SYNOPSIS

Plant cell wall is predominantly composed of complex structural polysaccharides such as celluloses and hemicelluloses. Polysaccharides of the primary cell wall components of plant are cellulose, hemicelluloses such as xyloglucans, mannans, galactomannans, glucomannans, laminarin, glucuronoarabinoxylans, arabinoxylan etc. Mannans are the polysaccharides with a backbone chain of β-(1→4)-linked mannose units. They constitute a major portion of hemicelluloses in hardwoods. The major distribution of mannan polysaccharides substituted with galactose and glucose units in plant hemicellulose reservoir is abundant in nature. Carob galactomannan from *Ceratonia siliqua* plant contains β-(1→4)-D-mannan backbone (78%) and α-(1→6)-linked galactose (22%) single units, whereas guar gum (from endosperm of guar seeds) backbone is a linear chain of β-(1→4)-linked mannose residues to which galactose residues are (1→6)-linked at every second mannose forming short side-branches. Glucomannan from *Amorphophallus konjac* is a water-soluble polysaccharide that is considered as a dietary fiber. The component sugars units in konjac glucomannan are β-(1→4)-linked D-mannose and D-glucose residues in a molar ratio of 1.6:1. β-D-mannanase (endo β-(1→4)-mannan mannohydrolase, E.C. 3.2.1.78) hydrolyzes β-(1→4)-D-mannopyranosyl linkages within the main chain of mannans, glucomannans, galactomannans and galactoglucomannans. Mannanases have been listed within glycoside hydrolase (GH) families *viz.* GH26, GH5 and GH113 in carbohydrate active enzyme database (http://www.cazy.org/Glycoside-Hydrolases.html) based on sequence similarity (Henrissat, 1991; Davis and Henrissat, 2002). β-D-mannanases belong to family GH5,

GH26, GH113 display a $(\beta/\alpha)_8$ barrel shaped protein folding pattern and the acid-base-assisted catalysis via a double displacement mechanism involving a covalent glycosyl-enzyme intermediate. The mechanism of glycosidic bond cleavage is found conserved within these families. These are the characteristic patterns of clan GH-A protein families and helped β-D-mannanases of family GH5, GH26 and GH113 to group into this clan. Due to the retaining double displacement mechanism, these enzymes can perform transglycosylation. Transglycosylation may leads to the synthesis of new glycosides or oligosaccharides longer than the original substrate. GH5 and GH113 mannanases have been described as able to catalyse transglycosylation reactions while to date no evidence of transglycosylation has been reported for GH26 mannanases. The non-catalytic polysaccharides recognizing modules were initially named cellulose binding domain. Later on the term CBM (carbohydrate binding module) evolved reflecting the diversity in ligand specificity of these proteins. Currently, CBMs have been classified into 67 distinct families based on sequence similarity (http://www.cazy.org/Carbohydrate-Binding-Module) and they display substantial diversity in terms of ligand specificity. The CBMs append to the glycoside hydrolases that degrade the soluble or insoluble polysaccharides. Family 35 carbohydrate binding module is often appended to glycoside hydrolase family 26 (GH26) and GH5 mannanases, xylanases (GH30) which significantly alter the polysaccharide specificity for plant cell wall polysaccharides such as galactomannan, glucomannan, mannan and glucouronoxylan. Out of 67 distinct families, CBMs are also included within 7 different fold families. The majority of the folded families are comprised of β-sandwich fold. Refinement of the CBM classification was improvised recently where Type A CBMs are those which can recognize the surface of the

crystalline polysaccharides and Type B CBMs are classified according to their endo acting mode of action internally to glycan chains. Type C CBMs are attributed by their exo-type binding either at the side chain or at the polysaccharide termini. Usually Type B CBMs are abundant in cellulases, xylanases and mannanases that bind to cellulose, xylan and mannan, respectively. Thus CBMs maintain the target substrate within the close proximity of their complete protein structures. In general, most family 35 carbohydrate binding module (CBM35) acquired a type B conformation in nature.

Present work

The present investigations are carried out on the "**Molecular cloning, expression, purification and biochemical characterization of β-mannanase (*Ct*Manf) of family 26 Glycoside Hydrolase and its truncated derivative (*Ct*ManT) and associated family 35 carbohydrate binding module (*Ct*CBM35) from *Clostridium thermocellum***". In the present study molecular characterization of endo-β-mannanae (*Ct*Manf) of family 26 glycoside hydrolase (GH26) and the associated carbohydrate binding module of family 35 (*Ct*CBM35) were carried out. The recombinant protein *Ct*Manf and its truncated derivatives *Ct*ManT, *Ct*CBM35 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. Reports of Type A having mannanse (Man26A and Man5A) activity are already available but to our knowledge this is the first report of any Type B (Man26B) having the ability to hydrolyze both mannan polysaccharides and cellulosic polysaccharides. 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 elaborately reviewed family 26 glycoside hydrolases and associated non-catalytic carbohydrate binding module (CBM), especially, CBM35. The chapter reviews about the type of core architecture seen in various family 26 GHs, their enzyme activities, substrate specificity, active site and active site residues. The various type of endo-β-mannanases belonging to family 26 GH, have been discussed and their biochemical properties have been highlighted. The various types of binding clefts or binding sites possible in CBM35 have also been elaborated. Previous literatures highlighting their ligand or polysaccharide binding capabilities have also been demonstrated. Previously characterized recombinant family 26 GHs and CBM35 from different bacteria have been illustrated. The chapter also reviews the potential applications of endo-β-mannanases.

Chapter 2 describes the full length gene *Ct*Manf and its truncated derivatives *Ct*ManT and *Ct*CBM35 were extracted from the genomic library of *Clostridium thermocellum* ATCC 27405 (nucleotide accession: CP000568.1). The molecular architecture revealed an N-terminal non catalytic family 35 carbohydrate binding module (*Ct*CBM35, 420 bp) followed by glycoside hydrolase module (*Ct*ManT, 1029

bp) at the C-terminus. PCR amplified fragment of full length gene *Ct*Manf showed a band of 1449 bp, whereas, the truncated catalytic module *Ct*ManT displayed a PCR amplified fragment of 1029 bp and *Ct*CBM35 420 bp, respectively. The above amplified products were purified from the agarose gel by gel extraction kit (Qiagen). *Ct*Manf and its truncated derivatives *Ct*ManT and *Ct*CBM35 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-28a(+) vector was also digested with *Nhe*I-*Xho*I restriction enzymes and the digested fragment of each was purified by gel extraction kit and followed by ligation.

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 kanamycin 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*Manf and its truncated derivatives *Ct*ManT and *Ct*CBM35 in pET-28a(+) vector was confirmed by restriction digestion analysis of the recombinant plasmid DNA. The *E. coli* BL-21 competent cells were transformed with recombinant plasmid DNA of full length *Ct*Manf as well as its truncated derivatives *Ct*ManT and *Ct*CBM35 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 hyperexpression 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 53 kDa (*Ct*Manf), 38 kDa (*Ct*ManT) on SDS-PAGE. The non-catalytic carbohydrate binding modules *Ct*CBM35 showed molecular sizes of 15 kDa on SDS-PAGE. The amount of protein obtained from 100 ml of cultures *Ct*Manf and its truncated derivatives *Ct*ManT, *Ct*CBM35 after IMAC purification were 0.16 mg/ml (*Ct*Manf), 0.15 mg/ml (*Ct*ManT), 0.12 mg/ml (*Ct*CBM35). These purified proteins were used for further biochemical and functional characterization.

Chapter 3 elaborates the substrate specificity and kinetic properties of catalytic modules *viz.* *Ct*Manf and *Ct*ManT from *C. thermocellum*. The catalytic modules *Ct*Manf and *Ct*ManT showed maximum enzyme activities at optimum pH 6.9 and 6.5, respectively. The optimal temperatures for *Ct*Manf and *Ct*ManT were 60°C and 50°C, respectively. *Ct*Manf and *Ct*ManT displayed more or less similar pH and thermal stability profile. Both *Ct*Manf and *Ct*ManT were stable over a pH range 5.0-7.5. The enzymes displayed significant loss of activity at 70°C or higher temperature. *Ct*Manf and *Ct*ManT showed maximum activity of 97.0±5.0 U/mg and 91.0±4.0 U/mg, respectively, with carob galactomannan. The catalytic enzymes *Ct*Manf and *Ct*ManT also displayed noticeable activities with soluble locust bean galactomannan (85.4±6.0 U/mg and 83.1±5.0 U/mg), konjac glucomannan (81.0±3.0 U/mg and 79.8±4.0 U/mg) and insoluble ivory nut mannan (50.0±2.0 U/mg and 26.5± 0.9 U/mg). Interestingly, both *Ct*Manf and *Ct*ManT displayed low but observable activity against barley β-

glucan, lichenan, carboxymethyl cellulose, hydroxyethyl cellulose, avicel and xyloglucan signifying multi-substrate specificty. The catalytic modules were unable to act upon synthetic substrates such as *p*NP-β-D-mannopyranoside and with *p*NP-α-D-mannopyranoside that ruling out any exo-enzyme activity. Therefore, both *Ct*Manf and *Ct*ManT can be considered as endo-β-mannanases. In addition the zymogram study of against both recombinant enzymes carob galactomannan confirmed their endo acting nature.

The enzyme activity of *Ct*Manf and *Ct*ManT increased significantly by approximately, 1.5 fold in presence of Ca^{2+} and Mg^{2+} salts, implying that these ions are needed as cofactors. Significant enzyme activities of both *Ct*Manf and *Ct*ManT were observed considerably at moderate concentrations of Mn^{2+}, Co^{2+} and Zn^{2+} ions. However, the enzyme activity was inhibited significantly in presence of Cu^{2+}, Al^{3+} and Ni^{2+} ions and cationic detergent SDS at their low concentrations. The enzyme activity of *Ct*Manf and *Ct*ManT, when assayed in the presence of Ca^{2+} ions, increased by two-fold, but in presence of EGTA and EDTA sharply decreased. This implied that Ca^{2+} ions might be involved in the catalysis and imparting stability to the *Ct*Manf and *Ct*ManT structures. Chaotropic agents, urea and GnHCl at higher concentrations also inactivated significantly the enzyme activity of recombinant enzymes.

Media composition played a significant role in production of recombinant *Ct*Manf and *Ct*ManT. The recombinant *Ct*Manf and *Ct*ManT showed highest cell density and concentration of protein in TY medium as compared with TB, 5xLB and LB medium. In LB medium moderate cell density was achieved with low protein concentration. The 5x LB medium did not support the growth due to higher

concentrations of yeast extract, tryptone and sodium chloride and as a result lowest protein concentration was achieved. The rich source of tryptone, yeast extract and phosphate salts facilitated to achieve highest cell densities in TY media as compared to other chemically defined media used. Higher concentration of phosphate is important for attaining high cell densities of *E. coli* harbouring *Ct*Manf and *Ct*ManT, as the lower concentrations of phosphate limits the growth. The low cell densities and lower production of recombinant proteins in LB and 5x LB medium were due to lack of buffering capacity

Protein-melting curves of full length protein *Ct*Manf showed that the catalytic module, *Ct*ManT and carbohydrate binding module, CBM35 melt independent of each other. The protein-melting peaks of *Ct*ManT and CBM35 shifted to higher temperature in the presence of Ca^{2+} ions. However, the addition of equimolar concentration of EDTA to the Ca^{2+} ion containing *Ct*Manf and *Ct*ManT, shifted back the melting temperature peaks to the original positions. This further corroborated the fact that Ca^{2+} ions impart thermal stability to the protein structures.

The structure and functional aspects of *Ct*ManT was examined by generating a model by homology modelling and validated successfully. The modelled *Ct*ManT displayed $(\beta/\alpha)_8$"TIM" barrel fold with extended loops around the active center that constructed the broader saddle-shaped, open active site cavity. The structure analysis showed that *Ct*ManT is structurally closer to the fungal mannanase as compared wth other structurally characterized bacterial mannanase available in PDB database. *Ct*ManT involved in retention mechanism where amino acid residue Glu181 acted as a catalytic acid/base, while Glu288 act as a catalytic nucleophile. Molecular docking

study with manno-configured ligands suggested wide binding site pocket accommodating larger oligosaccharides with higher affinity. 3D-structure superimposition of modeled *Ct*ManT with the template protein showed six potential subsite having two subsite towards aglycone whereas four subsite to glycon side for galactomannan.

Chapter 4 illustrates the binding analysis of *Ct*CBM35 from *Clostridium thermocellum*. The cloned family 35 Carbohydrate binding module (*Ct*CBM35) from *Clostridium thermocellum* preferred binding with manno-configured polysaccharides. *Ct*CBM35 discriminated during carbohydrate selection showing its affinity only with manno-configured ligands among the manno-, cello- and xylo-configured polysaccharides. The binding analysis of *Ct*CBM35 to soluble polysaccharides was evaluated using affinity gel electrophoresis. The equilibrium binding constant (K_a) of the *Ct*CBM35 was determined for mannotriose, konjac glucomannan, carob galactomannan and locust bean galactomannan by measuring the relative migration distance of proteins on native PAGE gels in the presence of above ligands. Both *Ct*CBM35 showed lesser affinity for carob galactomannan (for which the catalytic modules showed maximum activity) as compared to konjac glucomannan which was clear from greater retardation of *Ct*CBM35 at lower concentration of konjac glucomannan during affinity electrophoresis. Oligosaccharide and polysaccharide binding studies were also corroborated the findings of affinity electrophoresis where ligand binding was observed higher with mannotriose and konjac glucomannan as compared with carob and locust bean galactomannan. Ligand binding with *Ct*CBM35 was evidenced from the fluorescence

peack shift of 21 nm. In dynamic light scattering, the larger particle size of *Ct*CBM35 is due to the polysaccharide binding. The cationic interaction of aromatic residues with carob galactomannan and konjac glucomannan insists *Ct*CBM35 domain alteration to more compact form reducing the random diffusion of the particles between polysaccharide and amino acid residues. Due to simpler structure of oligosaccharide (mannotriose) and the polysaccharide ligand konjac glucomannan, the interaction with aromatic residues in the binding pocket of *Ct*CBM35 uphold strong binding as compared to carob galactomannan with substituted galactose side chain.

Instances of Ca^{2+} ion induced stability of *Ct*CBM35 was evident from protein melting study where the low melting temperature of the protein was shifted to higher temperature. The metal ion (Ca^{2+}) chelation by EDTA revealed the protein stability lost at very low temperature but the structure could be repaired while Ca^{2+} was added to minimize the effects of EDTA. Therefore, divalent metal cation Ca^{2+} plays significant role in holding the structure that is sensitive against temperature.

Higher temperature invoked lesser stability to *Ct*CBM35 and was evidenced from fluorescence spectrum at different temperatures. The loss in structural stability attributed to higher temperature was also hampered the ligand binding capacity of *Ct*CBM35 and isothermal titration calorimetry revealed that at the melting temperature *Ct*CBM35 lost its 100 fold ligand binding capacity. The highly substituted α-(1→6)-galactose on β-(1→4)-mannose backbone in locust bean galactomannan might be the another reason for lower binding by *Ct*CBM35. *Ct*CBM35 is sensitive towards thermal exposure and shows weak binding ability of

ligands at its melting temperature. Titrations against polysaccharides at 55°C, *Ct*CBM35 exhibited ~100 fold lower binding affinity against konjac glucomannan, locust bean galactomannan and carob galactomannan as compared to binding at 25°C. The interpretation of unfolding transition in presence of GnHCl and urea revealed that *Ct*CBM35 holds a significant structural stability and can resist denaturation at higher concentrations of these denaturants.

The 3-dimensional model of *Ct*CBM35 from *Clostridium thermocellum* generated by Modeller9v8 displayed predominance of β-sheets arranged as β-jelly-roll fold. The secondary structure of*Ct*CBM35 by PredictProtein showed presence of two α-helices (3%), twelve β-sheets (45%) and fifteen random coils (52%). Secondary structural element analysis of cloned, expressed and purified recombinant *Ct*CBM35 by Circular dichroism also corroborated the *in-silico* predicted secondary structure. Multiple sequence alignment of *Ct*CBM35 showed conserved residues (Tyr123, Gly124 and Phe125), which are commonly observed in mannan specific CBMs. The docking analysis of *Ct*CBM35 with manno-oligosaccharide displayed the involvement of Tyr26, Gln29, Asn43, Trp66, Tyr68, Leu69, Arg76 and Leu127 residues, making polar contact with the ligand molecules. Ligand docking analysis of *Ct*CBM35 exhibited higher binding affinity with mannotriose and galactomannan (Man-Gal-Man, moiety) substantiated the affinity binding and fluorescence results, displaying similar values of K_a.

Chapter 5 describes the TLC and HPAEC analysis of *Ct*Manf (full length endo-β-(1→4)-mannanase) hydrolyzed products of natural substrates. It was apparent from TLC that *Ct*Manf released mannotriose, mannobiose and mannose in prolonged

hydrolysis of carob galactomannan. But, at the earlier stages, the amount of mannobiose was less as compared to mannotriose and mannose. After complete hydrolysis of carob galactomannan *Ct*Manf was able to release mannose, mannobiose and mannotriose. The salient feature of *Ct*Manf catalysis involved only β-(1→4)-bond cleavage when mannobiose and mannotriose were used as substrate and liberated prominently mannose as main product. The release of large amount of mannose at early stage of enzymatic reaction by *Ct*Manf was seen against carob galactomannan and manno-oligosaccharides. HPAEC analysis corroborated the results of TLC analysis of hydrolysis of carob galactomannan products released by *Ct*Manf. The results of HPAEC showed that *Ct*Manf exclusively cleaves carob galactomannan into mannotriose, mannobiose and mannose. It was apparent from TLC and HPAEC analyses *Ct*Manf was able to hydrolyze only β-(1→4) bond cleavage and had potential to produce manno-oligosaccharides from carob galactomannan. The structure analysis of the pretreated copra meal revealed the efficiency of the pretreatment process resulting in significant structural deformity. FT-IR analysis showed the easy accessibility of the major functional groups of the abundant galactomannan for catalysis by *Ct*Manf. Both ^{1}H and ^{13}C NMR revealed that the polysaccharide content of the copra meal majorly galactomannan with α-(1→6)-linked galactose substituted in the β-(1→4)-linked mannose main chain. Time dependent *Ct*Manf catalyzed hydrolysis of the galactomannan of copra meal confirmed the presence of mannose, mannobiose and mannotriose using TLC, HPAEC and mass spectrum. The released manno-oligosaccharides were successfully purified using size exclusion chromatography with significant yield. The mixed treatment of mannobiose and mannotriose supported the enhanced growth of prebiotic

bacteria holding their potential as good prebiotic agent as compared to commercial inulin. The mixed manno-oligosaccharides are highly resistant against simulated gastric juice, intestinal fluid and α-amylase proving a stable prebiotic over inulin. *In vitro* cytotoxicity assay of MOS (500 μg mL^{-1}) on human epithelial colorectal adenocarcinoma cell line (HT-29) demonstrated 60% decreased viability of cells after 48h displaying anti-tumorigenic property.