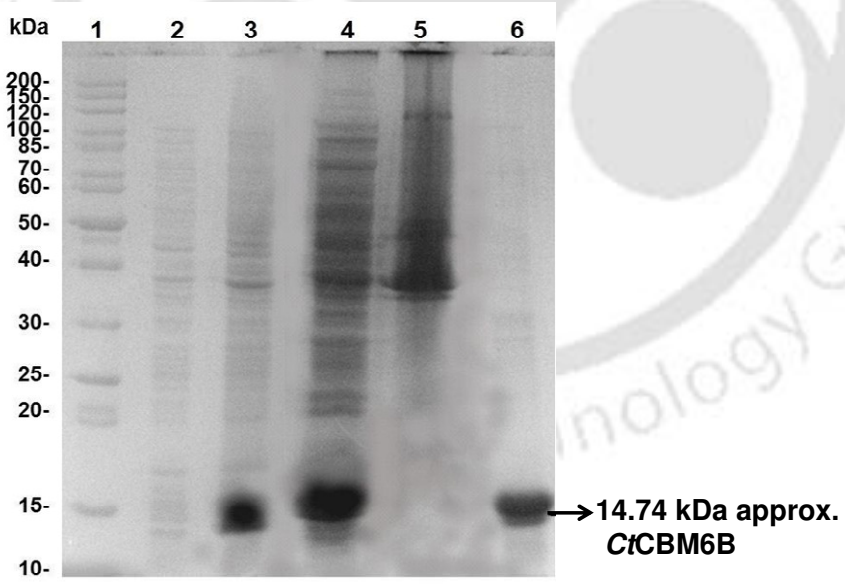


Fig. 2.3.18 SDS-PAGE (14.0%) gel showing expression and purification of *Ct*CBM6B. Lane1: Page Ruler protein marker (Fermentas), Lane 2: Uninduced *Ct*CBM6B cells, Lane 3: Induced *Ct*CBM6B cells, Lane 4: Cell free extract, Lane 5: Cell pellet (after sonication), Lane .6: Purified *Ct*CBM6B of 14.78 kDa approx.

2.3.4 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 2.3.2. The concentrations of recombinant proteins purified after HiTrap column and after dialysis against respective buffers were in the range of 0.4-0.6 mg/ml (Table 2.3.2). The total amount of recombinant enzymes present in eluted 2 ml of was 0.9 mg (*Ct*43Ara*f*), 1.08 (*Ct*GH43), 0.82 (*Ct*CBM6A) and 0.72 (*Ct*CBM6B) as displayed in Table 2.3.2.

Table 2.3.2 Amount of purified recombinant proteins obtained from 100 ml cultures.

Recombinant derivative	Protein concentration (mg/ml)	Volume of purified protein (ml)	Total amount of purified protein (mg)
Ct43Ara*f*	0.45±0.10	2.0	0.9±0.1
*Ct*GH43	0.54±0.12	2.0	1.08± 0.12
*Ct*CBM6A	0.41±0.05	2.0	0.82± 0.05
*Ct*CBM6B	0.36±0.08	2.0	0.72±0.08

2.4 Conclusions

The genes encoding *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were amplified from the genomic DNA of *Clostridium thermocellum* ATCC 27405 (GenBank Accession No: ABN52503.1). The molecular architecture revealed an N-terminal catalytic family 43 glycoside hydrolase module (*Ct*GH43, 903bp) followed by two carbohydrate binding modules (CBMs) *viz.* *Ct*CBM6A (405 bp) and *Ct*CBM6B (402 bp) at the C-terminus. PCR amplified fragment of full length gene *Ct*43Ara*f* showed a band of 1710 bp, whereas, the truncated catalytic module *Ct*GH43 displayed a PCR amplified fragment of 903 bp. *Ct*CBM6A and *Ct*CBM6B showed PCR amplified fragments of 405 bp and 402 bp, respectively. The above amplified products were purified from the agarose gel by gel extraction kit (Qiagen). *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were digested with *Nhe*I-*Xho*I restriction enzymes and the digested fragments extracted from gel and purified using gel extraction kit. Similarly, the pET vectors (pET-21a and pET-28a) were also digested with *Nhe*I-*Xho*I restriction enzymes and the digested fragment of each was purified by gel extraction kit.

The gene encoding catalytic module *Ct*43Ara*f* and its truncated derivatives *Ct*CBM6A and *Ct*CBM6B were ligated with pET-21a vector and *Ct*GH43 was ligated with pET-28a vector. *E. coli* (DH5α) competent cells were transformed with ligated samples of each derivative mentioned above. The transformed cells were spread on LB agar plates with appropriate antibiotics and grown at 37°C, which resulted in many colonies after 12h incubation. The plasmid DNA of each recombinant derivative was isolated from the colonies after growing in 5 ml LB medium supplemented with appropriate antibiotics at 37°C and at 180 rpm. The cloning of *Ct*43Ara*f* and its truncated derivatives in pET-21a and pET-28a vectors was

confirmed by restriction digestion analysis of the recombinant plasmid DNA as well as by DNA sequencing of the recombinant plasmids. The *E.coli* BL-21(DE3) competent cells were transformed with recombinant plasmid DNA of full length *Ct*43Ara*f* as well as its truncated derivatives *Ct*CBM6A and *Ct*CBM6B, whereas, *E.coli* BL-21(DE3) pLysS competent cells were transformed with recombinant plasmid DNA of *Ct*GH43, for expression of recombinant proteins. Over expression of recombinant proteins was achieved by using IPTG as inducer at a final concentration of 1 mM. The overexpression of the recombinant proteins were checked and confirmed by SDS-PAGE gel analysis.

After confirming the successful expression, the recombinant proteins were purified from their respective cell free extracts by immobilized metal ion chromatography (IMAC) using HiTrap chelating columns. The degree of purification obtained after this purification step was higher than 90%, and therefore these proteins were considered as suitable for further biochemical analysis. The catalytic enzymes purified by IMAC displayed molecular sizes of 63 kDa (*Ct*43Ara*f*), 34 kDa (*Ct*GH43) on SDS-PAGE. The non-catalytic carbohydrate binding modules (*Ct*CBM6A and *Ct*CBM6B) showed molecular sizes of 15 kDa and 14 kDa, respectively, on SDS-PAGE. The amount of protein obtained from 100 ml of cultures *Ct*43Ara*f* and its truncated derivatives *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B after IMAC purification were 0.9 mg (*Ct*43Ara*f*), 1.08 mg (*Ct*GH43), 0.82 mg (*Ct*CBM6A) and 0.72 mg (*Ct*CBM6B). These purified proteins were used for further biochemical and functional characterization.

References

Bayer, E.A., Belaich, J.P., Shoham, Y. and Lamed, R. (2004) The cellulosomes: multi-enzyme machines for degradation of plant cell wall polysaccharides. Annu. Rev. Microbiol. 58, 521-554.

Bayer, E.A., Shoham, Y. and Lamed, R., (2000a) Cellulose-decomposing prokaryotes and their enzyme systems. In (3rd ed.) Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., and Stackebrandt, E. (ed.), *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community*. 2, 578-617.

Bayer, E. A., Shoham, Y. and Lamed R. (2000b) The cellulosome-an exocellular organelle for degrading plant cell wall polysaccharides. In Doyle, R.J. (ed.). *Glycomicrobiology*. Kluwer Academic/Plenum Publishers, New York. 387-439.

Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol. 62, 293-300.

Bolam, D.N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B. and Gilbert, H.J. (2004) X4 modules represent a new family of carbohydrate-binding modules that display novel properties. J. Biol. Chem. 279, 22953-22963.

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

Bradford, M. (1976) A Rapid and Sensitive Method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248-254.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37, 233-238.

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

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. and Fontes, C.M.G.A. (2004) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates β-1,4 and β-1,3-1,4-mixed linked glucans at a single binding site. J. Biol. Chem. 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.

Davies, G.J. and Henrissat, B. (2002) Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era. Biochem. Soc. Trans. 30, 291-297.

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

Fontes, C.M.G.A. and Gilbert, H.J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu. Rev. Biochem. 79, 655-81.

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

Hanahan D. (1983) Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166, 557-580.

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

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

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

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

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

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

Neuhoff, V., Stamm, R. and 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. and Lee, S.K. (2012) The mechanism of sugar-mediated catabolite repression of the propionate catabolic genes in *Escherichia coli*. Gene, 504, 116-121.

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

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

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

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

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

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

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

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. and Davies GJ. (2005) How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides? Structure and activity of a *Clostridium thermocellum* lichenase, CtLic26A. J. Biol. Chem. 280, 32761-32767.

van Bueren, A.L., Morland, C., Gilbert, H.J. and Boraston, A.B. (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of β-1,3-linked glcans by presenting a unique ligand binding surface. J. Biol. Chem. 280, 530-537.

Chapter 3

Biochemical, functional and structural characterization of catalytic modules *Ct*43Ara*f* and *Ct*GH43 from *Clostridium thermocellum*

3.1 Introduction

Plant cell wall is known to be composed of complex structural polysaccharides like cellulose and hemicellulose (Fontes and Gilbert, 2010). The hetero-polymers of pentoses like D-xylose, L-arabinose and hexoses *viz*. D-mannose, D-glucose and D-galactose constitutes the hemicellulose. Often, xylans are hetero-polysaccharides with β-1,4-linked D-xylopyranose backbone containing arabinose, glucuronic acid, or its 4-O-methyl ether, acetic, ferulic, and p-coumaric acids side chains depending mainly on the source of xylans (Numan and Bhosle, 2006). Rye arabinoxylans contain xylose and arabinose in the A/X ratio of 0.49–0.82 and also ferulate residues attached to arabinose as esters at its O-5 position (Knudsen and Laerke, 2010) but in wheat arabinoxylans the arabinose to xylose ratios (A/X) varies from 0.47 to 0.58 (Ordaz-Ortiz and Saulnier, 2005). The L-arabinosyl residues are often found in hemicelluloses, such as arabinan, arabinoxylan, gum arabic and arabinogalactan. The cereal arabinoxylans are composed majorly of a backbone of 1,4-Linked-β-D-xylopyranosyl residues substituted with single *α*-arabinofuranosyl substituents attached to the O-2, O-3 or to both O-2,3 of the xylose residues (Matsuo

et al., 2000; Cartmell *et al*., 2011). It has been documented that α-L-arabinofuranosyl and to a lesser extent α-L-arabinopyranosyl side chains are attached to the β-D-galactopyranose main chain by 1,3- and 1,6- linkages in type II arabinogalactans (Pason *et al.*, 2010). The α-L-arabinofuranosidase (α-L-arabinofuranosidase, EC 3.2.1.55) are enzymes known to release terminal α-1,2- (Krog-Mikkelsen *et al.,* 2011), α-1,3- and α-1,5 (Numan and Bhosle, 2006) α-L-arabinofuranoside residues from hemicellulose such as arabinoxylan and other L-arabinose containing polysaccharides. Arabinofuranosidases have been reported from a few glycoside hydrolase families (GHs) *viz*., GH30 (Zhou *et al*., 2012), GH43 (Cartemell *et al*., 2011), GH51 (Sorenson *et al*., 2006), GH54 (Guais *et al.*, 2010) and GH62 (Hashimoto *et al*., 2011; Sokomoto *et al.*, 2011).

The α-L-arabinofuranosidases have been used synergistically with other cellulose degrading enzymes in agro-industrial processes (Saha *et al*., 2003). Lignocelluloses are known to have cellulose as major content but they also contain 20% hemicellulose which mainly consists of pentoses such as xylose and arabinose (Saha *et al.*, 2003; Ahmed *et al*., 2012). In the past α-L-arabinofuranosidases received less importance in the production of bio-ethanol because pentoses are less efficiently converted to ethanol than hexose sugars (Aristidou and Penttilä, 2000). But recently they have been used along with *Candida shehatae* that utilizes the pentose sugars for bio-ethanol production from cellulosic waste like mango and poplar leaves (Ahmed *et al.*, 2012). Also L-arabinose has been shown to inhibit intestinal sucrase and thereby it reducing the glycaemic response after sucrose ingestion in animals (Osaki *et al*., 2001).

Biochemical and functional characterization of *Ct*43Ara*f* and *Ct*GH43 is essential as all the α-L-arabinofuranosidases have the same inverting mechanism of

catalysis but the enzyme activities are different (http://www.cazy.org/GH43.html; 30). In the present study the catalytic modules *Ct*43Ara*f* and its truncated derivatives *Ct*GH43 were investigated and functionally characterized. The substrate specificity of *Ct*43Ara*f* and its truncated derivatives *Ct*GH43 against various natural polysaccharides and a few synthetic substrates *viz*. para-nitrophenyl glycosides (*p*NP-glycosides) were determined. The specific activity and kinetic parameters of full length *Ct*43Ara*f* (*Ct*GH43-*Ct*CBM6A-*Ct*CBM6B) and truncated *Ct*GH43 were compared to explore the influence of CBMs on enzyme activity. To our knowledge this is the first report of any α-L-arabinofuranosidase having the ability to hydrolyze both 4-nitrophenyl-α-arabinofuranoside (*p*NPA*f*) and 4-nitrophenyl-α-arabinopyranoside (*p*NPA*p*).

3.2 Materials and Methods

3.2.1 Substrates and reagents

Rye arabinoxylan, wheat arabinoxylan (soluble and insoluble), xyloglucan, arabinogalactan, sugar beet arabinan, pectic galactan, lupin galactan, pustulan, crudlan, mannan, galactomannan, rhamnogalactouronan were purchased from Megazyme International (Ireland). Oat spelt xylan, birchwood xylan, beechwood xylan; barley *β*-D-glucan, carboxy methylcellulose (CMC), carboxy ethylcellulose (CEC) and synthetic *p*NP-glycosides like *p*NP-α-L-arabinofuranoside (*p*NPA*f*) and *p*NP- α-L-arabinopyranoside (*p*NPA*p*), were purchased from Sigma-Aldrich Chemicals Pvt. Ltd., USA. D-xylose, L-arabinose, cellobiose, sodium hydroxide (NaOH) solution (50%, w/v), Bradford's reagent and ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid (EGTA) were 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 Enzyme activity assay

Initially, the assays of *Ct*43Ara*f* and *Ct*GH43 were carried out in 100 µl of a reaction mixture volume in 20 mM sodium phosphate buffer pH 7.0 containing 1% (w/v) substrate using 10 µl of purified enzyme (*Ct*43Ara*f* 0.45 mg/ml or *Ct*GH43 0.54 mg/ml). However, after optimization of pH and temperature, the assay mixture for *Ct*43Ara*f* contained 20 mM sodium phosphate buffer pH 5.7 while 20 mM sodium acetate buffer pH 5.4 was used for assay of *Ct*GH43. The reaction mixture was incubated at 50°C for 15 min. The enzyme activity was measured by estimating the liberated reducing sugar by the Nelson-Somogyi methodology as described in Section 3.2.2.4. To 100 µl of reaction mixture containing the reducing sugar, 100 µl of reagent D (Section 3.2.2.1) added. The solutions were mixed and heated for 20 min in the boiling water bath. After 20 min of boiling, the solution was ice cooled and 100 µl of reagent C (Section 3.2.2.1) was added. The colour developed rapidly and completed after the evolution of carbon dioxide stopped. The mixture was diluted by adding water to make up the volume to 1 ml and the absorbance was measured at 500 nm on a UV-Visible spectrophotometer (Varian, Carry 100 Bio). L-Arabinose in the range of 50-500 µg/ml was used for generating the standard plot as described in Section 3.2.2.3.

3.2.2.1 Preparation of reagents for reducing sugar estimation

Reagent A

Sodium carbonate anhydrous	6.25 g
Sodium potassium tatarate	6.25 g
Sodium bicarbonate	5.0 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.

3.2.2.3 Generation of standard plot of L-arabinose

The standard plot was prepared by varying the concentration of L-arabinose from 10-250 µg/ml. The reaction volume was kept to 100 µl with 50 mM sodium phosphate buffer pH 7.0. The reaction mixture containing buffer and L-arabinose was incubated (in a 1.5 ml microcentrifuge tube) at 50°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. Then 700 µl of deionized water was added to make the final volume to 1 ml. The absorbance was measured at 500 nm wavelength using UV-Visible spectrophotometer (Varian, Carry 100 Bio) against a buffer blank. A standard plot of optical density (OD) versus L-arabinose concentration (µg/ml) was generated and 1 OD equivalent of L-arabinose (µg/ml) was calculated. The 1 OD equivalent of L-arabinose (µg/ml) was converted in mg/ml terms for calculation of enzyme activity.

3.2.2.4 Calculation of enzyme activity of Ct43Araf and CtGH43

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*43Ara*f* and *Ct*GH43 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 L-arabinose concentration from standard plot

V = volume of the reaction mixture (ml)

t = time of reaction (min)

150 = molecular weight of L-arabinose

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

3.2.3 Substrate specificity of *Ct*43Ara*f* and *Ct*GH43

The recombinant proteins *Ct*43Ara*f* and *Ct*GH43 were analyzed against several natural polysaccharides. The major substrates (reported so far for family 43 GHs) were rye arabinoxylan, wheat arabinoxylan (soluble and insoluble), oat spelt xylan, birchwood xylan, beechwood xylan, galactan, arabinogalactan, arabinan. The activities against cellulosic substrates were also checked. Initially, the assays were performed to assess substrate specificities using 50 mM sodium phosphate buffer (pH 7.0) and taking 1% (w/v) final concentration of substrates in the assay mixture. The reaction was carried out at 50°C for 15 min. The enzyme activity was calculated as described in Section 3.2.2.

3.2.4 Determination of optimum pH and temperature of *Ct*43Ara*f* and *Ct*GH43

The pH profiles of *Ct*43Ara*f* and *Ct*GH43 were determined by performing enzyme assay in a wide range of pH using different buffers. Buffers of different pH range *viz.*, 20 mM sodium acetate (pH 4.0-5.6), 20 mM sodium phosphate (pH 5.7-7.5) and 20 Mm Tris/HCl (pH 7.5-8.0) were used for generating pH profile employing 1.0% (w/v) rye arabinoxylan as the substrate. Sodium phosphate and sodium acetate buffers were prepared as per the protocols described by Gomori (2010), whereas, Tris-HCl buffer was prepared following the method described by Sambrook and Russell (2001). The optimum temperature for enzyme activity of *Ct*43Ara*f* was determined by incubating the reaction mixture at varying temperatures in the range of 30-80°C, using 20 mM sodium phosphate buffer, pH 5.7. Similarly, the optimum temperature of *Ct*GH43 was determined using 20 mM sodium acetate buffer, pH 5.4. The enzyme activity was calculated by estimating the liberated sugar as described in Section 3.2.2.

3.2.5 pH stability of *Ct*43Ara*f* and *Ct*GH43

50 µl of *Ct*43Ara*f* (0.45 mg/ml) or *Ct*GH43 (0.45 mg/ml) were taken and lyophilized in a lyophillizer (Christ Griertrocknumgsanlagen GmbH, ALPHA 1-4). The protein was then dissolved in 50 µl of buffer (20 mM sodium phosphate or sodium acetate or Tris-HCl) and incubated at different pH varying from 4-8 for 15, 30 and 60 min. The lyophilized protein was dissolved in same amount of buffer so as to keep the protein concentration same in all tubes. The enzymes were then centrifuged at 10,000***g*** at 4°C for 5 min. 40 µl of supernatant was taken for determination of the enzyme activity that was calculated as per the description given in Section 3.2.2.

3.2.6 Thermal stability of *Ct*43Ara*f* and *Ct*GH43

*Ct*43Ara*f* (0.45 mg/ml) in 20 mM sodium phosphate buffer pH 5.7 and *Ct*GH43 (0.45 mg/ml) in 20 mM sodium acetate buffer pH 5.4 (50 µl) were incubated at various temperatures in the range of 4-80°C for 15, 30 and 60 min. The enzymes were then centrifuged at 10,000***g*** at 4°C for 5 min. 40 µl of supernatant was taken for estimation of the enzyme activity that was calculated as per description given in Section 3.2.2.

3.2.7 Substrate specificity of *Ct*43Ara*f* and *Ct*GH43 against natural polysaccharides

The enzyme assay for *Ct*43Ara*f* was performed using 20 mM sodium phosphate buffer (pH 5.7), whereas, the assay for *Ct*GH43 was carried out in 20 mM sodium acetate buffer (pH 5.4). The 100 µl reaction mixture contained 1.0% (w/v) substrate, 10 µl of enzyme (*Ct*43Ara*f,* 0.45 mg/ml or *Ct*GH43, 0.45 mg/ml). The reaction mixtures of both *Ct*43Ara*f* and *Ct*GH43 were incubated at 50ºC for 15 min. The assays were performed in duplicates. The concentration of reducing sugar was

estimated using a standard curve of arabinose as both *Ct*43Ara*f* and *Ct*GH43 predominantly showed α-L-arabinofuranosidase activity. 100 µl of reaction mixture was taken for estimation of enzyme activity that was calculated as per the description given in Section 3.2.2. One unit of activity was defined as the amount of enzyme which produced 1 µmole of arabinose per min. The kinetic parameters of *Ct*43Ara*f* and *Ct*GH43 were determined using various soluble substrates such as rye arabinoxylan, wheat arabinoxylan, oat spelt xylan, beechwood xylan and birchwood xylan and insoluble wheat arabinoxylan under above mentioned reaction conditions. The sodium acetate buffer was prepared following the method reported by Gomori (2010).

3.2.8 Substrate specificity of *Ct*43Ara*f* and *Ct*GH43 against synthetic *p*-nitrophenyl-glycosides (*p*NP-glycosides)

The assays of *Ct*43Ara*f* and *Ct*GH43 with synthetic substrates *p*-nitrophenyl glycoside (*p*NP-glycosides) *viz*., *p*-nitrophenyl-α-L-arabinofuranoside (*p*NPA*f*) and *p*-nitrophenyl-α-L-arabinopyranoside (*p*NPA*p*) were carried out by estimating the release of 4-nitrophenol (*p*NP) at 405 nm using a UV-Visible spectrophotometer (Varian, Carry 100 Bio). The enzyme reaction was performed in 1.0 ml reaction mixture containing *p*NPA*f* or *p*NPA*p* in 50 mM sodium phosphate buffer (pH 6.0), 20 µl of enzyme (*Ct*43Ara*f* 0.45 mg/ml or *Ct*GH43 0.45 mg/ml) incubated for 15 min at 50°C in a peltier temperature controller (Varian, Carry 100 Bio). The kinetic parameters of *Ct*43Ara*f* and *Ct*GH43 with *p*NPA*f* or *p*NPA*p* were determined by varying their concentrations in the range 20 to 500 µM. The reaction was stopped by adding 0.5 M sodium carbonate to make the reaction mixture highly alkaline (around pH 11.0). The assays were performed in triplicates. The released *p*NP was quantified using the molar extinction coefficient of 24150 litre/mole/cm as reported by Cartmell

et al. (2011). The preparation of stock solution of synthetic substrates required for enzyme assay is described in Section 3.2.8.1. The differences in chemical structure of *p*NPA*f* or *p*NPA*p* are depicted in Fig. 3.2.1A and 3.2.1B.

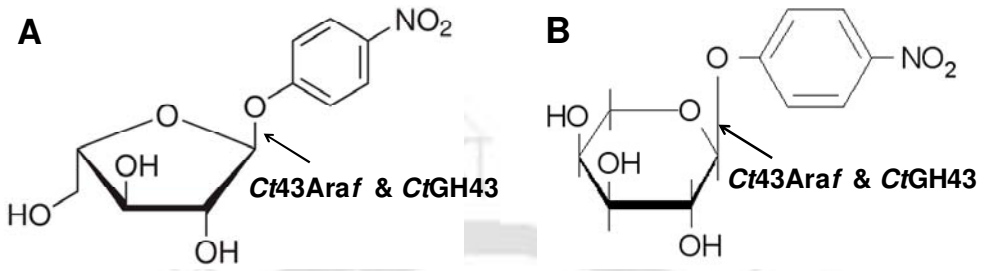


Fig. 3.2.1 Chemical structure of **A)** *p*-nitrophenyl-α-L-arabinofuranoside, **B)** *p*-nitrophenyl-α-L-arabinopyranoside. The arrows depict linkages / bonds where *Ct*43Ara*f* and *Ct*GH43 act to release *p*-nitrophenol.

3.2.8.1 Preparation of stock solution of pNPAf and pNPAp

2 mM stock solutions were prepared by taking 5.42 mg of *p*NPA*f* or *p*NPA*p* and first dissolving in 50-75 µl of dimethyl sulphoxide (DMSO) and then making up the final volume to 10 ml with 50 mM sodium phosphate buffer (pH 6.0). The solutions were kept on ice.

3.2.9 Effects of metal ions and chemical agents on enzyme activity of *Ct*43Ara*f* and *Ct*GH43

The effects of different metal cations on the activity of *Ct*43Ara*f* (0.45 mg/ml) and *Ct*GH43 (0.45 mg/ml) were determined using 100 µl reaction mixture (in duplicates) with oat spelt xylan (1%, w/v) as the substrate and adding respective metal salt at low molar concentrations (up to 20 mM). Assays for *Ct*43Ara*f* using 20 mM sodium phosphate buffer (pH 5.7) and for *Ct*GH43 using 20 mM sodium acetate buffer (pH 5.4) were performed at 50ºC. The reaction mixtures in both the case were subjected to 15 min of incubation. The blank with substrates having the respective salts were also

assayed in parallel. The effects of various salts of Na^{+}, Ca^{2+}, Mg^{2+}, Mn^{2+}, Zn^{2+}, Cu^{2+}, Co^{2+}, Ni^{2+}, Al^{3+} and chemical reagents like disodium EDTA and disodium EGTA were studied by varying their concentrations from 2.0-20.0 mM in the enzyme-substrate reaction mixture. The enzyme activity was calculated as described in Section 3.2.2. The control experiment contained no additives in the 100 µl reaction mixture.

3.2.10 Thin-Layer chromatography analysis of *Ct*43Ara*f* hydrolyzed products

The qualitative analysis of *Ct*43Ara*f* hydrolyzed products was performed by thin-Layer chromatography (TLC) on silica gel-coated aluminium plate (TLC Silica gel 60 $F_{254,}$ 20×20 cm, Merck) for detecting the sugars. The enzyme catalyzed reactions with 1% (w/v) substrate (rye arabinoxylan, wheat arabinoxylan and oat spelt xylan) were carried out in 100 µl reaction mixture maintaining the assay conditions as mentioned in Section 3.2.7. The 100 µl reaction mixture was then precipitated with 2 volumes of acetone and centrifuged at 4°C at 13,000***g*** for 5 min (Cote and Leathers, 2005). The supernatant was transferred to another micro-centrifuge tube and the reaction product precipitate was concentrated by evaporating the acetone. Then 2.0 µl of sample (enzyme-substrate reaction product) as well as standard (L-arabinose and D-xylose) solutions (1.0 mg/ml) were loaded on the TLC plate. The plate was dried for few min and kept in the developing chamber saturated with the developing solution (mobile phase). The mobile phase consisted of acetic acid-*n*-propanol-water-acetonitrile in the ratio 4:10:11:14 (Cote and Leathers, 2005). At the end of the run, migrated sugars were visualized by immersing the chromatogram in a solution (sulphuric acid: methanol 5:95, v/v; and α-napthol 5.0%, w/v). The TLC plates were then dried in a hot-air oven at 80°C for 20 min. The migrated sugars appeared as blue spots on the TLC plate.

3.2.11 High performance anion exchange chromatography (HPAEC) analysis of polysaccharide hydrolysis by *Ct*43Ara*f*

The *Ct*43Ara*f* (4.7 U/mg, 0.5 mg/ml) catalyzed reactions with 1% (w/v) substrate (rye arabinoxylan, wheat arabinoxylan and oat spelt xylan) was carried out in 100 µl reaction mixture maintaining the assay conditions as mentioned in Section 3.2.7. The 100 µl reaction mixture was incubated for 30 min at 50°C. The reaction was stopped by boiling the reaction mixture in a boiling water bath for 5 min. The 100 µl reaction mixtures were treated with 2 volumes of acetone to precipitate the remaining polysaccharides (substrates) and then centrifuged at 13,000***g*** for 10 min at 4°C. The supernatant containing the liberated sugar was transferred to another microcentrifuge tube and the acetone was removed by evaporation. The supernatant (50 µl) was diluted to 500 µl by adding ultra-pure (MilliQ) water and filtered through syringe filter using 0.2 µm membrane. The liberated sugars released due to the polysaccharide hydrolysis by enzyme reaction were analyzed by ion chromatography system or high pressure anion exchange chromatography (HPAEC) (Dionex, ICS-3000). From the filtered 500 µl, 25 µl of sample (liberated sugars) was run on CARBOPACK^{TM} PA-20 column (3 × 150 mm, Dionex), attached with CarboPac^{TM} PA20 guard column (3 × 30 mm) with Borate and Amino trap columns which removed impurities and provided high resolution. The instrument (Dionex, ICS-3000) was kept at constant temperature of 30°C during the analysis. The sample loop (sample loaded) size was kept to 25.0 µl and the flow rate was maintained at 0.3 ml/min. The elution of liberated sugars released due to enzymatic reaction was carried out with 50.0 mM sodium hydroxide (Sigma-Aldrich Chemicals Pvt Ltd., USA) using pulsed amperometric detector (PAD). L-arabinose and D-xylose (10 µg/ml) were used as standards. The solutions of standards were also filtered by passing through 0.2 µm filter before loading on the column.

3.2.12 Protein-melting study of *Ct*43Ara*f* and *Ct*GH43

Protein melting curves were generated by subjecting recombinant proteins (*Ct*43Ara*f* and *Ct*GH43) to varying temperatures and measuring the change in the absorbance at 280 nm by a UV-Visible spectrophotometer (Varian, Carry 100-Bio) following the method of Dvortsov *et al*. (2009). The column (IMAC) purified *Ct*43Ara*f* was dialyzed against 20 sodium phosphate buffer, pH 5.7, while, purified *Ct*GH43 was dialyzed against 20 mM sodium acetate bffer, pH 5.4 as described in Chapter 2, Section 2.2.21. The protein concentration for both *Ct*43Ara*f* and *Ct*GH43 were kept at 0.4 mg/ml. The absorbance at 280 nm was measured at different temperatures varying from 40-90°C using a peltier temperature controller. The protein solutions (1 ml) of *Ct*43Ara*f* and *Ct*GH43 were kept at the particular temperature for 10 min to attain the equilibrium. Similar experiment was carried out; with the addition of 10 mM $CaCl_2$ in the 1ml enzyme (0.4 mg/ml) solution and the temperature was then varied. The experiment was repeated with the addition of $CaCl_2$ and EGTA to 1 ml enzyme solution (0.4 mg/ml) in equi-molar concentrations of 10 mM, and finally the change in absorbance at 280 nm was measured. A curve of relative derivative absorption coefficient (first derivative coefficient) versus temperature was plotted to display the melting profile of the recombinant proteins, *Ct*43Ara*f* and *Ct*GH43, respectively. The first derivatives were calculated using Origin 8.0 software.

3.3 Results and Discussion

3.3.1 Substrate specificity of *Ct*43Ara*f* and *Ct*GH43 with natural polysaccharides

The enzyme assays with natural substrates were carried out using the optimized assay conditions as described in Section 3.2.7. The activities of full length *Ct*43Araf and its truncated derivative *Ct*GH43 with various natural polysaccharides are reported in Table 3.3.1. The maximum specific activity (U/mg) of *Ct*43Ara*f* and *Ct*GH43 were found to be 4.7 and 5.0, respectively, with rye arabinoxylan (Table 3.3.1). *Ct*43Ara*f* and *Ct*GH43 also displayed noticeable activity against wheat arabinoxylan (2.5 and 2.7 U/mg), oat spelt xylan (1.7 and 1.8 U/mg), beechwood xylan (1.0 and 0.9 U/mg) and birchwood xylan (0.7 and 0.8 U/mg) (Table 3.3.1). However, *Ct*43Ara*f* and *Ct*GH43 showed very less activity with arabinogalactan (0.25 and 0.32 U/mg) and α-L-arabinopyranosyl side chain containing polysaccharides *viz.* rhamnogalactouronan (0.22 and 0.23 U/mg) as displayed in Table 3.3.1.

In the last decade, few family 43 glycoside hydrolases have been reported from *Clostridium thermocellum* (galactanase), *Bacillus thetaiotaomicron* (α-1,2-arabinofuranosidase), *Celvibrio japonicas* (α-1,5-exoarabinanase) *Bacillus adolescentis* (only known arabinofuranosidase with ability to hydrolyze doubly substituted xylans) (Numan and Bhosle, 2006; Cartmell *et al.*, 2011). The catalytic modules (*Ct*43Ara*f* and *Ct*GH43) showed maximum activity against rye arabinoxylan, however, significant activities were also observed with wheat arabinoxylan, oat spelt xylan, birchwood xylan and beech wood xylan as displayed in Table 3.3.1. This indicated that the enzyme may be acting on α-1,3- as well as α-1,2- L-arabinosyl linked xylans. It has been previously reported that the rye arabinoxylans have nearly 50% (w/v) of the xylose residues substituted at O-3 and around 2% at both O-2 and O-3 by terminal L-arabinose (Bengtsson *et al.* 1992). This suggested that *Ct*43Ara*f*

displays α-1,3-arabinofuranosidase type of activity in exo-acting manner similar to the previous report by Bengtsson *et al*. (1992). *Ct*43Ara*f* and *Ct*GH43 also showed considerable activity with water soluble wheat arabinoxylan, which are rich in 2-mono and di-substituted xyloses and low in 3-mono and di-substituted xylose (Gruppen *et al*., 1992), showing that they are also having specificity to act on O-2 substituted xylose as also reported earlier (Numan and Bhosle, 2005; Saha, 2003). *Ct*43Ara*f* and *Ct*GH43 showed noticeable activities with beechwood and birchwood xylans comprising O-2 and O-3 substituted xylose. This further indicated that both the catalytic modules have the capacity to catalyze the hydrolysis of terminal non-reducing α-L-1,2- as well as α-L-1,3- arabinosyl residues, in an exo-acting manner as also reported by Bourgois *et al.* (2007). The activity of *Ct*43Ara*f* and *Ct*GH43 with oat spelt xylan was mainly due to the presence of 10% (w/w) arabinoxylans (arabinose at side chains). The low activity with arabinogalactan and rhamnogalactouronan was mainly due to the fact that the β-1,4-galactans are poorly substituted with α-L-arabinopyranose side-chains as reported by ØBro *et al.,* (2004). Therefore, it can be inferred that the enzymes (*Ct*43Ara*f* and *Ct*GH43*)* acted on the glycosidic linkage of α-arabinfuranosyl substituted main chain of β-1,4-xylose, occurring mainly in the arabinoxylans.

Table 3.3.1 Substrate specificity of *Ct*43Ara*f* and *Ct*GH43 from *C. thermocellum.*

Substrates	Specific Activity of [a] *Ct*43Ara*f* (U/mg)	Specific Activity of [b] *Ct*GH43 (U/mg)
Arabinoxylan, (Rye)	4.70±0.07	5.00±0.08
Arabinoxylan (wheat, soluble)	2.50±0.03	2.70±0.03
Xylan (Oat spelt)	1.70±0.08	1.80±0.07
Arabinoxylan (wheat, insoluble)	1.20±0.10	1.10±0.10
Xylan (Beechwood)	1.00±0.04	0.90±0.04
Xylan (Birchwood)	0.70±0.03	0.80±0.04
Arabinogalactan	0.25±0.05	0.32±0.05
Arabinan (sugar beet)	0.22±0.04	0.23±0.04
Rhamnogalactouronan	0.17±0.03	0.18±0.03
Galactan (Lupin)	0.10±0.03	0.10±0.03
β-D-Glucan (Barley)	0.10±0.02	0.10±0.04
Mannan	0.10±0.05	0.10±0.05
Carob galactomannan	0.10±0.03	0.1±0.02
Pectic galactan (Apple)	0.10±0.03	0.10±0.02
Pectic galactan (Citrus)	0.10±0.03	0.10±0.02
Pectic galactan (Lupin)	0.10±0.04	0.10±0.04
Carboxy methyl cellulose	ND	ND
Carboxy ethylcellulose	ND	ND

All the assays were performed at 50°C using 20 mM sodium phosphate (pH 5.7) buffer for [a]Ct43Araf and 20 mM sodium acetate (pH 5.4) buffer for [b]CtGH43.

The assays were performed in triplicates. The incubation time and other conditions for reducing sugar estimation were as same as described in the Section 3.2.7.

ND = *No activity detected.*

3.3.2 Substrate specificity and kinetic parameters of *Ct*43Ara*f* and *Ct*GH43 with natural and synthetic substrates

The kinetic properties *viz.* turn over number (k_{cat}, min^{-1}) and catalytic efficiency (k_{cat}/K_M, $min^{-1}mg^{-1}ml^{-1}$) of both the enzymes were determined with the natural as well as synthetic substrates (Table 3.3.2). *Ct*43Ara*f* and *Ct*GH43 showed maximum turnover number (k_{cat}) of 280 and 298 min^{-1} and also maximum catalytic efficiency (k_{cat}/K_M, $min^{-1}mg^{-1}ml^{-1}$) of 3.4×10^3 and 3.6×10^3 with rye arabinoxylan (Table 3.3.2). *Ct*43Ara*f* and *Ct*GH43 also displayed activity with insoluble wheat arabinoxylan, showing catalytic efficiencies of 7.1×10^2 $min^{-1}mg^{-1}ml^{-1}$ and 6.1×10^2 $min^{-1}mg^{-1}ml^{-1}$, respectively (Table 3.3.2). The catalytic efficiencies with other soluble substrates like oat spelt xylan (7.7×10^2 $min^{-1}mg^{-1}ml^{-1}$), beechwood (40 $min^{-1}mg^{-1}ml^{-1}$) and birchwood xylans (28 $min^{-1}mg^{-1}ml^{-1}$) were comparatively less as compared with rye arabinoxylan (Table 3.3.2).

The catalytic efficiencies of *Ct*43Ara*f* and *Ct*GH43 with *p*NP-*α*-L-arabinofuranoside (*p*NPA*f*) were found to be 5.6×10^3 $min^{-1}mg^{-1}ml^{-1}$ and 7.1×10^3 $min^{-1}mg^{-1}ml^{-1}$, respectively, while, the k_{cat}/K_M values of both the modules with *p*NP-*α*-L-arabinopyranoside (*p*NPA*p*) were 2.2×10^3 $min^{-1}mg^{-1}ml^{-1}$ (Table 3.3.2). This showed that the catalytic modules *Ct*43Ara*f* and *Ct*GH43 were able to release *p*-nitrophenol from both *p*NP-*α*-L-arabinofuranoside (*p*NPA*f*) as well as *p*NP-*α*-L-arabinopyranoside (*p*NPA*p*), but the catalytic efficiencies of both modules were approximately, 3-fold higher with *p*NP-*α*-L-arabinofuranoside as compared with *p*NP-*α*-L-arabinopyranoside (Table 3.3.2). Based on the catalytic efficiencies of *Ct*43Ara*f* and *Ct*GH43 with natural as well as synthetic substrates, it is evident that both these catalytic modules are predominantly *α*-L-arabinofuranosidase.

Table 3.3.2 Kinetic properties and catalytic efficiencies of *Ct*43Ara*f* and *Ct*GH43 from *C. thermocellum.*

Substrates	K_M (mg ml^{-1})	k_{cat} (min^{-1})	k_{cat} / K_M (min^{-1}mg^{-1}ml^{-1})
Ct*43Ara*f			
[a]**Natural polysaccharides**			
Rye arabinoxylan	0.082 ± 0.005	280.0. ± 4	3.4×10^3
Wheat arabinoxylan (soluble)	0.072 ± 0.003	190.0 ± 2	2.6×10^3
Wheat arabinoxylan(insoluble)	0.09 ± 0.015	63.0 ± 12	7.1×10^2
Oat spelt xylan	0.085 ± 0.005	65.0 ± 3.0	7.7×10^2
Birchwood xylan	0.95 ± 0.004	27.0 ± 0.9	2.8×10^1
Beechwood xylan	0.7± 0.004	28.0 ± 0.5	4.0×10^1
[c]**Synthetic *p*NP-glycosides**			
*p*NP-α-L-arabinofuranoside	0.05 ± 0.002	283.0 ± 2.0	5.6×10^3
*p*NP-α-L-arabinopyranoside	0.093 ± 0.003	210.0 ± 4.7	2.2×10^3
***Ct*GH43**			
[b]**Natural polysaccharides**			
Rye arabinoxylan	0.08 ± 0.002	298.0 ± 8.0	3.6×10^3
Wheat arabinoxylan (soluble)	0.078 ± 0.005	209.0 ± 2.0	2.7×10^3
Wheat arabinoxylan(insoluble)	0.1 ± 0.01	61.0 ± 18	6.1×10^2
Oat spelt xylan	0.08 ± 0.002	67.0 ± 2.0	8.3×10^2
Birchwood xylan	0.9 ± 0.002	29.0 ± 1.0	3.2×10^1
Beechwood xylan	0.8± 0.003	26.0 ± 2.0	3.3×10^1
[c]**Synthetic *p*NP-glycosides**			
*p*NP-α-L-arabinofuranoside	0.04 ± 0.004	287.0 ± 1.0	7.1×10^3
*p*NP-α-L-arabinopyranoside	0.097 ± 0.004	212.0 ± 2.0	2.2×10^3

The assays with natural substrates were carried out with 20 mM sodium phosphate buffer (pH 5.7) for [a]Ct43Araf and sodium acetate buffer (pH 5.4) for [b]CtGH43 at 50°C.

The assays were performed in triplicates. The incubation time and other conditions for reducing sugar estimation were as same as described in the Section 3.2.7.

[c] Assays with synthetic pNP-glycosides were carried out in 20 mM sodium phosphate buffer pH 5.7.

3.3.3 Effect of pH and temperature on the enzyme activity of *Ct*43Ara*f*

The enzyme activity of *Ct*43Ara*f* displayed significant dependence on the pH value of the buffer and the temperature at which the enzyme-substrate mixture was incubated. Initially, with rising pH, the enzyme activity increased to a maximum (pH 5.7) and then dropped to almost zero in the alkaline pH region (above pH 8.0) and the pH profile appeared as bell-shaped curve (Fig. 3.3.1). The full length catalytic module *Ct*43Ara*f* showed maximum enzyme activity at pH 5.7 (4.7 U/mg). The optimum temperature for *Ct*43Ara*f* was 50ºC and it displayed significant loss of activity at temperatures above 50°C (Fig. 3.3.1).

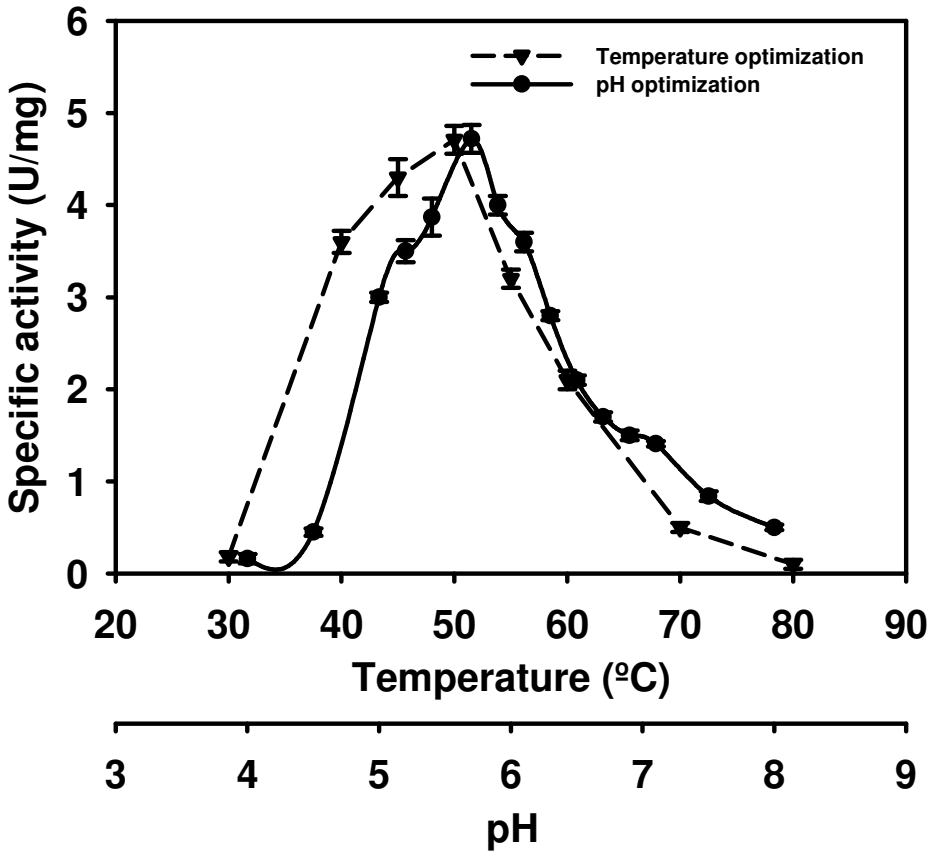


Fig. 3.3.1 Effect of pH and temperature on *Ct*43Ara*f* activity, (●) represents pH profile and (▼) represents temperature profile. *Ct*43Ara*f* (0.5 mg/ml) was incubated with rye arabinoxylan (1%, w/v) at 50°C for 15 min and the liberated sugar was estimated using Nelson and Somogyi's method as described in Section 3.2.7.

3.3.4 Effect of pH and temperature on the enzyme activity of *Ct*GH43

The enzyme activity of *Ct*GH43 increased to maximum (5 U/mg) at pH 5.4 and then it decreased to almost zero in the alkaline pH region (above pH 8.0) and the pH profile appeared as bell-shaped curve (Fig. 3.3.2) similar to that of *Ct*43Ara*f* (Fig. 3.3.2). The optimum temperature for *Ct*GH43 was found to be 50ºC (Fig. 3.3.2) same as that of *Ct*43Ara*f* (Fig. 3.3.2). Like *Ct*43Ara*f*, *Ct*GH43 also displayed loss of activity at higher temperatures than 50ºC (Fig. 3.3.2).

.

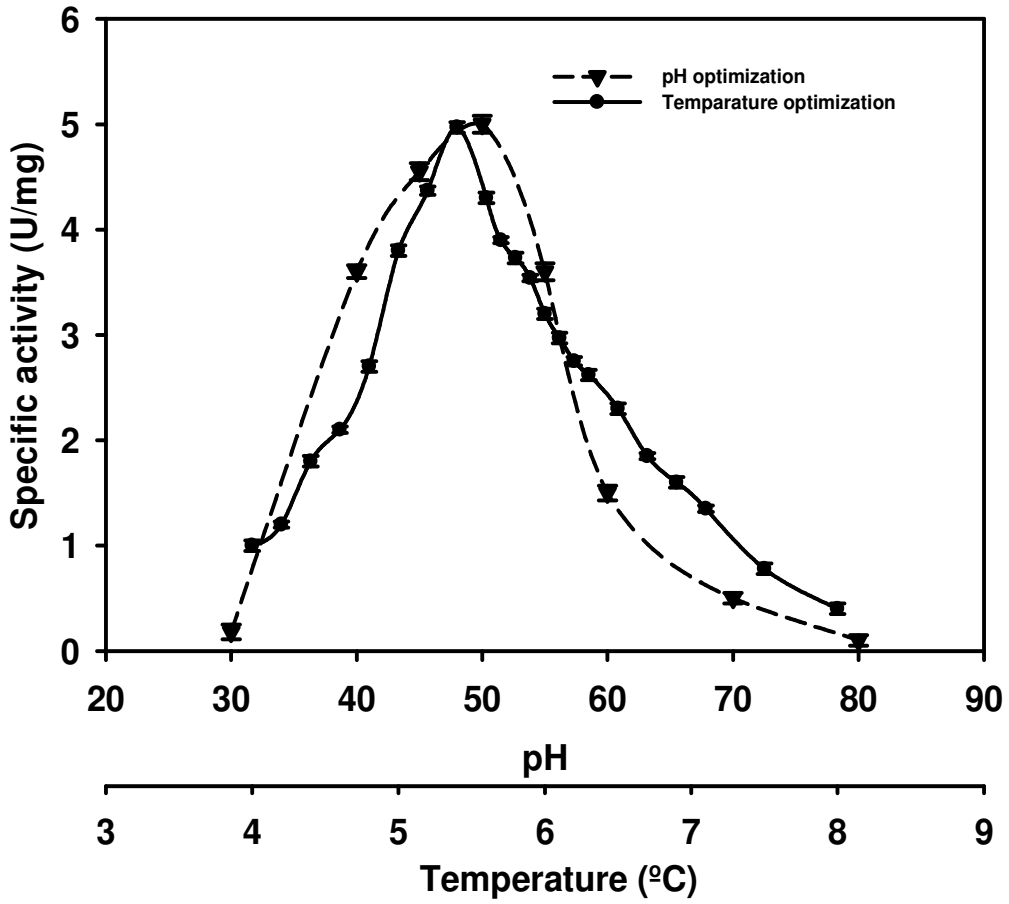


Fig. 3.3.2 Effect of pH and temperature on *Ct*GH43 activity, (●) represents pH profile and (▼) represents temperature profile. *Ct*GH43 (0.5 mg/ml) was incubated with rye arabinoxylan (1%, w/v) at 50°C for 15 min and the liberated sugar was estimated using Nelson and Somogyi's method as described in Section 3.2.7.

3.3.5 pH and thermal stability of *Ct*43Ara*f*

*Ct*43Ara*f* was found to be stable in the pH range of 5.0-6.5 (Fig. 3.3.3). The pH stability of an enzyme is governed by mainly two factors, i) type of enzyme and direct involvement of ionic groups in the catalytic mechanism and ii) participation of charged groups in the stabilization of the protein structure (Branden and Tooze, 1991). Ionic groups are often involved in enzymatic catalysis, and the state of these groups protonation is essential for the reaction. Deviations from the optimum pH value of *Ct*43Ara*f* may alter the state of protonation of the groups involved and perturb their involvement in the catalytic process. *Ct*43Ara*f* showed almost similar pH stability profile to that of optimum pH profile and this is because in both case, the same charged groups (similar amino acids) may be involved for stabilization of protein structure. *Ct*43Ara*f* was found to be stable in temperature range of 4-50°C (Fig. 3.3.3). However at higher temperature than 60°C, the enzyme stability was significantly decreased (Fig. 3.3.3). The most accepted explanation for the loss of enzyme activity at high temperature is the opening or uncoiling of the protein architecture, which results in decreased stability and consequently lower activity (Branden and Tooze, 1991, Creighton, 1993). The opening of protein tertiary structure changes the conformation of protein architecture that directly affects the velocity of enzyme catalyzed reaction.

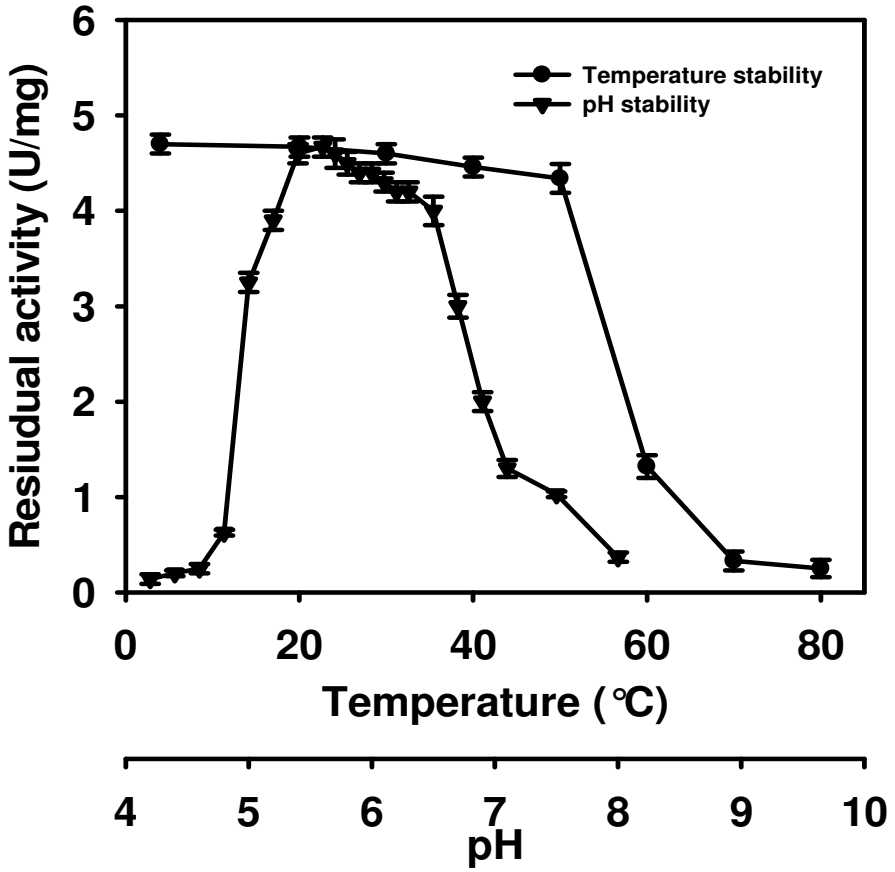


Fig. 3.3.3 pH and thermal stability analysis of *Ct*43Ara*f*. (▼) represents pH stability and (●) represents thermal stability. 50 µl *Ct*43Ara*f* (0.5 mg/ml) was incubated at different temperatures (4-50°C) for 30 min. After centrifugation, 40 µl aliquot of *Ct*43Ara*f* was incubated with rye arabinoxylan (1%, w/v) in a 100 µl reaction mixture at 50°C for 15 min. The liberated sugar was estimated using Nelson and Somogyi's method as described in Section 3.2.7.

3.3.6 pH and thermal stability of *Ct*GH43

*Ct*GH43 was stable in the pH range of 5-6.3 (Fig. 3.3.4) which is similar to the pH stability profile of *Ct*43Ara*f* (Fig. 3.3.4). This is mainly due to the fact that the charges involved in the stabilization of protein structure are same in both *Ct*43Ara*f* as well as *Ct*GH43. The thermal stability results indicated *Ct*GH43 remained almost unaffected in the temperature range 4-50°C (Fig. 3.3.4) which is similar to that *Ct*43Ara*f* (Fig. 3.3.3). But the thermal stability of both these enzymes was

significantly affected above 50°C as more than 70% decrease in enzyme activity was observed in both *Ct*43Ara*f* and *Ct*GH43 at temperatures above 60°C (Fig. 3.3.3 and Fig. 3.3.4, respectively). Similar to chemical reactions the velocity of the enzyme-catalysed reaction remained steady with increase in temperature up to 50°C (Fig. 3.3.3 and Fig. 3.3.4). From the temperature stability profile of *Ct*GH43 it is evident that at temperatures higher than 50°C, the enzyme activity declined mainly due loss of stability of protein structure. The thermostability profile indicates that 95% of the *Ct*GH43 activity was retained between 25- 50°C (Fig. 3.3.4).

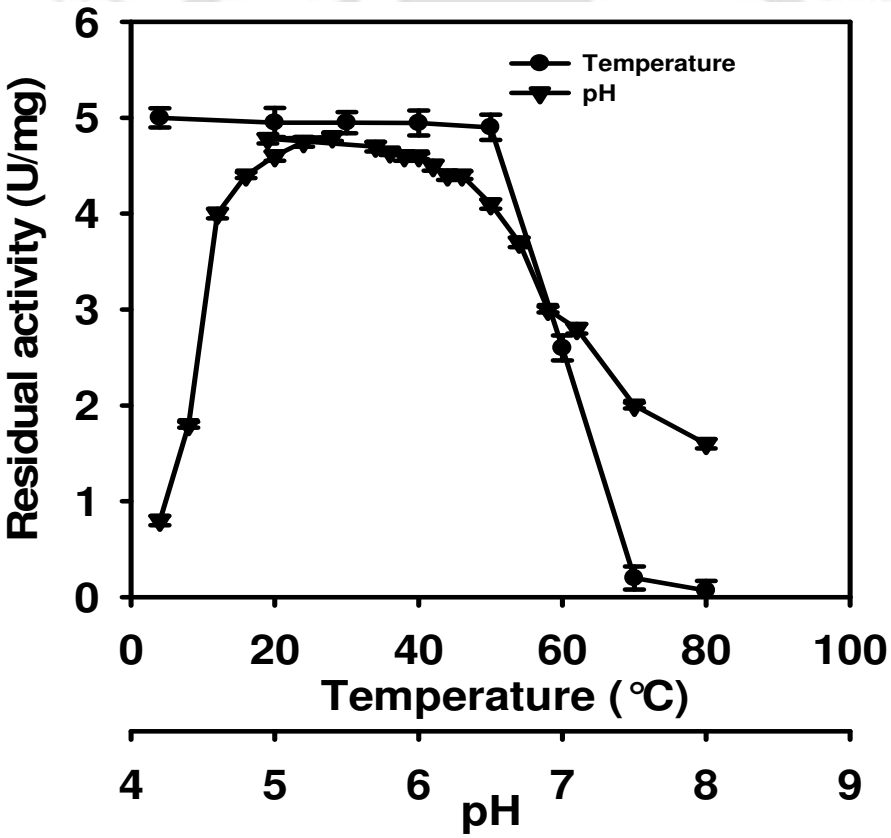


Fig. 3.3.4 pH and thermal stability analysis of *Ct*GH43. (▼) represents pH stability and (●) represents thermal stability. 50 µl *Ct*GH43 (0.5 mg/ml) was incubated at different temperatures (4-50°C) for 30 min. After centrifugation, 40 µl aliquot of *Ct*GH43 was incubated with rye arabinoxylan (1%, w/v) in a 100 µl reaction mixture at 50°C for 15 min. The liberated sugar was estimated using Nelson and Somogyi's method as described in Section 3.2.7.

3.3.7 Effects of metal ions and chemical agents on enzyme activity of *Ct*43Ara*f* and *Ct*GH43

The enzyme activity of *Ct*43Ara*f* and *Ct*GH43 remained largely unaffected up to 50 mM of sodium chloride but drastically decreased after 100 mM NaCl (Fig. 3.3.5A). The enzyme activities of *Ct*43Ara*f* and *Ct*GH43 increased 2-fold with lower concentrations of Ca^{2+} (8 mM) and Mg^{2+} (6 mM) ions, but sharply decreased after 10 mM (Fig. 3.3.5B and Fig. 3.3.5C, respectively). A marginal increase in activity of *Ct*43Ara*f* and *Ct*GH43 was also observed with Ni^{2+} salts (15 % and 47%; Fig. 3.3.5D), Zn^{2+} (8% and 24%; Fig. 3.3.5E) and Mn^{2+} (20% and 21%; Fig. 3.3.5F) salts for *Ct*43Ara*f* and *Ct*GH43, respectively. The enzyme activity of *Ct*43Ara*f* and *Ct*GH43 remained unaffected up to 5 mM concentrations of Cu^{2+}, Co^{2+} and Hg^{2+} salts (Fig. 3.3.5G, Fig. 3.3.5H and Fig. 3.3.5I, respectively). However, the enzyme activity of both *Ct*43Ara*f* and *Ct*GH43 significantly decreased at higher concentrations (20 mM) of Cu^{2+} (70%; Fig. 3.3.5G), Co^{2+} (75%; Fig. 3.3.5H), Hg^{2+} (80% and 70%, respectively; Fig. 3.3.5I) and Ag^{+} (80% and 75%, respectively; Fig. 3.3.5J). The enzyme activity of both the catalytic modules decreased to more than 90% in presence of 10 mM EDTA (Fig. 3.3.5K) as well as of 10 mM EGTA (Fig. 3.3.5L). The extent of decrease in enzyme activity in the presence of EGTA indicated that Ca^{2+} ions may be essential for their catalytic activity, as EGTA specifically binds and chelates Ca^{2+} ions in 1:1 molar ratio (Qin *et al*., 1999). The increase in catalytic activity of *Ct*43Ara*f* and *Ct*GH43 in the presence of Ca^{2+} and Mg^{2+} ions showed that these might be functioning as co-factors. The heavy metals especially Cu^{2+}, Co^{2+}, Hg^{2+}, and Ag^{+} ions at 10 mM concentration caused decrease in enzyme activity as shown previously for recombinant cellulases (Bharali *et al*., 2005; Ahmed *et al*., 2009). This decrease can be attributed to the fact that the above metal ions may be reacting with the sulfhydryl (-SH) groups of cysteine amino acid present in the catalytic modules (*Ct*43Ara*f* and

*Ct*GH43) and thereby causes hindrance for the substrate to bind with enzymes (Creighton, 1992). The effects of metal cation and chemical agents on the enzyme activity of *Ct*43Ara*f* and *Ct*GH43 have been summarized in Table 3.3.3. Table 3.3.3 showed that *Ct*43Ara*f* and *Ct*GH43 activities were 2-fold increased in presence of Ca^{2+} and Mg^{2+} ions, whereas, EDTA and EGTA at merely 10 mM concentration caused almost complete loss of activity. This suggested that *Ct*43Ara*f* and *Ct*GH43 may be metalloproteins and essentially require Ca^{2+} and Mg^{2+} for enzyme activity. Only a marginal increase in enzyme activity of *Ct*43Ara*f* and *Ct*GH43 were observed with Zn^{2+}, Mn^{2+} and Ni^{2+} ions, however, other heavy metals *viz.* Co^{2+}, Hg^{2+}, Cu^{2+} and Ag^{+} decreased the activity significantly at 10 mM concentration (Table 3.3.3). The control experiment contained no additives in the 100 µl reaction mixture.

Table 3.3.3 Maximum effect on enzyme activity of *Ct*43Ara*f* and *Ct*GH43 from *Clostridium thermocellum* at maximum concentration of metal ions and chelating agents.

Metal ion/Reagent	Concentration of additives (mM)	Relative activity (%)	
		Ct*43Ara*f	***Ct*GH43**
Control	--	100	100
Na^{+}	50.0	100	100
Ca^{2+}	8.0	216	217
Mg^{2+}	6.0	211	207
Ni^{2+}	4.0	115	147
Zn^{2+}	2.0	124	108
Mn^{2+}	4.0	120	121
Fe^{3+}	20.0	50	50
Al^{3+}	20.0	50	50
Cu^{2+}	10.0	30	30
Co^{2+}	20.0	25	25
Hg^{2+}	20.0	20	30
Ag^{+}	20.0	20	25
EDTA	20.0	05	05
EGTA	20.0	05	05

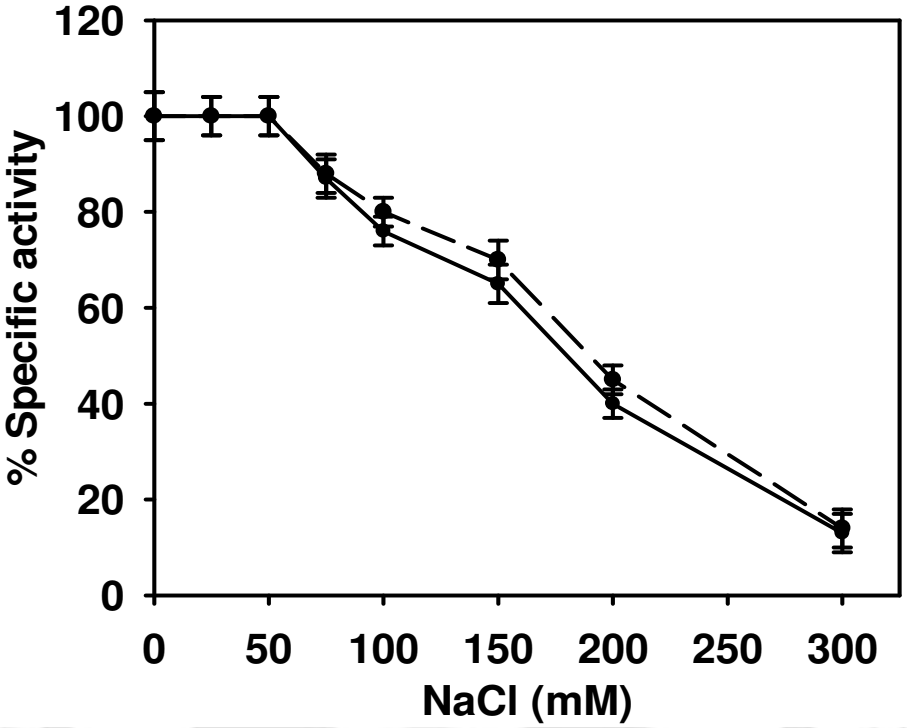


Fig. 3.3.5A Effect of NaCl on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (---) from *Clostridium thermocellum.*

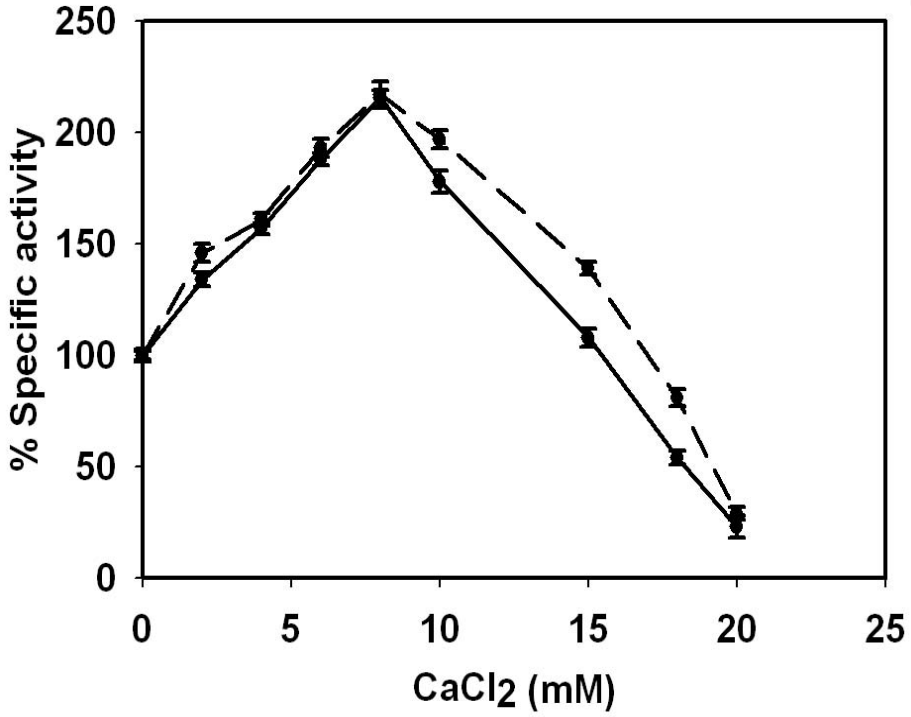


Fig. 3.3.5B Effect of $CaCl_2$ on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

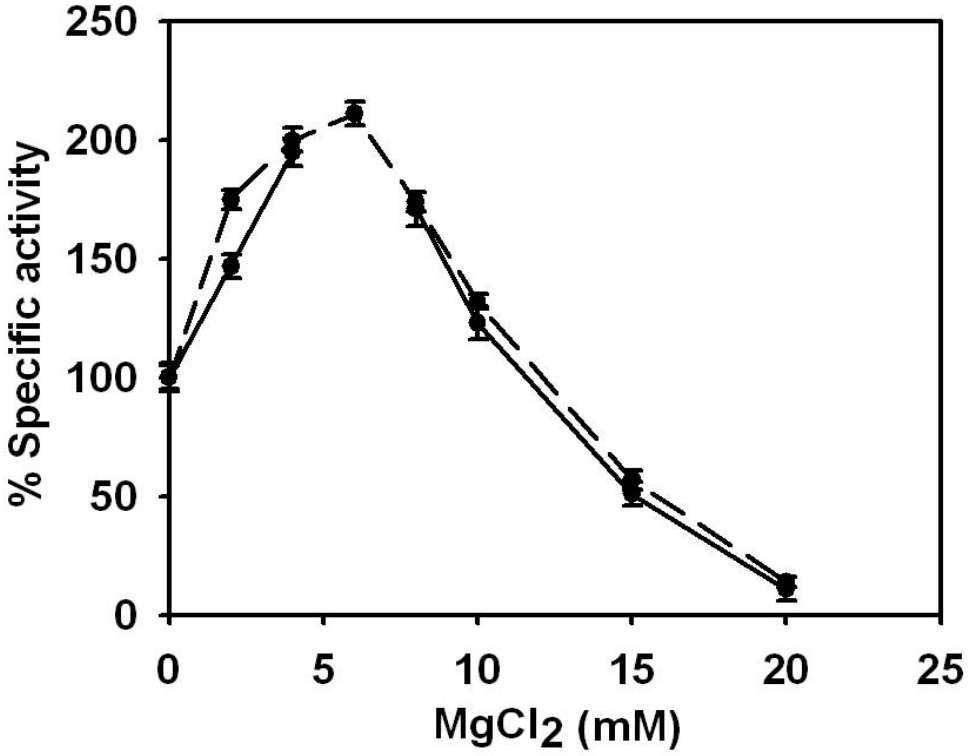


Fig. 3.3.5C Effect of $MgCl_2$ on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (---) from *Clostridium thermocellum.*

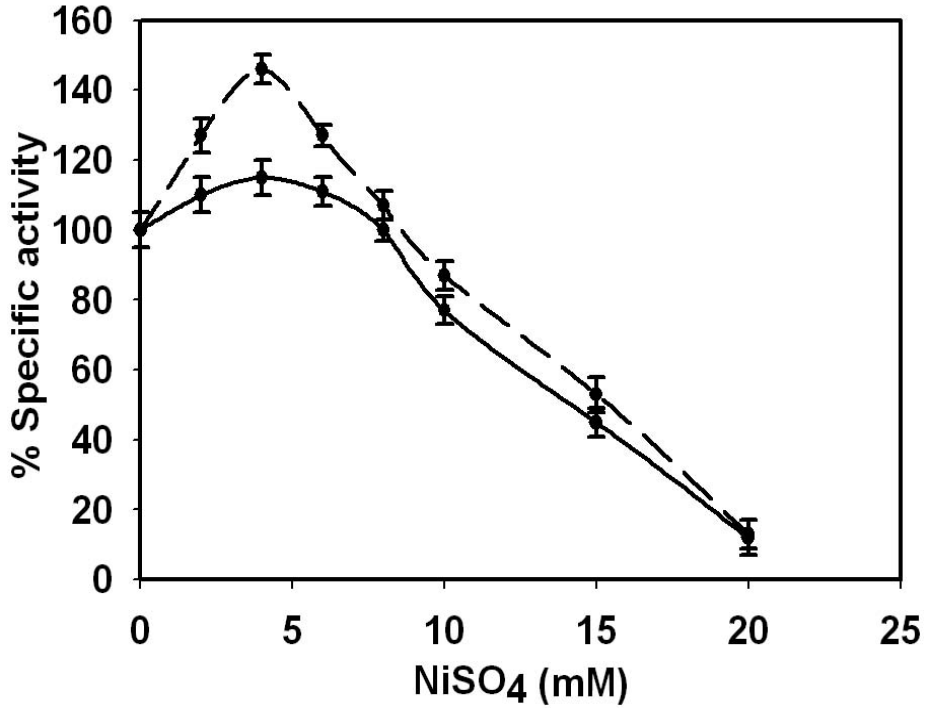


Fig. 3.3.5D Effect of $NiSO_4$ on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

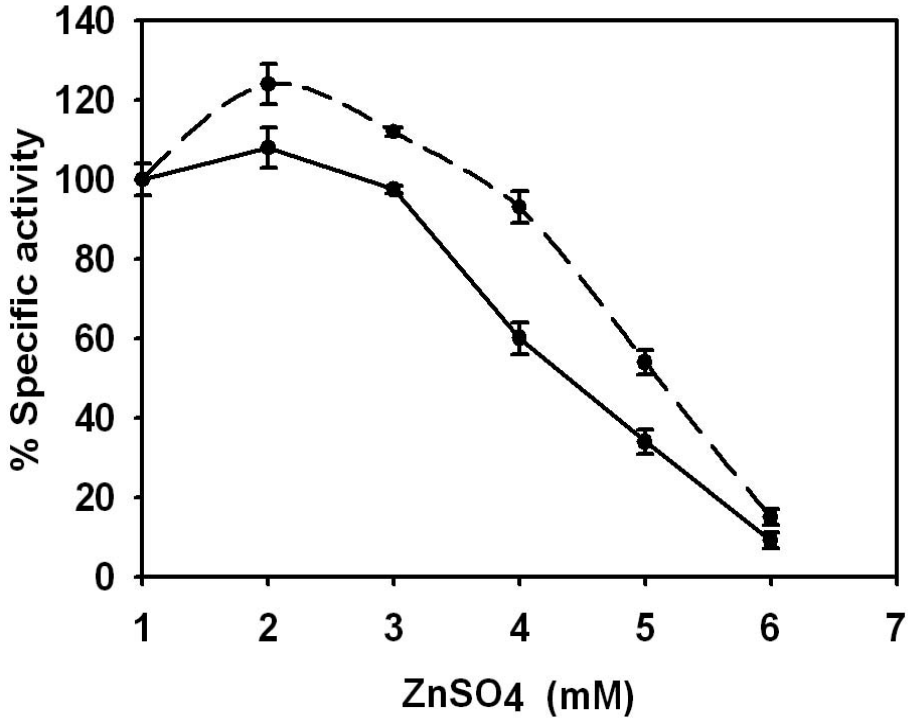


Fig. 3.3.5E Effect of $ZnSO_4$ on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

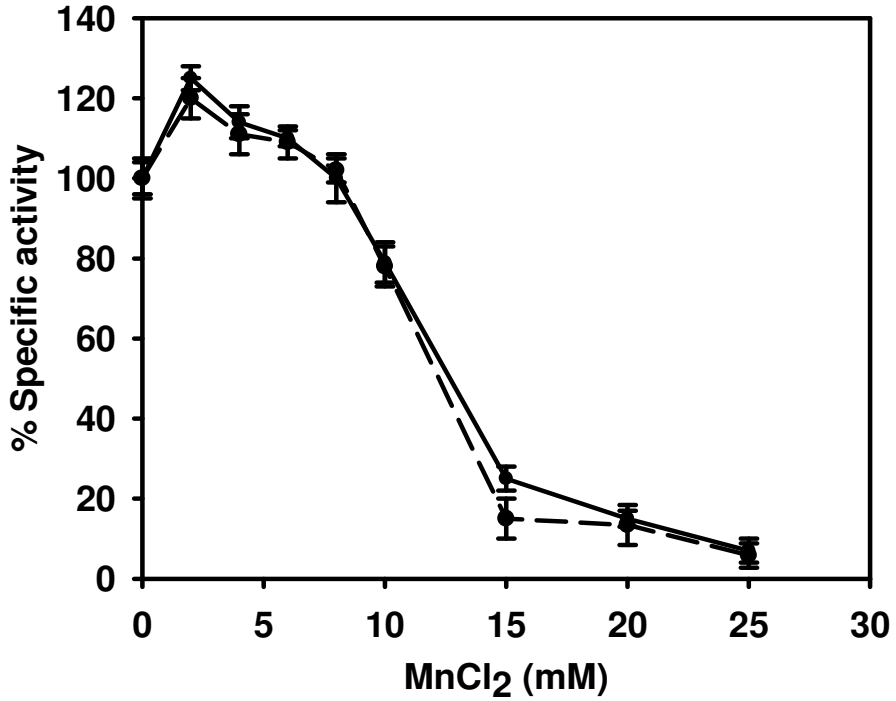


Fig. 3.3.5F Effect of $MnCl_2$ on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43(----) from *Clostridium thermocellum.*

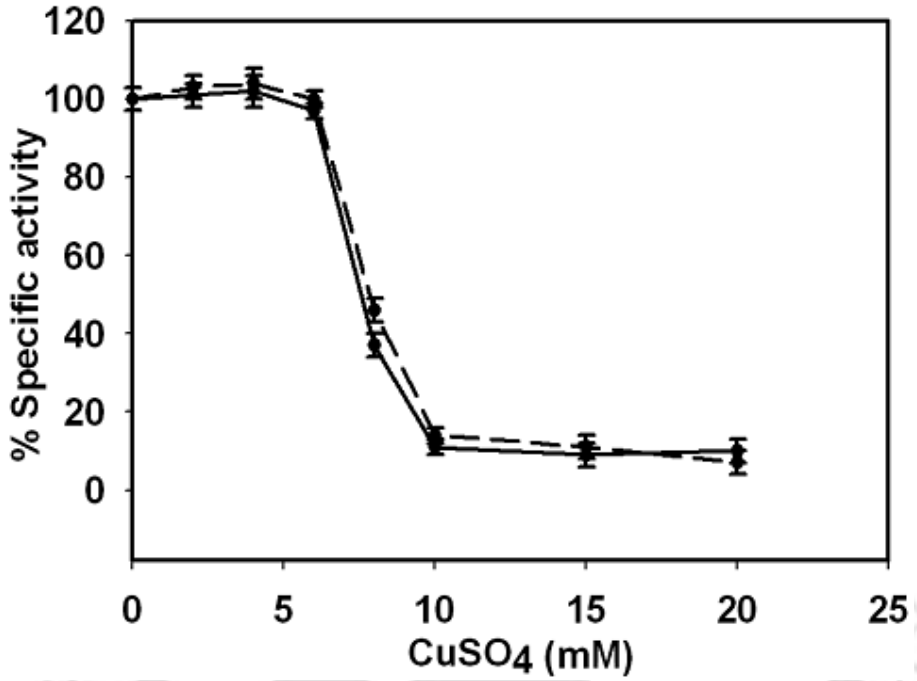


Fig. 3.3.5G Effect of $CuSO_4$ on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

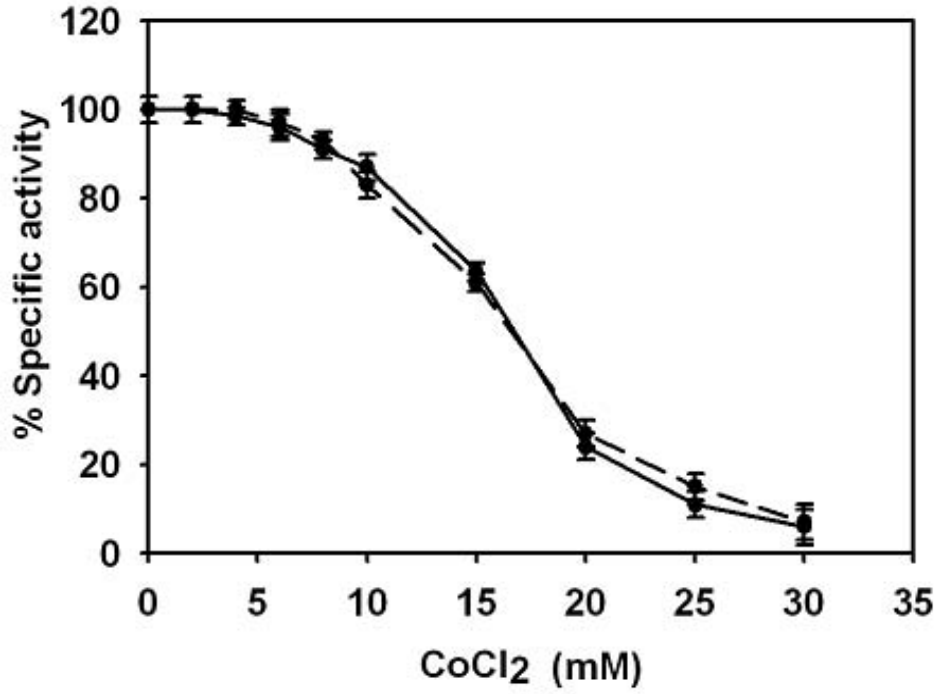


Fig. 3.3.5H Effect of $CoCl_2$ on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

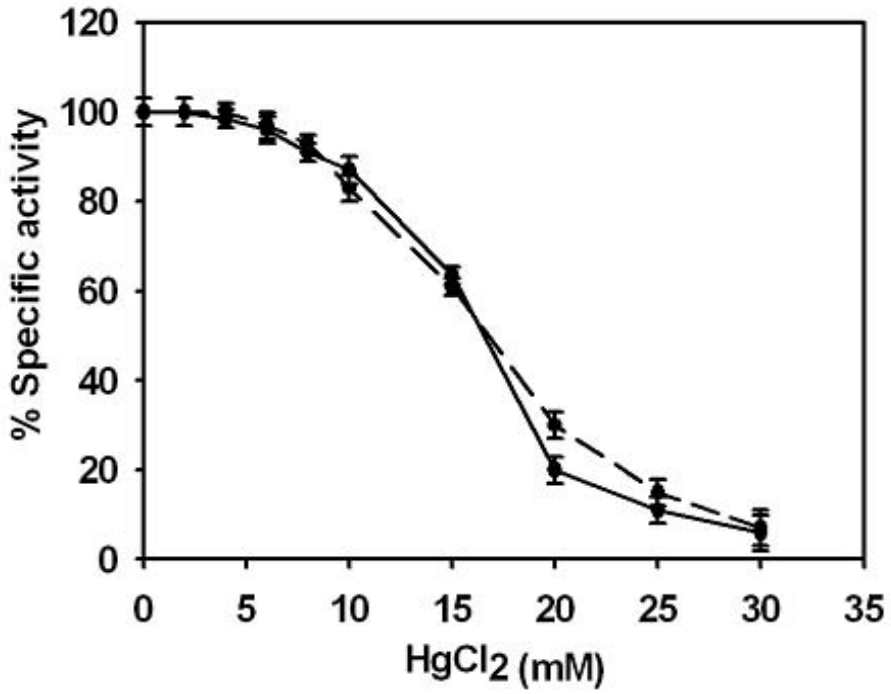


Fig. 3.3.5I Effect of $HgCl_2$ on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

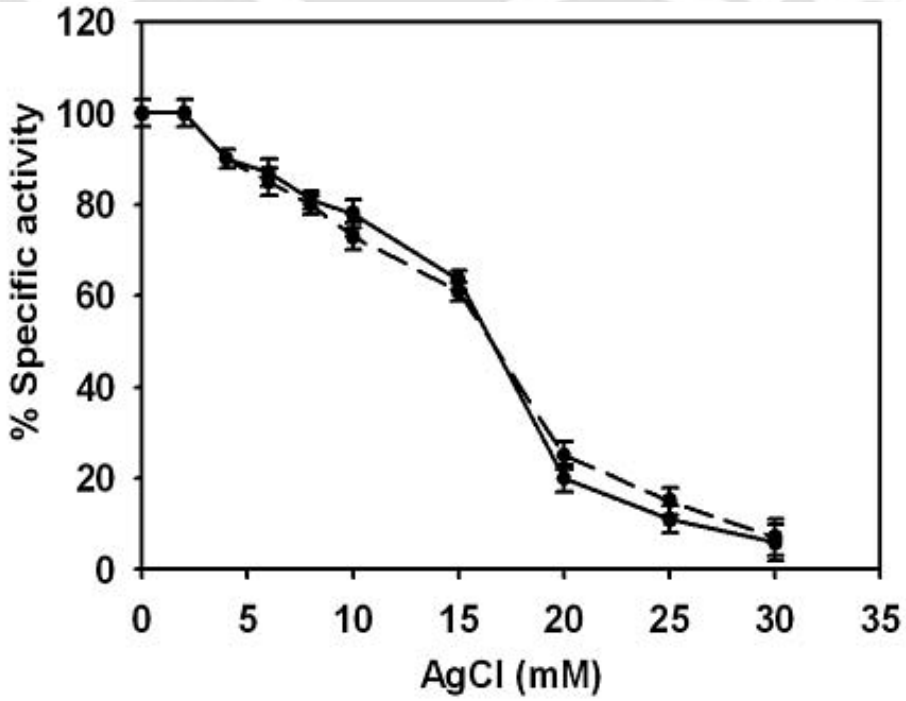


Fig. 3.3.5J Effect of AgCl on enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

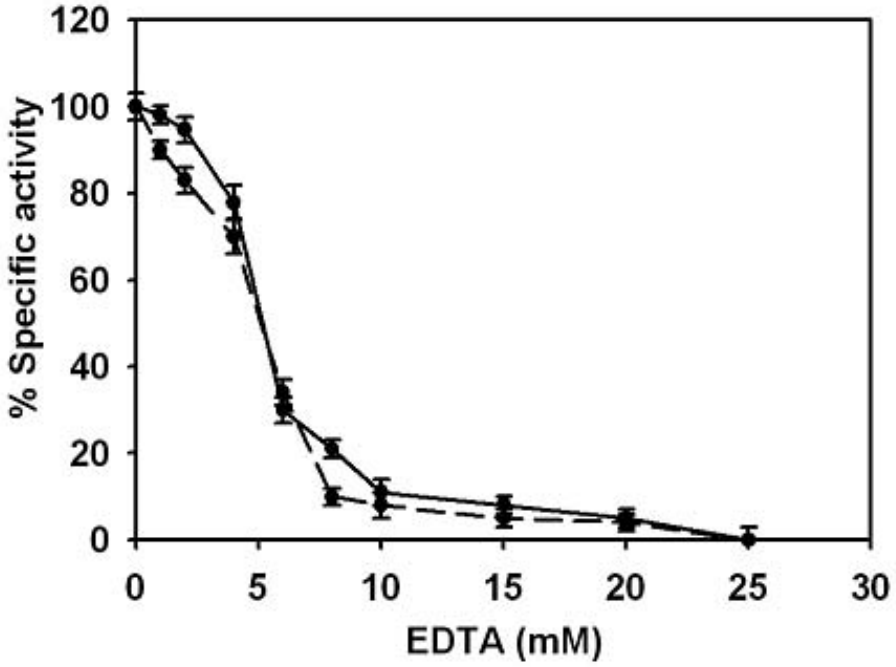


Fig. 3.3.5K Effect of EDTA on the enzyme activity of *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

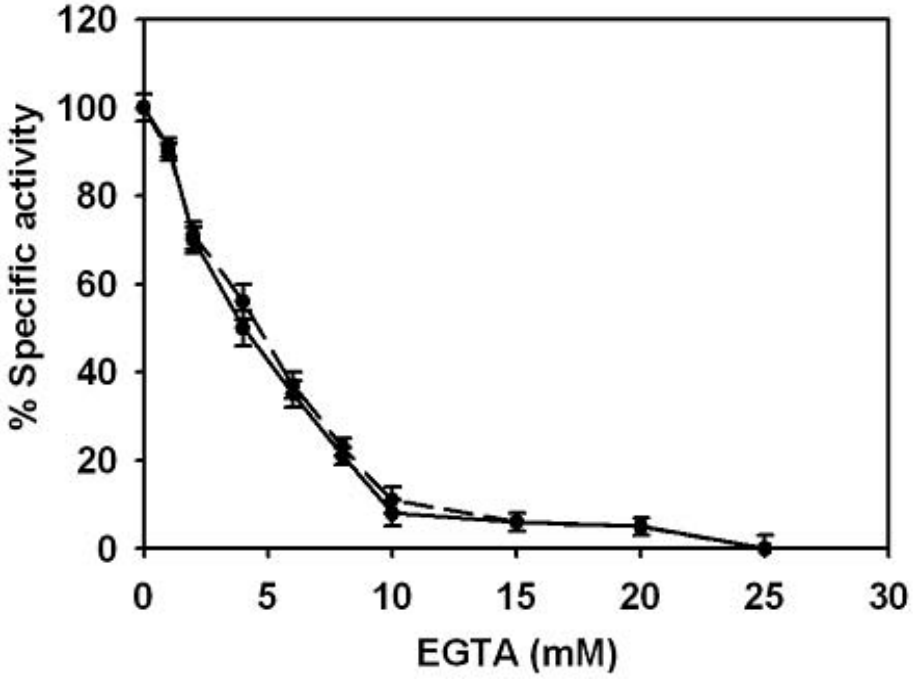


Fig. 3.3.5L Effect of EGTA on enzyme activity *Ct*43Ara*f* (—) and *Ct*GH43 (----) from *Clostridium thermocellum.*

3.3.8 Thin layer chromatography analysis of reaction products of *Ct*43Ara*f*

Thin layer chromatography (TLC) analysis of the enzyme reaction products (sugars) showed that *Ct*43Ara*f* released only L-arabinose from rye arabinoxylan, wheat arabinoxylan and oat spelt xylan (Fig. 3.3.6). The comparison of relative migration of *Ct*43Ara*f* hydrolyzed sugars with commercially available standards clearly indicated that arabinose was the major sugar released as no spots for xylose was observed on the TLC plate (Fig. 3.3.6). The lanes 4, 5 and 6 displayed only a single spot for arabinose identical to the arabinose standard shown in lane 1 (Fig. 3.3.6). Similar results were obtained with insoluble wheat arabinoxylan and beechwood xylan (data not shown). This further ascertained the fact that *Ct*43Ara*f* is predominantly α-L-arabinofuranosidase as the spots corresponded to the standard L-arabinose. These results were in agreement with the previous reports on α-L-arabinofuranosidase (Kaji, 1984; Kaneko *et al.*, 1995; Gilead and Shoham, 1995).

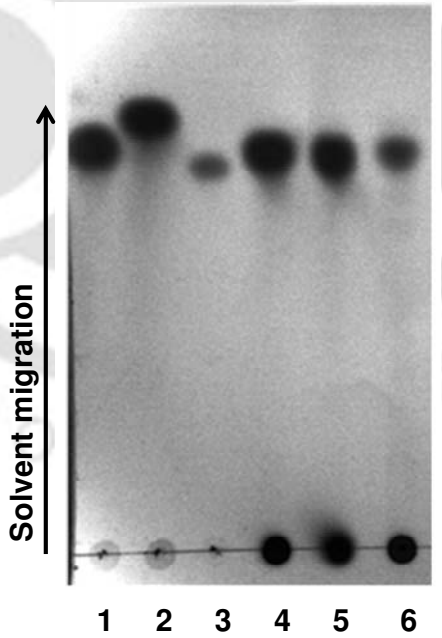


Fig. 3.3.6 TLC analysis of reaction products of *Ct*43Ara*f*. Dark spots on TLC plate show the standards L-arabinose, D-xylose and cellobiose (spots 1, 2 and 3, respectively) while spots 4, 5 and 6 represent hydrolyzed products from rye arabinoxylan, wheat arabinoxylan and oat spelt xylan, respectively, showing that only L-arabinose is released as the breakdown product.

3.3.9 High performance anion exchange chromatography (HPAEC) analysis of enzyme reaction products of *Ct*43Ara*f*

The reaction products of *Ct*43Ara*f* with rye arabinoxyalan, wheat arabinoxylan (soluble) and oat spelt xylan analyzed by high performance anion exchange chromatography (HPAEC) also showed only L-arabinose as the released sugar (Fig. 3.3.7). L-Arabinose and xylose used as standards for the HPEAC analysis of enzyme-substrate reaction products showed peaks at 10.4 min and at 13.6 min, respectively (Fig. 3.3.7A and Fig. 3.3.7B). The chromatogram of hydrolysis products of *Ct*43Ara*f* action on rye arabinoxylan, wheat arabinoxylan and oat spelt xylan shown in Fig. 3.3.7C, Fig. 3.3.7D and Fig. 3.3.7E, respectively, showed peaks at 10.4 min corresponding to L-Arabinose and showed no peak for xylose. The above results of hydrolysis products were similar to the results of previous reports on α-L-arabinofuranosidase by Xue *et al.* (2003), Tuncer and Ball (2003) and Ravanal *et al*. (2010). HPAEC analysis of hydrolyzed products by *Ct*43Ara*f* also strongly suggested that the enzyme is α-L-arabinofuranosidase.

The results of *Ct*43Ara*f* and *Ct*GH43 action on natural substrates and *p*NP-glycosides, TLC and HPAEC analysis of hydrolyzed products by *Ct*43Ara*f*, strongly suggested that these enzymes have α-L-arabinofuranosidase activity. The full length *Ct*43Ara*f* contains only one catalytic module (*Ct*GH43), so, it can be concluded that both the catalytic enzymes *Ct*43Ara*f* and *Ct*GH43 have α-L-arabinofuranosidase activity and they released α-L-arabinose (at side chains) from arabinoxylans in an exo-acting fashion as these are attached at the non-reducing end.

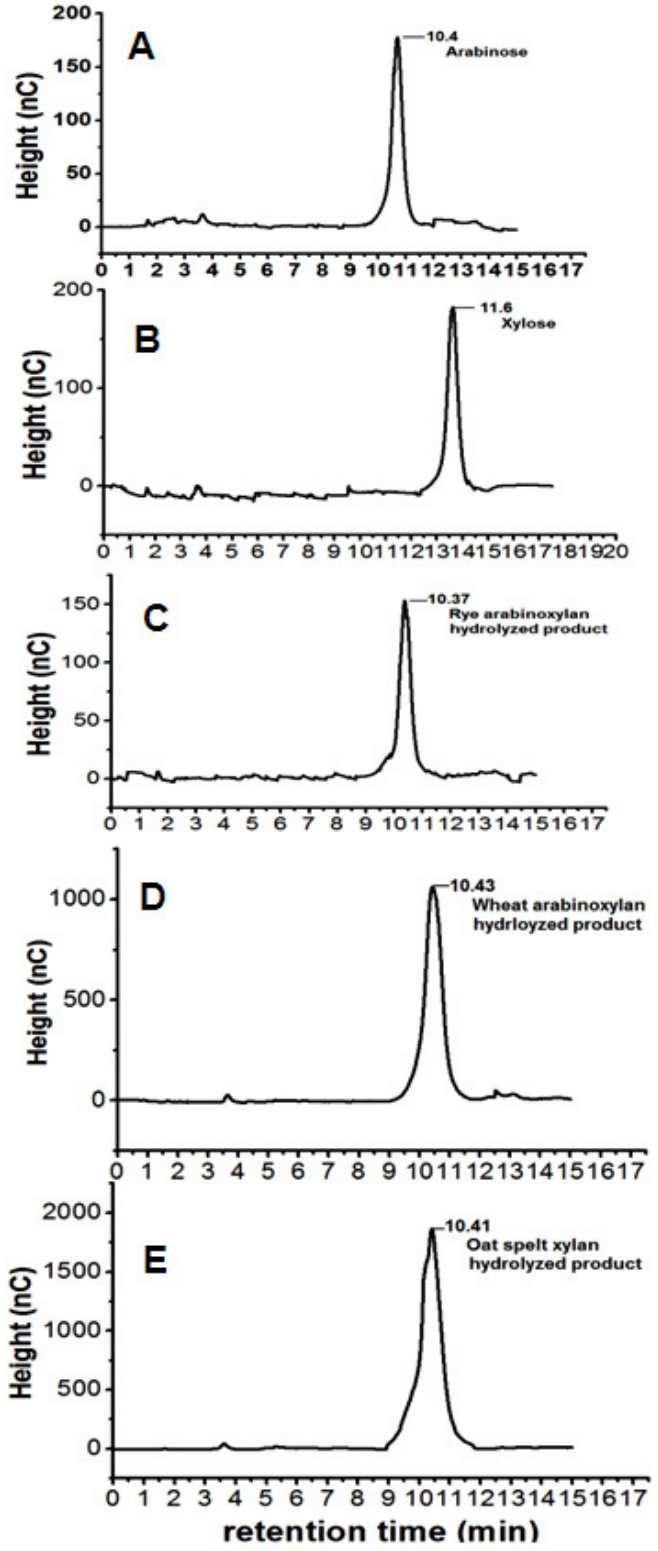


Fig. 3.3.7 HPAEC analysis of *Ct*43Ara*f* reaction mixture showing released sugars. **A**) arabinose, **B**) xylose, **C**) rye arabinoxylan, **D**) wheat arabinoxylan and **E**) oat spelt xylan. The reaction was carried out at pH 5.7, 50°C for 30 min.

3.3.10 Protein melting analysis of *Ct*43Ara*f* and *Ct*GH43

The full length module *Ct*43Ara*f* showed two separate melting peaks at 53°C and 78°C (Fig. 3.3.8A), whereas, *Ct*GH43 displayed a single melting peak at around 78°C (Fig. 3.3.8B). This suggested that the peak at 53°C is associated with CBMs (CBM6A-CBM6B) and the peak at 78°C was due to the catalytic module *Ct*GH43. This type of melting curve indicated that the catalytic module *Ct*GH43 and non-catalytic CBMs (CBM6A-CBM6B) are melting independently (Fig. 3.3.8A). The presence of Ca^{2+} ions (10 mM) caused significant changes in the protein-melting profiles of *Ct*43Ara*f* as well as of *Ct*GH43. The peaks in both the cases (shown as dotted lines) were shifted towards higher temperature i.e. to 55°C for CBM6A-CBM6B of *Ct*43Ara*f* (Fig. 3.3.8A) and to 83°C for catalytic modules *Ct*43Ara*f* and *Ct*GH43 (Fig. 3.3.8A and Fig. 3.3.8B, respectively). This shifting of melting-curve peaks suggested that Ca^{2+} ions have influence on melting rate and it may be doing so by imparting overall stability to both the protein structures as also previously reported (Dvortsov *et al*., 2009). When 10 mM EGTA salt was added to the solutions of *Ct*43Ara*f* and *Ct*GH43 containing Ca^{2+} (10 mM), the melting peaks of catalytic *Ct*GH43 and non-catalytic CBMs (CBM6A-CBM6B) were shifted back to the original positions of 78°C and 53°C, respectively, for *Ct*GH43 and CBMs, as evident from Fig. 3.3.8A and Fig. 3.3.8B (dash-dots).

The shifting back of melting peaks in the presence of EGTA can be explained by the fact it chelates the Ca^{2+} ions making them not available for the enzymes (Qin *et al.,* 1999). The catalytic activity of *Ct*43Ara*f* and *Ct*GH43 were also increased to 2-fold in presence of Ca^{2+} ions (Fig. 3.3.5B), whereas, in presence of EGTA, the enzyme activity of both these enzymes decreased (Fig. 3.3.5L). This fact in conjunction with the protein-melting curve analysis demonstrated that Ca^{2+} ions are

imparting stability to the protein structures and also increasing their catalytic efficiencies. Similar observations were reported with a β-1,3-glucanase and associated CBMX module (Dvortsov *et al*., 2009). Similar observations were also made for protein folding studies of lysozyme by Dobson *et al*. (1994) and Finkelstein and Galzitskaya, (2004).

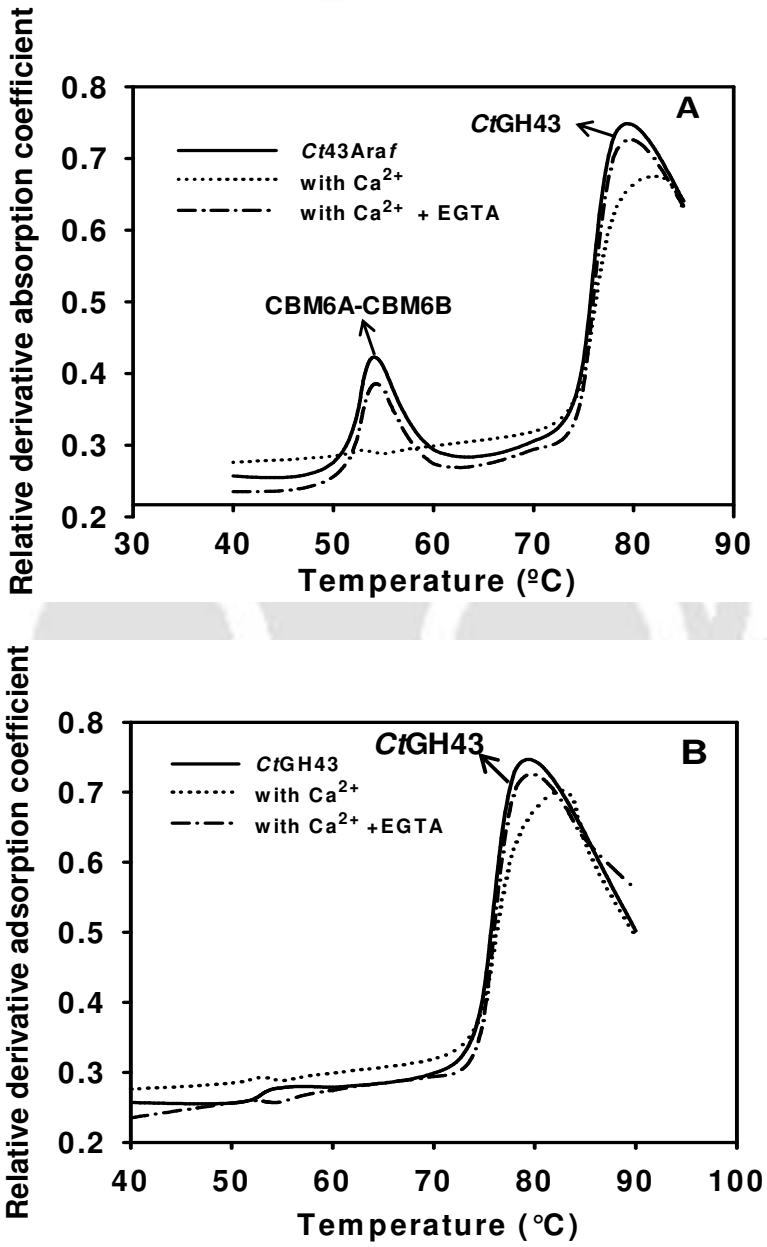


Fig. 3.3.8 Protein-melting analysis displaying normal melting curve (—), melting curve in presence of 10 mM Ca^{2+} ions (**---**), and melting curve in presence of 10 mM Ca^{2+} ions and 10 mM EGTA (**–•–**), **A**) Melting-profile of *Ct*43Ara*f* and **B)** melting profile of truncated derivative, *Ct*GH43.

3.4 Conclusions

The catalytic modules *Ct*43Ara*f* and *Ct*GH43 showed maximum enzyme activities at optimum pH 5.7 and 5.4, respectively. The optimal temperature for both was 50ºC. *Ct*43Ara*f* and *Ct*GH43 displayed more or less similar pH and thermal stability profile. *Ct*43Ara*f* was stable over a pH range of 5-6.8 while *Ct*GH43 was stable in the range 5-6.3. Both the enzymes displayed significant loss of activity when enzyme substrate reaction was carried out at 60°C or higher temperature. *Ct*43Ara*f* and *Ct*GH43 showed maximum activity 4.7 U/mg and 5.0 U/mg, respectively, with rye arabinoxylan. The catalytic enzymes *Ct*43Ara*f* and *Ct*GH43 also displayed noticeable activities with soluble wheat arabinoxylan (2.5U/mg and 2.7 U/mg), oat spelt xylan (1.7 U/mg and 1.8 U/mg) and insoluble wheat arabinoxylan (1.2 U/mg and 1.1 U/mg). They also displayed considerable activities with beechwood xylan (1.0 U/mg and 0.9 U/mg) and birchwood xylan (0.7 U/mg and 0.8 U/mg). Both *Ct*43Ara*f* and *Ct*GH43 primarily displayed activity against arabinoxylans with the ability to act on both O-2 and O-3 substituted xyloses to release L-arabinose, implying that they possess both α-1,2- and α-1,3- type of α-arabinofuranosidase activities. The catalytic modules were able to act and degrade synthetic substrates pNP-α-L-arabinofuranoside as well as *p*NP-*α*-L-arabinopyranoside. Both catalytic modules displayed high catalytic efficiencies against *p*NPA*f* and *p*NPA*p,* elucidating their bi-functional nature. But close inspection of catalytic efficiency data revealed that *Ct*43Ara*f* is primarily *α*-L-arabinofuranosidase as the k_{cat} /K_M was 3-fold higher in case of *p*NPA*f* as compared to *p*NPA*p*.

The enzyme activity of *Ct*43Ara*f* and *Ct*GH43 increased significantly by more than two-fold in presence of Ca^{2+} and Mg^{2+} salts, implying that these ions may be needed as cofactors. Marginal increase in the enzyme activities of both was also

observed with Ni^{2+}, Mn^{2+} and Zn^{2+} ions at lower concentrations. However, the enzyme activity was unaffected by lower concentration of Cu^{2+}, Co^{2+}, Hg^{2} and Ag^{+} ions but it reduced at higher concentrations. The enzyme activity of *Ct*43Ara*f* and *Ct*GH43, when assayed in the presence of Ca^{2+} ions, increased by two-fold, but in presence of EGTA it sharply decreased. This implied that Ca^{2+} ions might be involved in the catalysis and imparting stability to the structures of *Ct*43Ara*f* and *Ct*GH43.

The TLC analysis of the hydrolysis products of *Ct*43Ara*f* with rye arabinoxylan, wheat arabinoxylan and oat spelt xylan, revealed that L-arabinose was the main sugar that was released as a result of enzyme substrate reaction. HPAEC analysis of the hydrolysis products of *Ct*43Ara*f* with rye arabinoxylan, wheat arabinoxylan and oat spelt xylan also confirmed that L-arabinose was the only monosaccharide released after the hydrolysis. The TLC and HPAEC analyses confirmed that *Ct*43Ara*f* and *Ct*GH43 released only *α*-L-arabinosyl side chain sugars from arabinoxylans. The result of TLC and HPAEC combined with results obtained with synthetic pNP-glycosides confirmed that both *Ct*43Ara*f* and *Ct*GH43 are predominantly α-L-arabinofuranosidase.

Protein-melting curves of full length protein *Ct*43Ara*f* and *Ct*GH43 showed that the catalytic module (GH43) and CBMs (CBM6A-CBM6B) melt independent of each other. The protein-melting peaks of *Ct*GH43 and CBMs shifted to higher temperature in the presence of Ca^{2+} ions. However, on addition of equimolar concentration of EGTA to the solutions of *Ct*43Ara*f* and *Ct*GH43, the melting temperature peaks were shifted back to the original positions. This further corroborated the fact that Ca^{2+} ions impart thermal stability to the protein structures.

References

Ahmed, S., Bharali, S., Purama, R.K., Majumder, A., Fontes, C.M.G.A. and Goyal, A. (2009) Structural and biochemical properties of lichenase from *Clostridium thermocellum*. Indian J. Microbiol. 49, 72-76.

Aristidou, A. and Penttilä, M. (2000) Metabolic engineering applications to renewable resource utilization. Curr. Opin. Biotechnol. 11, 187-198.

Bengtsson, S., Aman, P. and Andersson, R.E. (1992) Structural studies on water-soluble arabinoxylans in rye grain using enzymatic hydrolysis. Carbohyd. Polym. 17, 277-284.

Bharali, S., Purama, R.K., Majumder, A., Fontes, C.M.G.A. and Goyal, A. (2005) Molecular cloning and biochemical properties of family 5 glycoside hydrolase of bi-functional cellulase from *Clostridium thermocellum.* Indian J. Microbiol. 45, 317-321.

Bourgois, T.M., Van Craeyveld, V., Van Campenhout, S., Courtin. C.M., Delcour, J.A., Robben J. and Volckaert, G. (2007) Recombinant expression and characterization of XynD from *Bacillus subtilis* subsp. subtilis ATCC 6051: a GH43 arabinoxylan arabinofuranohydrolase. Appl. Microbiol. Biotechnol. 75, 1309-1317.

Branden, C. and Tooze, J. (1991) In *Introduction to Protein Structure*, (2nd ed.), Garland publishing, Taylor and Francis Group, Ney York, NY.

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

Cote, G.L. and Leathers T.D. (2005) A method for surveying and classifying *Leuconostocs* pp. glucansucrases according to strain-dependent acceptor product patterns. J. Ind. Microbiol. Biotechnol. 32, 53-60.

Creighton, T.E. (1992) In *Proteins: Structures and Molecular Properties*, (2nd ed.), Freeman, W. H. & Company, Macmillan Higher Education, Ney York, NY.

Dobson, C.M., Evans, P.A. and Radford SE. (1994) Understanding how protein folds: the lysozyme story so far, Trends Bid. Sci. 19, 31-37.

Das, S.P., Ravindran, R., Ahmed, S., Das, D., Goyal, D., Fontes, C.M.G.A. and Goyal, A. (2012) Bioethanol production involving recombinant *C. thermocellum* hydrolytic hemicellulase and fermentative microbes. Appl. Biochem. Biotechnol. 167, 1475-1488.

Dvortsov, I,A., Lunina, NA, Chekanovskaya, L.A., Schwarz, W.H., Zverlov, V.V. and 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.

Finkelstein, A.V. and Galzitskaya, O.V. (2004) Physics of protein folding. Phys. Life Rev. 1, 23-56.

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

Gomori, G. (2010) Handbook of Biochemistry and Molecular Biology. In *Preparation of Buffers for Use in Enzyme Studies* (eds. Lundblad, R.L. and MacDonald, F.M.), CRC Press, Taylor and Francis Group, LLC, USA, 721–724.

Gruppen, H., Hamer, R.J. and Voragen, A.G.J. (1992) Water-unextractable cell-wall material from wheat-flour, fractionation of alkali-extracted polymers and comparison with water-extractable arabinoxylans. J. Cereal Sci. 16: 53-67.

Guais, O., Tourrasse, O., Dourdoigne, M., Parrou, J.L. and Francois, J.M. (2010) Characterization of the family GH54 α-L-arabinofuranosidases in *Penicillium funiculosum*, including a novel protein bearing a cellulose-binding domain. Appl. Microbiol. Biotechnol. 87, 1007-1021.

Hashimoto, K., Yoshida, M. and Hasumi, K. (2011) Isolation and characterization of CcAbf62A, a GH62 α-L-arabinofuranosidase, from the basidiomycete *Coprinopsis cinerea*. Biosci. Biotechnol. Biochem. 75, 342-345.

Kaneko, S., Sano, M. and Kusakabe, I. (1995) Purification and some properties of alpha-L-arabinofuranosidase from *Bacillus subtilis* 3-6. Appl. Environ. Microbiol. 60, 3425-3428.

Kaji, A. (1984) L-Arabinosidases. Adv. Carbohydr. Chem. Biochem. 42, 383-394.

Knudsen, K.E.B. and Laerke, H.N. (2010) Rye arabinoxylans, molecular structure, physicochemical properties and physiological effects in the gastrointestinal tract. Cereal Chem. 87, 353-362.

Krog-Mikkelsen, I., Ole, H., Inge, T., Holst, J.J., Andersen, J.R. and Bukhave, K. (2011) The effects of L-arabinose on intestinal sucrase activity: dose-response studies in vitro and in humans. Am. J. Clin. Nutr. 94, 472-478.

Matsuo, N., Kaneko, S., Kuno, A., Kobayashi, H. and Kusakabe, I. (2000) Purification, characterization and gene cloning of two alpha-L-arabinofuranosidases from *Streptomyces chartreusis* GS901. Biochem. J. 346, 9-15.

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

Numan, M.T. and Bhosle, M.B. (2006) α-L-Arabinofuranosidases, the potentials applications in biotechnology. J. Ind. Microbiol. Biotechnol. 33, 247-260.

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

Ordaz-Ortiz, J.J. and Saulnier. L. (2005) Structural variability of arabinoxylans from wheat flour, comparison of water-extractable and xylanase-extractable arabinoxylans. J. Cereal Sci. 42, 119-125.

Pason, P., Kosugi, A., Waeonukul, R., Tachaapaikoon, C., Ratanakhanokchai, K., Arai, T., Murata, Y., Nakajima, J. and Mori, Y. (2010) Purification and characterization of a multienzyme complex produced by *Paenibacillus curdlanolyticus* B-6. Appl. Microbiol. Biotechnol. 85, 573-580.

Qin, N., Olcese, R., Bransby, M., Lin, T. and Birnbaumer, L. (1999) Ca^{2+}-induced inhibition of the cardiac Ca^{2+} channel depends on calmodulin. Proc. Natl. Acad. Sci. (USA). 96, 2435-2438.

Ravanal, M.C., Callegari, E. and Eyzaguirre, J. (2010) Novel bifunctional alpha-L-arabinofuranosidase/xylobiohydrolase (ABF3) from *Penicillium purpurogenum*. Appl. Environ. Microbiol. 76, 5247-5253.

Saha, B.C. (2003) Hemicellulose bioconversion. J. Ind. Microbiol. Biotechnol. 30, 279-291.

Sakamoto, T., Ogura, A., Inui, M., Tokuda, S., Hosokawa, S., Ihara, H. and Kasai, N. (2011) Identification of a GH62 α-L-arabinofuranosidase specific for

arabinoxylan produced by *Penicillium chrysogenum.* Appl. Microbiol. Biotechnol. 90, 137-146.

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

Somogyi, M. 1945. A new reagent for the determination of sugars. J. Biol. Chem. 160, 61-68.

Sørensen, H.R., Jørgensen, C.T., Hansen, C.H., Jørgensen, C.I., Pedersen, S. and Meyer, A.S. (2006) A novel GH43 alpha-L-arabinofuranosidase from *Humicola insolens*, mode of action and synergy with GH51 alpha-L-arabinofuranosidases on wheat arabinoxylan. Appl. Microbiol. Biotechnol. 73, 850-861.

Tateishi, A., Mori, H., Watari, J., Nagashima, K., Yamaki, S. and Inoue, H. (2005) Isolation, characterization and cloning of α-L-arabinofuranosidase expressed during fruit ripening of Japanese pear. Plant Physiol. 138, 1653-1664.

Tuncer, M. and Ball, A.S. (2003) Purification and partial characterization of alpha-L-arabinofuranosidase produced by *Thermomonospora fusca*. Folia Microbiol. 48, 168-172.

Zhou, J., Bao, L., Chang, L., Zhou, Y. and Lu, H. (2012) Biochemical and kinetic characterization of GH43 β-d-xylosidase/α-L-arabinofuranosidase and GH30 α-L-arabinofuranosidase/β-d-xylosidase from rumen metagenome. J. Ind. Microbiol. Biotechnol. 39, 143-152.

Chapter 4

Binding analysis of *Ct*CBM6A and *Ct*CBM6B of *Ct*43Ara*f* from *Clostridium thermocellum*

4.1 Introduction

Carbohydrate-binding modules (CBMs) are non-catalytic components usually found associated with the carbohydrate-active enzymes glycoside hydrolases (GHs). Their primary function is to assist in targeting the appended catalytic modules to the ligand and concentrating them on the surface of the ligand. Similar to GHs the CBMs are also categorized into different families. There are four popular protocols for investigating and quantifying the CBM-polysaccharide binding relationship. These methods include (i) solid state depletion assay, (ii) affinity gel electrophoresis, (iii) UV difference and fluorescence spectroscopy and (iv) isothermal titration calorimetry (Abbot and Boraston, 2012). Although, affinity gel electrophoresis has been the most popular method for the last one decade (Takeo, 2000; Tomme *et al*., 2000; Takeo *et al*., 2002), now-a-days, isothermal titration calorimetry analysis is gaining importance mainly due to its high sensitivity and less time consumption. There are several CBM families that display diversity in ligand binding affinity (Pires et al., 2004; Cejzek et al., 2004; van Bueren et al., 2005; Ficko-Blean *et al.,* 2012.), which is mainly because of the flexibility in the recognition of sugar ring

configuration, type of linkage and degree of polymerization. Based on the nature of the ligands recognized by CBMs, these modules have been categorized into three functional classes: Types A, B, and C (Boraston *et al*., 2004). Type A CBMs bind to highly crystalline polysaccharides, such as cellulose or chitin; Type B CBMs recognize soluble oligosaccharide chains, while Type C CBMs bind to smaller oligosaccharides, generally recognizing the terminal sugars of these ligands (Fontes and Gilbert, 2012; Abbot and Boraston, 2012).

The family 6 CBMs have been previously demonstrated to bind with amorphous cellulose, xylans and a few modules also bind to β-1,3-glucan, β-1,3-1,4-glucan, and β-1,4-glucan (http://www.cazy.org/CBM6.html). CBM6s are found in a range of GHs displaying activity against xylan, mannan, β-glucans, agarose, and arabinans. There are more than a dozen Protein Data Bank (PDB) entries for CBM6s, including many in complex with different ligands. Several complementary biochemical studies have been performed in past demonstrating the ability of family 6 CBMs to interact with diverse carbohydrate targets (Abbot and Boraston, 2012). Because of the ligand diversity and continually expanding knowledge of sequence-based information of carbohydrate-modifying enzymes, experimental determination of CBM binding specificity and identification of novel mechanisms in carbohydrate recognition by these proteins have become time-consuming and complicated processes. For this reason Abbott *et al*. (2009) deduced a model for family 6 CBMs (CBM6) based upon several factors including phylogenetic relatedness and structural and functional properties. They demonstrated that five regions within the binding site, named A-E, play key roles in ligand selection and affinity (Fig. 4.2.1). The regions A to C are located in a primary sub site and contribute mainly to binding energy and selection for O2, O3, and O4 equatorial hydroxyls. Region D determines

the CBM will interact with internal or terminal structures of the carbohydrate ligand. Region E displays largest degree of variation and is thus predicted to make the most significant contribution to specificity.

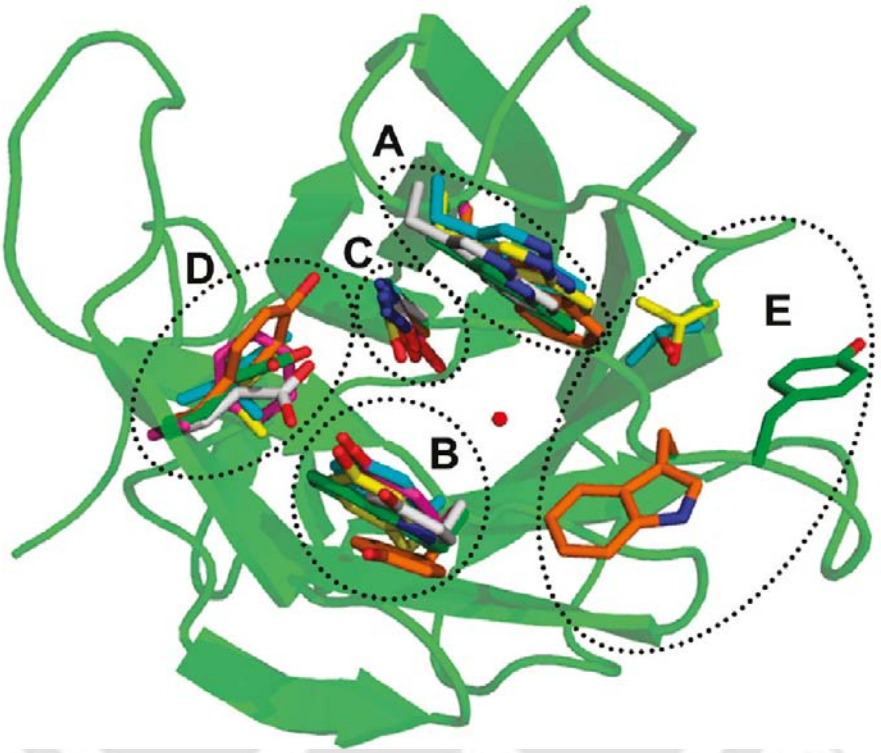


Fig. 4.1 Regions affecting ligand binding and affinity of CBM6s. (A) It shows structural alignments of known CBM6 structures. Secondary structure elements extracted from the *Bacillus halodurans Bh*CBM6 coordinates are displayed above the sequences. Regions important for ligand binding within site I (A-E loop) are shown within the black dotted lines (oval or circular in shape) and labeled below the alignment. (B) Structural superimposition of amino acids from regions A-E contributing to formation of the complex in CBM6s. The transparent ribbon diagram displayed as a backbone colored green is from *Bh*CBM6. The amino acid side chains are displayed in different colors for the sake of clarity: green for *Bh*CBM6, yellow for *Clostridium stercorarium Cs*CBM6-1, magenta for *Cs*CBM6-3, cyan for *Clostridium thermocellum Ct*CBM6, gray for *Cellvibrio mixtus Cm*CBM-2, and orange for *Saccharophagus degradans Sd*CBM6-2. The conserved water molecule found within the core of all CBM6 binding sites is shown as a red sphere (Abbott *et al*., 2009).

In the Fig. 4.1 the regions A, B and C are spatially and functionally conserved, whereas, regions D and E are responsible for apparent diversity in structure that confer functional variation within the family.

Affinity electrophoresis is a simple and effective tool for the analysis of protein-binding affinities to soluble polysaccharides (Takeo, 1985; Takeo, 2000; Tomme *et al*., 2000). This approach is particularly popular for the characterization of a carbohydrate-binding module. Knowledge of the binding characteristics of these enzymes can be helpful in elucidating their functional properties. Moreover, in some cases CBMs are able to bind polysaccharides targets other than their specified ligand and this information can be essential to understand the basic mechanism of these proteins in their natural environment (Morais *et al.*, 2012).

In the present study the ligand-binding affinity of *Ct*CBM6A and *Ct*CBM6B against soluble polysaccharides was demonstrated by carrying out the affinity gel electrophoresis. Rye arabinoxylan, wheat arabinoxylan and oat spelt xylan were used as soluble ligands. The binding affinity of *Ct*CBM6A and *Ct*CBM6B was also determined against the insoluble ligand, wheat arabinoxylan.

4.2 Materials and Methods

4.2.1 Reagents, chemicals and ligands

The ingredients 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. Glycerol, methanol, concentrated hydrochloric acid and glacial acetic acid were supplied by Merck, India. Bradford reagent was purchased from Sigma-Aldrich Pvt. Ltd. USA. Tween 20 and $CaCl_2$ were procured from Merck Limited, India. Oat spelt xylan was purchased from Sigma-Aldrich Pvt. Ltd. USA. Rye arabinoxylan and wheat arabinoxylan (insoluble) were obtained from Megazyme International, Ireland.

4.2.2 Quantitative binding analysis of *Ct*CBM6A and *Ct*CBM6B against soluble polysaccharides

4.2.2.1 Preparation and running of native- PAGE with soluble ligands

Binding of both the CBMs (*Ct*CBM6A and *Ct*CBM6B) to soluble polysaccharides was evaluated by affinity electrophoresis following the protocol of Tomme *et al.* (2000) on native-PAGE in absence and presence of varying amount of polysaccharide. The native gel was prepared using stock solutions *viz.* 30% (w/v) acrylamide, 1.0 % (w/v) polysaccharide, 1M Tris-HCl (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. The only difference in native-PAGE was the absence of SDS in resolving gel. So, 7.5% resolving (native) gel was prepared following the protocol described in Chapter 2, Section 2.2.20 without SDS. 1.0% (w/v) stock solution of rye arabinoxylan and oat spelt xylan were made by dissolving in sterile deionized water until a clear solution was obtained. Thereafter, necessary aliquot of ligand solution (from 1.0 %, w/v stock) was added to the resolving gel prior to the

polymerization to make gels containing varying concentration of ligands. The remaining components of native-PAGE, described above were added and the gel was polymerized. Bovine serum albumin (BSA) (10 µg) was loaded on the gels as an internal standard for reference. Electrophoresis was carried out for 2h at 4°C in a Mini PROTEAN Tetra pack (Bio-Rad, USA) at a constant current of 10 mA. The gels were stained with Coomassie Brilliant Blue R250 (Sigma) and destained as described in Chapter 2, Section 2.2.20.6 to detect protein. Qualitative binding of the CBMs was assessed visually on 7.5% native-PAGE. Quantitative analysis of ligand binding was assessed by measuring the migration distance of *Ct*CBM6A and *Ct*CBM6B for calculating the equilibrium association constant (K_a). Relative mobility for *Ct*CBM6A and *Ct*CBM6B was calculated as the migration distance of native protein divided by the migration distance of bromophenol blue dye front.

4.2.2.2 Preparation of native-PAGE running buffer

The native-PAGE buffer was prepared using components as described in Table 4.2.1. The native-PAGE running buffer did not contain SDS as a component unlike SDS-PAGE running buffer but the final pH was same (pH 8.3). A 5x stock of running buffer was prepared and diluted to 1x before use (Table 4.2.1).

Table 4.2.1 Composition of 5x native-PAGE running buffer.

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

4.2.2.3 Preparation of sample buffer

A 5x sample loading buffer was prepared by dissolving the components maintaining the concentration of components as described in Table 4.2.2 and pH of the buffer was adjusted to 6.8. The components were dissolved in the order as

mentioned in Table 4.2.2 to make 5x sample buffer. However, the final concentration while loading to a native-PAGE gel was always kept to 1x by mixing 4 volumes of sample (protein) with 1 volume of 5x sample buffer.

Table 4.2.2 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)
Bromophenol Blue	0.025 (%, w/v)

4.2.2.4 Calculation of equilibrium association constant

The equilibrium association constant values were calculated by methods described earlier by Takeo (1984) and Tomme *et al.* (2000). The interaction of CBM with a polysaccharide can be expressed in Eq. (1) and (2):

$$P + L \leftrightarrow PL \quad (1)$$

$$K_a = [PL]/[P][L] \quad (2)$$

where, P and L are free CBM and polysaccharide (ligand), respectively, PL is the complex and [P], [L], and [PL] are their respective concentrations. K_a is the equilibrium association constant for the interaction. The expressions for the mole fractions of the CBMs in the free (X_P) and complexed form (X_{PL}) are given as:

$$X_P = 1/Q \quad (3a)$$

$$X_{PL} = K_a [L]/[Q] \quad (3b)$$

Where [Q] is the partition function, given as

$$Q = 1 + K_a [Q] \quad (4)$$

The mobility of a CBM (*Ct*CBM6A and *Ct*CBM6B) relative to the standard can be expressed as:

$$R = X_P R_P + X_{PL} R_{PL} \quad (5)$$

where, R is the relative mobility of *Ct*CBM6A or *Ct*CBM6B at a given concentration of polysaccharide (R= mobility of the polypeptide (mm)/mobility of the reference (mm), R_P is the relative mobility of *Ct*CBM6A or *Ct*CBM6B in the absence of the polysaccharide, and R_{PL} is the relative mobility of *Ct*CBM6A or *Ct*CBM6B in the presence of polysaccharide. Substitution of Eq. (3a), (3b), and (4) into Eq. (5) gives:

$$R = (R_P + R_{PL} K_a [L])/(1 + K_a[L] \quad (6)$$

Eq. (6) can be used to regression analysis K_a and R_{PL} from affinity data plotted as the relative mobility of *Ct*CBM6A or *Ct*CBM6B vs. the polysaccharide concentration. R_P is set as a constant equal to the relative mobility of *Ct*CBM6A or *Ct*CBM6B in the absence of polysaccharide. Normalizing the initial relative mobility in the absence of polysaccharide to 1.0, changes the Eq. (6) to

$$R = (1 + R_M K_a [L]) / (1 + K_a[L]) \quad (7)$$

where, R_R ($R_R = R/R_P$) is the relative mobility normalized to the initial relative mobility in the absence of polysaccharide. R_M ($R_M = R_{PL}/R_P$) is the relative mobility of *Ct*CBM6A or *Ct*CBM6B fully complexed with polysaccharide normalized to the relative mobility in the absence of polysaccharide. Eq. 7 can be used for regression analysis using K_a and R_M from affinity data plotted as R_R of *Ct*CBM6A or *Ct*CBM6B against the polysaccharide concentration.

4.2.3 Binding analysis of *Ct*CBM6A and *Ct*CBM6B with insoluble ligands

Ligand binding analysis of *Ct*CBM6A and *Ct*CBM6B was carried out by taking 30.0 µg of each of above proteins in 50 mM Tris-HCl buffer (pH 7.5) containing 0.05% (v/v) Tween 20 and 5.0 mM $CaCl_2$ mixed with 1.0 mg of insoluble wheat arabinoxylan and the final reaction volume was kept to 200 µl (Carvalho *et al*., 2004). This reaction mixture was shaken on a vortex (Vibrax VXR B, IKA) continuously at 4°C for 2h. The mixture was centrifuged at 13000***g*** and at 4°C for 5 min. The supernatant containing the unbound fraction was removed and its protein content was determined by Bradford's method as described in Chapter 2, Section 2.2.22 (Bradford, 1976). The pellet comprising mainly of precipitated insoluble ligand and the protein bound to it was washed three times with 200 µl of buffer (50.0 mM Tris-HCl, pH 7.5). It was re-suspended in 200 µl of 10% (w/v) SDS containing 10% (w/v) β-mercaptoethanol and boiled for 10 min. The bound protein fractions from the pellet and unbound protein in supernatant were analyzed by SDS-PAGE on a 14.0 % acrylamide gel prepared using protocol as described in Chapter 2, Section 2.2.20. The controls containing recombinant *Ct*CBM6A and *Ct*CBM6B without insoluble ligands were also performed in parallel to ensure that no protein precipitation occurred during reaction.

4.2.3.1 Equilibrium binding kinetics of CtCBM6A and CtCBM6B with insoluble wheat arabinoxylan

When a ligand (polysaccharide) interacts with a protein, the binding follows the law of mass action and at equilibrium the rate of forward reaction is equal to rate of backward reaction. Based on this, equilibrium association constants (K_a) were obtained from the depletion isotherms (plot of [B] versus [F]) after fitting (nonlinear regression) of the raw data to a equilibrium binding type or adsorption model. The

equation 1 for adsorption reaction given below was modified from Din *et al*. (1994) as described by Gilkes *et al*. (1992) and represented as,

$$K_a = \frac{[B]}{[N][F]} \qquad (1)$$

where, N is the concentration of available binding sites on the ligand (the number of polysaccharide lattice units) [moles/g of CBM], [B] is the concentration of bound CBM (mol/g of ligand), [F] is the molar concentration of free CBM and K_a is the equilibrium association constant [M^{-1}]. Now, when a single CBM interacts with only one lattice unit of polysaccharide and there are no positive or negative co-operative effects,

$$[N] = [N_0] - [B] \qquad (2)$$

where, N_0 is the total concentration of binding sites of ligand in the absence of CBM. Substitution of Equation 2 into Equation 1 and rearrangement gives Langmuir equation,

$$[B] = \frac{[N_0]\, K_a\, [F]}{(1+K_a\, [F])} \qquad (3)$$

Equation 4 is the widely used for non-linear regression analysis of one binding site model, following the previous reports of Gilkes *et al.* (1992) and Creagh *et al.* (1996). However, if binding involves CBM occupying several lattice units of ligands, then the surface is considered as an array of overlapping potential binding sites. Under these conditions N_0 is considered as a probability function which depends on bound protein [B] and its configuration (Pilz *et al*., 1990; Shen *et al.,* 1991). Under these conditions Equation 2 can be expressed as,

$$[N] = [N_0] - a[B] \qquad (4)$$

where, *a* is the number of lattice unit occupied by a single CBM molecule. Substituting Equation 4 into Equation 1 gives,

$$[B] = \frac{[N_0]\, K_a\, [F]}{(1 + aK_a\, [F])} \qquad (5)$$

Rearrangement of Equation 5 yields Euation 6,

$$\frac{1}{[B]} = \frac{1}{K_a\, [N_0]} \times \frac{1}{[F]} + \frac{a}{[N_0]} \qquad (6)$$

The slope ($1/K_a[N_0]$) and intercept ($a/[N_0]$) of a plot of 1/[B] versus 1/[F] were calculated by fitting the straight line for low values of [B]. Now the relative equilibrium association constant, K_r (litres/g of ligand) was calculated as,

$$K_r = K_a\, [N_0] \qquad (7)$$

The equation 7 can be used for comparing affinities of different CBMs for a given ligand. GraphPad Prism 5.0 software was used for calculating binding constants, binding capacity and generating graphs using single binding site model (McLean *et al*., 2000).

4.3 Results and Discussion

4.3.1 Binding analysis of CBMs to soluble polysaccharides

The binding analysis of CBMs (*Ct*CBM6A and *Ct*CBM6B) to soluble polysaccharides was evaluated using affinity gel electrophoresis. The equilibrium binding constant (K_a) of the CBMs was determined for rye arabinoxyalan and oat spelt xylan by measuring the relative migration distance of proteins on native PAGE gels in the presence of above ligands. Both *Ct*CBM6A and *Ct*CBM6B showed lesser affinity for rye arabinoxylan (for which the catalytic modules showed maximum activity) as compared to oat spelt xylan which was clear from greater retardation of CBMs at lower concentration of oat spelt xylan during affinity electrophoresis (Fig. 4.3.1 and 4.3.12). Fig. 4.3.3 shows a plot of normalized relative mobilities of CBMs (*Ct*CBM6A and *Ct*CBM6B) with rye arabinoxylan and oat spelt xylan concentrations which, clearly demonstrate that *Ct*CBM6B has lower mobility as compared to *Ct*CBM6A with oat spelt xylan. Both CBMs showed much less affinity to rye arabinoxylan. The K_a values for *Ct*CBM6A and *Ct*CBM6B were determined from a plot of 1/r versus ligand concentration (rye arabinoxylan and oat spelt xylan). The point where regression line (made by joining 1/r values) cut x-axis (ligand concentration in %, w/v) gave $-1/K_a$ and from this K_a value was calculated as described by Tomme *et al*. (2000). The equilibrium binding constant (K_a) of both CBMs with rye arabinoxylan was 4.0 % (w/v) showing very low affinity against this ligand (Fig. 4.3.4A and Fig. 4.3.4B). However, affinity electrophoresis analysis of CBMs showed that *Ct*CBM6B had much higher affinity (19.6, % w/v) as compared to *Ct*CBM6A (2.63, % w/v) with oat spelt xylan (Fig. 4.3.4A and Fig. 4.3.4B). This implies that *Ct*CBM6B had high affinity for xylans (oat spelt xylan) as evident from the above mentioned K_a values, whereas, *Ct*CBM6A had very low affinity for xylan

as well rye arabinoxylans (Table 4.3.1). The CBMs of family 6 are known to contain modules of diverse specificity and variation in the location of ligand binding site with respect to its 3-dimensional structure (Shoseyov *et al*., 2006; Abbot *et al*., 2009). A few recombinant family 6 carbohydrate binding proteins (CBM6) have been reported in past and were found to bind with variety of ligands like cellulose (Henshaw *et al*., 2004), xylans (Cejzek *et al*., 2004; Pires *et al*., 2004) and mixed β-(1,3)(1,4)glucans (van Bueren *et al.,* 2005; Pires *et al*., 2004). In all CBM6 crystal structures characterized till now, the conserved residues Tyr-33, Trp-92 and Asn-120 have been reported to play a critical role in ligand binding (Abbot and Boraston, 2012; Pires *et al*., 2004). The affinity gel electrophoretic analysis of *Ct*CBM6A showed that it had very low affinity for both rye arabinoxylan as well as oat spelt xylan, indicating that, it is different from the xylan binding CBM6 reported by Czjzek *et al*. (2004). *Ct*CBM6B also showed lower affinity for rye arabinoxylan, but had significantly higher affinity (9-fold) to oat spelt xylan as compared to *Ct*CBM6A (Table 4.3.1). This implied that *Ct*CBM6B had higher affinity for xylans as compared to *Ct*CBM6A. Both the modules *Ct*CBM6A and *Ct*CBM6B did not show binding with other hemicellulosic ligands such as beechwood xylan, birchwood xylan, xyloglucan, arabinan, mannan and arabinogalactan. Both the CBMs did not show any binding with cellulosic ligands. The binding affinities of *Ct*CBM6A and *Ct*CBM6B with other cellulosic and hemicellulosic ligands were also analyzed and it was observed that both the CBMs did not show any affinity for arabinan, lichenan, birchwood xylan and beechwood xylan (Fig. 4.3.5).

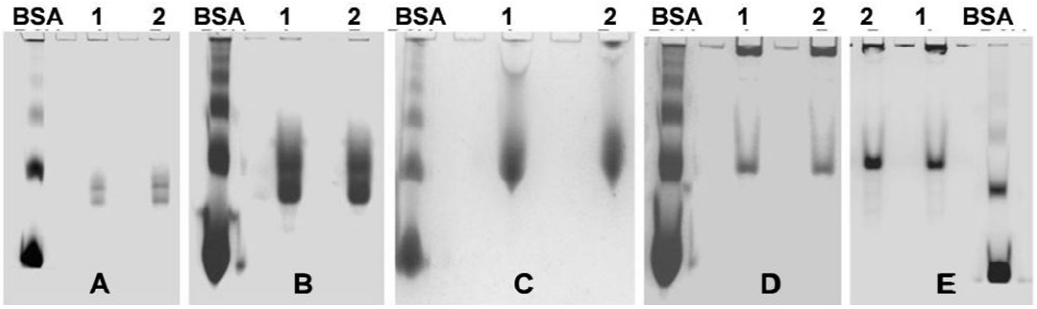


Fig. 4.3.1 Quantitative binding analysis by affinity gel electrophoresis showing binding of *Ct*CBM6A and *Ct*CBM6B to rye arabinoxylan varying in concentration from 0.00% (A), 0.02% (B), 0.04% (C), 0.1% (D) and to 0.2% (E) on 7.5% Native-PAGE gels run at 4°C to which ligands were added prior to polymerization. Lanes 1 and 2, represent *Ct*CBM6A and *Ct*CBM6B, respectively.

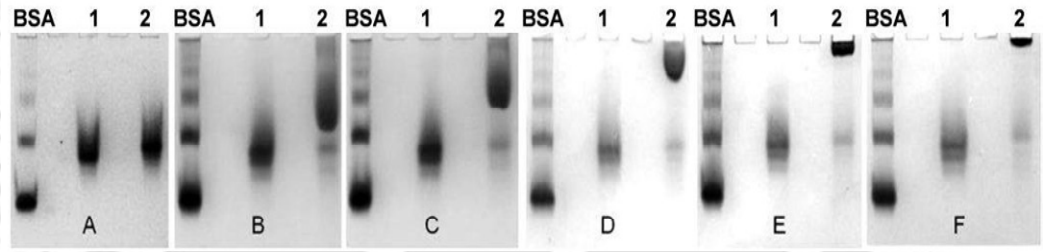


Fig. 4.3.2 Quantitative binding analysis by affinity gel electrophoresis showing binding of *Ct*CBM6A and *Ct*CBM6B to oat spelt xylan 0.0% (A), 0.02% (B), 0.04% (C), 0.08% (D), 0.1% (E) and 0.12% (F)) on 7.5% Native-PAGE gels run at 4°C to which ligands were added prior to polymerization. Lane 1 and Lane 2 represent *Ct*CBM6A and *Ct*CBM6B, respectively.

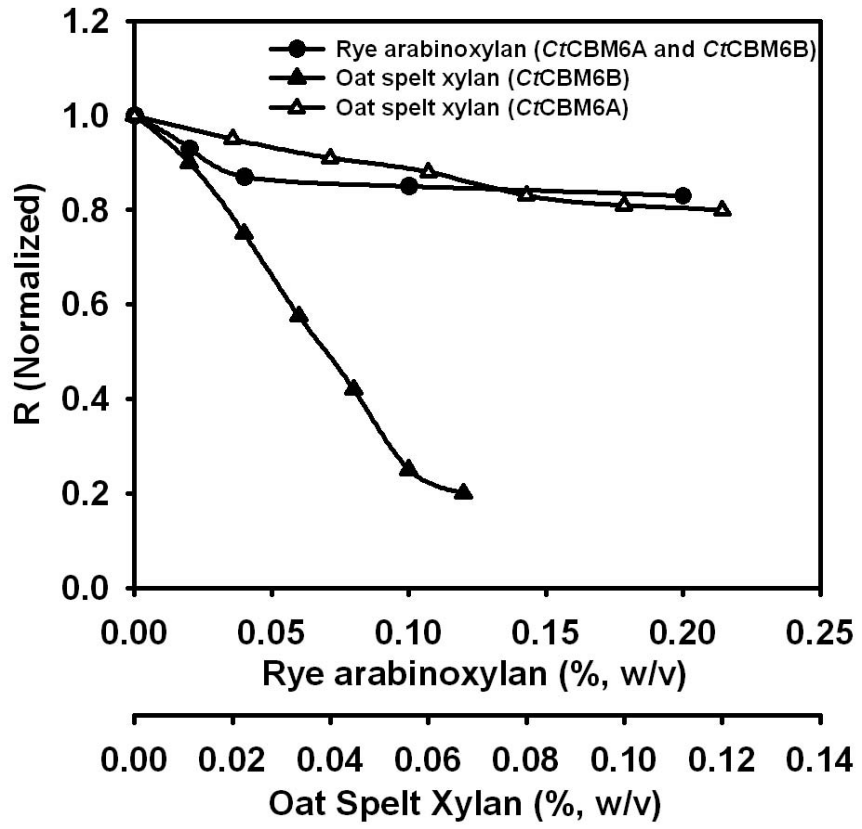


Fig. 4.3.3 Normalized relative mobility of *Ct*CBM6A and *Ct*CBM6B on native-PAGE gels containing varying concentrations of rye arabinoxylan and oat spelt xylan.

Table 4.3.1 Quantitative binding analysis of *Ct*CBM6A and *Ct*CBM6B with soluble polysaccharides *viz.* rye arabinoxylan and oat spelt xylan quantified by affinity gel electrophoresis.

Ligand	*Ct*CBM6A, K_a (%, w/v)	*Ct*CBM6B, K_a (%, w/v)
Rye arabinoxylan	4.0	4.0
Oat spelt xylan	2.63	19.6

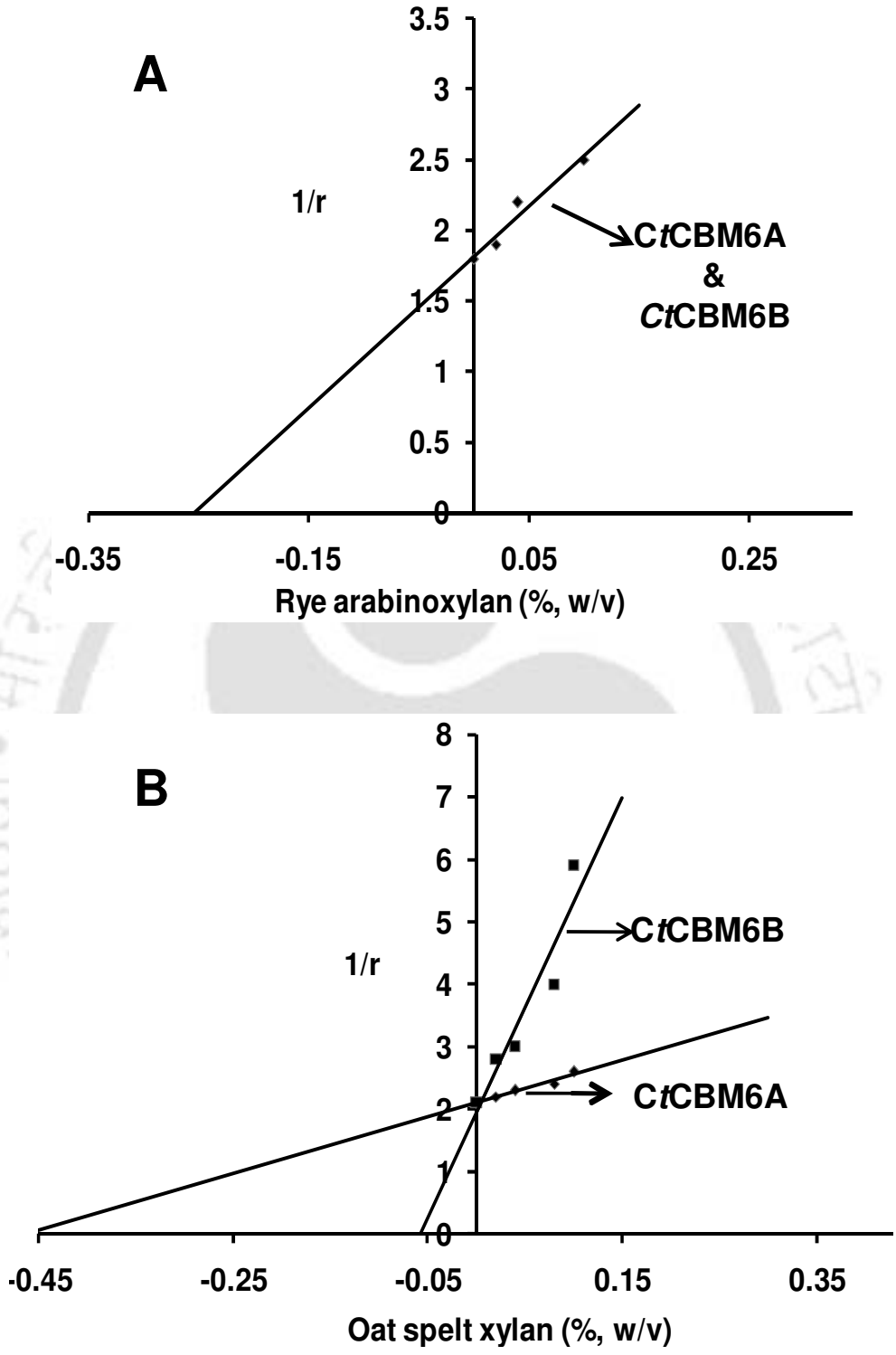


Fig. 4.3.4 A plot of 1/r versus ligand concentration for determination of K_a by the regression analysis, where r is the relative migration distance of CBMs (*Ct*CBM6A and *Ct*CBM6B) in presence of (**A**) rye arabinoxylan and (**B**) oat spelt xylan.

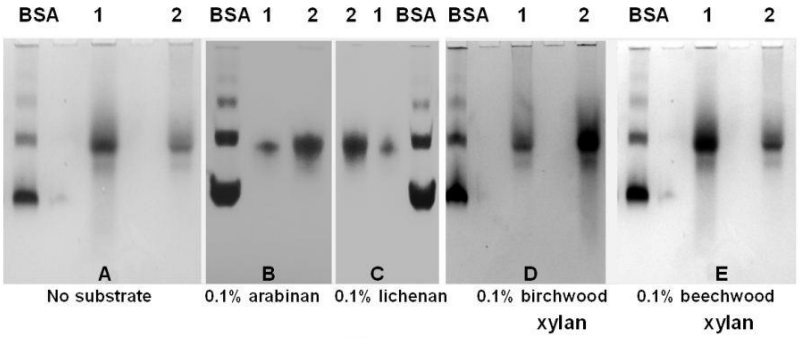


Fig. 4.3.2 Quantitative binding analysis of *Ct*CBM6A and *Ct*CBM6B by affinity gel electrophoresis without any ligand and with 1.0%, w/v of arabinan (B), lichenan (C), birchwood xylan (D) and beechwood xylan (E), on 7.5% Native-PAGE gels run at 4°C to which ligands were added prior to polymerization. Lane 1 and Lane 2 represent *Ct*CBM6A and *Ct*CBM6B, respectively.

4.3.2 Binding of CBMs to insoluble polysaccharides

The qualitative binding of *Ct*CBM6A and *Ct*CBM6A with insoluble wheat arabinoxylan as analyzed on SDS-PAGE revealed significant binding of both the CBMs (Fig. 4.3.6A, Lane 6; Fig. 4.3.6B, Lane 3). In Fig. 4.3.6A, lane 4 & lane 6 show the bound and unbound protein fractions, of *Ct*CBM6A with insoluble wheat arabinoxylan, whereas, lanes 2 and 3 in Fig. 4.3.6B show the unbound and bound fractions of *Ct*CBM6B. The initial protein fractions were also loaded on the same gel as displayed in lane 2 for *Ct*CBM6A (Fig. 4.3.6A) and lane 4 for *Ct*CBM6B (Fig. 4.3.6B). The SDS-PAGE gels clearly indicated that both the CBMs have affinity for the insoluble wheat arabinoxylan. GraphPad Prism 5.0 software was used for non-linear regression analysis using one binding-site equation (Tomme *et al*., 1998). The binding capacity (N_o) of *Ct*CBM6A and *Ct*CBM6B was found to be 11 µmole g^{-1} and 17 µmole g^{-1}, respectively (Table 4.3.2). The K_a for *Ct*CBM6B (9.4×10^3 M^{-1}) was much higher as compared to *Ct*CBM6A (2.3×10^2 M^{-1}), indicating that *Ct*CBM6B had 1.5 times higher affinity for insoluble wheat arabinoxylan (Table 4.3.1). The qualitative binding assays together with quantitative binding isotherm analysis (Fig.

4.3.6A and Fig. 4.3.6B) demonstrated that, both the CBMs were able to bind insoluble wheat arabinoxylan though with different binding capacities (Table 4.3.2). A close inspection of nonlinear one binding site model data revealed that *Ct*CBM6B had higher binding capacity as well as equilibrium association constant as compared to *Ct*CBM6A and thus displayed better binding capacity (Table 4.3.2). CBMs belonging to family 6 have been previously reported to bind with insoluble ligand like acid-swollen cellulose (Pires *et al*., 2004). This showed that CBM6 have binding abilities towards diverse nature of ligands.

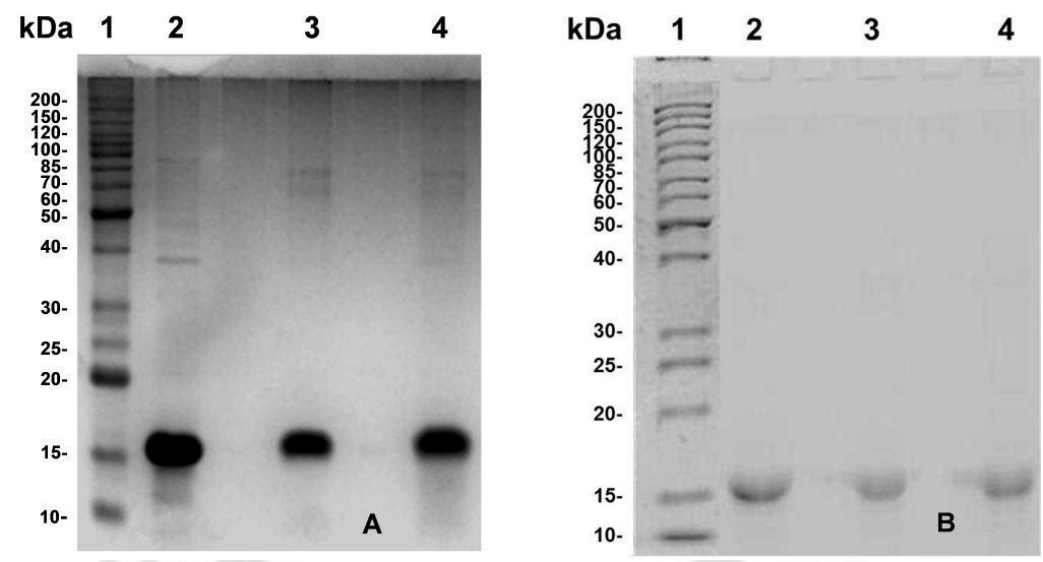


Fig. 4.3.6 Qualitative ligand (polysaccharide) binding analysis of *Ct*CBM6A and *Ct*CBM6B against insoluble wheat arabinoxylan on SDS-PAGE. **A**) Lane 1: protein marker, lane 2: purified *Ct*CBM6A, lane 3: bound *Ct*CBM6A fraction and lane 4: unbound *Ct*CBM6A fraction. **B**) Lane 1: protein marker, lane 2: unbound *Ct*CBM6B fraction, lane 3: bound *Ct*CBM6B fraction and lane 4: purified *Ct*CBM6B.

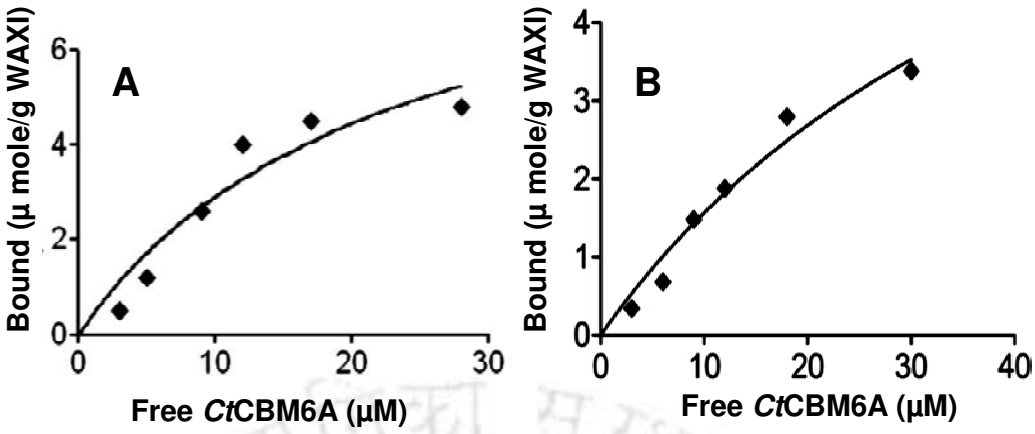


Fig. 4.3.6 Quantitative ligand (polysaccharide) binding analysis of recombinant *Ct*CBM6A and *Ct*CBM6B with insoluble polysaccharides. **A**) Showing adsorption isotherm for *Ct*CBM6Awith insoluble wheat arabinoxylan and, **B**) showing adsorption isotherm for *Ct*CBM6B binding to insoluble wheat arabinoxylan. The adsorption isotherm was generated by plotting bound (µ mole g^{-1}) versus free fractions (µM) of recombinant CBMs (*Ct*CBM6A and *Ct*CBM6B) determined by Bradford's method as described in Chapter 2, Section 2.2.22.

Table 4.3.2 Adsorption isotherm analysis of *Ct*CBM6A and *Ct*CBM6B with insoluble wheat arabinoxylan.

Carbohydrate binding module (CBM)	K_r (l/g)	N_0 (µmole g^{-1})	K_a (M^{-1})
*Ct*CBM6A	0.26±0.06	11±1.0	2.3±0.4 $\times 10^2$
*Ct*CBM6B	0.16±0.02	17± 0.7	9.4±0.25 $\times 10^3$

4.4 Conclusions

The present work illustrated the binding analysis of *Ct*CBM6A and *Ct*CBM6B of *Ct*Ara*f*43 from *Clostridium thermocellum*. The binding analysis of CBMs (*Ct*CBM6A and *Ct*CBM6B) to soluble polysaccharides was evaluated using affinity gel electrophoresis. The equilibrium binding constant (K_a) of the CBMs was determined for rye arabinoxyalan and oat spelt xylan by measuring the relative migration distance of proteins on native PAGE gels in the presence of above ligands. Both *Ct*CBM6A and *Ct*CBM6B showed lesser affinity for rye arabinoxylan (for which the catalytic modules showed maximum activity) as compared to oat spelt xylan which was clear from greater retardation of CBMs at lower concentration of oat spelt xylan during affinity electrophoresis. The normalized curve displayed that the relative mobility of CBMs (*Ct*CBM6A and *Ct*CBM6B) with rye arabinoxylan and oat spelt xylan concentrations. The curve clearly demonstrated that *Ct*CBM6B is more retarded as compared to *Ct*CBM6A with oat spelt xylan.

The equilibrium binding constant (K_a) values for *Ct*CBM6A and *Ct*CBM6B were determined from a plot of 1/r versus ligand concentrations (rye arabinoxylan and oat spelt xylan). The K_a of both CBMs with rye arabinoxylan was similar 4.0 (%, w/v) indicating very less affinity for this ligand. *Ct*CBM6B had much higher affinity 19.6 (%, w/v) as compared to *Ct*CBM6A 2.63 (%, w/v) with oat spelt xylan. This implies that *Ct*CBM6B has high affinity for oat spelt xylan. *Ct*CBM6A displayed very low affinity for both rye arabinoxylan as well as oat spelt xylan, indicating that, it is different from the xylan binding CBM6 reported by Czjzek *et al*. (2004). *Ct*CBM6B also showed lower affinity for rye arabinoxylan, but showed significantly higher affinity (9-fold) with oat spelt xylan as compared to *Ct*CBM6A. Both *Ct*CBM6A and *Ct*CBM6B did not show affinity for beechwood xylan, birchwood

xylan, xyloglucan, arabinan, mannan and arabinogalactan. Both the CBMs (*Ct*CBM6A and *Ct*CBM6B) did not show any binding with carboxy methylcellulose.

The qualitative and quantitative binding analysis of *Ct*CBM6A and *Ct*CBM6B displayed that they have significant binding affinity for insoluble wheat arabinoxylan. The binding capacity (N_o) of *Ct*CBM6B was found to be 1.5 times more than *Ct*CBM6A. The adsorption isotherm analysis demonstrated that equilibrium association constant (K_a) of *Ct*CBM6B for insoluble wheat arabinoxylan was also higher as compared to *Ct*CBM6A.

References

Abbott, D.W. and Boraston, A.B. (2012) Quantitative approaches to the analysis of carbohydrate-binding module function. Methods. Enzymol. 510, 211-231.

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

Bharali, S., Purama, R.K., Majumder, A., Fontes, C.M.G.A. and Goyal. A. (2007) Functional characterization and mutation analysis of family 11, carbohydrate-binding module (*Ct*CBM11) of cellulosomal bifunctional cellulase from *Clostridium thermocellum*. Indian J. Microbiol. 47, 109-118.

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

Bradford, M. (1976) A Rapid and Sensitive Method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 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. and Fontes, C.M.G.A. (2004) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates β-1,4 and β-1,3-1,4-mixed linked glucans at a single binding site. J. Biol. Chem. 279, 34785-34793.

Creagh, A.L., Ong, E., Jervis, E., Kilburn, D.G.and Haynes, C.A. (1996) Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. Proc. Natl. Acad. Sci. (USA) 93, 12229-12234.

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

Darbon, H., Smith, N.L., Black, G.W., Henrissat, B. and Gilbert, H.J. (2001) The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. J. Biol. Chem. 276, 48580-48587.

Din, N., Forsythe, I.J., Burtnick, L.D., Gilkes, N.R., Miller, R.C. Jr., Warren, R.A., Kilburn, D.G. (1994) The cellulose-binding domain of endoglucanase A (CenA) from *Cellulomonas fimi*: evidence for the involvement of tryptophan residues in binding. Mol Microbiol. 11, 747-55.

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

Ficko-Blean, E., Stuart, C.P., Suits, M.D., Cid, M., Tessier, M., Woods, R.J. and Boraston, A.B (2012) Carbohydrate recognition by an architecturally complex α-N-acetylglucosaminidase from *Clostridium perfringens*. PLoS One, 7(3), e33524.

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

Morais, S., Lamed, R. and Bayer, E. (2012) Affinity electrophoresis as a method for determining ligand-binding specificity for carbohydrate active enzymes for soluble polysaccharides. Himmel, M.E (ed.), In *Biomass conversion*: Method and protocol, Springer publication, 908, 119-122.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning:* a laboratory manual. 2nd ed., Vol. 1. Plainview, Cold Spring Harbor Laboratory Press, New York.

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

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

Pilz, I., Schwarz, E., Kilburn, D.G., Miller, R.C. Jr., Warren, R.A. and Gilkes, N.R. (1990) The tertiary structure of a bacterial cellulase determined by small-angle X-ray-scattering analysis. 271, 277-280.

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

Shen, H., Schmuck, M., Pilz, I., Gilkes, N.R., Kilburn, D.G., Miller, R.C. Jr. and Warren, R.A. (1991). Deletion of the linker connecting the catalytic and cellulose-binding domains of endoglucanase A (CenA) of *Cellulomonas fimi* alters its conformation and catalytic activity. J. Biol. Chem. 266, 11335-11340.

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

Tomme, P., Boraston A, Kormos JM, Warren RA, Kilburn DG. 2000. Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules. Enzyme. Microb. Technol. 27, 453-458.

Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A.L., Sturch, K., Gilkes, N.R., Haynes, C.A., Warren, R.A. and Kilburn, D.G. (1998) Characterization and affinity applications of cellulose-binding domains. J. Chromatogr. B. 715, 283–96.

Takeo, K. (1985) Affinity electrophoresis. Electrophoresis. 5, 187–95.

Takeo, K. (2000) Advances in affinity electrophoresis. J. Chromatogr. A. 698, 89-105.

van Bueren, Al., Morland, C., Gilbert, H.J. and Boraston, A.B. (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of β-1,3-linked glucans by presenting a unique ligand binding surface. J. Biol. Chem. 280, 530-537.

Chapter 5

In silico analyses of substrate-function and ligand binding sites of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B from *Clostridium thermocellum*

5.1 Introduction

At present, more than 17,000 glyocoside hydrolase (GH) sequences comprising of around 130 families have been reported in the CAZy database (http://www.cazy.org/fam/acc_GH.html). The sequence-based classifications of the catalytic modules as glycoside hydrolase (GH) families and into different clans are available on the continuously updated carbohydrate-active enzymes (CAZy) server (www.cazy.org). Glycoside hydrolase belonging to family 43 (GH43) were found to have β-xylosidase, β-1,3-xylosidase, α-L-arabinofuranosidase, arabinanase, xylanase, galactan 1,3-β-galactosidase etc activities (http://www.cazy.org/GH43.html). Hydrolysis of the glycosidic bond takes place by either overall inversion of the anomeric configuration of the substrate as shown in Chapter 1, sub-section 1.3.4.3 (Davies and Henrissat, 1995; Lairson *et al*., 2008).

The non-catalytic polysaccharide-recognizing modules of glycoside hydrolases were originally designated as CBDs (cellulose-binding domains), because the first examples of these protein domains bound crystalline cellulose as their primary ligand (Gilkes *et al.*, 1988). Subsequently, the more inclusive term CBM

(carbohydrate-binding module) evolved to reflect the diverse ligand specificity of these modules (Michel *et al.*, 2009). CBMs have been grouped into 64 families according to the data in Carbohydrate Active enZYmes (CAZy) database, based upon amino acid sequence similarity (http://www.cazy.org/Carbohydrate-Binding-Modules.html). Within several of these sequence-based families, subtle differences in structure lead to diverse ligand specificity (Tomme *et al.*, 1998; Czjzek *et al.*, 2001; Boraston *et al.*, 2002). Many CBMs have now been identified experimentally, and several hundred putative CBMs can be further identified on the basis of amino acid similarity. The elucidation of CBM structures in complex with oligosaccharide ligands helps to provide particularly valuable insights into how these proteins recognize their target ligands. Furthermore, this diversity in ligand specificity determines the exploitation of these protein modules in many biotechnological applications (Kraulis *et al.*, 1989).

The protein sequences of *Ct*GH43, CtCBM6A and *Ct*CBM6B were subjected to BLAST to search for homology. Based on the homology search results, multiple sequence analysis (MSA) was performed to identify the conserved residues in *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B. The secondary structures of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were determined to know about the presence of helices, strands and coils (Martin, 2004; Rost and Sanders, 1993). The secondary structural elements present in *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were analyzed from the secondary structure prediction. MODELLER was used for homology or comparative modeling of three dimensional (3-D) structures of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B by using closest resembling structures from PDB as templates. Now-a-days, MODELLER is the most popular program for comparative model building (Sanchez and Sali, 2000; Marti-Renom *et al.*, 2000; Sali and Blundell, 1993). Recent

versions of MODELLER automatically derive the restraints from the known related structures and their alignment with the target sequence. The 3-D structures of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were validated based on the Ramachandran plot using RAMPAGE, PROCHECK and MolProbity analysis. In the Ramachandran (RC) plot first deduced by G.N. Ramachandran, systematically the torsion angles around Cα-N binding phi (φ) and the Cα-C psi (ψ) of the constituent amino acid residues are varied to find the most stable conformations. Most of the pairs must be in favoured regions of the plot and only a few shall be in "disallowed" regions (Lovel *et al*., 2003; Ramachandran and Sasishekaran, 1968). The globular proteins are composed of amino acids which have the potential to form favorable interactions; but intrinsically unstructured proteins (IUPs) adopt no stable structure as their amino acid composition does not allow enough favorable interactions to take place (Dosztanyi *et al*., 2005). IUPred was used to find out the intrinsic disorders present in *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B.

The identification of ligand-binding sites is often the starting point for protein function annotation and structure-based studies. Many computational methods for the prediction of ligand-binding sites have been developed in recent decades. We utilized the metaPocket software package, in which the predicted sites from four methods *viz.* LIGSITE(cs), PASS, Q-SiteFinder, and SURFNET were combined together to improve the prediction success rate. The metaPocket was used for putative catalytic or ligand binding site prediction of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B. THEMATICS (http://pfweb.chem.neu.edu/thematics/submit.html) is a method which is used for predicting the catalytic site & binding sites from 3-D structure where the charge on an ionizable residue is assumed to be on a particular atom of the ionizable group, designated as the 'charge center'; for example in the case of Asp or Glu

residues, the charge is assumed to be associated to the central C atom of the carboxyl group (Ondrechen *et al*., 2001). Using the coordinates of these charge centres, clusters of ionizable residues are defined such that the distance of the 'charge centre' of a particular residue must be within a 9Å of at least one other charge centre from another residue in the cluster. These clusters are the *THEMATICS positive clusters.* Clusters with two or more members are considered predictive. The presence or absence of unstructured elements in *Ct*GH43 was assessed using IUPred program. The globular proteins are composed of amino acids which have the potential to form a large number of favorable interactions, whereas intrinsically unstructured proteins (IUPs) adopt no stable structure because their amino acid composition does not allow sufficient favorable interactions to form (Dosztanyi *et al*., 2005).

Bioinformatics investigations were carried on the protein sequences of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B. This chapter deals in details about the sequence analysis, homology modelling, active site residue identification and docking analysis of recombinant proteins namely *Ct*GH43, CtCBM6A and *Ct*CBM6B. The homology and conserved sequences of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were identified and based on this templates were selected for homology modelling. The secondary structure prediction revealed the number of alpha-helices, beta-helices and coils present in *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B. 3-D structures of *Ct*GH43, CtCBM6A and *Ct*CBM6B were generated by homology modelling and the respective structures were validated. The putative ligand binding sites of *Ct*GH43, CtCBM6A and *Ct*CBM6B and residues were determined using Molegro virtual docker and ZDOCK from Discovery studio 2.5. Various ligands were docked with *Ct*GH43, CtCBM6A and *Ct*CBM6B were performed. Based on the docking score, the recombinant proteins having maximum tendency to bind with particular substrate or ligand was identified.

Circular Dichroism (CD) is a technique which measures the differential absorption of left and right circularly polarized radiation by chromophores which either possess intrinsic chirality or are placed in chiral environments. The chromophores are the chemical groups capable of selective light absorption resulting in the coloration of certain organic compounds. Proteins possess a number of chromophores which can give rise to CD signals. The peptide bond formation corresponds to far UV region (240-180 nm). The CD spectrum in this region can be analyzed to give the content of regular secondary structural features such as α-helix, β-sheet and coils (Kelly *et al*., 2005). The α-helix displays a strong and characteristic CD spectrum in the far UV region (Creighton, 1997). The secondary structure content from CD data assumes that spectrum is a linear combination of CD spectra of each contributing secondary structure type (α-helix and β-sheets) weighted by its abundance in the polypeptide conformation (Andrade *et al.,* 1993; Creighton, 1997). The CD spectrum in the near UV region (320-260 nm) corresponds to aromatic amino acid side chains (Creighton, 1997). Because of its relatively modest resource demands, CD has been used extensively to give useful information about protein structure, the extent and rate of structural changes and ligand binding (Kelly *et al*., 2005). In the protein design field, CD is used to assess the structure and stability of the designed protein fragments. The purified proteins, *Ct*GH43 and *Ct*CBM6B were subjected to CD analysis to determine the percentage of α-helix and β-sheets and coils.

5.2 Materials and Methods

5.2.1 Multiple sequence analysis of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B

Basic Local Alignment Search Tool (BLAST) is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences (Altschul *et al.*, 1990). BLAST compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. The protein sequences of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were fed to BLAST server (http://blast.ncbi.nlm.nih.gov/) and sequences showing similarity were noted. Multiple sequence alignment (MSA) is often used to assess sequence conservation of protein domains, tertiary and secondary structures, and even individual amino acids or nucleotides (Thompson *et al.*, 1994; Nuin *et al.*, 2006). MSA (ttp://www.ch.embnet.org/software/ClustalW.html) helps in the identification of the conserved domains among the closely related sequences, which may also be at active sites and involved in binding. Therefore, MSA of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were performed to get information about the conserved sequences. The MSA of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B was carried out using ClustalW (http://www.ch.embnet.org/software/ClustalW.html; Chenna *et al.*, 2003). The distances between the sequences were calculated using point accepted mutation (PAM) matrix (Deyhoff *et al.*, 1978).

5.2.2 Phylogenetic analysis of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B

Phylogenetic analysis was performed for finding evolutionary relations of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B. The phylogenetic analysis of above proteins

revealed information about the common ancestors' they have evolved from. It was based on blastp from Max Planck Institute for developmental biology (MPI) toolkit server (http://toolkit.tuebingen.mpg.de/), Germany. The input set of query sequences were assumed to have an evolutionary relationship by which they share a lineage and are descended from a common ancestor. Based on MSA results of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B, the sequence homology can be inferred and phylogenetic analysis can be conducted to assess the sequences' shared evolutionary origins. This can be further expressed in rooted (known common ancestor) or unrooted (unknown common ancestor) tree form showing their evolutionary relation. Different type methods for generating phylogenetic tree exists *viz.* i) distance methods like UPGMA (Sokal and Michener, 1958) and Neighbour joining and ii) character state methods like parsimony and maximum likelihood. The distances between the sequences were calculated using point accepted mutation (PAM) and blocks substitution matrix (BLOSUM) matrices.

5.2.3 Secondary structure prediction of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B

The secondary structure prediction is a set of techniques in bioinformatics that aim to predict the secondary structures of proteins and nucleic acid sequences based only on knowledge of their primary structure. For proteins, this means predicting the formation of protein structures such as alpha helices and beta strands. PSI-PRED VIEW (McGuffin et al., 1999; McGuffin and Jones, 2003) was used for the secondary structure prediction of various turns, helices and coils that may be present in *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B. PSIPRED is a highly accurate secondary structure prediction method employing two-feed forward neural networks which perform analysis on output obtained from PSI-BLAST. The neural networks

employed in PSI-PRED are parallel, distributed information processing structures. A training phase is used to teach the network to recognize the relation between secondary structure and amino acid sequences based on the proteins whose structure is already known (McGuffin *et al*., 1999). The secondary structural analysis was carried out by programs available at Phyre servers (http://www.sbg.bio.ic.ac.uk/phyre).

5.2.4 Homology modeling and structure validation of 3-dimensional models of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B

For model building, the *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B sequences were fed to homology detection & structure prediction by HMM-HMM comparison, HHpred (Soding, 2005) for finding the homologs or suitable templates. HHpred is the first server that is based on the pair wise comparison of hidden Markov model (HMM) for homology (Krogh *et al*., 1994). Multiple sequence alignments (MSA) of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B gave sequences suitable to be used as templates for 3-dimensional (3-D) model building. Templates with significant expectation value (E-value $\leq$ 0.001) were used for model building of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B, respectively. The 3-dimensional structure was modelled by using a versatile program called MODELLER (Marti-Renom *et al*., 2000; Eswar *et al*., 2010). It first finds one or more template (s) and generates an alignment between template and the target sequence. Molecular dynamics (MD) force field are used also integrating Newton's laws of motion but some restraints were imposed additionally named as probability density function (PDF) of Cα distances calculated from templates (Sanchez and Sali, 2000; Marti-Renom *et al*., 2000; Sali and Blundell, 1993; Fiser *et al*., 2000). If more than one template is used then highly conserved regions have greater restraints compare to those which vary more. The predicted

model was refined by energy minimization (Sali and Blundell, 1993; Fiser *et al*., 2000). The most fitting model based on energy minimization using SWISS-MODEL Swiss-Pdb Viewer (earlier known as Deep View) (http://www.expasy.org/spdbv/) having lowest value of MODELLER objective function was selected and visualized using a softwares, PyMOL (http://www.pymol.org) and RasMol (http://www.biomed.curtin.edu.au/biochem/ help/download. html). The predicted 3-dimensional models of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B were validated by Ramachandran plot (RC plot) analysis using RAMPAGE (Lovell *et al.,* 2002), PROCHECK (Laskowski *et al*., 1993) and MolProbity (Davies *et al*., 2010) and also by Verify3D (Luthy *et al*., 1992). The RC plot with the help of RAMPAGE server (www.cryst.bioc.cam.ac.uk/rampage) and RC plot server (http://dicsoft1.physics.iisc.ernet.in/rp). RC plot is basically a conformation chart which tells us the values of φ and ψ that are sterically possible and so gives permissible areas and forbidden areas. In the plot the core or allowed regions are the permissible areas for φ and ψ angle pairs for residues in a protein (Ramachandran *et al.,* 1963, Lovel *et al*., 2003; Ramachandran and Sasishekaran, 1968). It showed various residues falling under most favoured, favoured and in disallowed regions based on residues falling in permissible and forbidden areas of plot as per their φ and ψ torsion angle values (Ramachandran *et al*., 1963; Lovel *et al*., 2003; Ramachandran and Sasishekaran, 1968).

5.2.4.1 Protein structure validation

The stereochemical validation of model structures of proteins is an important part of the comparative molecular modeling process. Validation tools help to assess the compatibility of an amino acid sequence with a proposed three-dimensional structure. Several widely-used structure validation tools, including RAMPAGE

(Lovell *et al.,* 2002), PROCHECK (Laskowski *et al*., 1993), MolProbity (Davies *et al*., 2010) and Verify3D (Luthy *et al*., 1992) were employed for 3-dimensional structure validation.

5.2.5 Catalytic and ligand binding site prediction of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B

Q-SiteFinder and theoretical microscopic titration curves (THEMATICS) were used for catalytic site and ligand binding site prediction of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B. These tools help to compute and predict the cavities in the protein structure. The identification of catalytic residues is possible and this is a key step in understanding the function of enzymes. Identifying the region of the protein most likely to be involved in function is useful in order to gain information about its potential role. The metaPocket is a meta-server used to identify pockets or cavities on protein surface to predict ligand-binding sites (Bingding, 2009). metaPocket uses the predicted sites from four methods: LIGSITE PASS (Hendlich *et al*., 1997), Q-SiteFinder (Larie and Jackson, 2005), and SURFNET (Glaser *et al.,* 2006) are combined together to improve the prediction success rate (Bingding, 2009). In Q-SiteFinder, the protein surface was coated with a layer of methyl (-CH3) probes to calculate van der Waals interaction energies between the protein and probes. Probes with favourable interaction energies were retained and further, clusters of these probes were ranked based on the number of probes in a cluster. The largest or energetically most favourable cluster is then ranked first and considered as a potential ligand-binding site. THEMATICS method was based on the calculation of the theoretical microscopic titration curves (Ondrechen *et al.*, 2001). The experimental determination of the macroscopic pH dependence of kinetic data was used to get the pKa of candidate residues involved in the catalyzed reaction. Theoretical calculations of the

type used in this work have been extensively used to predict the pKa values of residues and to explain experimentally observed perturbations in pKa. In a few cases, a small fraction of the theoretical titration curves have been reported to have perturbed shapes (Ondrechen *et al.,* 2001,). It is now established that these shapes can be used as a diagnostic tool to determine the location of the active site.

5.2.5.1 Analysis of active site residues of CtGH43

Analysis of catalytic active site followed by the molecular docking was done by using Molecular Virtual Docker (MVD), which includes the MolDock algorithm (Thomsen and Christensen, 2006). The docking scoring function of MolDock is an extension of the piecewise linear potential (PLP) including new hydrogen bonding and electrostatic terms. To further improve docking accuracy, a re-ranking scoring function is introduced, which identifies the most promising docking solution from the solutions obtained by the docking algorithm. MolDock efficiently identified correct binding mode of 87% of the complexes. In comparison, the accuracy of other protein-ligand docking software packages like Glide and Surflex are 82% and 75%, respectively (Thomsen and Christensen, 2006). Similarly, the docking software FlexX gave 58% and GOLD 78% accuracy on subsets containing 76 and 55 cases, respectively. Out of many binding sites predicted, active site was selected on the basis of maximum surface area and volume. Important residues involved in active sites were identified and compared for the modelled proteins. This algorithm is based on a new heuristic search algorithm that combines differential evolution with a cavity prediction algorithm. The general basis for identification of active site residues was calculated based on the distances between ligand binding sites and all the protein residues and if the distance is within a threshold value (5.0 Å was used), the residue is judged to be a functional residue.

5.2.6 Hydrogen bonding plot of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B

Hydrogen bonding (HB) plot uses the hydrogen bonds calculated by the HBPLUS program (MacDonald and Thornton, 1994) using the definition of hydrogen bonding parameters of Baker and Hubbard (1984). In representations of the HB plot characteristic patterns of secondary structure elements can be seen, similarly as observed in the case of distance plot and contact maps (Bikadi *et al.,* 2007). Helices can be identified as strips directly adjacent to the diagonal, antiparallel beta strands by strips perpendicular to the diagonal, and parallel beta strands by off-diagonal strips parallel to the diagonal. Hydrogen bonding network reveal the following features:

5.2.6.1 The α helix

The hydrogen bonds occurring between the backbone carbonyl (C=O) oxygen (acceptor) of one residue '*i*' and the amide hydrogen (donor) of residue four ahead in the polypeptide chain are designated as *i*+4 (Fig. 5.2.1). The 3_{10} designation refers to numbers of backbone atoms located between the donor and acceptor atoms (10) and the fact that there are three residues per turn (Fig. 5.2.2). The 3_{10} helices show hydrogen bonds formed between residues *i* and *i*+3. The 3_{10} helices are often seen in proteins when regular α-helix is distorted by the presence of unfavourable residues, near a turn region or when short sequences fold into helical conformations. Another type of hydrogen bond formed between carbonyl and NH groups separated by five residues is designated as *i+5* and this structure is a π helix (Fig. 5.2.3). These are rare form of secondary structure as Φ/Ψ angles of π helix, lie at the edge of the allowed minimum energy region of Ramachandran plot. HB plot being symmetrical can differentiate between these types of helices (MacDonald and Thornton, 1994; Bikadi *et al*., 2007), since the distance between the diagonals depends on the type of helices as shown in Figs. 5.2.1- 5.2.3.

5.2.6.2 Parallel β-sheet

In a parallel arrangement, all of the N-termini of successive strands are oriented in the same direction. Parallel β sheet appears in the HB plot as parallel to the diagonal (Fig. 5.2.4).

5.2.6.3 Antiparallel β-sheet

In antiparallel arrangement, the successive beta strands alternate directions so that the N-terminus of one strand is adjacent to the C-terminus of the next strand. Antiparallel arrangement appears in HB plot as cross-diagonal (Fig. 5.2.4) (MacDonald and Thornton, 1994; Bikadi *et al*., 2007).

5.2.6.4 Helix capping

The ends of helices have N–H and C=O groups that are necessarily lacking backbone hydrogen bond partners, but nearby side chains from within the helix could reach these backbone polar atoms and cap them off. These motifs can be seen in HB plots by narrowed or capped motifs at the end of the helices (MacDonald and Thornton, 1994; Bikadi *et al*., 2007).

5.2.6.5 Unfolded region/loops

Unfolded regions can be recognized by the irregularities or even the lack of the hydrogen bonds. Loops appear as breaks in the diagonal between the cross-diagonal beta-sheet motifs (Utas *et al*., 2006; Bikadi *et al*., 2007).

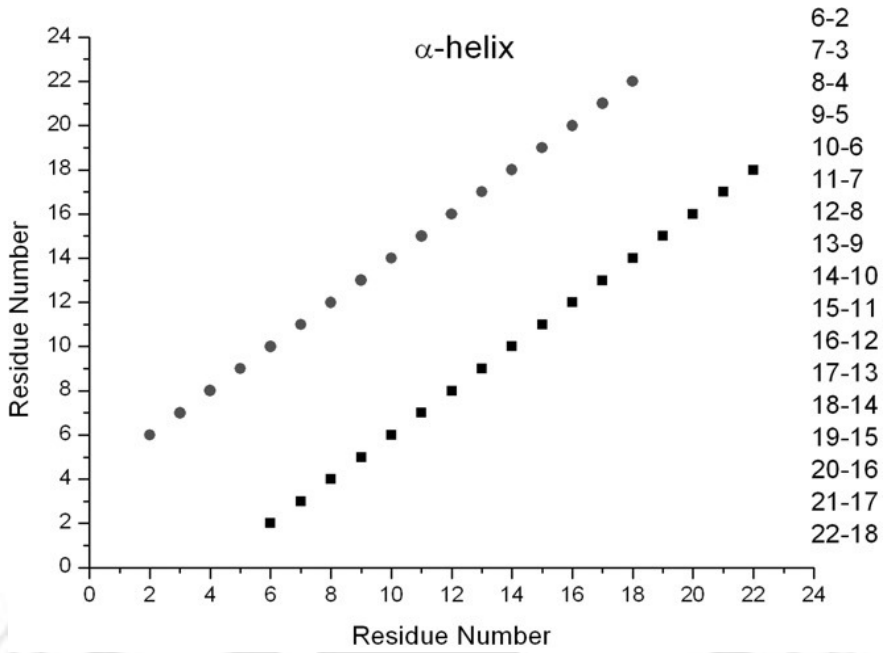


Fig. 5.2.1 HB plot representation of alpha helix (*i*+4).

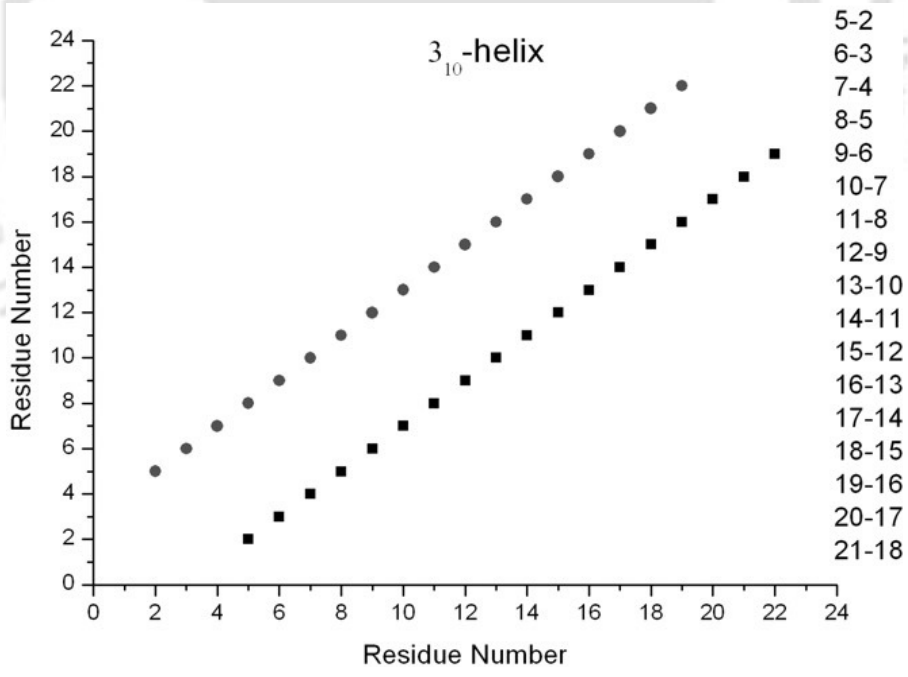


Fig. 5.2.2 HB plot representation of 3_{10} helix (*i*+3).

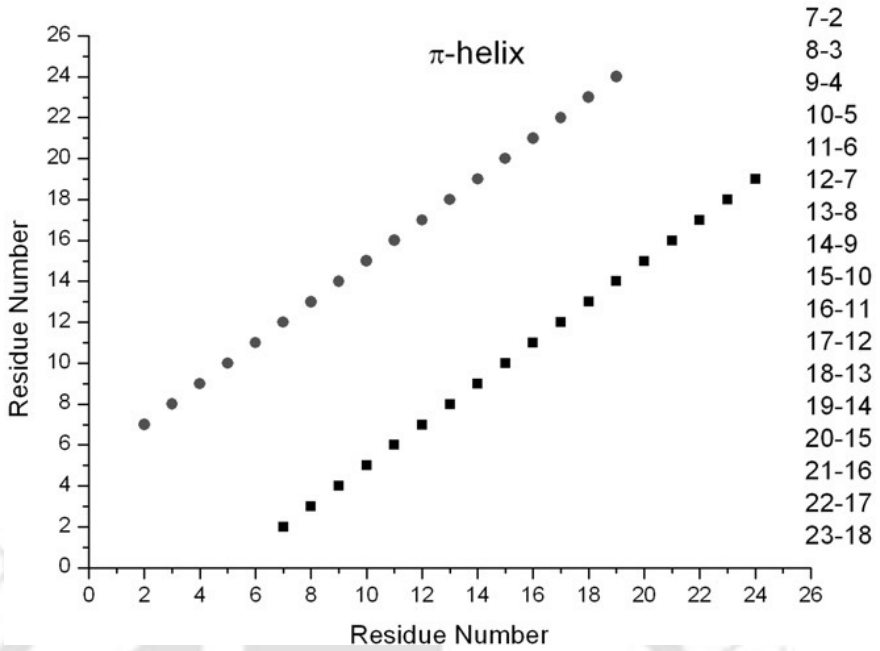


Fig. 5.2.3 HB plot representation of pi helix (*i*+5).

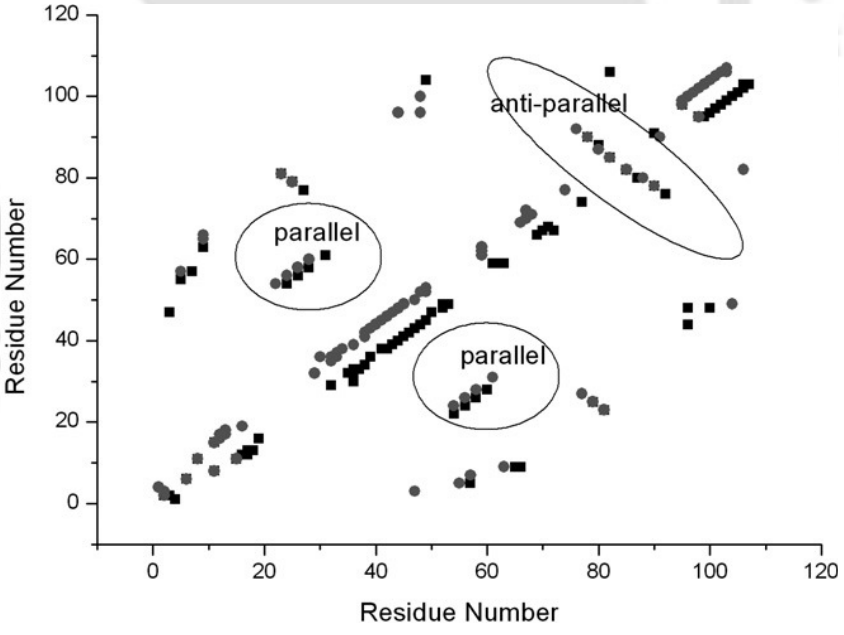


Fig. 5.2.4 HB Plot representation of a protein containing parallel and antiparallel beta sheets. Parallel beta sheets run parallel to diagonal, whereas antiparallel beta sheets appear as cross-diagonals.

5.2.7 Molecular Docking of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Comparative docking study helps to know the common residues involved in both the structures and the ligands involved in this. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using for example scoring functions. The docking studies were performed by making use of Molecular Virtual Docker or MolDock (Thomsen and Christensen, 2006) for *Ct*GH43 and ZDOCK from Accelrys Discovery Studio 3.5 (Pierce *et al*., 2011) for *Ct*CBM6A and *Ct*CBM6B. Both MolDock and ZDOCK are well-known software packages for simulating small molecule and macromolecule systems. Discovery Studio makes it easier to examine the properties of large and small molecules, study systems, identify leads and optimize candidates.

5.2.7.1 Docking analysis of CtGH43 using Molecular Virtual Docker

The docking strategy is based on semi-empirical free energy force field to evaluate conformations during docking simulations. The force field was parameterized using a large number of protein-inhibitor complexes for which both structure and inhibition constants (K_i) are known. The force field evaluates the binding in two steps. The ligand and protein start in an unbound conformation. In the first step, the intermolecular energetic was estimated for the transition from these unbound states to the conformation of the ligand and protein in the bound state. The second step then evaluated the intermolecular energetic of combining the ligand and protein in their bound conformation. The force field includes six pair-wise evaluations (V) and an estimate of the conformational entropy lost upon binding (ΔSconf):

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf})$$

where, L refers to the "ligand" and P refers to the "protein" in a ligand-protein docking calculation. Each of the pair-wise energetic terms included evaluations for dispersion/repulsion, hydrogen bonding, electrostatics and desolvation,

$$V = W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}} + W_{sol}\sum_{i,j}\left(S_i V_j + S_j V_i\right)e^{(-r_{ij}^2/2\sigma^2)}$$

The weighting constants W, have been optimized to calibrate the empirical free energy based on a set of experimentally-determined binding constants. The first term is a typical 6/12 potential for dispersion/repulsion interactions. The parameters are based on the Amber force field for the simulation of biomolecules *viz.* protein (Case *et al*., 2005). The second term is a directional H-bond term based on a 10/12 potential. The parameters C and D are assigned to give maximum well depth of 5 kcal/mol, at 1.9Å for hydrogen bonds with oxygen and nitrogen, and a well depth of 1.0 kcal/mol at 2.5Å for hydrogen bonds with sulfur. The function E (*t*) provides directionality based on the angle *t* from ideal H-bonding geometry. The third term is a screened Coulomb potential for electrostatics. The final term is a desolvation potential based on the volume of atoms (V) that surround a given atom and shelter it from solvent, weighted by a salvation parameter (S) and exponential term with distance-weighting factor σ=3.5Å. MolDock (Thomsen and Christensen, 2006) uses a standard set of parameters and weights for the force field.

5.2.7.2 Docking analysis of CtCBM6A and CtCBM6B using Accelrys Discovery Studio 3.5

Accelrys Discovery Studio 3.5 uses ZDOCK software to perform docking and to predict protein binding partners based on initial-stage rigid-body docking algorithm described by Pierce *et al.* (2011). ZDOCK (version 3.0.2) utilized a novel pair wise statistical potential and a recently optimized fast Fourier transform (FFT) to search efficiently the best position of adjacent subunits and then generated the multimers of *Ct*CBM6A and *Ct*CBM6B. It also used a new protocol called Dock Fragments, to place fragments in a receptor active site of *Ct*CBM6A and *Ct*CBM6B**.** This protocol used the well validated and published Multiple Copy Simultaneous Search (MCSS) algorithm that generated a collection of positioned and oriented chemical functional groups that interacted in some way with the binding-site region of a molecule. It also has the option to perform fragment minimization using Chemistry at HARvard Molecular mechanics (CHARMm) (Brooks *et al.*, 1983). ZRANK was used for re-ranking protein docking predictions with an optimized energy function to increase the accuracy of docked conformations as described by Pierce and Weng (2007). ZRANK significantly improved the success rate over the initial ZDOCK rankings. The amount of test cases with No. 1 ranked hits increased when predictions from ZDOCK versions were considered (Pirece and Weng, 2007).

5.2.8 Intrinsic structural disorder analysis of *Ct*GH43

Intrinsically unstructured Prediction (IUPred) software was employed to analyze for intrinsically unstructured regions in protein (Dosztanyi *el al*., 2005). The underlying assumption is that globular proteins are composed of amino acids which have the potential to form a large number of favorable interactions, whereas,

intrinsically unstructured proteins (IUPs) adopt no stable structure because their amino acid composition does not allow sufficient favorable interactions to form.

5.2.9 Structural analysis of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B by Circular Dichroism

5.2.9.1 Instrument calibration

Far-UV Circular Dichroism (CD) spectra were recorded on spectropolarimeter (Jasco Corporation, JASCO J-815), equipped with a peltier system for temperature control at 25°C using a sample cell with a path length of 0.1 cm. The CD instrument was operated similar to a single beam spectrophotometer. Camphorsulfonic acid which has the negative CD peak at 192.5 nm with molar ellipticity of 15,840 deg cm^2 $dmol^{-1}$, can be used to check instrument performance in far UV region (Creighton, 1997). The machine was allowed to warm up to 30-60 min as extremely high stability of CD signal is required because measured CD spectra are generally small.

5.2.9.2 Slit width time constant and scanning rate

Slit width and time constants are the two major variable parameters that determine the signal to noise ratio of CD signals (Kelly *et al*., 2005). The slit width was determined by observing the maximum value of CD spectra that did not give distortion of CD spectrum. At a given scanning speed a low time constant (0.5 s) is maintained and CD spectrum is recorded and then it is varied to get the maximum values till no distortion was seen. The spectral accumulation parameters used for each far UV CD spectrum were, a scanning rate of 50 nm/min and 1 nm bandwidth in the far UV region 195 to 250 nm with an average of six scans.

5.2.9.3 Sample preparation

A protein (*Ct*GH43 or *Ct*CBM6B) concentration of 0.1% to 0.001% (w/v) is suitable for CD measurements with 0.01 and 0.1 cm path length, respectively (Creighton, 1997). The protein samples in (10-15 µM) in 20 mM sodium phosphate buffer, pH 7.0 were passed through a 0.45 µm filter before subjecting to spectropolarimeter, and the spectra was recorded following the method described by Kelly *et al.* (2005).

5.2.9.4 Circular Dichroism spectrum

The Circular Dichroism CD spectrum in the far UV region or data are generally presented in terms of mean residue ellipticity (MRE, expressed as deg cm^2 $dmol^{-1}$) as a function of wavelength, calculated according to the procedure described earlier (Kelly *et al*. 2005; Greenfield *et al*., 2006). All CD spectra were corrected for buffer contributions and secondary structures were calculated by using web based K2d neural network software package (http://www.embl.de/~andrade/k2d.html) as described by Andrade *et al*. (1993) and Greenfield *et al*. (2006).

5.2.9.5 Circular dichroism data evaluation

The CD data are recorded either as differences in the absorbance of left (ΔL) or right (ΔR) handed polarized light given as,

$$\Delta\varepsilon = \Delta_L - \Delta_R \qquad (1)$$

or they can be expressed in terms of molar ellipticity,

$$[\theta] = 3300 \times \Delta\varepsilon. \qquad (2)$$

Applying Beer-Lambert law to equation 1 yields,

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R = \frac{\Delta_L - \Delta_R}{c \times l} \qquad (3)$$

where, c is the concentration of protein in moles/litre and l is the path length in cm. $\Delta\varepsilon$ has unit of liter × mol^{-1} × cm^{-1} or liter × $(mol\ residues)^{-1}$ × cm^{-1}.

So, $[\theta]$ (molar ellipticity) can be calculated as

$$[\theta] = \frac{\theta \times 100 \times M_r}{c \times l} \qquad (4)$$

where, M_r is the protein molecular weight, θ is ellipticity measured in degrees.

Now if the protein concentration is in mg/ml, the concentration of amino acids can be calculated by assuming a mean residue weight (MRW) of 110 amino acid residue (Creighton, 1997).

The mean molar ellipticity $[\theta]_{MRE}$ can be calculated as,

$$[\theta]_{MRE} = \frac{\theta \times 100 \times M_r}{c \times l \times N_A} \qquad (5)$$

where, N_A is the number of amino acids per protein and M_r is the protein molecular weight. The molar ellipticity was expressed in terms of deg cm^2 $dmol^{-1}$ (Greenfield *et al*., 2006). The factor 100 is added in the above equations for converting the molar concentration to $dmol/cm^2$ concentration units. The relation between $[\theta]$ and $\Delta\varepsilon$ are given in equation 2 which allows immediate transformation of raw θ data into $\Delta\varepsilon$ values and vice versa using the relation

$$\theta = 33\ (\Delta_L - \Delta_R) \qquad (6)$$

This implies that a measured ellipticity of 10 mdeg is equivalent to a $\Delta\varepsilon$ of only 0.0003 absorbance units.

5.3 Results and Discussion

5.3.1 Structure and docking analysis of *Ct*GH43

5.3.1.1 Multiple sequence alignment of CtGH43

The multiple sequence alignment (MSA) was carried out using Multalin by considering the similar sequences extracted from the protein BLAST (blastp) analysis of *Ct*GH43 (Corpet, 1988). MSA of *Ct*GH43 showed consensus portions of aligned sequences highlighted in red which are identical amino acids in all 5 sequences (Fig. 5.3.1). Conserved amino acids in at least 4 out of 5 sequences are highlighted in blue and the non-identical portions are in black colour and the gaps were also included for improving the alignment (Fig. 5.3.1). *Ct*GH43 showed conserved regions of amino acid residues with family 6 carbohydrate binding modules (CBM6s) from *Clostridium thermocellum* JW10, *Clostridium papyrosolvens, Clostridium cellulolyticum* H10 and *Acetovibrio cellulyticus* (Fig. 5.3.1).

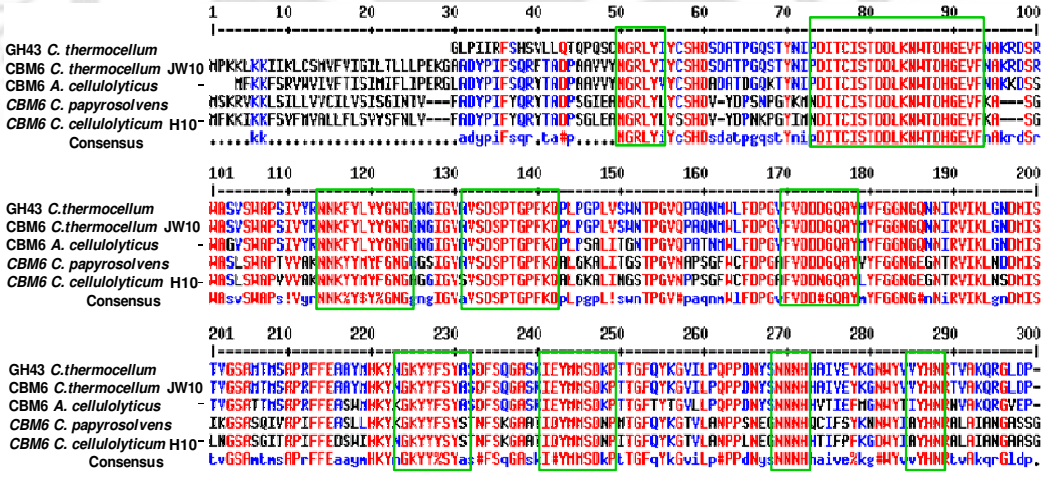


Fig. 5.3.1 Multiple sequence alignment analysis of *Ct*GH43 using ClustalW. The analysis showed similarities with amino acid sequences of family 6 CBM from *Clostridium thermocellum* JW10, *Clostridium cellulolyticum* H10, *Clostridium papyrosolvens, Clostridium cellulyticum* H10 and *Acetovibrio cellulyticus*. Identical amino acids in all 5 sequences are shown in red in green boxes. In blue are the conserved amino acids in at least 4 out of 5 sequences and the least matching residues are shown in black.

5.3.1.2 Phylogenetic analysis of CtGH43

A phylogenetic tree of *Ct*GH43 was constructed using maximum likelihood tree (MLtree) based on the blastp results by considering the sequences showing maximum similarity to find out the evolutionary inter-relationships (Christin *et al*., 2008). The blastp analysis of *Ct*GH43 showed nearly 90% similarity with CBM6 from *Clostridium thermocellum* JW10, 88% similarities with CBM6 from *Acetovibrio cellulyticus,* 63% similarity with CBM6 from *Clostridium papyrosolvens* DSM 2782 and 62% with CBM6 from *Clostridium cellulolyticum* H10 (Fig. 5.3.2). The Phylogenetic tree showed evolutionary resemblance of *Ct*GH43 with family 6 carbohydrate binding proteins or modules of *Clostridium thermocellum* JW10*, Clostridium cellulolyticum* H10*, Clostridium papyrosolvens* DSM2782 and *Acetovibrio cellulolyticus* (Fig. 5.3.2).

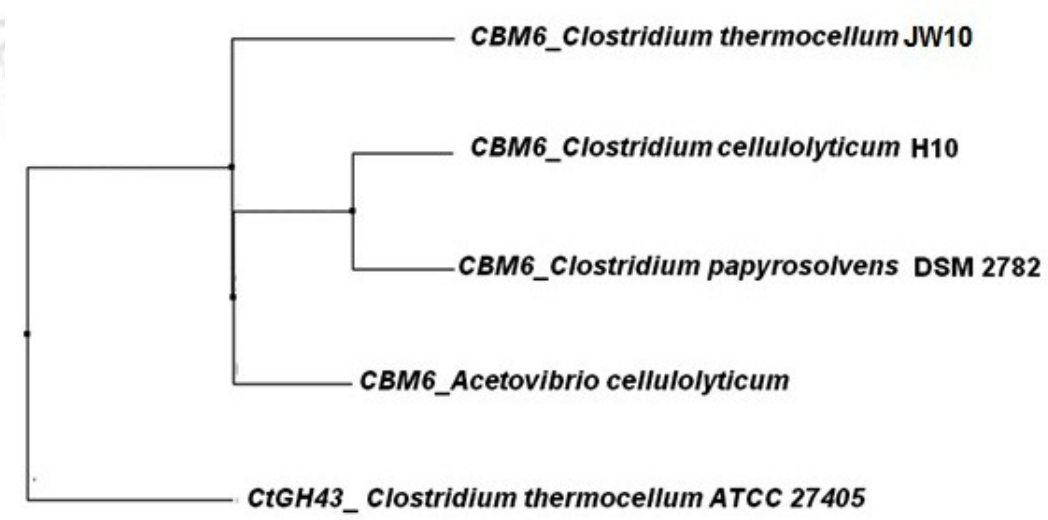


Fig. 5.3.2 Phylogenetic tree constructed by MLtree based on the blastp. The protein sequence of *Ct*GH43 showed close resemblance with carbohydrate binding module of family 6 (CBM6s) from *Clostridium thermocellum* JW10, *Clostridium cellulolyticum* H10, *Clostridium papyrosolvens* DSM2782 and *Acetovibrio cellulolyticus*.

5.3.1.3 Secondary structure of CtGH43

PSI-PRED analysis of *Ct*GH43 sequence showed many strands and coils along with their confidence level for such occurrence (Fig. 5.3.3). The secondary structure of *Ct*GH43 displayed very few or no helices similar to Clan F structures belonging to family 43 glycoside hydrolase as reported earlier (Henrissat, 1991; Davies and Henrissat, 1995; Henrissat and Baroch, 1996).

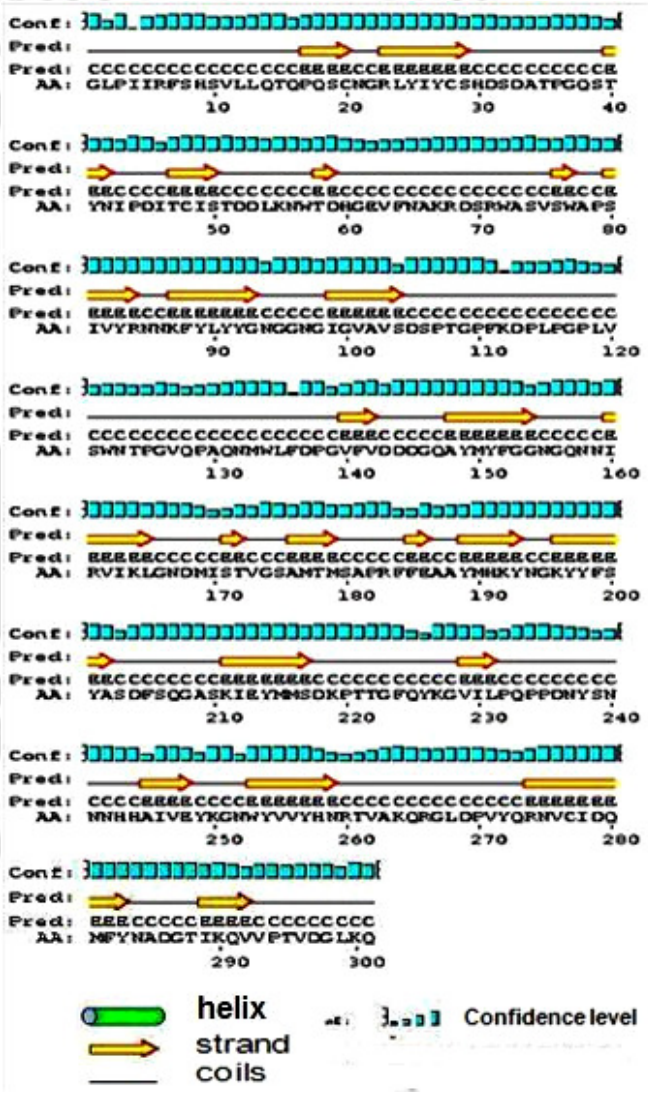


Fig. 5.3.3 Secondary structure of *Ct*GH43 using PSI-PRED VIEW software showing β-strands (shown as arrow region) and coils (as continuous black line) with confidence level of prediction. Cylinders denote helices and yellow arrow denotes β-strands.

5.3.1.4 Homology modeling of CtGH43

The HHpred analysis for homology search gave 9 hits having significant probability score and better sequence similarity as compared to others. The 3-dimensional structure prediction of *Ct*GH43 using the templates from protein data bank (PDB) was done (Fig. 5.3.4) by utilizing MODELLER 9v2 for homology modeling. Comparative or homology modeling was generated by using the templates *viz.* arabinoxylan arabinofuranohydrolases (3C7F, Vandermarliere *et al.,* 2009), β-1,4-xylosidase (1YIF, unpublished), xylan β-1,4-xylosidase (1YRZ, unpublished), β-D-xylosidase (2EXH, Brux *et al*., 2006) and endo-1,5-α-L-arabinanase (1WL7, Yamaguchi *et al*., 2005) utilizing the crystal structures from the PDB (Berman, Westbrook and Feng, 2000; Berman *et al*., 2007; Marchler-Bauer and Anderson, 2011). The characteristic 5-fold β–propeller core was easily identified from 3-dimesional model of *Ct*GH43 (Fig. 5.3.4).

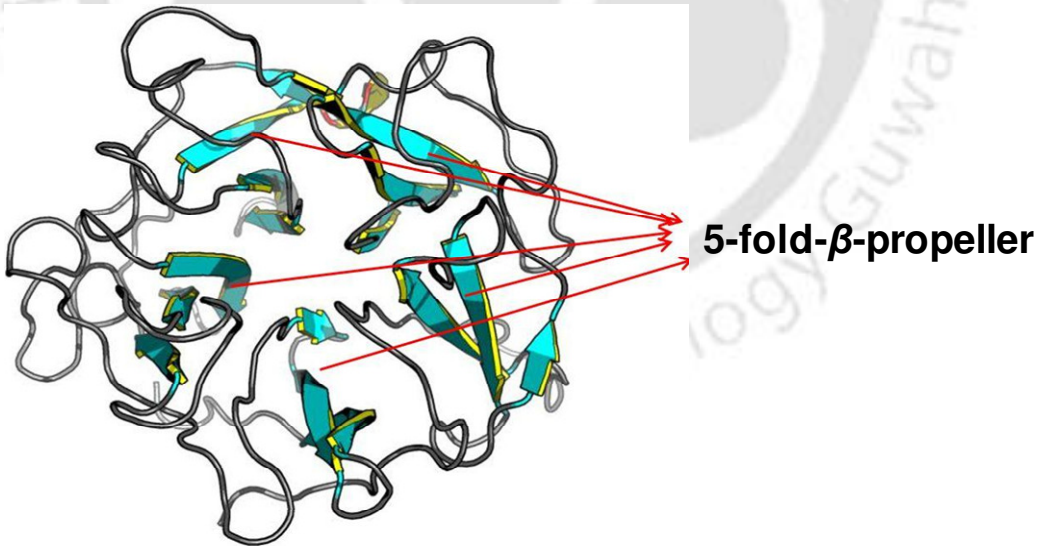


Fig. 5.3.4 Three-dimensional structure of *Ct*GH43 developed utilizing MODELLER 9v2 and visualized using PyMOL software showing the characteristic fold (5 fold β-sheets in green, coils in greyish black).

5.3.1.4.1 Structure validation using VERIFY 3D

The energy minimization of pdb file (generated by MODELLER) of *Ct*GH43 was done using Swiss-Pdb Viewer version 4.0 before feeding the pdb file for analysis by VERIFY3D for validation (Guex and Peitsch, 1997). The validation of predicted structure of *Ct*GH43 by VERIFY 3D confirmed the model is acceptable (Fig. 5.3.5). Negligible portion of *Ct*GH43 were found to show score below 0.2 (very few residues cross the light grayish line) indicating a stable conformation of the predicted model.

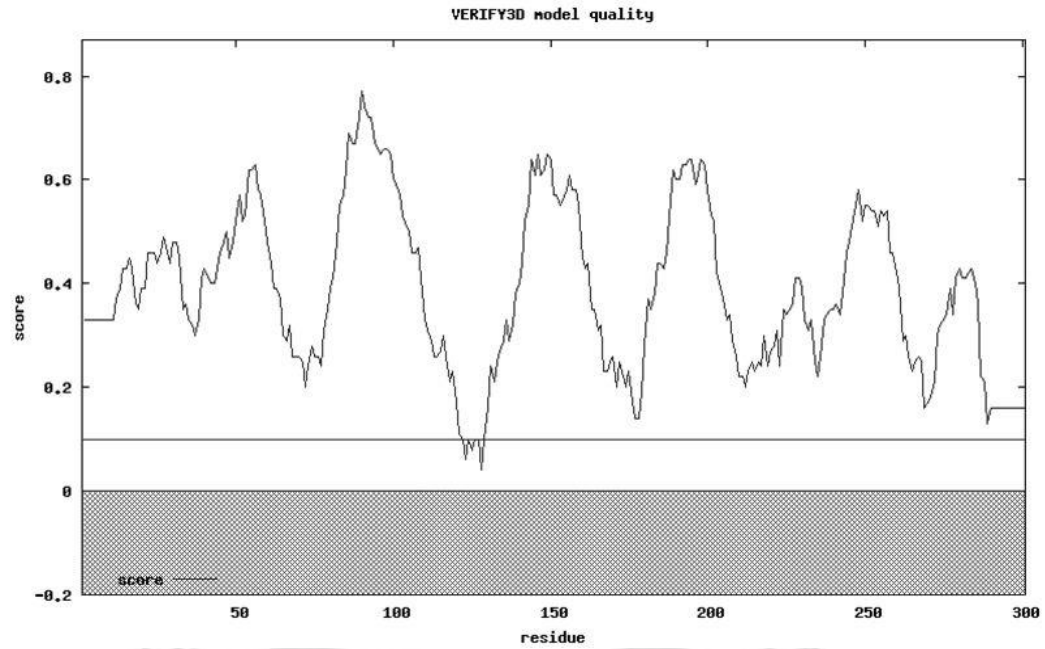


Fig. 5.3.5 Quality assessment of predicted 3-dimensional model of *Ct*GH43 by VERIFY3D software. Very few residues cross the light greyish line having negative value implying predicted model is acceptable.

5.3.1.4.2 Ramachandran plot analysis of 3-D structure of CtGH43

The 3-D structure analysis of *Ct*GH43 by MODELLER showed little or no segments of helices (Fig. 5.3.4). Similarly, repetitive values in the region of phi (Φ) = -110 to -140 and psi (Ψ) = +110 to +130 gave extended chains with conformations

that allow interactions between closely folded parallel segments (β-sheet structures) (Fig. 5.3.6 and Fig. 5.3.7). β-sheets in folded proteins are twisted rather than planar, with a right-handed twist of 0°-30° between strands. β-sheets can consists entirely of parallel or mixture of parallel and anti-parallel sheets. The structure of *Ct*GH43 is composed mostly of β-sheets with 178 residues and the RC plot showed a broad range of values in the (-110°, +130°) regions (Fig. 5.3.6 and Fig. 5.3.7). The same figure shows only 4 residues in 3_{10} helical regions, which are present in first quadrant of conformational map (Fig 5.3.6 and Fig. 5.3.7). The RAMPAGE plot indicates that out of the 301 residues, 267 (89%) were in most favoured region, 23 (8%) were in favoured region and8 (nearly 3%) were in disallowed region implying that the predicted model is quite acceptable (Fig. 5.3.6 and Fig. 5.3.7).

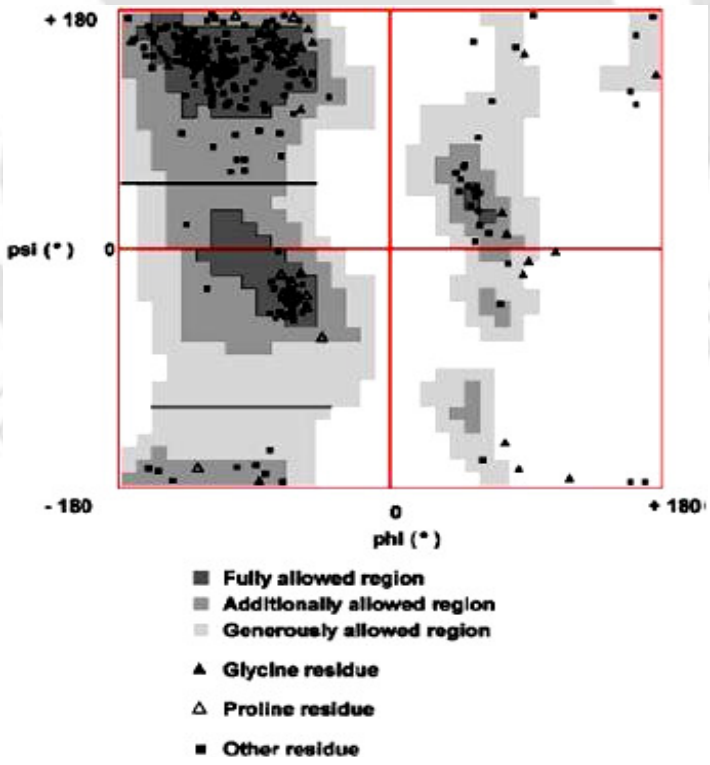


Fig. 5.3.6 Ramachandran plot for *Ct*GH43 from the online server available from Indian Institute of Science (IISc), Bangalore. It shows the left-handed as well as right-handed helices, beta sheets and disallowed and allowed areas.

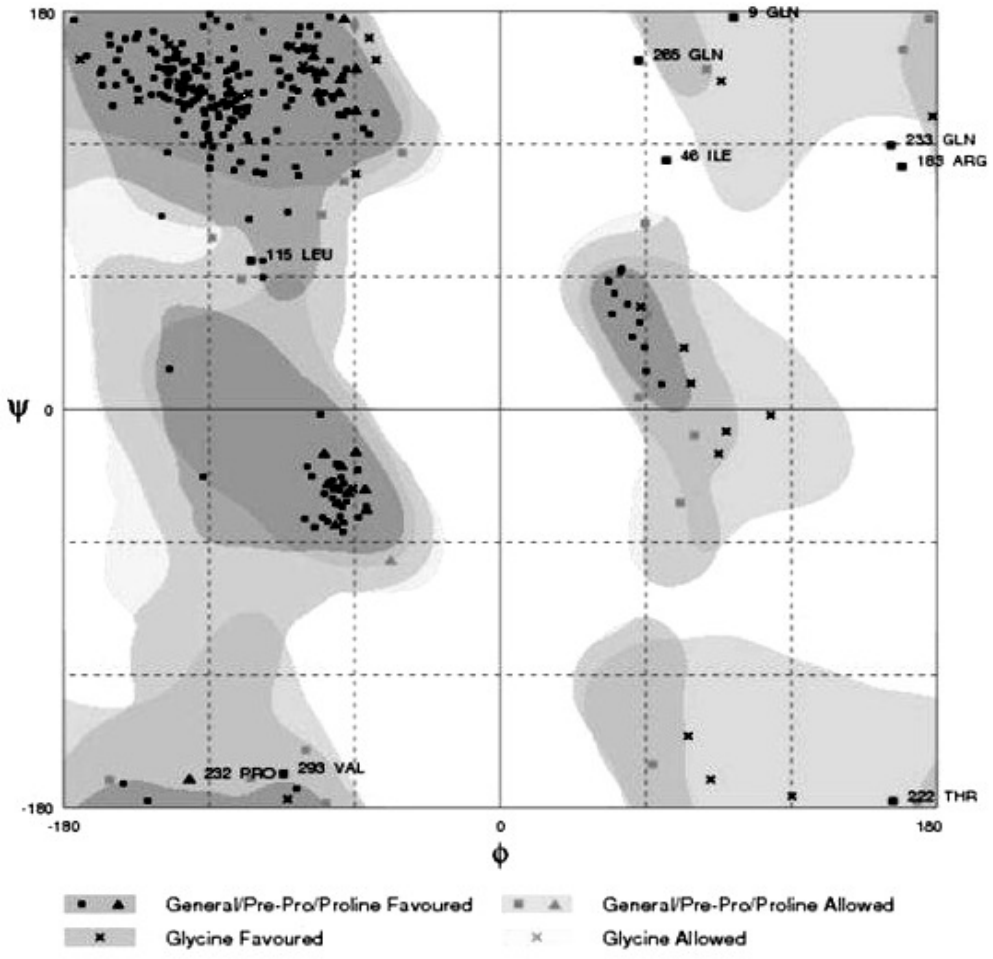


Fig. 5.3.7 Ramachandran plot analysis of *Ct*GH43 by RAMPAGE software shows the residues falling in most favoured, allowed and disallowed region and the glycine residues (267 residues are in favoured region, 23 in allowed region and 8 in disallowed region) so 97 % residues have favoured conformations.

RC plot for general, glycine, pre-proline and proline was also generated and it showed that the glycine, pre-proline and proline of *Ct*GH43 are falling under allowed regions and also those glycine residues falling in disallowed region (Fig. 5.3.8). The Ramachandran Z-score (-1.706) indicated stable backbone conformations of all residues corresponding to the known allowed areas in the RC plot is within the expected range (Ramachandran, Sasisekharan and Ramakrishnan, 1966). The MolProbity RC plot analysis displayed that 97.3% of residues are in the favoured

region and only 2.7% falls is the outliers. This further confirmed the acceptability of the 3-D model of *Ct*GH43 (Fig. 5.3.9).

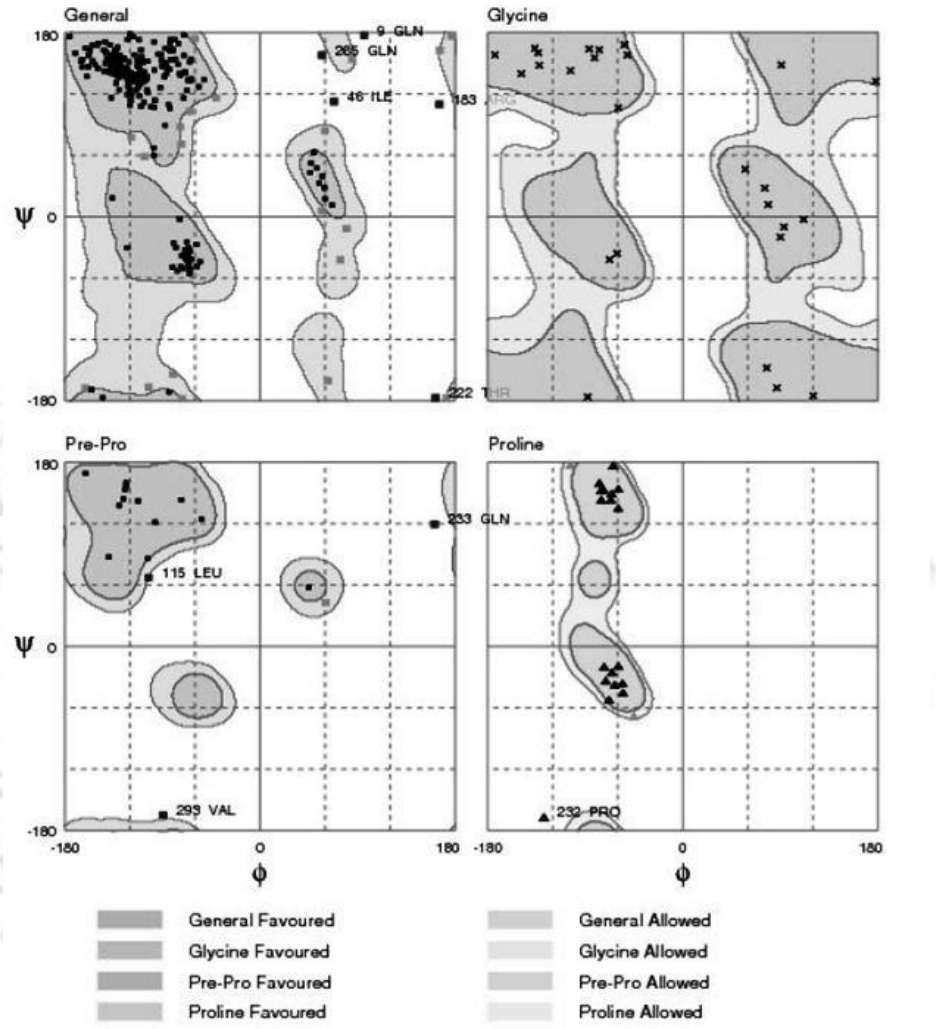
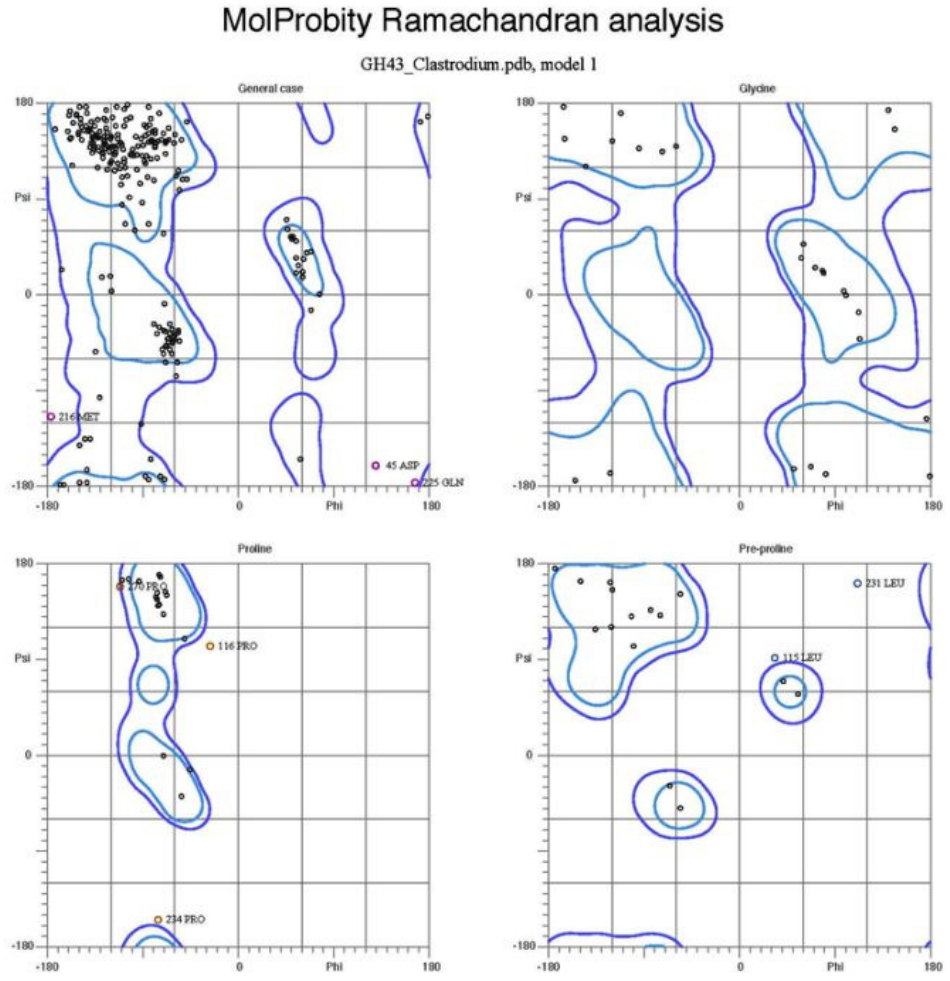


Fig. 5.3.8 Ramachandran plot analysis of *Ct*GH43 for general, gly, Pre-Pro, Pro using RAMPAGE. The conformations and location of each of the above is shown in individual plots having heading as general, Glycine, pre-Pro and Pro.



87.0% (260/299) of all residues were in favored (98%) regions.
97.3% (291/299) of all residues were in allowed (>99.8%) regions.

There were 8 outliers (phi, psi):
45 ASP (129.2, -160.9)
115 LEU (33.7, 92.7)
116 PRO (-27.5, 103.1)
216 MET (-178.0, -114.8)
225 GLN (167.0, -176.1)
231 LEU (111.6, 162.5)
234 PRO (-76.4, -154.5)
270 PRO (-112.8, 159.2)

Fig. 5.3.9 Structure validation of *Ct*GH43 by Ramachandran Plot using MolProbity.

5.3.1.5 Hydrogen bonding plot of CtGH43

The hydrogen bonding plot (HB plot) was generated from the *Ct*GH43 3D coordinate pdb file from HB plot server (www.virtuadrug.com/hbplot). The characteristic patterns of secondary structure elements were displayed by HB plot (Fig. 5.3.10). Anti parallel beta strands were represented by strips perpendicular to the diagonal. Visual inspection of the HB plots revealed the regions where the density of the hydrogen bridges are high and the regions, which have very H-bonds. The HB plot of *Ct*GH43 mainly displayed type I (distance smaller than 2.5 Å between donor and acceptor) and type II and very few types III class (distance greater than 3.2 Å) of hydrogen bonds (Fig. 5.3.10).

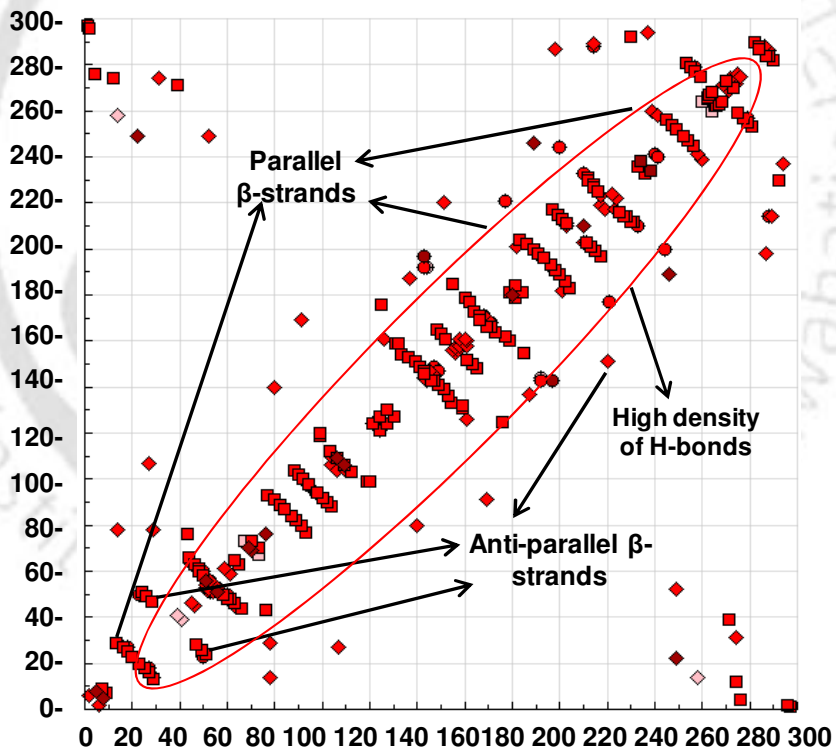


Fig. 5.3.10 Hydrogen bonding plot of *Ct*GH43 which showed the anti-parallel β-strands, being represented by strips perpendicular to the diagonal. Three classes of hydrogen bonding distinguished by colour coding:
(a) Type I (brown): Short (distance smaller than 2.5 Å between donor and acceptor),
(b) Type II (red): intermediate (between 2.5 Å and 3.2 Å) and
(c) Type III (pink): long hydrogen bonds (greater than 3.2 Å).

5.3.1.6 Prediction of catalytic or binding sites of CtGH43

The active site residues were identified using a combination of software packages or servers *viz.* MetaPocket, QSiteFinder and THEMATICS. Fig. 5.3.11 displayed the active site residues of *Ct*GH43 involved in ligand binding, which may play important role in the enzyme catalysis. The residues found to be present at the active site were Gln14, Gln38, Trp77, Trp134, Phe 136, Asp137, Gly153, Gly154, Asn155, Glu186, Asp204, Lys211 and Ile212.

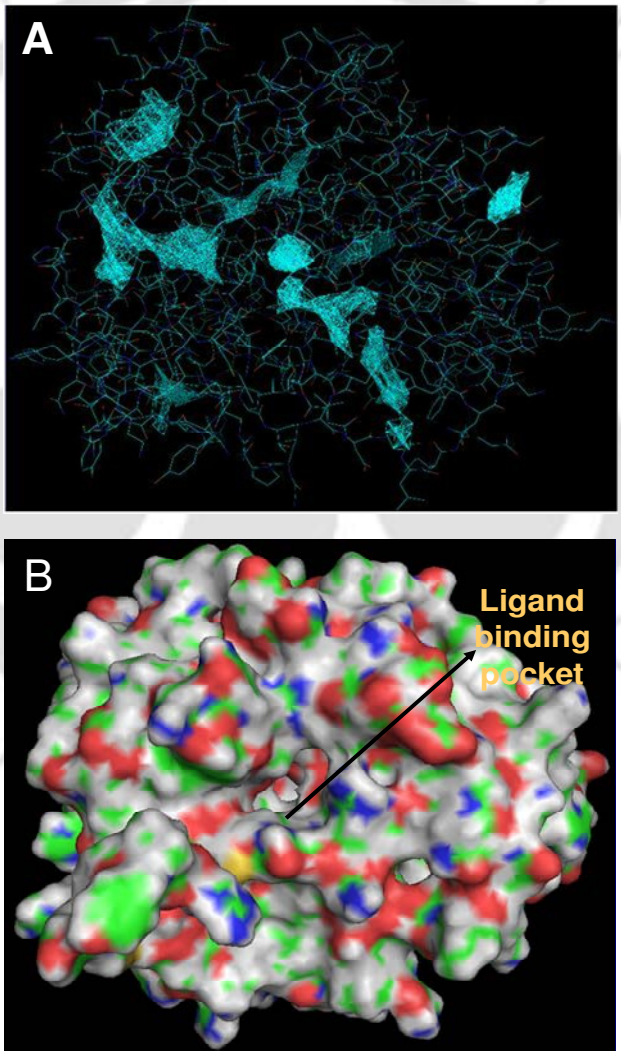


Fig. 5.3.11 Three dimensional structure of *Ct*GH43 A) revealing the entire Positive cluster or cavities (binding sites) by THEMATICS and B) showing the possible ligand biding sites based on QSiteFinder and metaPocket.

5.3.1.7 Active site and docking analysis of CtGH43

Molegro Vertual Docker (MVD) displayed five cavities with the volume 186.36, 107.008, 84.48, 51.712 and 33.28 Å (Fig. 5.3.12A). Each cavity was analyzed with the ligand xyloteatrose for binding. A grid resolution of 0.3Å was used and maximum iteration number was set to 1500 with maximum population size of 50. The ligand was set into the cavity and a sphere of 15Å was considered for flexible docking with *Ct*GH43 (Fig. 5.3.12B).

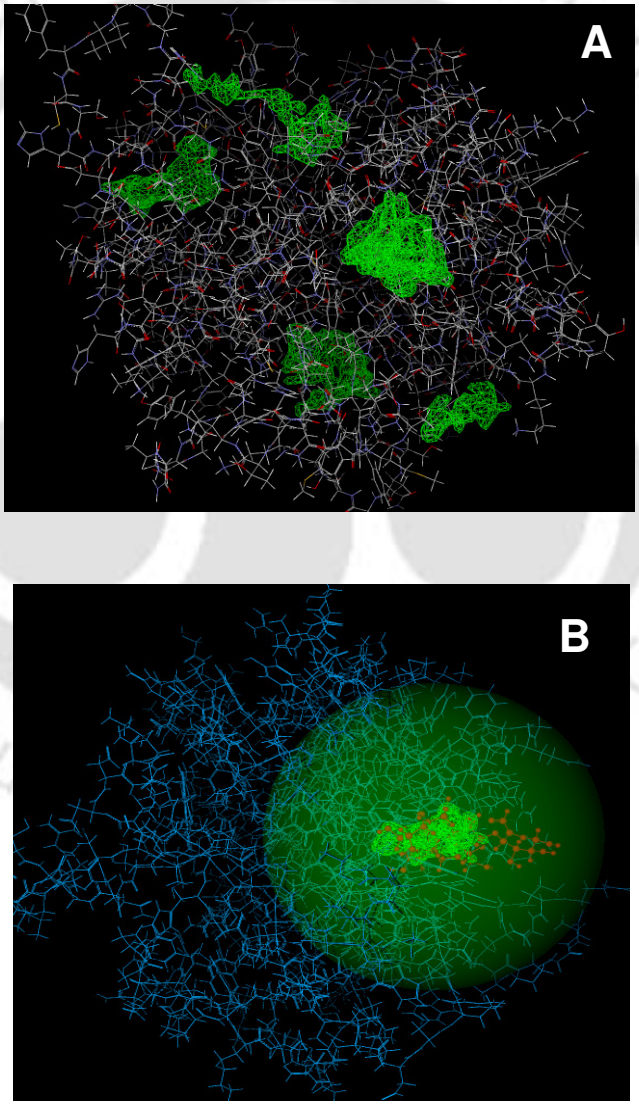


Fig. 5.3.12 3-dimensional structure of *Ct*GH43 **A**) showing ligand binding cavities in green patches and **B**) showing ligand (in red) at one of the cavities with spherical constraints is depicted.

The catalytic site of *Ct*GH43 was discovered to be situated at one end of the funnel-shaped cavity of the β-propeller domain that is formed as a result of the β-sheet packing (Fig. 5.3.13).

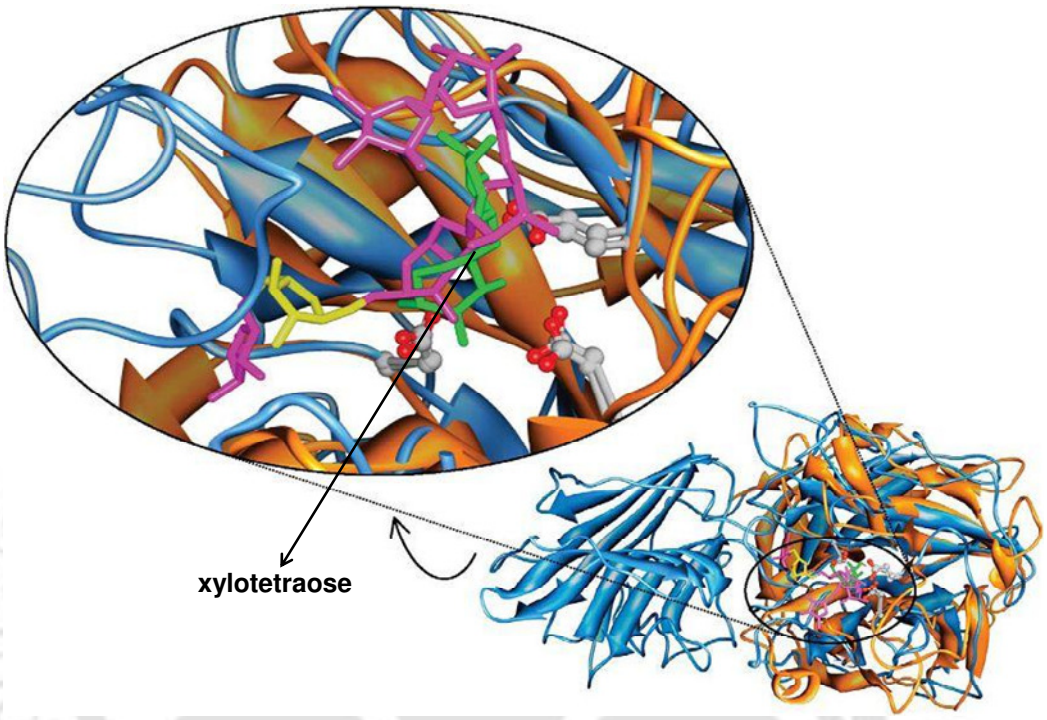


Fig. 5.3.13 Funnel-shaped cavity of the β-propeller domain of *Ct*GH43 (zoomed region shows the catalytic active site of *Ct*GH43.

Like other GH43 members, three acidic residues are critical for the catalytic activity of *Ct*GH43; Asp137, the general base; Glu186, the general acid; and Asp204, which was thought to play a role in pKa modulation, and in maintaining the correct orientation of the general-acid residue (Fig. 5.3.14). These three residues are pointed toward the centre of the cavity and located on blades I, III, and IV on the innermost β-strands. The active site possesses a pocket topology, which is mainly constructed from β-propeller domain residues.

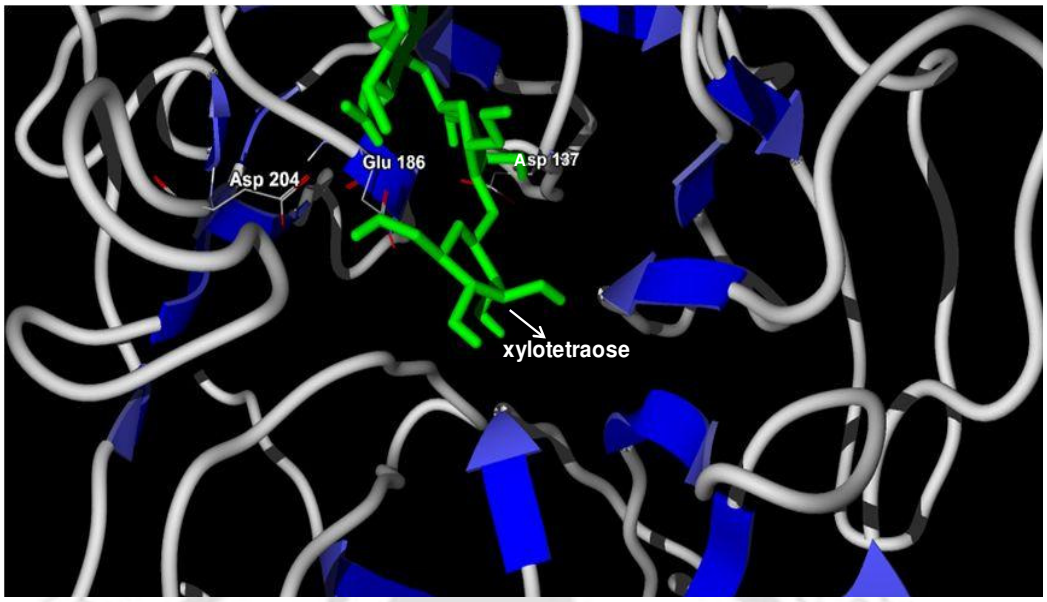


Fig. 5.3.14 Three dimensional structure of *Ct*GH43 revealing all the residues that are present at active site. It shows typical architecture where an Asp is the catalytic general base, a Glu as the catalytic general acid an Asp as pKa modulator. The ligand is shown in green.

5.3.1.7.1 Enzyme-ligand docking score analysis

Docking process resulted into the 5 ligand orientations with the energies and scores shown in Table 5.3.1. All forms of orientations were analyzed with the different cavities until the optimized results were obtained, both in terms of energy and as well as orientation of molecules. Further, two Asp residues and one glutamate residue interacting with xylotetraose were discovered. Table 5.3.2 shows the values of distance in angstrom (Å) and the interaction between *Ct*GH43 and xylotetraose during catalysis. After the ligand bound to the enzyme, the ligand energy inspector was used from the MVD suite to analyze energy terms (Table 5.3.3).

Table 5.3.1 Energy of different xylotetraose (Xyl) orientations.

Name	MolDock Score	Rerank	HBond
(00) xylotetraose	-185.45	-102.883	-24.6033
(01) xylotetraose	-169.414	-60.3592	-20.4303
(02) xylotetraose	-160.474	-113.846	-20.0904
(04) xylotetraose	-158.525	-35.5969	-21.6551
(03) xylotetraose	-155.171	-105.87	-18.2158

Table 5.3.2 Interaction between *Ct*GH43 and xylotetraose during catalysis.

*Ct*GH43 Atom	Ligand Atom	Distance (Å)
Asp-137-OD1	Xyl-H18	1.80
Asp-137-OD2	Xyl-H16	1.57
Glu-186-E1	Xyl-H12	2.12
Asp-204-OD2	Xyl-H12	2.95

Table 5.3.3 Ligand energy distribution after binding in bound to *Ct*GH43.

Descriptors		Value	MolDock Score
Total Energy		-129.546	-145.03
External Ligand interactions	Steric (PLP)	-136.272	-136.272
	Steric (LJ12-6)	-15.708	0
	Hydrogen Bonds	-26.188	-26.188
	Hydrogen Bonds* (no direct.)	-39.129	0
Internal Ligand interactions	Torsional Strain	18.376	18.376
	Torsional Strain	0	0
	Hydrogen Bonds	0	0
	Steric (PLP)	-0.947	0
	Steric (LJ12-6)	70.322	0

* No directionality

5.3.1.9 Probable reaction mechanism based on the CtGH43 docking results

The docking analysis showed presence of a catalytic triad of two Asp and a Glu residue at the active site. Based on the interaction of these amino acids at the active site with the ligand xylotetraose hydrogen atoms and on previous reports of family 43 glycoside hydrolase, a plausible model for reaction mechanism model could be deduced. The general base, Asp137, is located near the anomeric centre in the sugar ring (5.3.15A). This residue is situated between two conserved residues that provide the exact orientation required for the catalytic activity of this acidic residue. The position of Asp137 allowed it to activate a water molecule for a single displacement attacks on the anomeric carbon from the opposite side of the glycosidic bond, resulting in inversion of the anomeric configuration (OD1, OD 2 conformations, Fig. 5.3.15A). The general acid residue, Glu186, is 2.12 Å formed the glycosidic oxygen, allowing it to protonate the leaving aglycon (Fig. 5.3.15A). For this residue to function as a general acid, its pKa must be considerably higher than normal (4.07). Asp204 is positioned about 4Å from Glu186, a distance that enables it to modulate the pKa of this residue and raises it close to neutral pH keeping it protonated and allowed it to function as acid for the reaction as shown in Fig. 5.3.15B.

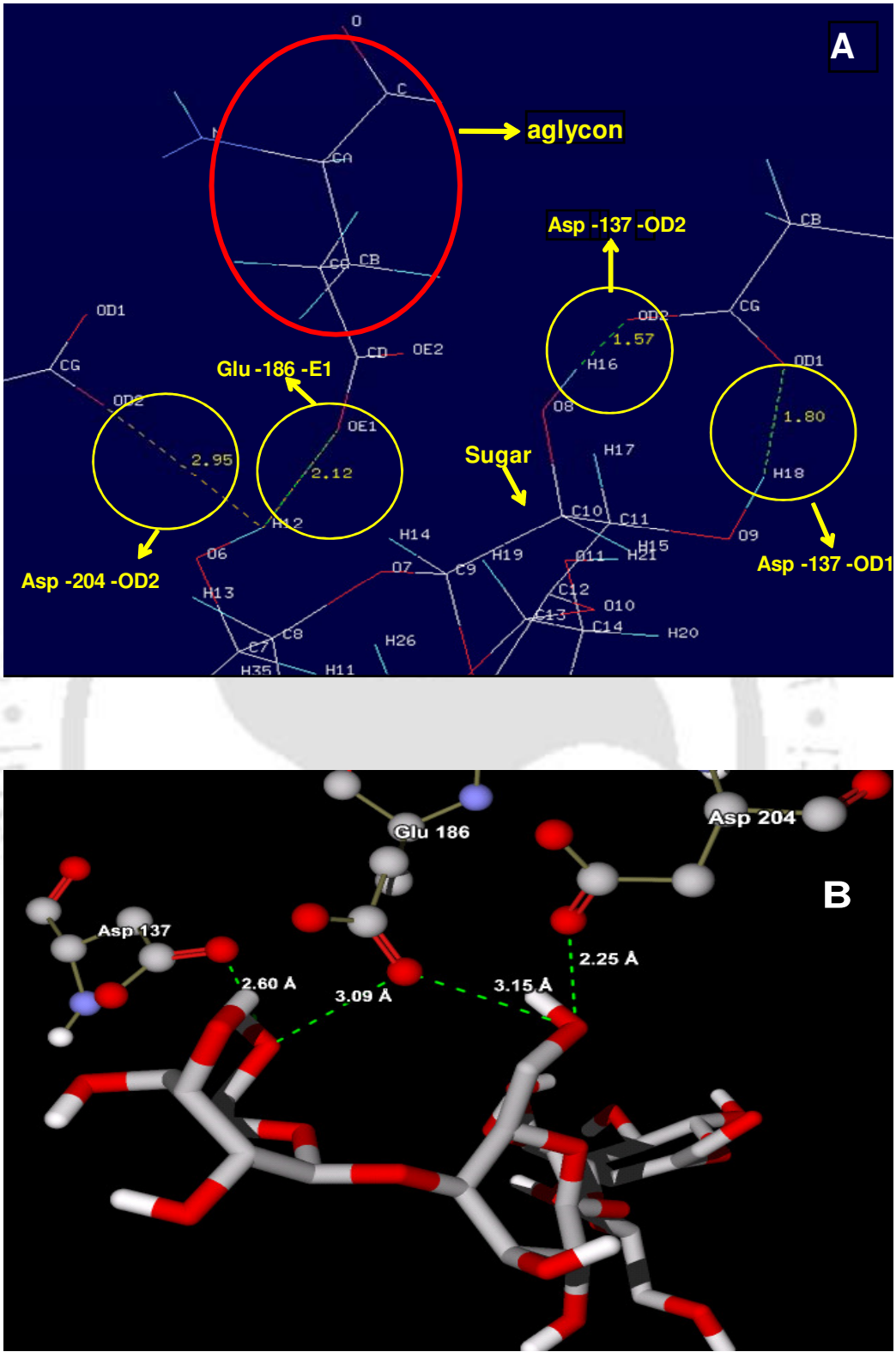


Fig. 5.3.15 The reaction mechanism of *Ct*GH43 showing **A)** interaction of two Asp and a Glu residues with ligand (xylotetraose) at the active site and **B)** the distances (Å) and residues involved in the interaction of *Ct*GH43 with the xylotetraose.

5.3.1.10 Prediction of structural disorder in protein

IUPred was used for the prediction of intrinsically unstructured regions of proteins based on estimated energy content (Dosztányi *et al*., 2005). The IUPred result shows that there are only few residues which are showing disorder tendency index more than 0.5, concluding that protein does not have any significant disordered region (Fig. 5.3.16). The plot, therefore, confirms the fact that there are very few or no intrinsically unstructured regions in *Ct*GH43.

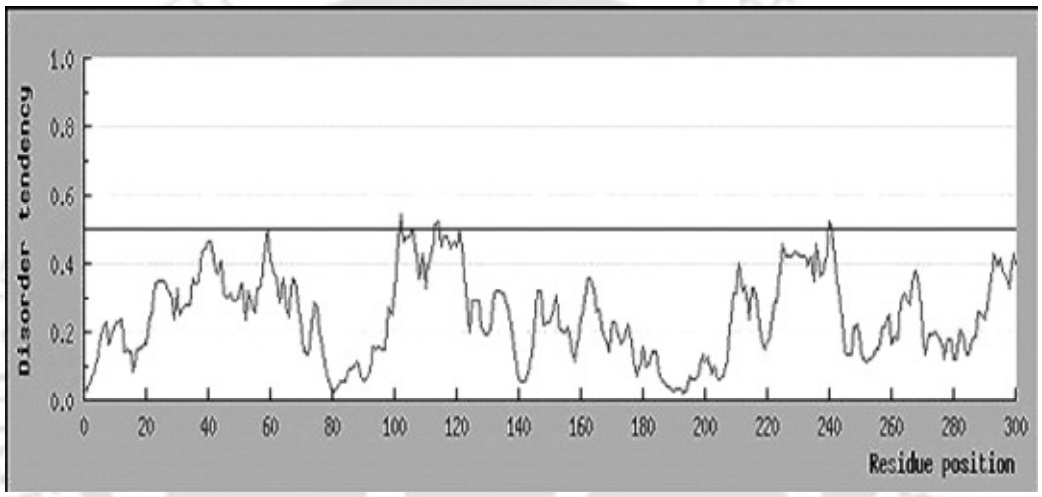


Fig. 5.3.16 IUpred analysis of *Ct*GH43 for intrinsically unstructured regions.

5.3.2 Structure and binding analyses of *Ct*CBM6A and *Ct*CBM6B

5.3.2.1 Multiple sequence alignment analysis of CtCBM6A and CtCBM6B

The multiple sequence analysis (MSA) of *Ct*CBM6A displayed many conserved aromatic residues when compared to the family 6 carbohydrate binding modules from *Clostridium thermocellum* (1gmm), *Celvibrio mixtus* (1uxx) and *C. cellulolyticum* (2v4v) as shown in Fig. 5.3.17A. MSA results of *Ct*CBM6B also displayed similar conserved aromatic amino acids when were compared with family 6 carbohydrate binding modules (CBM6s) from *C. thermocellum* (1gmm), *Celvibrio mixtus* (1uxx) and *Clostridium stercorarium* (1uy3) (Fig. 5.3.17B). Therefore, both the CBMs showed close conservation of sequences with other family 6 carbohydrate binding module. This also illustrated their close evolutionary relationship and indicated that they might have evolved from common ancestor.

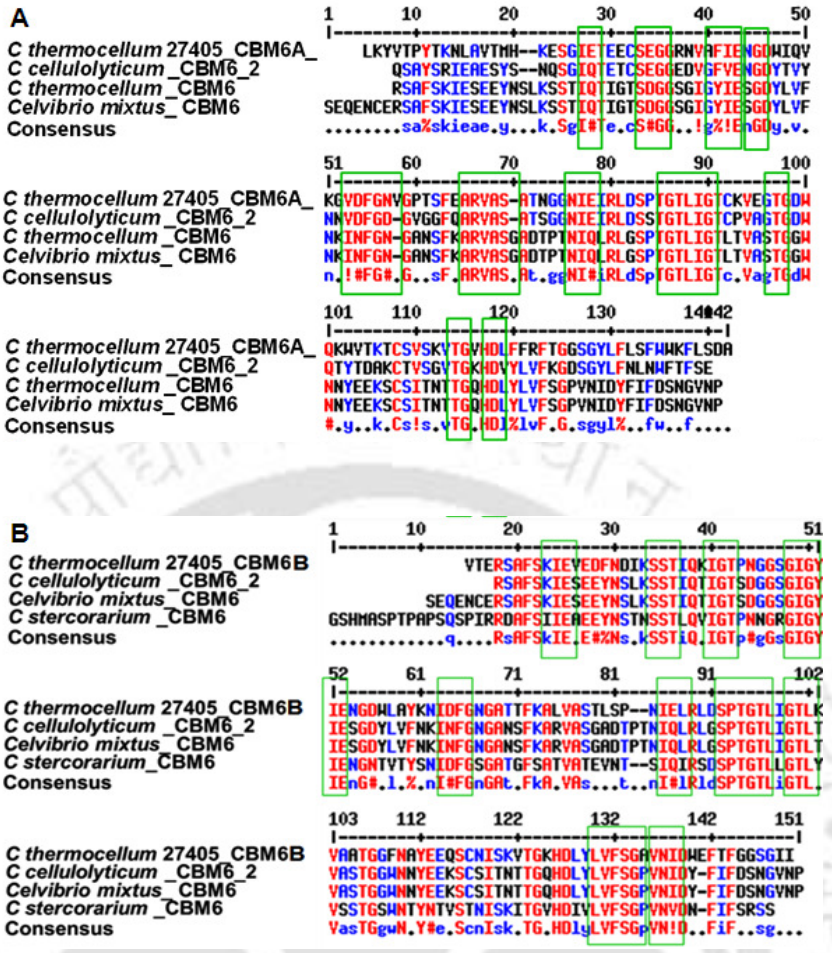


Fig. 5.3.17 Multiple sequence alignment of **A)** *Ct*CBM6A showing matching sequences with other CBM6 from *Clostridium thermocellum, C. cellulolyticum* and *Celvibrio mixtus* and **B)** *Ct*CBM6B showing matching sequences with other CBM6 from *Clostridium thermocellum, C. stercorarium* and *Celvibrio mixtus*. Identical amino acids in all 5 sequences are shown in red. In blue are the conserved amino acids in at least 4 out of 5 sequences and non-identical residues are shown in black.

5.3.2.2 Secondary structure of CtCBM6A and CtCBM6B

The secondary structural analysis of *Ct*CBM6A ad *Ct*CBM6B revealed that they mainly contain β-strands and coils (Fig. 5.3.18). Both CBMs showed β-strands mainly because of the large aromatic residues (tryptophan, tyrosine and phenylalanine) and C^{β}-branched amino acids (isoleucine, valine, and threonine) are known to adopt β-strand conformations (Branden and Tooze, 1999). Phyre server

analysis of *Ct*CBM6A showed 4.4 % of α-helices, 46.6% β-sheets and the rest 49% as random coils. However, these preferences are not strong enough to produce a reliable method of predicting secondary structure from the sequence alone.

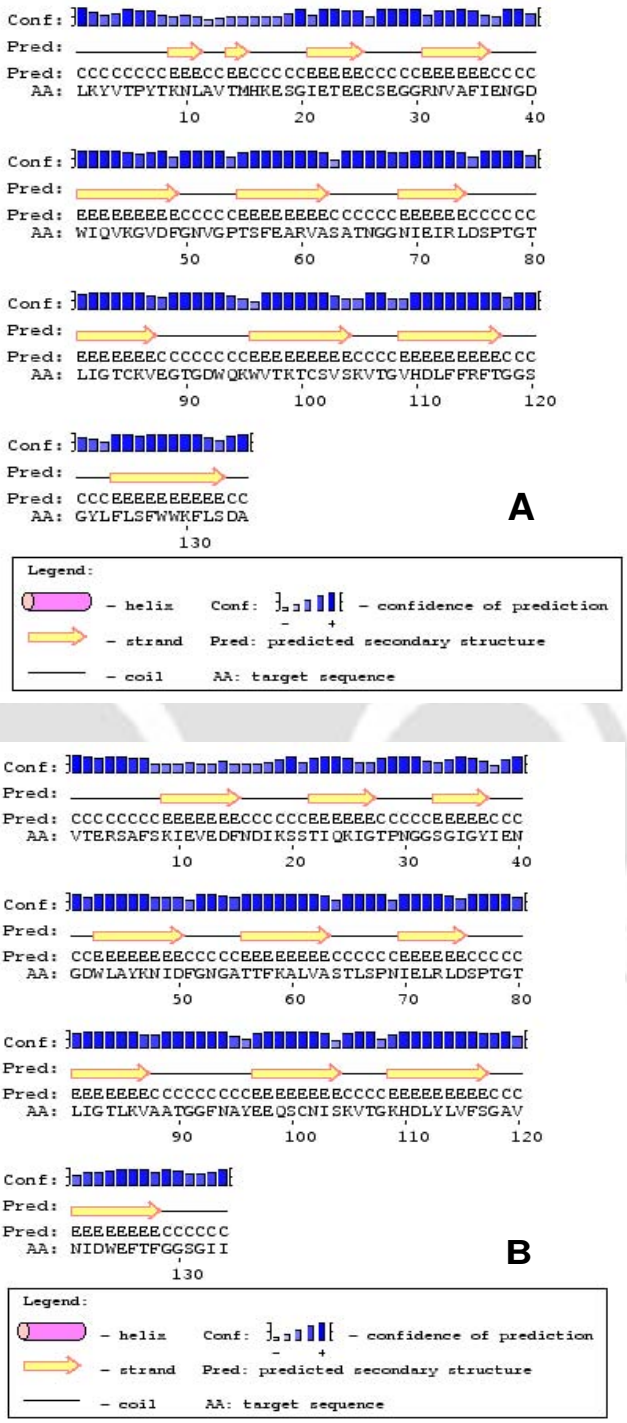


Fig. 5.3.18 Secondary structure prediction of **A**) *Ct*CBM6A and **B**) *Ct*CBM6B using PSIPRED software.

5.3.2.3 Homology modelling of CtCBM6A and CtCBM6B

The homology modelling of *Ct*CBM6A and *Ct*CBM6B using MODELLER showed multiple β-strands and coils, similar to the secondary structures predicted for these proteins by PSIPRED. Again the existence of β-strands might be due the presence of large aromatic acid residues like tyrosine, tryptophan and phenylalanine. The β-sandwich core architecture of the predicted three dimensional structures of *Ct*CBM6A and *Ct*CBM6B were quite similar to already known structure of CBM6 family (Fig. 5.3.19). *Ct*CBM6A and *Ct*CBM6B belonged to tmpCBM-A clan as per CAZy classification (http://www.cazy.org/CBM6.html) since they show a typical β-sandwich fold which is conserved among CBM6s (Cantarel *et al*., 2009).

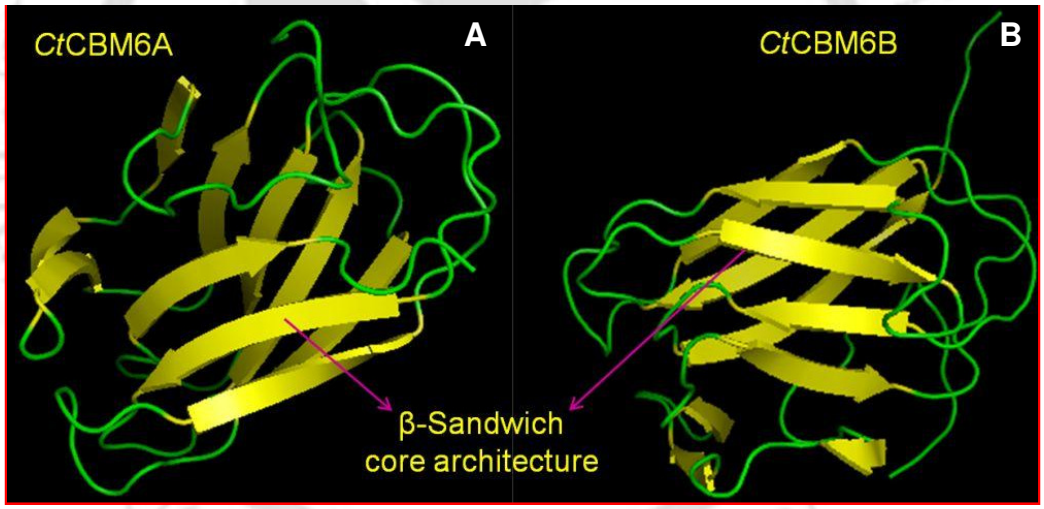


Fig. 5.3.19 3-dimensional structure of **A**) *Ct*CBM6A and **B)** *Ct*CBM6B showing *β*-strands in yellow and the loop regions (green).

5.3.2.3.1 Structure validation of CtCBM6A and CtCBM6B by Ramachandran plot

Structure validation of predicted structures of *Ct*CBM6A and *Ct*CBM6B was performed by analyzing the Ramachandran Plot using PROCHECK (Laskowski *et al*., 1993). Fig. 5.3.20A for *Ct*CBM6A and Fig. 5.3.20B for *Ct*CBM6B showed that nearly 90% of the residues fall in the most favoured region and no residues were found in disallowed region. The analysis of various parameters revealed that the 89.4% residues of *Ct*CBM6A and 92.9% residues of *Ct*CBM6B fall in the most favoured region (Table 5.3.4). 9.7% of *Ct*CBM6A and 6.7% residues of *Ct*CBM6B fall in generously allowed region and only 0.9% of residues in both cases were in additionally allowed region (Table 5.3.4). No residues in both the modules were found in disallowed region (Table 5.3.4). The 3-dimensional structure of *Ct*CBM6A and *Ct*CBM6B were subjected to analysis by PROCHECK G, Verify3D, Prosall (-ve) and MolProbity to check for stable conformations of phi and psi angles. The analysis was done based on the Z score which signified the standard deviations away from the mean. The Z scores of *Ct*CBM6A and *Ct*CBM6B (-1.49 and -1.30) determined by PROCHECK G (phi and psi angles) analysis showed that of the both the proteins have very stable conformation (Table 5.3.5). Similarly, Z scores calculated by Molprobity and Prosall (-ve) showed that the 3-dimensional models of *Ct*CBM6A and *Ct*CBM6B had acceptable conformation (Table 5.3.5). So, the overall assessment based on various validation programs showed clearly that the proposed model is acceptable, as described in Table 5.3.5.

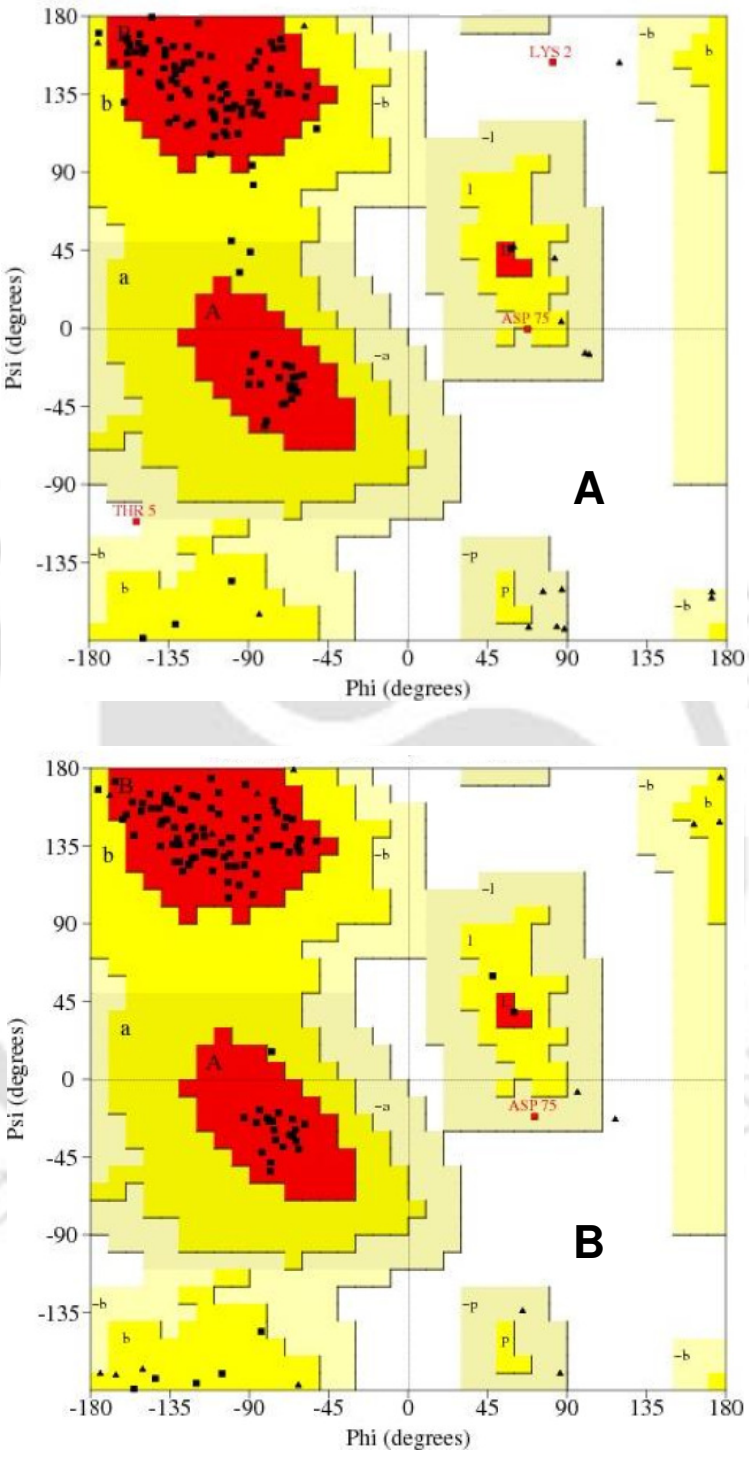


Fig. 5.3.20 Ramachandran plot of **A**) *Ct*CBM6A and **B)** *Ct*CBM6B showing that most of the residues fall in the favoured region.

Table 5.3.4 Ramachandran plot analysis of 3-dimensional of *Ct*CBM6A and *Ct*CBM6B from PROCHECK program.

CBM Type	Most favoured regions	Additionally allowed regions	Generously allowed regions	Disallowed regions
*Ct*CBM6A	89.4%	9.7%	0.9%	0.0%
*Ct*CBM6B	92.9%	6.2%	0.9%	0.0%

Table 5.3.5 Overall statistics of structure quality factors for 3-dimensional model of *Ct*CBM6A and *Ct*CBM6B.

CBM6 TYPE	*Ct*CBM6A			*Ct*CBM6B		
Parameters	Mean Score	Standard Deviation	Z Score	Mean Score	Standard Deviation	Z Score
Procheck G factor (phi/psi only)	-0.46	N/A	-1.49	-0.41	N/A	-1.30
Procheck G factor (all dihedral angles)	-0.29	N/A	-1.71	-0.28	N/A	-1.66
Verify3D	0.50	0.0000	0.64	N/A	N/A	N/A
Prosall (-ve)	0.48	0.0000	-0.70	0.47	0.0000	-0.74
Molprobity score	115.01	0.0000	-18.21	95.12	0.0000	-14.80

5.3.2.4 Hydrogen bonding plots of CtCBM6A and CtCBM6B

The hydrogen bonding plots of *Ct*CBM6A and *Ct*CBM6B displayed the characteristic patterns of secondary structure elements (Fig. 5.3.21). Both *Ct*CBM6A and *Ct*CBM6B showed many hydrogen bonds in many segments of anti-parallel beta strands, represented by strips perpendicular to the diagonal (Fig. 5.3.21A and Fig. 5.3.21B, respectively). The HB plot clearly demonstrated the fact that the structure mainly contains β-sheets which is a common feature of family 6 carbohydrate binding modules as reported in the CAZy server database (http://www.cazy.org/Carbohydrate-Binding-Modules.html). Both *Ct*CBM6A and *Ct*CBM6B showed large number of type I and type II hydrogen bonds (Fig. 5.3.21A and Fig. 5.3.21B). Both *Ct*CBM6A

and *Ct*CBM6B displayed almost equal number of parallel and anti-parallel β-strands (Fig. 5.3.21A and Fig. 5.3.21B).

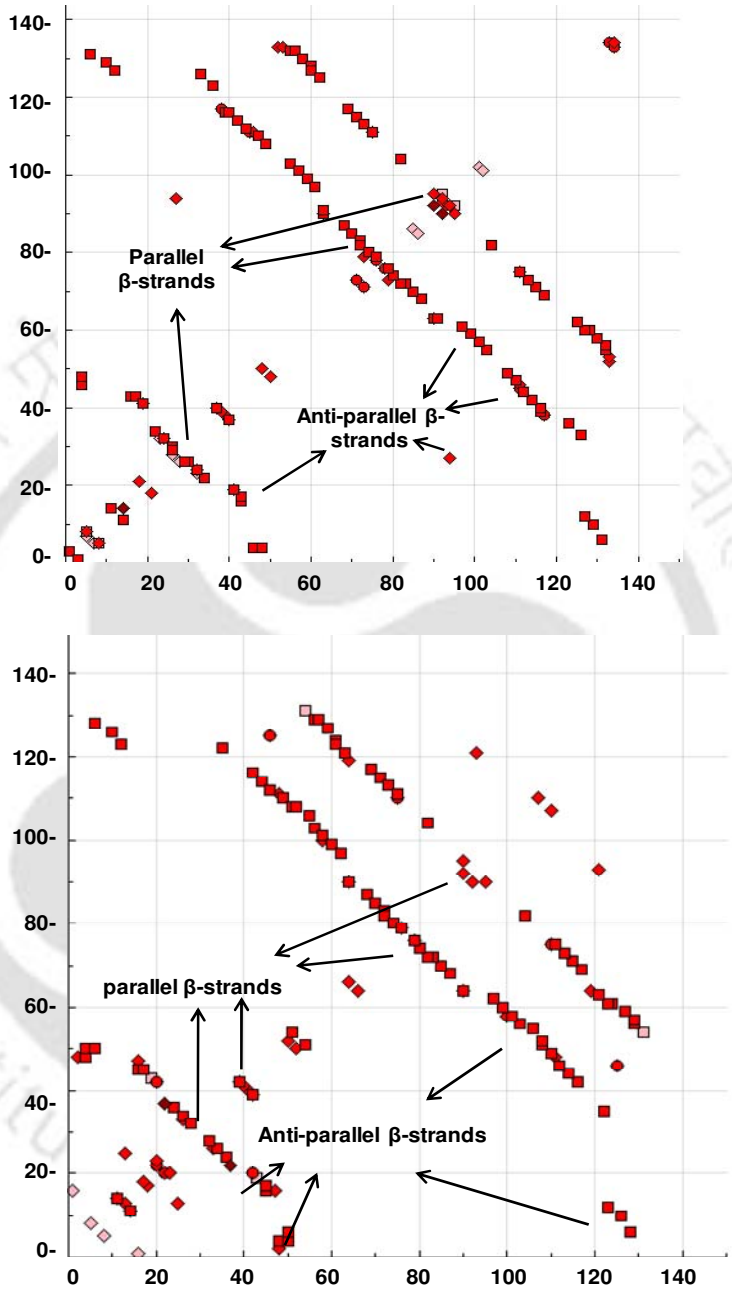


Fig. 5.3.21 Hydrogen bonding plot of **A**) *Ct*CBM6A and **B**) *Ct*CBM6B is showing anti-parallel β-strands, being represented by strips perpendicular to the diagonal. Three classes of hydrogen bonding distinguished by colour coding based on the distance between donor and acceptor molecules:
(a) Type I (brown): Short (2.5 Å between donor and acceptor),
(b) Type II (red): intermediate (between 2.5 Å and 3.2 Å) and
(c) Type III (pink): long hydrogen bonds (greater than 3.2 Å).

5.3.2.5 Ligand or catalytic binding site residues of CtCBM6A and CtCBM6B

The ligand binding sites of *Ct*CBM6A and *Ct*CBM6B were predicted using MetaPocket displayed the clefts or cavities as ligand binding regions as depicted by green patches in Fig. 5.3.21A and Fig. 5.3.21B. Gly, Asn, Phe and Ala were found at the ligand binding site of *Ct*CBM6A. Tyr, Phe, Gly and Asn residues were found to in the ligand binding site of *Ct*CBM6B. Similar residues have been also reported in the previous studies of family 6 carbohydrate binding module from *Cellvibrio mixtus* (Pires *et al,*. 2004).

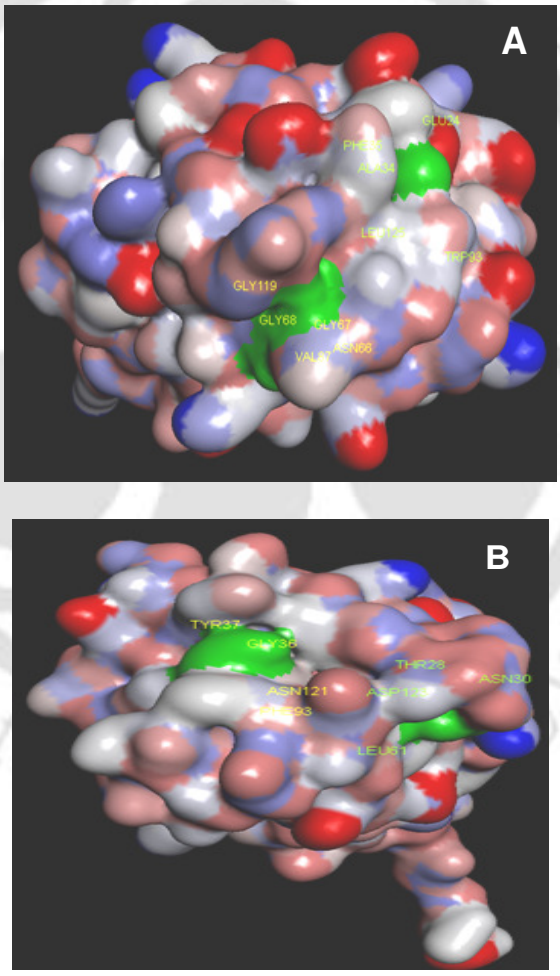


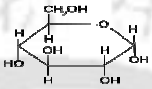
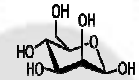
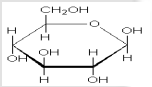
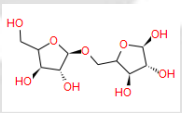
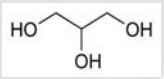
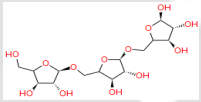
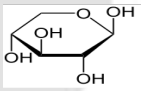
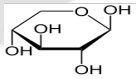
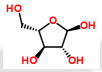
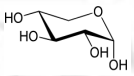
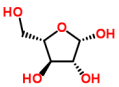
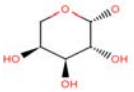
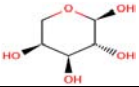
Fig. 5.3.21 MetaPocket results showing ligand binding cavities on **A)** *Ct*CBM6A showing Gly, Asn, Phe and Ala residues and **B)** *Ct*CBM6B showing Tyr, Phe, Gly and Asn residues. The green patches in the above figures represent binding cavities.

5.3.2.6 Docking analysis of CtCBM6A and CtCBM6B

The docking analysis for finding putative ligand binding sites of CBMs (*Ct*CBM6A and *Ct*CBM6B) was carried out using Accelrys Discovery studio 3.5, which utilized ZDOCK protein-ligand docking. The protein-ligand binding energy (kcal/mol) were noted with each substrate and the based on this docking scores (CDOCKER) were given (Table 5.3.6). *Ct*CBM6A (-18.487 and -14.232, kcal/mol) and *Ct*CBM6B (-14.287 and -3.43, kcal/mol) displayed maximum docking scores (kcal/mol) with arabinobiose and arabinotriose, respectively (Table 5.3.6). Significant docking scores for *Ct*CBM6A and *Ct*CBM6B, respectively, was also observed with α-D-xylopyranose (-9.384 and -10.806, kcal/mol), β-D-xylopyranose (-8.502 and -11.331, kcal/mol), β-L-arabinopyranose (-4.34 and -10.292, kcal/mol) and α-L-arabinofuranose (-3.06 and -10.04, kcal/mol). This indicated that the CBMs have affinity mainly for xylans and substituted xylans (Table 5.3.6). *Ct*CBM6B (-8.7 kcal/mol) showed slightly higher binding with D-xylose as compared to *Ct*CBM6A (-5.8 kcal/mol). These results were in agreement with the affinity gel electrophoresis results of the CBMs on native PAGE gels described in Chapter 4, Section 4.3.1, where, *Ct*CBM6B showed greater affinity (with oat spelt xylan) as compared to *Ct*CBM6A. The data obtained of *Ct*CBM6A and *Ct*CBM6B demonstrated that the binding clefts and protein surfaces of CBM6s confer the extensive range of specificities displayed by this protein family. This is in sharp contrast to other families of CBMs where variation in specificity between different members is conferred by differences in the topology of a single binding site. The range of ligand recognition observed in CBM6 is the result of variation in the location of the ligand binding site in different members of this family (Abbott *et al.*, 2009). Even though CBM6s share a

common ancestor (high structural similarity) and yet they have ligand-binding sites in different locations on the protein scaffold.

Table 5.3.6 Molecular docking scores for *Ct*CBM6A and *Ct*CBM6B obtained using Discovery Studio 3.5.

S. No.	Ligand Name	Ligand Structure	CBM6-Ligand Binding-Energy (kcal/mol)	
			*Ct*CBM6A	*Ct*CBM6B
1	α-glucose		9.393	9.158
2	d-mannose		9.648	11.153
3	β-glucose		7.247	12.254
4	Arabinobiose		-18.487	-14.287
5	Glycerol		-14.919	-6.403
6	Arabinotriose		-14.232	-3.433
7	α-D-xylopyranose		-9.384	-10.806
8	β-D-xylopyranose		-8.502	-11.331
9	α-L-arabinofuranose		-6.012	-8.751
10	D-xylose		-5.804	-8.789
11	β-L-arabinofuranose		-4.799	-12.576
12	β-L-arabinopyranose		-4.340	-10.292
13	α-L-arabinopyranose		-3.064	-10.043

5.3.3 Structural analysis of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B by Circular Dichroism

5.3.3.1 Secondary structure analysis of CtGH43 by Circular Dichroism

The protein solutions (15 µM) were analyzed on a spectropolarimeter and the percentage of α-helices, β-sheets and random coils present in *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B structures were calculated using K2d software. The analysis of CD spectra of *Ct*GH43 for detecting the secondary structural elements was based on the previous reports of CD spectra of proteins by Kelly *et al.* (2005). The comparison of CD spectra of *Ct*GH43 with the reports of Kelly *et al.* (2005) showed that it mostly contained β-sheets and coils (Fig. 5.3.22). The CD spectra *Ct*GH43 was analyzed using K2d as described by Andrade *et al.* (1993) revealed that it contains 48% β-sheets, 49% random coils and only 3% α-helices (Table 5.3.7).

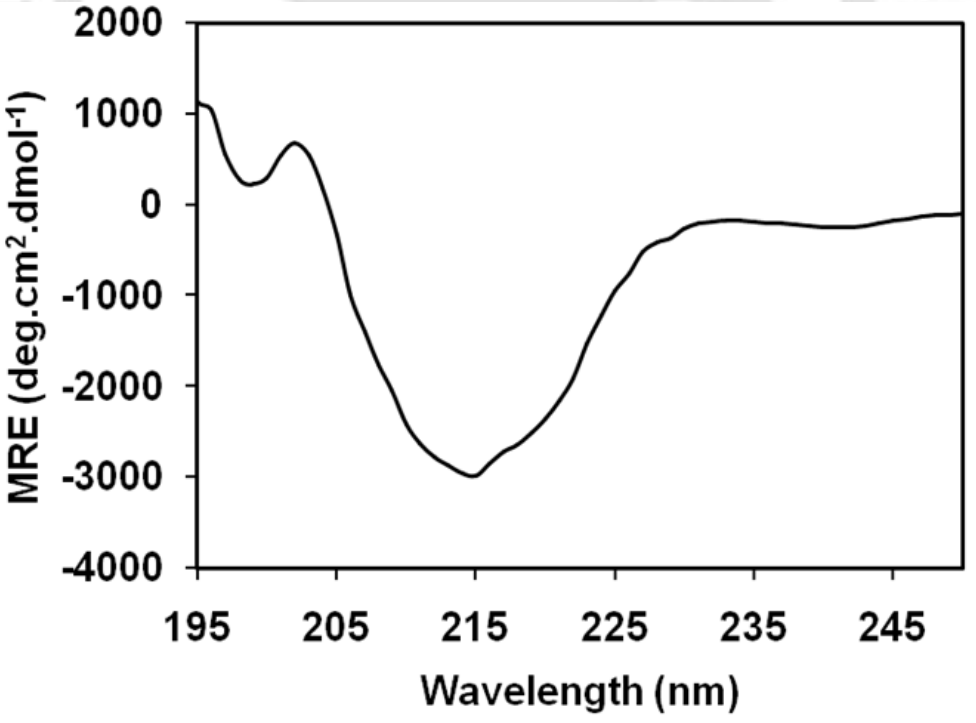


Fig. 5.3.22 Far-UV CD spectra of *Ct*GH43 (15 µM) from *Clostridium thermocellum* in 20 mM sodium phosphate buffer, pH 7.0.

The secondary structure prediction via PSI-PRED VIEW (as described in Section 5.3.1.3) did not show any α-helix like CD spectra. The dominance of β-sheets and random coils was observed in both cases indicated that the prediction was in agreement with the CD spectra (Table 5.3.7). The presence of high percentages of β-sheets and random coils are in agreement with the three dimensional structure prediction of *Ct*GH43, which also revealed 5-fold β-propeller architecture as described in Section 5.3.1.4.

Table 5.3.7 The percentage of secondary structure contents of *Ct*GH43 protein as estimated from far-UV CD spectra

Secondary structure contents of *Ct*GH43	Percentage (%) by CD analysis	Percentage (%) by PSIPRED VIEW*
α-helix	03	00
β-sheet	48	51
Random Coil	49	49

* *Secondary structure prediction using PSIPRED VIEW software.*

5.3.3.2 Circular Dichroism analysis of CtCBM6A

*Ct*CBM6B showed a CD spectra corresponding to α-helices, β-sheets and random coils (Fig. 5.3.23) when compared with CD spectra of proteins reported by Kelly *et al*. (2005) and Creighton (1997). The analysis of CD spectra of *Ct*CBM6B by K2d software as described by Andrade *et al.* (1993) showed 25% α-helices, 30% β-sheets and 45% random coils (Table 5.3.8). The secondary structure predictions using PSIPRED VIEW and Phyre servers (as described in Section 5.3.2.2) showed β-sheets (48%) and random coils (52%) but no α-helix. However, Phyre server prediction showed 4.4 % α-helices. The difference in CD analysis results and prediction method might be because the recombinant proteins never folds as well as the native protein (de Marco *et al.*, 2005). Also the predictions are based on the available database of

three dimensional structures which do not contain huge number of solved crystal structure of family 6 CBMs and moreover, the accuracy of these predictions is about 77%.

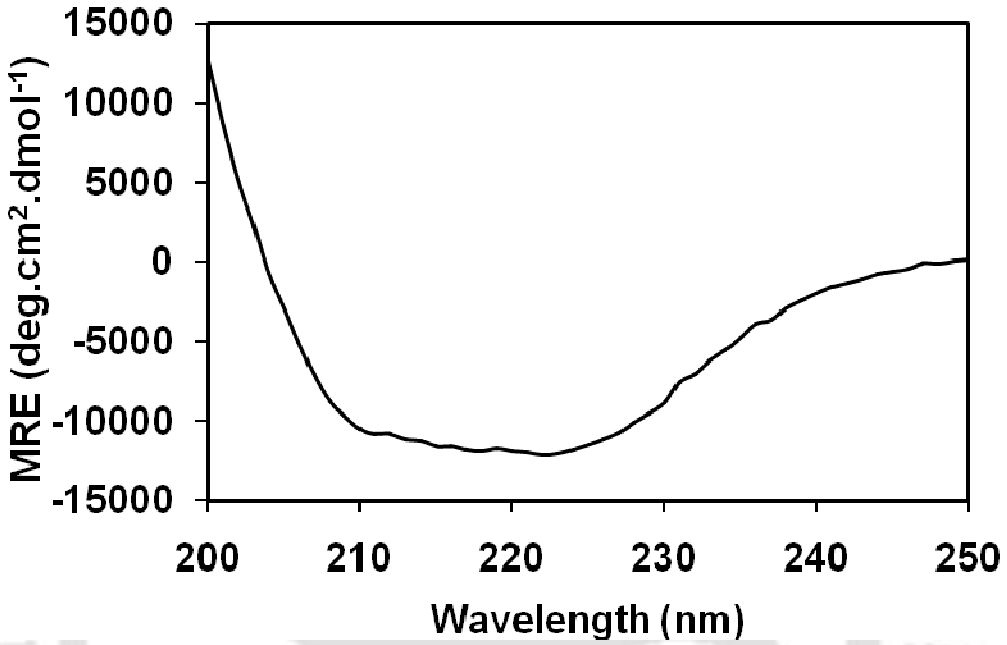


Fig. 5.3.23 Far-UV CD spectra of *Ct*CBM6A (15 µM) from *Clostridium thermocellum* in 20 mM sodium phosphate buffer, pH 7.0.

Table 5.3.8 The percentage of secondary structure contents of *Ct*CBM6A protein as estimated from far-UV CD spectra

Secondary structure contents of *Ct*CBM6A	Percentage (%) by CD analysis	Percentage (%) by PSIPRED VIEW*	Percentage (%) by Phyre**
α-helix	25	00	4.4
β-sheet	30	48	46.6
Random Coil	45	52	49

** Secondary structure prediction using PSIPRED VIEW software.*

*** Secondary structure prediction by jnet from Phyre server (http://www.sbg.bio.ac.uk/phyre).*

5.3.3.3 Circular Dichroism analysis of CtCBM6B

*Ct*CBM6B also showed a CD spectra corresponding to only β-sheets and random coils (Fig. 5.3.24) when compared with CD spectra of proteins reported by Kelly *et al*. (2005) and Creighton (1997). The analysis of CD spectra of *Ct*CBM6B by K2d software as described by Andrade *et al.* (1993) showed 44% β-sheets, 54% random coils and only 1% α-helix (Table 5.3.9). The secondary structure prediction using PSIPRED VIEW (as described in Section 5.3.2.2) showed β-sheets (48%) and random coils (52%) but no α-helix.

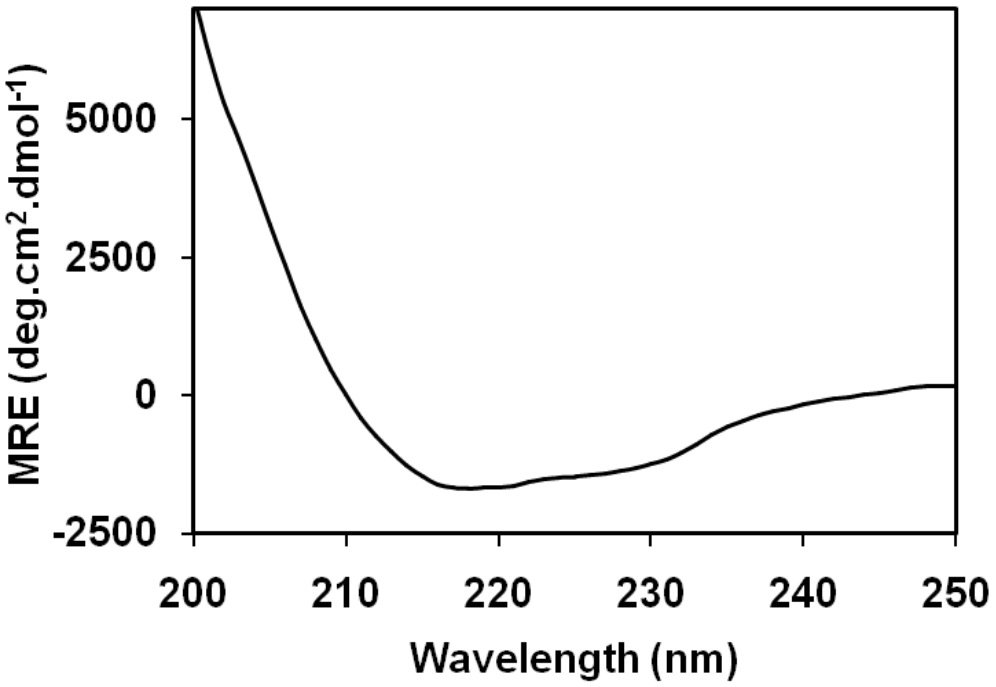


Fig. 5.3.24 Far-UV CD spectra of *Ct*CBM6B (15 μM) from *Clostridium thermocellum* in 20 mM sodium phosphate buffer, pH 7.0.

The percentage of β-sheets determined by CD analysis and predicted by PSIPRED VIEW were (44% and 48%, respectively) and random coils were (54% and 52%, respectively) nearly similar (Table 5.3.9). The three dimensional structure of

*Ct*CBM6B predicted by MODELLER was found to contain β-sandwich architecture as described in 5.3.2.3. The CD spectral data also corroborates the above statement.

Table 5.3.9 The percentage of secondary structure contents of *Ct*CBM6B protein as estimated from far-UV CD spectra.

Secondary structure contents of *Ct*CBM6B	Percentage (%) CD analysis	Percentage (%) by PSIPRED VIEW*
α-helix	01	00
β-sheet	44	48
Random Coil	54	52

** Secondary structure prediction using PSIPRED VIEW software.*

5.4 Conclusions

The *in silico* analyses of putative ligand binding sites of *Ct*GH43, *Ct*CBM6A and *Ct*CBM6B from *Clostridium thermocellum* was carried out. The phylogenetic analysis of *Ct*GH43 showed close evolutionary relationship with family 6 Carbohydrate Binding Modules (CBM6) from *Clostridium thermocellum* JW20, *Clostridium Cellulolyticum* H10, *Clostridium papyrosolvens, Clostridium cellulyticum* and *Acetovibrio cellulyticus*. The comparative modeling of family 43 glycoside hydrolase (*Ct*GH43) from *Clostridium thermocellum* was performed based on crystal structures of arabinoxylan arabinofuranohydrolases (*Bacillus subtillis*, 3C7F), β-1,4-xylosidase (*Bacillus subtillis*, 1YIF), xylan β-1,4-xylosidase (*Bacillus halodurans*, 1YRZ), β-D-xylosidase (*Geobacillus stearothermophilus*, 2EXH) and endo-1,5-α-L-arabinanase (*Bacillus thermodenitrificans* TS-3, 1WL7) from PDB. The secondary structure analysis of *Ct*GH43 established that in most of the sequences, β-sheets and coils dominated followed by alpha helix, beta strands and beta turns. The structure having lowest MODELLER objective function was selected. The 3-D structure revealed typical 5-fold β–propeller. Energy minimization and validation with Verify3D indicated acceptability of the predicted 3-D structure. The Ramachandran plot analysis by RAMPAGE confirmed that family 43 glycoside hydrolase (*Ct*GH43) contains less or negligible segments of helices. It also showed that out of 301 residues, 267 (89.3%) were in most favoured region, 23 (7.7%) were in allowed region and 9 (3.0%) were in outlier region. MolProbity analysis of Ramachandran also confirmed that the predicted 3-D structure of *Ct*GH43 is acceptable. The hydrogen bonding plot of *Ct*GH43 showed that mostly, type I and type II hydrogen bonds are present, with the high occurrence of parallel β-sheets. The ligand binding analysis by QSiteFinder and THEMATICS gave the active site

residues and the regions. Molecular Virtual docker (MVD), which showed a catalytic triad, where, one Asp acts as the catalytic general base, the Glu acts as the catalytic general acid and the other Asp is responsible for pKa modulation. IUPred analysis of *Ct*GH43 showed no disordered region. The best docking scores were obtained with xylotetraose as the ligand which gave a clue that the enzyme catalyzes xylans and related structures.

The sequence analyses and structure prediction of *Ct*CBM6A and *Ct*CBM6B were carried out. The secondary structures of *Ct*CBM6A and *Ct*CBM6B showed that they are mainly composed of β-strands. The 3-dimensional structures of *Ct*CBM6A and *Ct*CBM6B were generated using *Clostridium thermocellum* (1GMM), *Celvibrio mixtus* (1UXX), *Clostridium cellulolyticum* (2v4v) and *Clostridium stercorarium* (1UY3) as template crystal structures from PDB database. The predicted structure of *Ct*CBM6A and *Ct*CBM6B revealed typical β-sandwich structure, common to family 6 CBMs. Structure validations was done using an array of validation programs, which involved mainly the plotting and analysis of the Ramachandran Plot and other parameters. The final outcome of the above analysis confirmed that the 3-D models of *Ct*CBM6A and *Ct*CBM6B were acceptable and had stable conformation. The HB plot of *Ct*CBM6A and *Ct*CBM6B displayed equal number of parallel and anti-parallel β-sheets and majorly type I and type II hydrogen bonds. The docking energies analysis for identifying the putative ligand binding residues indicated that the binding of both CBMs is stronger with xylose-based ligands such as xylopyranose, arabinopyranose and arabinofuranose. However, *Ct*CBM6B showed better docking scores with arabinopyranose and arabinofuranose sugars than *Ct*CBM6A. The docking results of CBM6B were in agreement with experimental results of affinity gel electrophoresis

result with oat spelt xylan which contains xylopyranose and arabinofuranose sugar residues.

The CD analysis confirmed the fact that β-sheets and random coils were the main secondary structural elements present in the recombinant protein *Ct*GH43. Only 3% α-helices were present in the *Ct*GH43 structure. The CD spectra of *Ct*CBM6B showed only 1% of α-helix while most of the secondary structure consisted of random coils and β-sheets. The results of CD analysis of both *Ct*GH43 and *Ct*CBM6B were in agreement with the secondary structure prediction of the same protein using PSIPRED VIEW software with slight difference in number of α-helices. These observations confirm the fact the secondary structural elements of *Ct*GH43 and *Ct*CBM6B are mostly made up of β-sheets and random coils and very less number of α-helices, whereas, *Ct*CBM6A showed appreciable α-helical content.

References

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

Baker, E.N., Hubbard, R.E. (1984) Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44, 97-179.

Bayer, E.A. Kenig, R. & Lamed, R. (1983) Adherence of *Clostridium thermocellum* to cellulose, J. Bacteriol. 156, 828-836.

Bernstein, F.C. Koetzle, T.F. and Williams G.J.B. (1977) The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112, 535-542.

Berman, H.M. Westbrook, J. and Feng, Z. (2000) The protein data bank. Nucleic Acids Res. 28, 235-242.

Berman, H.M., Henrick, K., Nakamura, H., Markley, J., Bourne, P.E. and Westbrook, J. (2007) Realism about PDB. Nat. Biotechnol. 25, 845 - 846.

Bikadi, Z., Demko, L. and Hazai, E. (2007) Functional and structural characterization of a protein based on analysis of its hydrogen bonding network by hydrogen bonding plot. Arch. Biochem. Biophys. 461, 225-234.

Branden, C. and Tooze, J. (1999) In *Introduction to protein structure* (2nd ed). Garland Publishing: New York, NY.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (2004) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4, 187-217.

Case, D.A., Cheatham, T.E., Darden, T. III., Gohlke, H., Luo, R., Merz, K.M. Jr., Onufriev, A., Simmerling, C., Wang, C. and Woods, R. (2005) The Amber biomolecular simulation programs. J. Computat. Chem. 26, 1668-1688.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31, 3497-3500.

Corrall, O.L. and Ortega, F.V. (2006) In *Xylanases in advances in agricultural and food biotechnology*. (eds. Guevara-González, R.G. and Torres-Pacheco, I.), Research Signpost, Trivandrum, Kerala, 305-322.

Davies, G. and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure, 3, 853-857.

Demain, A.L. Newcomb, M. and Wu, J.H. (2005) Cellulase, clostridia, and ethanol. Microbiol. Mol. Biol. Rev. 69, 124-154.

Davis, I.W., Murray, L.W., Richardson, J.S., Richardson, D.C. (2010) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. 32, W615-619.

Dayhoff, M.O., Schwartz, R. and Orcutt, B.C. (1978) A model of Evolutionary Change in Proteins. In *Atlas of protein sequence and structure* (vol 5, supplement 3 ed.). National Biomedical Research Foundation, 345–358.

de Marco, A., Vigh, L., Diamant, S. and Goloubinoff, P. (2005) Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid or benzyl alcohol-over expressed molecular chaperones. Cell Stress Chaperones, 10, 329-339.

Dosztányi, Z. Csizmók, V. Tompa, P. and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics, 21**,** 3433-3434.

Eswar, N. Marti-Renom, M.A., Webb, B., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U. and Sali, A. (2006) Comparative protein structure modeling with MODELLER. In *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., 15, 5.6.1-5.6.30.

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

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

Glaser, F., Morris, R.J., Najmanovich, R.J., Laskowski, R.A. and Thornton, J.M. (2006) A method for localizing ligand binding pockets in protein structures. Proteins, 62, 479-488.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis, 18, 2714-2723.

Hendlich, M., Rippmann, F. and Barnickel, G. (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J. Mol. Graph. Model. 15, 359-363, 389.

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

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

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

Huang, B. (2009) metaPocket: a meta approach to improve protein ligand binding site prediction. Omics, 13, 325-330.

Johnson, E.A. Sakajoh, M. Halliwell, G. Madia, A. and Demain, A.L. (1982) Saccharification of complex cellulosic substrates by cellulase system from *Clostridium thermocellum*. Appl. Environ. Microbiol. 43, 1125–1132.

Karplus, M. (1983) CHARMM: A Program for macromolecular energy, minimization, and dynamics calculations, J. Comp. Chem. 4, 187-217.

Krogh, A. Brown, M. Mian, I.S. Sjolander, K. and Haussler, D. (1994) Hidden Markov models in computational biology: Applications to protein modeling. J. Mol. Biol. 235, 1501-1531.

Lairson, L.L. Henrissat, B. Davies, G.J. and Withers, S.G. (2008) Glycosyltransferases: structures, functions and mechanisms. Annu. Rev. Biochem. 77, 521-555.

Lamed, R. Setter, E. and Bayer, E.A. (1983) Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*, J. Bacteriol. 156, 828-836.

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

Laurie, A.T. and, Jackson, R.M. (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics, 21, 1908-1916.

Lovell, S.C., Davis, I.W., Arendall, W.B., III., de Bakker, P.I.W. Word, J.M., Prisant, M.G., Richardson, J.S. and Richardson, D.C. (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins, 50, 437-450.

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

Lynd, L.R. Weimer, P.J. van Zyl, W.H. and Pretorious, I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. 66, 506-577.

Martin, A.C.R. (2004) Comparative modelling. In *Bioinformatics: Genes, proteins and computers* (eds. Orengo CA, Jones DT, Thornton JM), New York, BIOS Scientific Publishers, 135-150.

Marti-Renom, M.A., Stuart, A. Fiser, A. Sánchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325.

Marchler-Bauer, A. Lu, S. and Anderson, J.B. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res., 39, 225-229.

McDonald, I.K. and Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. 238, 777-793.

McGuffin, L.J., Bryson, K. and Jones, D.T. (1999) The PSIPRED protein structure prediction server. BBSRC, CCP11 Newsletter, 8.

McGuffin, L.J., and Jones, D.T. (2003) Improvement of the GenTHREADER method for genomic fold recognition. Bioinformatics, 19, 874-881.

Ondrechen, M.J., Clifton, J.G. and Ringe, D. (2001) THEMATICS: A simple computational predictor of enzyme function from structure. Proc. Nat. Acad. Sci. (USA), 98, 12473-12478.

Pierce, B. and Weng, Z. (2007) ZRANK: re-ranking protein docking predictions with an optimized energy function. Proteins, 67, 1078-1086.

Pierce, B.G., Hourai, Y. and Weng, Z. (2011) Accelerating protein docking in ZDOCK using an advanced 3D convolution library. PLoS One, 6, e24657.

Ramachandran, G.N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. Adv. Protein Chem. 23, 283-437.

Ramachandran, G.N. Ramakrishnan, C. and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 95-99.

Ramachandran, G.N. Sasisekharan, V. and Ramakrishnan, C. (1966) Molecular structure of polyglycine II. Biochemi. Biophys. Acta. 112, 168-170.

Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. 202, 865-884.

Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815.

Sánchez, R. and Sali, A. (2000) Comparative protein structure modeling: In *Introduction and practical examples with MODELLER in Protein Structure Prediction: Methods and Protocols*. (ed. D. M. Webster), New Jersey, Totowa, Humana Press, 97-129.

Söding, J. (2005) HHpred: Protein homology detection by HMM-HMM comparison. Bioinformatics, 21, 951-960**.**

Thomsen, R. and Christensen, M.H. (2006) MolDock: a new technique for high-accuracy molecular docking. J. Med. Chem. 49, 3315-3321.

Utas, J.E., Kritikos, M., Sandström, D., Akermark, B. (2006) Water as a hydrogen bonding bridge between a phenol and imidazole. A simple model for water binding in enzymes. Biochim. Biophys. Acta. 1757, 1592-1596.

List of Publications

1. Saprativ P. Das[†1], Rajeev Ravindran[†1], **Shadab Ahmed**[†1], Debasish Das, Dinesh Goyal, Carlos M.G.A. Fontes and Arun Goyal (**2012**) Bioethanol production involving recombinant *C. thermocellum* hydrolytic hemicellulase and fermentative microbes. *Applied Biochemistry and Biotechnology*. 167, 1475-1488. ([†] equal contribution).
2. **Shadab Ahmed**, Rahul Charan, Arabinda Ghosh and Arun Goyal (**2012**) Comparative modeling and ligand binding site prediction of a family 43 glycoside hydrolase from *Clostridium thermocellum. Journal of Proteins and Proteomics* 3(1) 31-38.
3. **Shadab Ahmed**, Tushar Saraf and Arun Goyal (**2009**) Prediction of catalytic and ligand binding sites and hydrogen bonding plot from protein sequence of family 39 glycoside hydrolase *(Ct*GH39) from *Clostridium thermocellum. Journal of Applied Bioscience and Biotechnology*. 3, 633-640.
4. **Shadab Ahmed**, Tushar Saraf and Arun Goyal (**2009**) Homology modeling of family 39 glycoside hydrolase from *Clostridium thermocellum. Current Trends in Biotechnology and Pharmacy* 3, 210-218.
5. **Shadab Ahmed**, Sangeeta Bharali, Ravi Kiran Purama, Avishek Majumder, Carlos M.G.A. Fontes and Arun Goyal (**2009**) Structural and biochemical properties of Lichenase from *Clostridium thermocellum. Indian Journal of Microbiology* 49, 72-76.
6. **Shadab Ahmed**, Deepmoni Deka, M. Jawed, Dinesh Goyal, Carlos M.G.A. Fontes and Arun Goyal (**2009**) Biochemical characterization of a recombinant derivative (*Ct*Lic26A-Cel5) of a cellulosomal cellulase from *Clostridium thermocellum. Current Trends in Biotechnology and Pharmacy* 3, 633-640.
7. **Shadab Ahmed**, Carlos M.G.A. Fontes and Arun Goyal (**2008**) Recombinant Lichenase from *Clostridium thermocellum* binds glucomannan but not to lichenan: Analysis by affinity electrophoresis. *Annals of Microbiology*, 58(4) 723-725.
8. Deepmoni Deka, **Shadab Ahmed**, Nadeem Akhtar, Sangeeta Bharali, M. Jawed, Carlos M.G.A. Fontes, Dinesh Goyal and Arun Goyal (**2009**) Determining substrate specificity and biochemical characterization of a full length recombinant cellulase (Lic26A-Cel5-CBM11) of *Clostridium thermocellum. Journal of Applied Bioscience and Biotechnology.* 5, 13-18.
9. Ravi Kiran Purama, Mayur Agrawal, Avishek Majumder, **Shadab Ahmed** and Arun Goyal (**2008**) Antibiotic sensitivity, carbohydrate fermentation and plasmid profiles of glucansucrase producing four *Leuconostoc* strains. *Journal of Pure and Applied Microbiology* 2, 139-146.

To be communicated

1. **Shadab Ahmed**, Arabinda Ghosh, Carlos M.G.A. Fontes, and Arun Goyal (**2012**) A novel α-L-arabinofuranosidase of family 43 glycoside hydrolase (GH43) from Clostridium thermocellum active over a wide range of α-L-arabinosyl substituted xylans.

Conferences/Symposia/Meetings

1. **Shadab Ahmed**, Arabinda Ghosh, Carlos M.G.A. Fontes and Arun Goyal (**2011**) Biochemical characterization of family 43 glycoside hydrolase (GH43F) and its truncated derivative (GH43) from *Clostridium thermocellum*. (8th New Horizons in Biotechnology (Annual convention of Convention of Biotech Research Society of India), November 21-24, NIIST (CSIR), Trivandrum, India, IB111.

2. **Shadab Ahmed**, Carlos M.G.A. Fontes and Arun Goyal (**2011**) Cloning and substrate binding studies on Carbohydrate Binding Modules (CBM6A and CBM6B) from *Clostridium thermocellum.* (8th New Horizons in Biotechnology (Annual convention of Convention of Biotech Research Society of India), November 21-24, NIIST (CSIR), Trivandrum, India, IB112.

3. **Shadab Ahmed**, Arabinda Ghosh, Carlos MGA Fontes and Arun Goyal (**2011**) Cloning, expression, purification and biochemical characterization of family 43 glycoside hydrolase (GH43) from *Clostridium thermocellum*. 9th Carbohydrate Bioengineering Meeting (CBM9), Lisbon, Portugal.

4. Arabinda Ghosh, **Shadab Ahmed**, Anil Kumar Verma, Carlos MGA 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 (CBM9), Lisbon, Portugal.

5. **Shadab Ahmed**, Arabinda Ghosh, Arun Goyal (**2010**) Cloning and expression of family 43 glycoside hydrolase from *Clostridium thermocellum.* International Symposium on Recent Advances in Cross-Disciplinary Microbiology: Avenues & International Workshop on rRNA Sequencing, Phylogeny & next generation Genome Sequencing (51th Annual Conference of Association of Microbiologists of India), December 14-17, Birla Institute of Technology, Mesra, Ranchi, India.

6. **Shadab Ahmed**, Vikas Gupta, Carlos M.G.A. Fontes and Arun Goyal (**2009**) Investigating the 3-dimensional structure of family 43 glycoside hydrolase (*Ct*GH43), a cellulase from *Clostrdium thermocellum* structure for possible interactions using

molecular docking and other bioinformatics tools. IIIrd International Conference on Environmental, Industrial and Applied Microbiology (BioMicroWorld 2009), Dec 2-4, University of Lisbon, Lisbon, Portugal, pp 409.

7. **Shadab Ahmed**, Rahul Charan and Arun Goyal (**2009**) Prediction of 3-D structure, catalytic and ligand binding sites of family 43 glycoside hydrolase (GH43) from *Clostridium thermocellum.* 8th Carbohydrate Bioengineering Meeting, May, 10-13, Ischia Island, Naples, Italy, p11.

8. **Shadab Ahmed**, Rahul Charan and Arun Goyal (**2008**) Prediction of 3-Dimensional structure of family 43 glycoside hydrolase from *Clostridium thermocellum* by sequence homology based modelling. International Symposium on Microbial Biotechnology: Diversity, Genomics and Metagenomics, Dec 18-20, University of Delhi, India. p7-8.

9. **Shadab Ahmed**, Rishikesh Shukla and Arun Goyal (**2008**) Purification and biochemical characterization of bi-functional recombinant derivative (Lic26A-GH5) of cellulosomal cellulase from *Clostridium thermocellum*. International Congress on Bioprocesses in Food Industries (5th Convention of Biotech Research Society of India, November 6-8, Osmania University, Hyderabad, India. p172.

VITAE

The author was born on December 15, 1979 in Adra, District Purulia (West Bengal). He passed the Secondary School Examination (Madhyamik Pariksha) conducted by West Bengal Board of Secondary Education (10th Class), West Bengal in 1996, and All India Senior Secondary School Certificate Examination (AISSCE) conducted by Central Board of Secondary Education (12th Class), New Delhi in 1998. He completed B.Pharm. from J.S.S. College of Pharmacy affiliated to Rajiv Gandhi University of Health Sciences, Bengalore (Karnataka) in October, 2003. He did his M.Tech. (Biotechnology) from Vellore Institute of Technology, Vellore, in 2006. In 2007 (March) he joined as Senior Research Fellow in the Department of Biotechnology under Prof. Arun Goyal in a CSIR funded project.

Mr. Shadab Ahmed joined the Ph.D. program in January 2008 at the Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati 781 039, Assam, India. He successfully completed the course work with 9.25/10 CPI (highest in his batch). He received Institute Fellowship (IIT Guwahati) from Jan 2008 to Dec 2011, under the scheme run by the Ministry of Human Resource and Development (MHRD), New Delhi. From Jan 2012 onwards he received Maulana Azad National Fellowship from University Grants Commission, New Delhi. He has given the open (PhD Synopsis) Seminar on August 24, 2012 and presented his thesis work before the Doctoral Committee and his performance was satisfactory. He submitted the PhD thesis in September 2012. Presently, he is continuing as Senior Research Fellow in the same Department.