JPP JOURNAL OF PROTEINS AND PROTEOMICS
4(1), June 2013, pp. 27-34

Research Article

# ANALYSIS OF STRUCTURAL ELEMENT OF FAMILY 6 CARBOHYDRATE BINDING MODULE (*CT*CBM6B) OF ALPHA-L-ARABINOFURANOSIDASE FROM *CLOSTRIDIUM THERMOCELLUM*

**Shadab Ahmed[1], Saurabh Gautam[2], Munishwar N. Gupta[2] and Arun Goyal[1*]**

[1]*Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati-781039, Assam, India*
[2]*Department of Chemistry, Indian Institute of Technology Delhi, Hauz Khas, New Delhi- 110006, India*

***Abstract:*** The amino acid sequence of a family 6 carbohydrate binding module (*Ct*CBM6B) from *Clostridium thermocellum* α-L-arabinofuranosidase showed close evolutionary relationship with some other member of family 6 carbohydrate binding modules. The CD spectrum analysis confirmed the secondary structure prediction of *Ct*CBM6B as both showed β-sheets (44-48%) and random coils (52-54%) and no α-helix. The hydrogen bonding plot of *Ct*CBM6B showed many segments of parallel and anti-parallel β-strands which was similar to the secondary structure prediction by PSIPRED VIEW. The three dimensional structure of *Ct*CBM6B generated by MODELLER revealed a typical β-sandwich architecture at its core, characteristic of β-jelly roll CBM superfamily. The Ramachandran plot analysis by PROCHECK showed that out of 134 residues, 92.9% were in most favoured region, 6.2% in additionally allowed region and only 0.9% in generously allowed region which indicated a stable conformation of 3D model of *Ct*CBM6B. The docking analysis of *Ct*CBM6B for finding putative ligand binding sites showed that it has high binding affinity for arabinobiose, β-L-arabinofuranose and β-D-xylopyranose indicated by lower ligand binding energy (-14.28 kcal $mol^{-1}$, -12.5 kcal $mol^{-1}$ and -11.3 kcal $mol^{-1}$, respectively). *Ct*CBM6B also showed appreciable binding affinity with α-D-xylopyranose (–10.8 kcal $mol^{-1}$), β-L-arabinopyranose (–10.2 kcal $mol^{-1}$), α-L-arabinopyranose (–10.0 kcal $mol^{-1}$) and α-L-arabinofuranose (–8.75 kcal $mol^{-1}$). The results indicated that *Ct*CBM6B has high potential for binding arabinan, xylans and substituted xylans.

***Keywords***: *Ct*CBM6B; homology modeling; docking; arabinofuranose; arabinopyranose.

## Introduction

Family 6 carbohydrate binding module (CBMs) are different from other CBM families in that, these modules are known to contain multiple distinct ligand binding sites. The family 6 carbohydrate binding module (CBM6s) are known to contain modules of diverse specificity and variation in the location of substrate binding site with respect to their 3-dimensional structure as shown in previous reports (Henshaw *et al.*, 2004; Hashimoto, 2006; Shoseyov *et al.*, 2006). This variation in ligand recognition is exemplified in CBM family 6 (CBM6), which contains proteins that recognize xylan (Cezjek *et al.*, 2001), 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 al.*, 2004; Boraston *et al.*, 2004). 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

Corresponding Author: **Arun Goyal**
*E-mail: arungoyl@iitg.ernet.in*

Received: May 24, 2013
Accepted: June 20, 2013
Published: July 30, 2013

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 (Czjzek *et al*., 2001; Hashimoto, 2006). MODELLER is the most commonly used program for homology modeling studies (Sali and Blundell, 1993; Eswar *et al*., 2006). It first generates an alignment between one or more template and the target sequence and applies greater satisfaction of restraints compared to those that vary. The predicted model is generated and refined by energy minimization (Sali and Blundell, 1993; Fiser *et al*., 2000). The PROCHECK analysis of Ramachandran is performed basically to check the values of φ and ψ that are sterically possible (Laskowski *et al*., 2001). The secondary structure content from circular dichroism (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 (Branden and Tooze, 1991; Andrade *et al.,* 1993; Creighton, 1994).

Since previous studies on family 6 CBMs have shown diverse ligand binding ability it is essential to determine the putative ligands having maximum tendency to bind with *Ct*CBM6B. Discovery Studio 3.5 software package provides a free, feature-rich molecular modeling environment, for both small molecule and macromolecule applications (*http://accelrys.com/events/webinars/discovery-studio-25/abstracts.html*). This study describes the *in silico* sequence analysis of a gene encoding *Ct*CBM6B located at C-terminal of α-L-arabinofuranosidase (family 43 glycoside hydrolase) from *C. thermocellum*. 3-dimensional model of *Ct*CBM6B was built and the putative ligands showing significant affinity for this protein were identified. The three dimensional model of *Ct*CBM6B was generated by comparative modeling by MODELLER using the crystal structures of closely related sequences as templates.

## Materials and Methods

***Phylogenetic analysis*** - Phylogenetic analysis for finding evolutionary relation was based on BLAST*p* (*http://blast.ncbi.nlm.nih.gov/Blast.cgi*) program against PDB database in NCBI.

***Secondary structure prediction*** - PSI-PRED VIEW was used for the secondary structure prediction of various turns, helices and coils that may be present in *Ct*GH43 (*http://bioinf2.cs.ucl.ac.uk/psiout/a43f2d904ff1ff82.psi.pdf*).

***Hydrogen bonding plot*** - Hydrogen bonding (HB) plot is a new tool used for exploring protein structure describing structure as a network of hydrogen bonding interaction (Bikadi *et al.,* 2007). HB Plot offers a simple way of analyzing protein secondary and tertiary structure. Hydrogen bonds stabilizing secondary structural elements and those formed between distant amino residues - defined as tertiary hydrogen bonds can be easily identified in a HB Plot (McDonald and Thornton, 1994; Bikadi *et al*., 2007). The 3D model of *Ct*CBM6B was used as input file for generating HB plot from the HB plot server (*http://dept.phy.bme.hu/virtuadrug/hbplot/bin/pdb_upload.php*).

***Homology modeling of CtCBM6B*** - The BLAST*p* analysis of *Ct*CBM6B amino acid sequence was used as input for identifying the homologous protein sequences with available crystal structure against PDB database. Templates or sequences with significant expectation value (E-value ≤ 0.001) were used for model building of *Ct*CBM6B. The three dimensional (3D) structure was generated by using a versatile program called MODELLER v9.10 (Sali and Blundell, 1993; Eshwar *et al.,* 2006; Madhusudhan *et al*., 2009). It first identifies one or more and generates an alignment between template and the target sequence. CBM6s from *Clostridium thermocellum* (1gmm), *Celvibrio mixtus* (1uxx) and *Clostridium stercorarium* (1uy3) having close homology with *Ct*CBM6B served as templates for *Ct*CBM6B model building (Ahmed *et al*., 2013). If more than one template is used then highly conserved regions have greater restraints compare to those which vary more. The initial models generated for *Ct*CBM6B were prioritized on the basis of MOLPDF, discrete optimized protein energy (DOPE) and GA341 score (Eshwar *et al*., 2006). The most fitting model based on energy minimization having lowest value of MODELLER objective function (MOLPD and DOPE) was selected and visualized using PyMOL

(*http://www.pymol.org*). The predicted 3D model of *Ct*CBM6B was validated by Ramachandran plot (RC plot) analysis using PROCHECK (Laskowski *et al*., 1993). RC plot is basically displays 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; 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 (Lovel *et al*., 2003).

***Docking analysis of 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*CBM6B. It also used a new protocol called Dock Fragments, to place fragments in a receptor active site of *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). Finally, the docking scores were expressed in terms of ligand binding energy (-kcal $mol^{-1}$) with higher negative value indicating better protein-ligand affinity.

***Circular dichroism analysis of CtCBM6B-*** Far-UV circular dichroism (CD) spectra of *Ct*CBM6B were recorded on a spectropolarimeter (Jasco Corporation, Tokyo, JASCO J-815), equipped with a peltier system for temperature control at 25°C using a cell with a path length of 0.1 cm. The spectral accumulation parameters were carried out using a scan-rate of 50 nm $min^{-1}$, a 1 nm bandwidth in the wavelength range of 195-250 nm with an average of six scans for each far-UV spectrum. The CD data of *Ct*CBM6B is presented in terms of mean residue ellipticity (MRE, expressed as deg $cm^2$ $dmol^{-1}$) as a function of wavelength, calculated following the procedure described earlier (Kelly *et al*., 2005) using a protein concentration of 15 μM in 10 mM Tris-HCl, pH 7.5. All CD spectra were corrected for buffer contributions and secondary structures were calculated by using web based K2d neural network software package (*http://www.ogic.ca/projects/k2d2/*) as described by Perez-Iratxeta and Andrade-Navarro (2008) and Greenfield *et al*. (2006).

## Results and Discussion

### *In silico sequence analysis and homology modeling of* Ct*CBM6B*

The multiple sequence alignment (MSA) of *Ct*CBM6B reported elsewhere had shown close relation with some of the other family 6 carbohydrate binding modules (CBM6s) from *C. thermocellum* (1gmm), *Celvibrio mixtus* (1uxx) and *Clostridium stercorarium* (1uy3) (Ahmed *et al*., 2013). The phylogenetic tree analysis of *Ct*CBM6B displayed the evolutionary relationship with other family 6 carbohydrate binding modules from *Clostridium thermocellum* JW20, *Clostridium thermocellum* DSM 260, *Clostridium stercorarium, Celvibrio mixtus,* and *Acetovibrio cellulolyticus* and (Figure 1). *Ct*CBM6B also showed the somewhat distant evolutionary relationship with a few catalytic proteins like putative xylanse from *Clostridium thermocellum,* α-N-arabinofuranosidase from *Clostridium papyrosolvens* and glucouronoarabinoxylan-β-1.4-endo xylanse from *Clostridium papyrosolvens* (Figure 1). This indicated that there exists an evolutionary relationship between *Ct*CBM6B and members of family 6 CBM. The comparison of

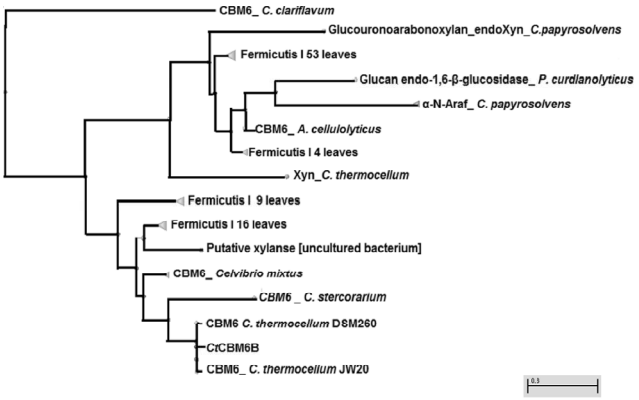



***Figure 1:*** **Phylogenetic tree constructed by MLtree based on the BLASTp. The protein sequence of *Ct*CBM6B showed close resemblance with some other family 6 carbohydrate binding module (CBM6s) from *C. thermocellum* JW20, *C. thermocellum* DSM 260, *C. stercorarium, Celvibrio mixtus*. *Ct*CBM6B also shared the common ancestor as some of the catalytic proteins like xylanse from *C. thermocellum* and α-N-arabinofuranisidase and glucouronoarabinoxylan β-1,4-endo-xylanase from *Clostridium papyrosolvens***

above results of *Ct*CBM6B with previous reports of *Ct*GH43 (Ahmed *et al.*, 2012) showed that both shared evolutionary relationship with CBM6 from *Celvibrio mixtus* and *Acetovibrio cellulolyticus*. The distant evolutionary relationship with catalytic proteins may be due *Ct*CBM6B and these proteins having the same ancestor. The amino acid sequence analysis of *Ct*CBM6B also revealed the absence of typical ligand binding site residues like Tyr-33 or Tyr-34, Trp-92 and Asn 120, associated with soluble ligands in β-jelly roll sub families like CBM1, CBM2, CBM5,CBM6, CBM10 and CBM35, which constitutes a superfamily as reported previously (Czjzek *et al*, 2001; Hashimoto, 2006).

The secondary structure prediction of *Ct*CBM6B using PSIPRED VIEW showed β-sheets (48%) and random coils (52%) but no α-helix (Figure 2A). Similar results were observed CD spectrum of *Ct*CBM6B corresponding to only β-sheets and random coils (Figure 2B) when compared with CD spectra of proteins described by Kelly *et al*. (2005) and Creighton (1994). The analysis of CD spectrum 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 (Figure 2A and Figure 2B, Table 1). The secondary structure of *Ct*CBM6B showed predominantly β-sheets and random coils but negligible α-helix content. These results are similar to the catalytic family 43 glycoside hydrolase (*Ct*GH43) of *Clostridium thermocellum* reported previously (Ahmed *et al.*, 2012). The summation of above results indicated that the full length family 43 glycoside hydrolase gene has a frequent occurrence of β-sheets and coils in its structure. The hydrogen bonding plot of *Ct*CBM6B displayed the characteristic patterns of secondary structure elements (Figure 2C). *Ct*CBM6B showed hydrogen bonds in many segments of parallel as well as anti-parallel beta strands, represented by strips perpendicular to the diagonal (Figure 2C). The HB plot of *Ct*CBM6B also indicated the fact that the structure contains β-sheets predominantly over α-helices. *Ct*CBM6B showed large number of type I and type II hydrogen bonds (Figure 2C). *Ct*CBM6B displayed almost equal number of parallel and anti-parallel β-strands (Figure 2C).

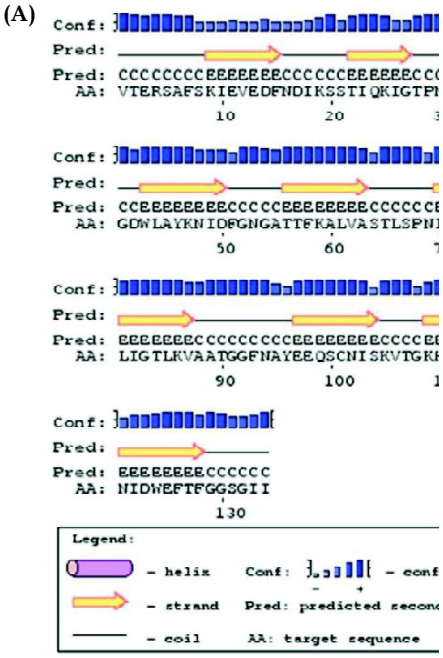
**Table 1**
**The percentage of secondary structural contents of *Ct*CBM6B protein as estimated from far-UV CD spectra**

| *Secondary structure contents of CtCBM6B* | *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.*

The homology modeling of *Ct*CBM6B by MODELLER showed multiple β-strands and coils (Figure 3A). The characteristic β-sandwich core architecture of the predicted three dimensional structures of *Ct*CBM6B was a significant feature of the 3D model (Figure 3A). *Ct*CBM6B belongs to tmpCBM-A clan as per CAZy database classification (*http://www.cazy.org/CBM6.html*) since it shows a typical β-sandwich fold which is conserved among CBM6s a sub family of β-jelly roll type of protein (Cantarel *et al.*, 2009). The β-sandwich type of architecture is common to many sub families of β-jelly roll CBM families like 2, 3, 4, 6, 9, 15, 17, 22, 27, 28, 29, 32, 34, 36 (Boraston *et al.*, 2004). Based on the reported characterized CBMs, in the CAZy databases, the predominant fold among the various CBMs is the β-sandwich (fold family 1) as reported earlier (Jamal-Talabani *et al.*, 2004). The reason behind *Ct*CBM6B

predominantly showing β-strands as compared to α-helices can be attributed mainly to the large aromatic residues (tryptophan, tyrosine and phenylalanine) and $C^{\beta}$-branched amino acids (isoleucine, valine, and threonine) which are known to adopt β-strand conformations as described previously (Creighton, 1994; Branden and Tooze, 1991). The Ramachandran plot analysis of 3D model of *Ct*CBM6B by PROCHECK (Laskowski *et al*., 1993) showed that the 89.4% residues fall in the most favoured region, 9.7% fall in generously allowed region and only 0.9% of residues were in additionally allowed region, indicating that the model has very stable conformation (Figure 3B). The acceptability of the model was further confirmed by VERY3D (Luthy



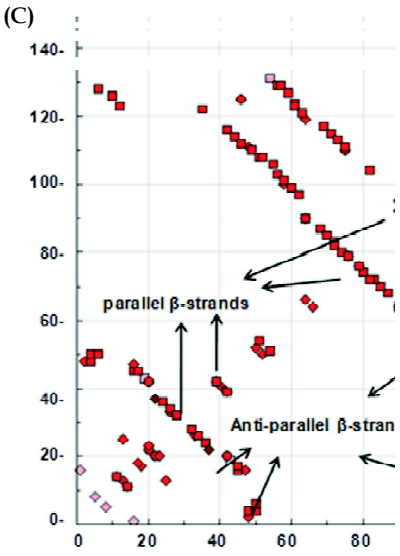



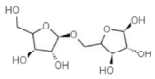
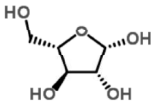
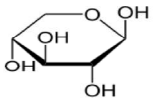
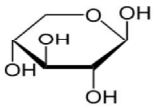
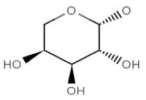
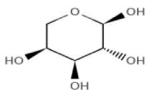
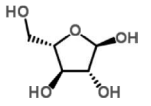
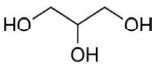
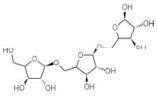

*Figure 2:* **Secondary structure prediction and analysis of *Ct*CBM6B by A) PSIPRED VIEW which showed β-strands and coils only. B) CD spectra of purified *Ct*CBM6B (15.0 mM) in sodium phosphate buffer (pH 7.5).** C) **Hydrogen bonding plot showing anti-parallel β-strands, being represented by strips perpendicular to the diagonal. Three classes of hydrogen bonding distinguished by colour coding based on the distance between donor and acceptor molecules: (i) 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)**

*et al*., 1992) analysis which showed no segment of 3D model of *Ct*CBM6B with bad score (Figure 3C). VERIFY 3D is basically analyzed the compatibility of the 3D model of *Ct*CBM6B with its own amino acid sequence (1D) and based on this it generated a plot as shown in Figure 3C. The evaluation of 3D model using VERIFY 3D has been reported previously (Ahmed *et al*., 2012).

### *Docking analysis of CtCBM6B*

The docking analysis for finding putative ligands of *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 ligand and the based on this docking scores (CDOCKER) in terms of ligand binding energy (kcal $mol^{-1}$) are given (Table 2). *Ct*CBM6B displayed maximum docking score or lowest binding energy with arabinobiose, β-L-arabinofuranose and β-D-xylopyranose indicated by lower ligand binding energy (-14.28 kcal $mol^{-1}$, -12.5 kcal $mol^{-1}$ and 11.3 kcal $mol^{-1}$,

**Table 2**
**Molecular docking scores for *Ct*CBM6B obtained using Discovery Studio 3.5**

| S. No. | Ligand Name | Ligand Structure | CBM6-Ligand Binding- Energy (kcal/mol) CtCBM6B |
|---|---|---|---|
| 1. | Arabinobiose |  | -14.287 |
| 2. | β-L-arabinofuranose |  | -12.576 |
| 3. | β-D-xylopyranose |  | -11.331 |
| 4. | α-D-xylopyranose |  | -10.806 |
| 5. | β-L-arabinopyranose |  | -10.292 |
| 6. | α-L-arabinopyranose |  | -10.043 |
| 7. | α-L-arabinofuranose |  | -8.751 |
| 8. | Glycerol |  | -6.403 |
| 9. | Arabinotriose |  | -3.433 |

respectively) (Table 2). *Ct*CBM6B also showed noticeable ligand binding affinity with α-D-xylopyranose (-10.8 kcal $mol^{-1}$), β-L-arabinopyranose (-10.2 kcal $mol^{-1}$), α-L-arabinopyranose (-10.0 kcal $mol^{-1}$) and α-L-arabinofuranose (-8.75 kcal $mol^{-1}$) as shown in Table 2. These above mentioned a result gives a clue that arabinan, xylans and substituted xylans could be the putative ligands for *Ct*CBM6B.

The molecular docking data of *Ct*CBM6B observed with these ligands showed that the

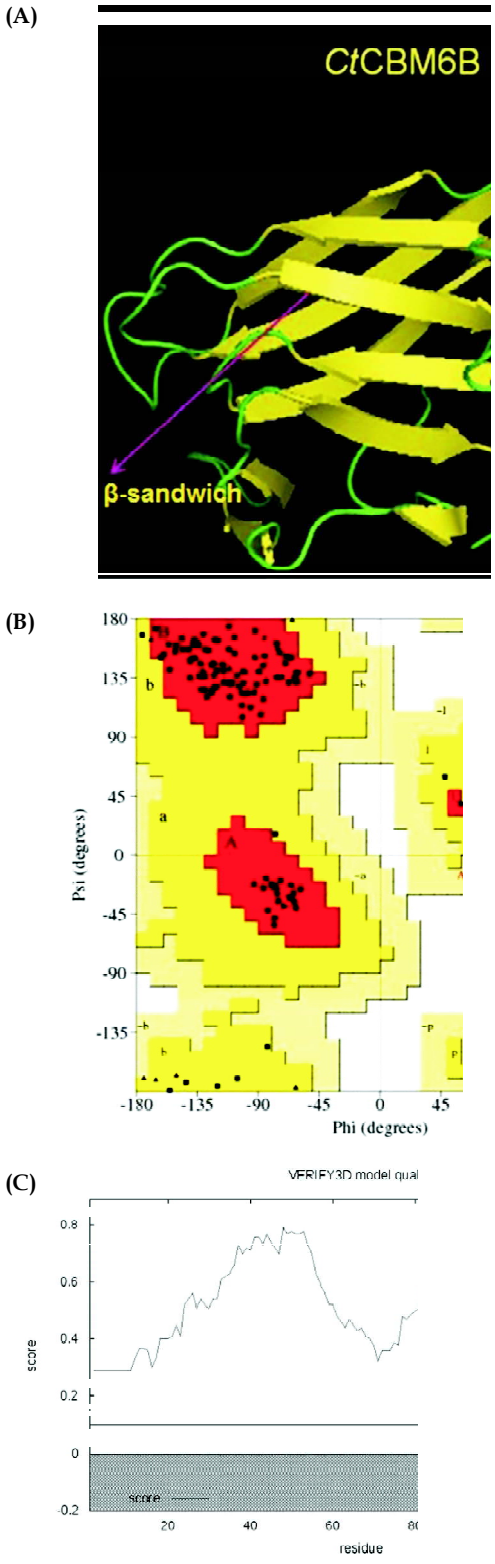



*Figure 3:* **(A) MODELLER generated 3D model of *Ct*CBM6B showing distinct β-strands in sandwich network. (B) Ramachandran plot of 3D model of *Ct*CBM6B analyzed using PROCHECK showing that most of the residues fall in the favoured regions. (C) VERIFY 3D plot showing that no segment falls below 0.2 (indicated by a thin grey line) score indicating a stable conformation**

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. Whereas, all members of family 6 CBMs even though contain the typical â-sandwich architecture and yet displays a very diverse ligand binding tendencies and this has been reported very often in the past by Czjzek *et al*., (2001); Boraston *et al*., (2004); Hashimoto, (2006) and recently by Abbot *et al*., (2009). 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). So, the present study, like previous reports further puts emphasis on the fact that even though CBM6s share a common ancestor (high structural similarity) and yet they have ligand-binding sites in different locations on the protein scaffold.

## Conclusions

*Ct*CBM6B displayed evolutionary relationship with other members of family 6 carbohydrate binding modules. The PSIPRED VIEW and HB plot of *Ct*CBM6B revealed many segments of parallel and anti-parallel β-strands which were similar to secondary structure analysis by CD spectra. The three dimensional structure of *Ct*CBM6B showed β-sandwich architecture at its core. The Ramachandran plot analysis of *Ct*CBM6B by PROCHECK displayed emphasized the acceptability of the 3D model of *Ct*CBM6B. The docking analysis of *Ct*CBM6B for finding putative ligand binding sites showed that it has high binding affinity for arabinobiose, β-L-arabinofuranose and β-D-xylopyranose followed by α-D-xylopyranose, β-L-arabinopyranose, α-L-arabinopyranose and α-L-arabinofuranose. These results indicated that *Ct*CBM6B could have affinity for binding with arabinan, xylans and substituted xylans.

## *Acknowledgment*

The research work was supported by a project grant (Grant No: BT/23/NE/TBP/2010) from Department of Biotechnology, Ministry of Science and Technology, New Delhi, Govt. of India to AG. Institute scholarships from IIT Guwahati and Moulana Azad National Fellowship (UGC-MANF) to SA are gratefully acknowledged.

## *Abbreviations*

*Ct*CBM6B, family 6 carbohydrate binding module from *C. thermocellum*; 3D, three dimensional; RC plot, Ramachandran plot; Z-score, standard deviations away from the mean; CD, circular dichroism; HB plot, hydrogen bonding plot; E- value, expectation value; BLAST*p*, Basic local alignment search tool (p stands for protein).

## *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.

Ahmed, S., Charan, R., Ghosh, A. and Goyal, A. (2012). Comparative modeling and ligand binding site prediction of a family 43 glycoside hydrolase from *Clostridium thermocellum*. J. Proteins Proteom. 3, 31-38.

Ahmed, S., Louis, S. A., Bras, L. A. J., Fontes, C. M. G. A. and Goyal A. (2013). The family 6 carbohydrate binding module (*Ct*CBM6B) of *Clostridium thermocellum* alpha-L-arabinofuranosidase binds xylans and thermally stabilized by $Ca^{2+}$ ions. Biocatal. Biotransform. In Press.

Andrade, M. A., Chacón, P., Merelo, J. J. and Morán, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. Protein Eng. 6, 383-390.

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.

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.

Branden, C. and Tooze, J. (1991). Introduction to protein structure, Chapters 1 and 2, 2nd ed. Garland Publishing, New York, pp. 1-31.

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.

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, D233-238.

Creighton, T. E. (1994). Proteins. In Conformational Properties of Polypeptide Chains, 2nd ed. W.H. Freeman, New York pp. 171-1196.

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.

Eswar, N., Marti-Renom, M. A., Webb, B. M., Madhusudhan, S., Eramian, D., Shen, M., Pieper, U. and Sali, A. (2006). Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 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.

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. 1, 2876-2890.

Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. Cell Mol. Life Sci. 63, 2954-2967.

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.

Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M. and Davies, G. J. (2004). *Ab initio* structure determination and functional characterization of CBM36; a new family of calcium-dependent carbohydrate binding modules. Structure, 12, 1177–1187.

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

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. App. Cryst., 26, 283-291.

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 Cα geometry: phi, psi and Cβ 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.

Madhusudhan, M. S., Webb, B. M., Marti-Renom, M. A., Eswar, N. and Sali, A. (2009). Alignment of multiple protein structures based on sequence and structure features. Protein Eng. Des. Sel. 22, 569-574.

McDonald, I. K. and Thornton, J. M. (1994). Satisfying Hydrogen Bonding Potential in Proteins. J. Mol. Biol. 238, 777-793.

Perez-Iratxeta, C. and Andrade-Navarro, M. A. (2008). K2D2: estimation of protein secondary structure from circular dichroism spectra. BMC Struct. Biol. 13, 8-25.

Pierce, B. G., Hourai, Y. and Weng, Z. (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. PLoS One, 6, e24657.

Pierce, B. G. and Weng, Z. (2007). ZRANK: re-ranking protein docking predictions with an optimized energy function. Proteins, 67, 1078-1086.

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.

Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 95-99.

Ramachandran, G. N. and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. Adv. Protein Chem. 23, 283-437.

Sali, A. and Blundell, T. L. (1993). Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815.

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

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 â-1,3-linked glucans by presenting a unique ligand binding surface J. Biol. Chem. 280, 530-537.