

Research Article

STRUCTURE MODELING AND CHARACTERIZATION OF
A RHAMNOGALACTURONAN LYASE (*CtRGL*) FROM
Clostridium thermocellum

Arun Dhillon and Arun Goyal*

Carbohydrate Biotechnology Laboratory, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

Abstract: The truncated N-terminal catalytic module (*CtRGL*) of molecular size ~ 64 kDa of cellulosomal rhamnogalacturonan lyase, (*CtRGLf*) from *Clostridium thermocellum* belonging to family 11 polysaccharide lyase(PL11) was structurally characterized. Multiple sequence alignment revealed that *CtRGL* has conserved active site as well as Ca²⁺ binding sites. The secondary structure prediction by PsiPred and Circular Dichroism analysis showed the presence of approximately, 2% α -helix, 30% β -sheet and 65% loops. The structure of *CtRGL* based on homology modeling showed a β -propeller fold. Structure validation by Ramachandran plot revealed 99.6% residues in allowed region. Comparison of *CtRGL* structure with that of YesW and YesX proteins from *Bacillus subtilis* showed conserved catalytic cleft and endo-lytic mode of action. Docking analysis of *CtRGL* established Arg398, Thr475, Lys476 and Tyr536 as key residues of the active site. Lys476 was predicted to act as a catalytic base and Thr475 as a catalytic acid during α -elimination. This study provides an insight into the structural determinants for mode of action and substrate recognition.

Keywords: Rhamnogalacturonan lyase; Rhamnogalacturonan I; Pectin; Polysaccharide lyase; Circular Dichroism; Homology modeling; Docking; *Clostridium thermocellum*

Note : Coloured Figures available on Journal Website in "Archives" Section

Introduction

Pectin is a complex plant polysaccharide present in the primary cell wall and middle lamella (O'Neill *et al.*, 1996). It has three major component polysaccharides, homogalacturonan (HG), rhamnogalacturonan I (RG I) and rhamnogalacturonan II (RG II) (O'Neill *et al.*, 1996). HG is assembled from α -(1,4) linked D-galactopyranosyl uronic acid (GalpA) residues. RG I main chain contains alternating GalpA and L-rhamnopyranosyl (L-Rhap) residues. The

monomeric unit of RG I main chain is a disaccharide, [→4)- α -D-GalpA-(1→2)- α -L-Rhap-(1→]. The RG I main chain is decorated with linear or branched chains of α -L-arabinofuranosyl (α -L-Araf) and β -D-galactopyranosyl residues (β -D-Galp) residues. The length of these side chains varies from 1 to more than 50 residues (O'Neill *et al.*, 1996). RG II is composed of α -(1,4) linked GalpA residues substituted with various monosaccharide residues including L-Rhap (O'Neill *et al.*, 1996).

Degradation of plant cell wall (PCW) polysaccharides by enzymes is crucial for recycling of carbon in nature. PCW polysaccharide degrading enzymes are utilized by plants for cell wall modification and by plant pathogenic microbes for invasion (Naran *et al.*, 2007; Laatu and Condemeine,

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

Received: July 21, 2017

Accepted: September 8, 2017

Published: September 10, 2017

2003). RG I degrading enzyme may cleave RG I main chain by hydrolysis (glycoside hydrolases) or α -elimination (polysaccharide lyases) mechanism (Silva *et al.*, 2016). These enzymes are referred to as exo-acting when they cleave the polysaccharide chain from terminals and endo-acting when their cleavage site is randomly distributed along the polysaccharide chain (Silva *et al.*, 2016). In the Carbohydrate-Active Enzymes database (CAZy; <http://www.cazy.org>) enzymes that synthesize, modify and cleave oligo/polysaccharides have been classified into sequence based families (Lombard *et al.*, 2014). In the CAZy database rhamnogalacturonan lyases (RG lyases) have been categorized under polysaccharide lyase (PL) family 4 (PL4) and 11 (PL11).

Collectively, reports are available for seven members of family PL4 from *Erwinia chrysanthemi* (renamed as *Dickeyadadantii*), *Aspergillus aculeatus*, *A. nidulans*, *Penicillium chrysogenum* and *Arabidopsis thaliana* (Laatu and Condemine, 2003; Kofod *et al.*, 1994; Bauer *et al.*, 2006; Iwai *et al.*, 2015; Lin *et al.*, 1999). A total of six members of family PL11, one each from *Cellvibrio japonicus*, *Clostridium cellulolyticum*, *Bacillus licheniformis*, *Clostridium thermocellum* and two from *Bacillus subtilis* have been characterized so far (McKie *et al.*, 2001; Pages *et al.*, 2003; Silva *et al.*, 2014; Dhillon *et al.*, 2016; Ochiai *et al.*, 2007). PL11 RG lyases, Rgl11A (*Cellvibrio japonicas*) and CtRGLf (*Clostridium thermocellum*) are unique as they possess a Carbohydrate Binding Module (CBM) associated with the catalytic module (McKie *et al.*, 2001; Ochiai *et al.*, 2007). The structures of three PL11 RG lyases, YesW (*B. subtilis*), YesX (*Bacillus subtilis*) and YesW_B1 (*Bacilluslicheniformis*) have been solved (Ochiai *et al.*, 2007; Ochiai *et al.*, 2009; Silva *et al.*, 2014). PL11 RG lyases take an 8-bladed β -propeller fold.

Clostridium thermocellum is a thermophilic, anaerobic bacterium that secretes a wide repertoire of PCW polysaccharide degrading enzymes. These enzymes assemble by means of protein-protein (Cohesion-Dockerin) interactions into a large multi-enzyme complex called, cellulosome (Fontes and Gilbert, 2010). *Clostridium thermocellum* genome codes for multiple types of cellulases, hemicellulases and pectatelyases (Fontes and Gilbert, 2010). However, only one RG I degrading enzyme, CtRGLf, an RG lyase has been characterized (Dhillon *et al.*, 2016). This study reports the structural characterization of the catalytic module, CtRGL, present at the N-terminal

of multi-modular RG lyase, CtRGLf by homology modeling, circular dichroism and its docking analysis.

Materials and Methods

Amino acid sequence analysis of CtRGL—The amino acid sequence of CtRGL, the catalytic module of CtRGLf (GenBankAccession no.ABN51485.1) was retrieved from the Protein database of NCBI (<https://www.ncbi.nlm.nih.gov/protein/ABN51485.1>). BLAST tool was used to detect the presence of putative domains (Altschul *et al.*, 1990). The amino acid sequences of rhamnogalacturonan lyases; YesW (*B. subtilis*), YesX (*B. subtilis*), Rgl11Y (*C. cellulolyticum*) and Rgl11A (*P. cellulosa*) were obtained using the CAZy database (<http://www.cazy.org/>) (Lombard. *Et al.*, 2014). Multiple Sequence Alignment (MSA) was carried out using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) (Sievers *et al.*, 2011).

Secondary structure analysis of CtRGL—The secondary structure of CtRGL was predicted using PsiPred tool (Jones, 1999). The secondary structure composition of CtRGL was determined by circular dichroism (CD). The gene encoding CtRGL from *Clostridium thermocellum* cloned, expressed and purified earlier (Dhillon *et al.*, 2016) was used in the present study. 15iM of purified recombinant CtRGL dissolved in 1 ml of 50 mM TrisHCl buffer (pH 8.5) was used for the CD analysis in the Far UV range (190–250nm). The CD spectrum was recorded on a spectro-polarimeter (J-815Jasco Corporation, Tokyo) at 25°C. The CD spectrum of CtRGL was expressed in terms of molar residual ellipticity (MRE deg cm²dmol⁻¹) and plotted as a function of wavelength (Kelly *et al.*, 2005). The CD spectrum was normalized for buffer contributions and the secondary structure was predicted using K2D3 server (Perez-Iratxeta and Andrade-Navarro, 2008).

Homology modeling of CtRGL—TheModeller9.14 program was used to build the 3-dimensional structure of CtRGL (Eswar *et al.*, 2006). Multiple templates were used to model the 3D-structure of CtRGL. The structures and sequences of RG lyase YesW (PDB id: 2Z8R) from *B. subtilis*, RG lyase YesX (PDB id: 2ZUY) from *B. subtilis* and RG lyase (PDB id: 4CAG) from *B. licheniformis* were selected. The ‘salign()’ command was used to generate multiple sequence alignment of these sequences. The sequence of CtRGL was then aligned to the sequence

of template structures. The 'automodel' class was used for building *CtRGL* models. The resulting models were evaluated based on DOPE score. The model with lowest DOPE score was then chosen for addition of Ca^{2+} ions as ligands. The energy of modelled structure was minimized on the YASARA server (Krieger *et al.*, 2009). Quality of the modelled structure after energy minimization was analyzed by developing its Ramachandran plot using PDBSum (Laskowski, 2001). The modelled structure was also validated using Verify 3D program.

Docking analysis of CtRGL - Molecular docking study was performed using AutoDock (Morris *et al.*, 2007). The modeled *CtRGL* structure (containing Ca^{2+} ions) after energy minimization was used. Autodock did not assign any charge to the Ca^{2+} ions while preparation of *CtRGL* structure for docking. Rhamnogalacturonan I (RG I) ligands used were: RG I disaccharide- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2), RG I trisaccharide- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap, RG I tetrasaccharide- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2), RG I pentasaccharide- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap and RG I hexasaccharide- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2). Homogalacturonan ligands used were, digalacturonic acid- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA, trigalacturonic acid- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA, tetragalacturonic acid- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA and pentagalacturonic acid- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 4)- α -D-GalpA. PDB files for the ligands were prepared using the GLYCAM server (Kirschner *et al.*, 2008). The ligands and enzyme (*CtRGL*) were saved in PDBQT format. A grid box was created around the active site residues to accommodate the ligands. Grid point spacing was

set at 0.375 \AA . The X, Y, Z co-ordinates of the grid box were 25.359, -50.188 and 86.36, respectively. 30 GA runs were used to obtain 30 different docked conformations of ligands. The docked conformations were then ranked according to the binding free energy. The docked conformation of ligand with lowest binding energy was chosen to generate the protein-ligand complex, which was visualized in Chimera and PyMol. LigPlot program was used to generate the 2D schematic representation of the protein ligand interaction.

Results and Discussion

Sequence analysis of CtRGL

BLAST result revealed that the protein with GenBank Accession No. ABN51485.1 is a multi-modular protein. Presence of three major modules namely, Dockerin I, family 35 CBM and a family 11 polysaccharide lyase was predicted. The N-terminal family 11 polysaccharide lyase module (*CtRGL*) shared amino acid sequence similarity with RG lyases, YesW (62% identity) from *B. subtilis*, YesX (59% identity) from *B. subtilis*, RG lyase (62% identity) from *B. licheniformis*, Rgl11A (66% identity) from *Cellvibrio japonicas* and Rgl11Y (66% identity) from *Clostridium cellulolyticum* (Table 1).

Multiple sequence alignment of *CtRGL* with these sequences provided information about the residues involved in catalysis (Fig. 1). Arg398, Thr475, Lys476 and Tyr536 residues were conserved in all the aligned sequences. These residues are involved in substrate catalysis in RG lyases, YesW and YesX, the closest homologues of *CtRGL*. The residues, Thr475 and Tyr536 make hydrogen bonds with rhamnose present at -1 subsite and are conserved in *CtRGL*. The residues, Asn103, Glu123, Asn473 and Gly474 making hydrogen bonds with rhamnose residue present at +2 subsite are also conserved. *CtRGL* also contains conserved rhamnose binding residues (Asn138, Arg158,

Table 1
BLAST analysis of *CtRGL* to identify its homologues.

Organism	PDB ID	Query coverage (%)	Identity (%)	e-Value	Total Score
<i>Bacillus licheniformis</i>	4CAG	99	62	0	743
<i>Bacillus subtilis</i>	2Z8R	99	62	0	742
<i>Bacillus subtilis</i>	2ZUY	99	59	0	714
<i>Bacillus pumilius</i>	5BV9	19	27	2.2	32
<i>Cellvibrio japonicus</i>	-	99	66	0	700
<i>Clostridium cellulolyticum</i>	-	99	66	0	788

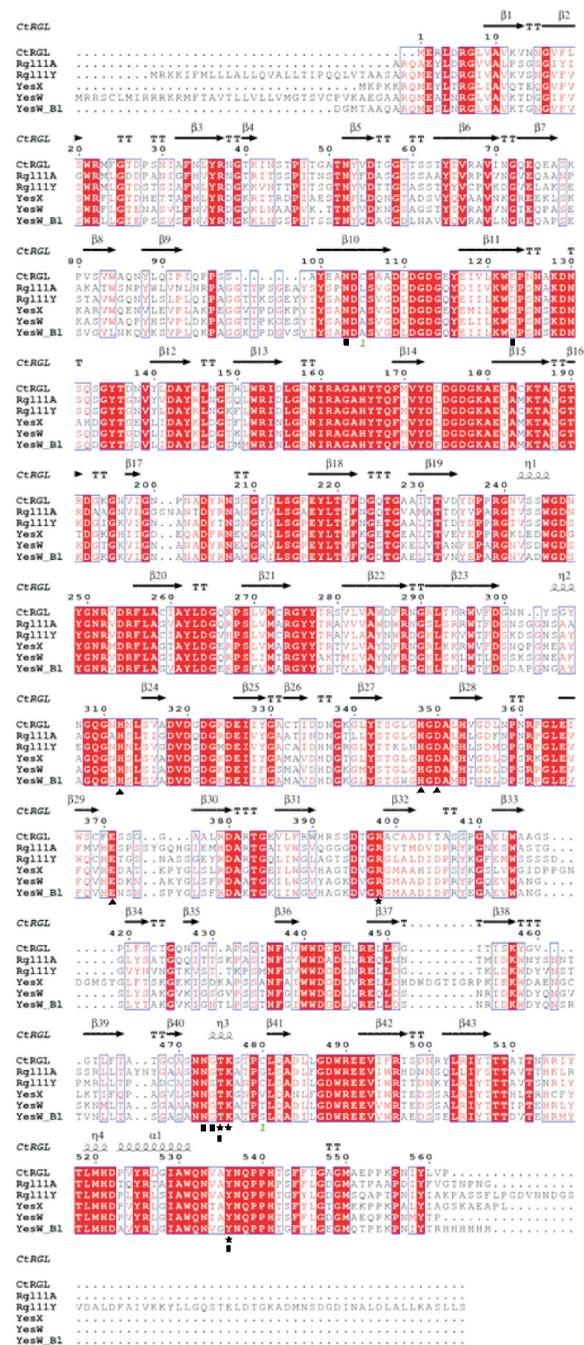


Figure 1: Multiple sequence analysis of CtRGL from *Clostridium thermocellum* with rhamnogalacturonan lyases from *Cellvibrio japonicus* (Rg11A), *Bacillus subtilis* (YesW & YesX), *Bacillus licheniformis* (YesW_B1) and *Clostridium cellulolyticum* (Rg11Y). The conserved residues are shown against red background and the semi conserved residues are enclosed in boxes. The conserved catalytic amino acid residues have been indicated by a star symbol. The amino acid residues making hydrogen bonds with α-L-rhamnose residues at the active site have been indicated by (■) symbol. Amino acid residues interacting with the Ca²⁺ ion at the active site have been indicated by symbol (○). The secondary structure elements of CtRGL have been shown above the sequences. ('α'-alpha helix, 'β'- beta strand, 'T'-turn and 'η'- 3₁₀-helix).

Table 2
Secondary structure composition of CtRGL.

Secondary structure element	PsiPred (%)	CD analysis (%)
α-helix	1.7	2.75
β-strand	31.7	30.1
Loops	66.6	67.15

Gly189 and Arg206) in addition to the residues present at the active site. The amino acid residues of YesW forming coordinate bonds with the Ca²⁺ ion involved in catalysis are conserved in CtRGL as His312, His348, Asp350 and Glu371 (Fig. 1).

Secondary structure of CtRGL

PsiPred server predicted that 1.7% of CtRGL residues form α-helix, 31.7% residues are arranged as β-strands and 66.6% of the residues occur as loops (Table 2, Fig. 2A). The secondary structure composition of CtRGL was confirmed by CD analysis. The CD analysis result showed that 2.75% of the residues form α-helix, 30.1% residues give rise to β-strands and 67.15% of the residues form loops (Table 2, Fig. 2B). The CD results corroborated with those predicted by PsiPred.

3D-structure of CtRGL

Multiple sequence alignment showed stretches of semi-conserved residues among the aligned sequences (Fig. 1). Some stretches of amino acid sequence of CtRGL were identical to one sequence while some stretches were identical to other sequence. Therefore, multiple sequences were used for modelling even though they had similar identity and coverage. Structure of protein can influence the mode of enzyme action (Ochiai *et al.*, 2009). Therefore, to avoid any biased model towards any particular structure during modelling both YesX (exo acting lyase) and YesW (endo acting lyase) were used.

The modeled CtRGL structure was organized into an N-terminal β-sheet domain (Val9-Val69) and a β-propeller domain (Tyr88-Pro562) (Fig. 3A). The Ramachandran plot developed for CtRGL showed that 89.7% residues lie in most favored regions, 9.7% residues are in additionally allowed regions, 0.2% residues reside in generously allowed region and only 0.4% residues are present in disallowed region (Fig. 3B). Verify3D results showed that 100% of the residues had an average 3D/1Dsore ≥ 0.2. The

DALI server was used to find structural homologues of *CtRGL*. The results showed that modeled *CtRGL* structure is most similar to that of RG lyase, YesW with an rmsd of 0.5 Å over 559 C_α atoms (Holm and Rosenström, 2010).

The N-terminal β-sheet domain of *CtRGL* is composed of five β-strands arranged as two anti-parallel β-sheets (Fig. 2A, Fig. 3A). A similar β-sheet rich structure called the side β-sheet structure is evident in family 30 glycoside hydrolase (GH30) members (Verma and Goyal, 2014). The side β-sheet structure is involved in substrate recognition and is indispensable for catalysis (Verma and Goyal, 2014). However, the N-terminal β-sheet domain was not found to bind ligands in case of homologues of *CtRGL* (Ochiai *et al.*, 2009). The β-propeller domain consists of anti-parallel β-sheets arranged as eight blades (A-H) of a propeller around a central α-helix (Fig. 2A, Fig. 3A). Each blade, except blade C and G is made of four anti-parallel β-sheets. Blade C has six β-sheets while blade G has three β-sheets. Each blade has conserved Ca²⁺ ion binding sites and binds one Ca²⁺ ion except blade D and E. Blade D does not bind any Ca²⁺ ion while blade E binds two Ca²⁺ ions. Two Ca²⁺ ions are present in central cavity of the β-propeller. The modelled *CtRGL* structure was superposed with its closest homologue, YesW. It was observed that a Ca²⁺ ion is present at the active site (Fig. 3C). The active site residues, Arg398, Thr475, Lys476 and Tyr536 of *CtRGL* were oriented

in the same manner as YesW active site residues (Fig. 3D). The Ca²⁺ ion present in the active cleft of YesW was shown to interact with the substrate (Ochiai *et al.*, 2009). The Ca²⁺ ion present at the active site of *CtRGL* is therefore crucial and it is coordinated by conserved residues, His312, His348, Asp350 and Glu371 (Fig. 3E).

The other critical information that could be gained from the multiple sequence alignment and structural comparison was the mode of enzyme action. In the exo-acting RG lyase (YesX), the residues, Pro439, Pro440, Gly441, Asn442, Asp443, Gly444, Met445, Ser446 and Tyr447, form a loop over its active site (Fig. 1 and 4A). This loop prevents the active site of YesX to accommodate large polysaccharide chains by creating steric-hindrance. These residues are not present in YesW (which acts endo-lytically) and other RG lyases including *CtRGL* (Fig. 1). This indicated that *CtRGL* is also an endo-RG lyase. Moreover, the endo-lytic mode of *CtRGL* activity was demonstrated earlier (Dhillon *et al.*, 2016).

Docking analysis and ligand binding by *CtRGL*

The interaction of *CtRGL* with its ligands was analyzed by docking studies. The binding energies for these ligands are reported in Table 3. The binding energies of rhamnogalacturonan I oligosaccharides of upto 5 residues in length was ~ -5 kcal/mol, but the binding of RG I hexa-

Table 3
Docking analysis of interaction between *CtRGL* and various ligands

Ligand	Predicted binding energy (kcal/mol)	Residues making H-bonds	Residues making hydrophobic interactions
RG I disaccharide	-5.12	Arg398, Lys476, Thr475, Ile435	Gly397, Asp451, Ser477, Asn436
RG I trisaccharide	-5.06	Tyr536, Lys476, Asp350, Arg398, Asp395, Thr475	Thr396, Gly397, Ile435
RG I tetrasaccharide	-5.32	Asp350, Lys476, Asp129, Ala164, Tyr536, Thr475	Arg398, Ala127, Gly474, Glu123, Gly163, Lys128, Asn 473
RG I pentasaccharide	-5.16	Asp129, Asp350, Arg397, Gly397, Arg398, Ala416, Thr475, Lys476, Tyr536	Ala164, His165, Thr396, Ala415
RG I hexasaccharide	-1.08	Tyr536, Asp350, Thr475	Gly474, Lys476, Asn473, Arg398, Gly397
Digalacturonic acid	-5.70	Arg398, Thr475, Lys476	Gly397, Asp451, Gly434
Trigalacturonic acid	-3.70	Thr475, Lys476	Arg398, Tyr536
Tetragalacturonic acid	-3.50	Thr475, Asn473, Ala164, Asp129, His165, Tyr536	Arg398, Lys476, Lys128
Pentagalacturonic acid	-2.35	Asn473, Thr475	Arg398, Ile435, Asp350, Tyr536, Gly474, Lys476, Gly397

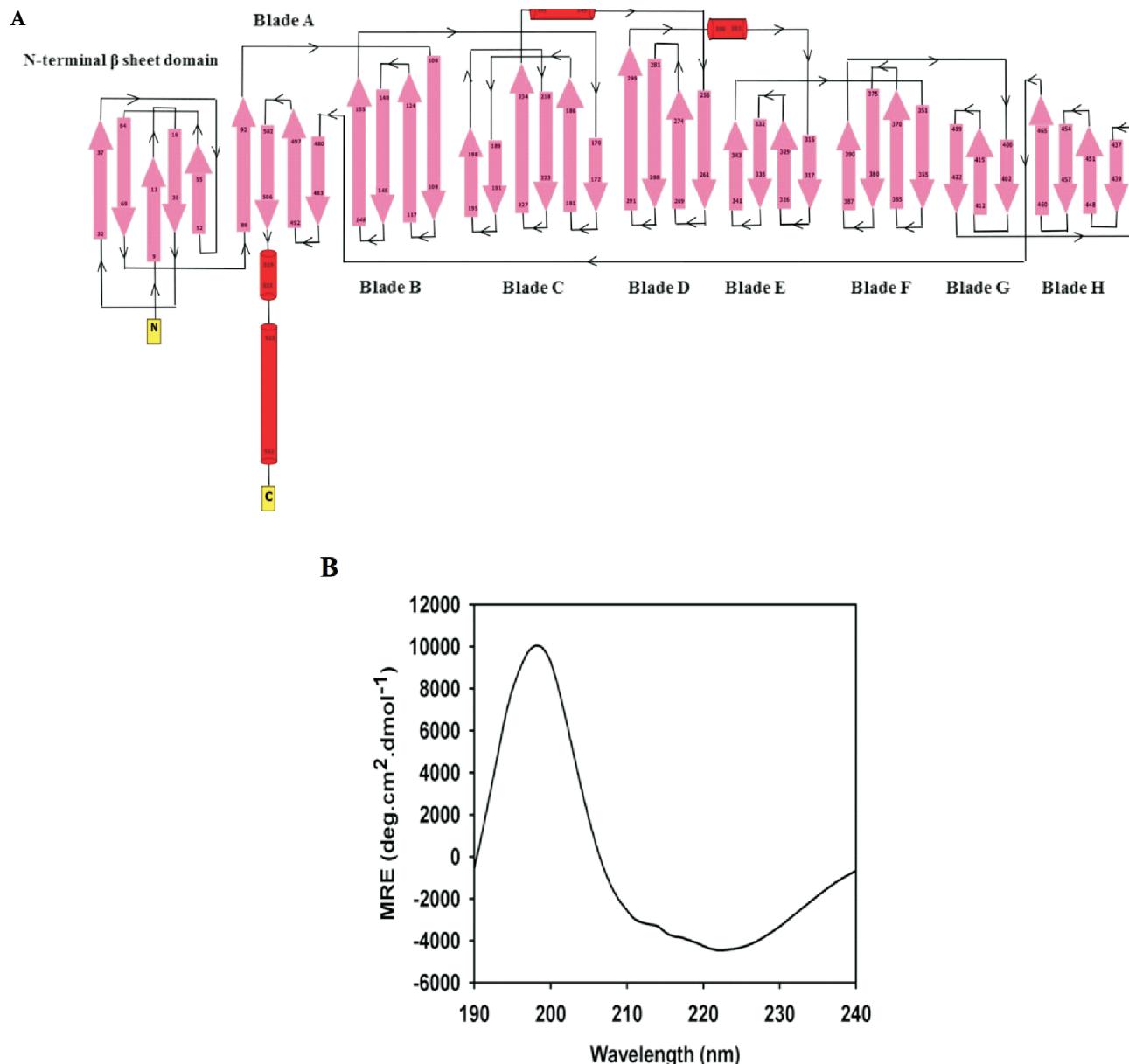


Figure 2: (A) Topology diagram of modelled CtRGL where β -sheets are shown in pink arrows and α -helices are shown as red cylinders. (B) Far UV-CD spectrum of CtRGL.

saccharide (α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA-(1 \rightarrow 2)) was comparatively weaker (-1 kcal/mol). This indicated that the active site of CtRGL could only accommodate five residues. The RG I pentasaccharide docked on the surface of CtRGL is shown in Fig. 5. The RG I pentasaccharide is held at the active cleft by Asp129, Asp350, Arg397, Gly397, Arg398, Ala416, Thr475, Lys476, Tyr536 through formation of hydrogen bonds and by Ala164, His165, Thr396, Ala415 through hydrophobic interaction (Fig. 5B). CtRGL made eight H-bonds with α -L-Rhap residues of RG I

pentasaccharide and only 2 H-bonds with α -D-GalpA residues. This indicated the importance of α -L-Rhap residues in ligand recognition. The pentagalacturonic acid oligosaccharide is held at the active site of CtRGL by only 3 hydrogen bonds, whereas, RG I pentasaccharide is held by 12 hydrogen bonds (Fig. 5C). The weak binding of pentagalacturonic acid showed that CtRGL has greater specificity for α -L-Rhap residues than α -D-GalpA residues and that α -L-Rhap are crucial for the enzyme (Table 3). A comparison of docked poses of ligands was made (Fig. 6). It was found that RG I di-, tri-, tetra- and penta-saccharide were held at

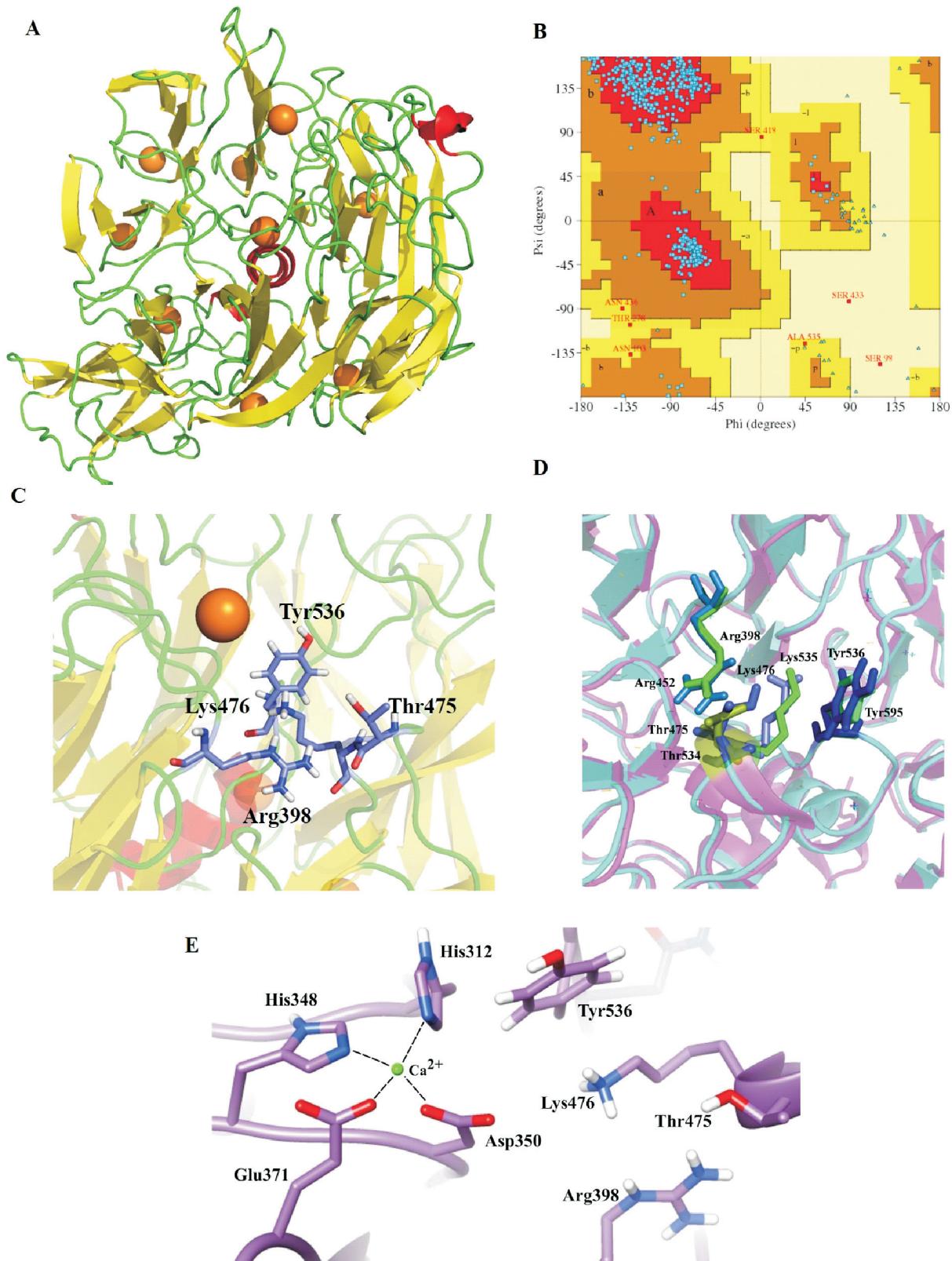


Figure 3: (A) Modelled CtRGL structure showing eight bladed β -propeller fold. Ca^{2+} ions are shown as orange spheres, (B) Ramachandran plot for modeled CtRGL, (C) active site residues of CtRGL where Ca^{2+} ion is shown as a sphere, (D) superposition of modeled CtRGL structure (cyan) over structure of YesW (magenta) (PDB id: 2Z8S) shows similar orientation of active residues. CtRGL residues are shown in blue while YesW residues are in light green and (E) Coordination of Ca^{2+} present at the active site by CtRGL residues.

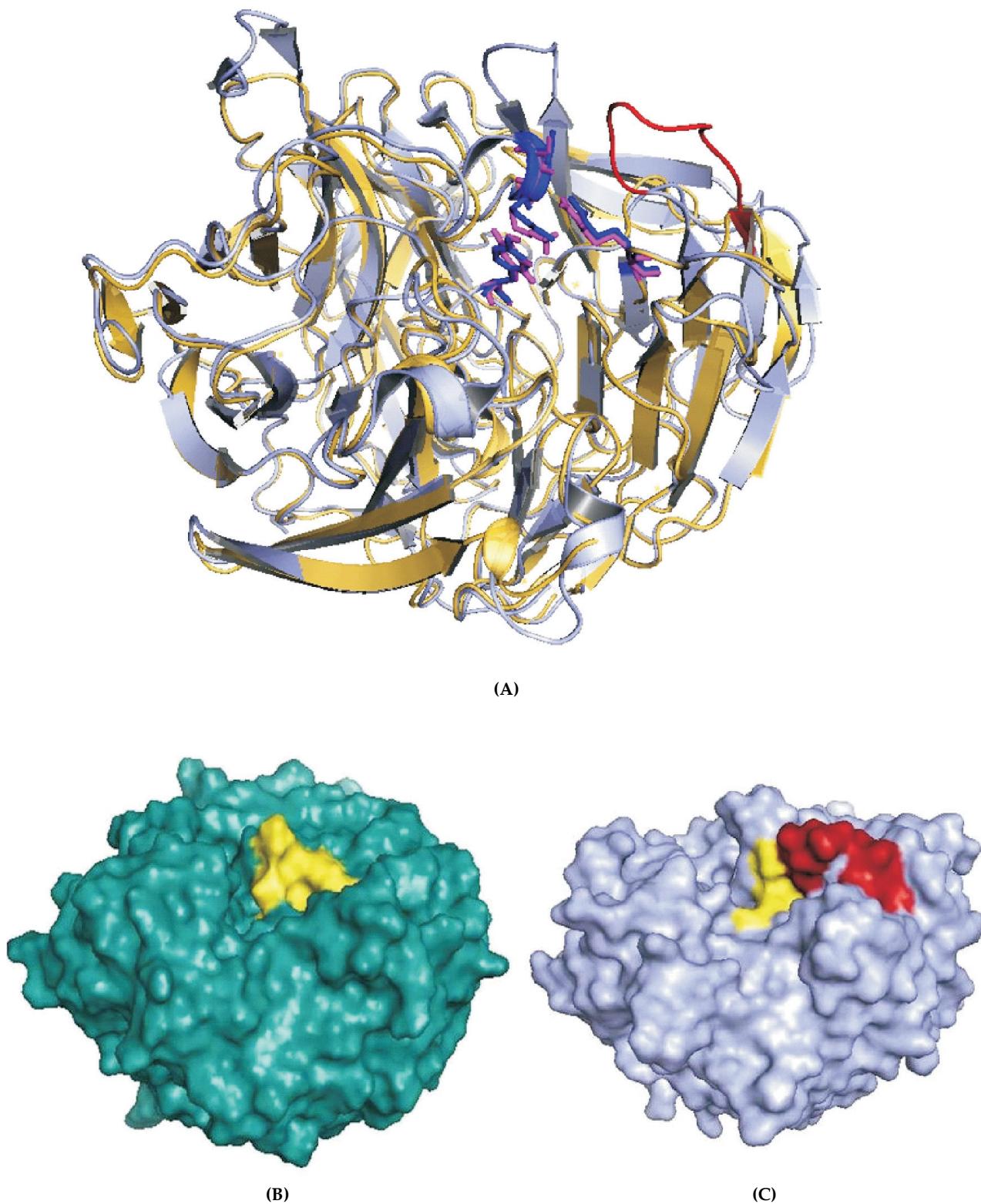


Figure 4: (A) Superposition of modeled *CtRGL* structure (yellow) over structure of exo-RG lyase, *YesX* (blue)(PDB id: 2ZUY). The loop, (⁴³⁹PPGNDGMSY⁴⁴⁷), responsible for exo-type activity of *YesX* is shown in red. *CtRGL* residues are shown in pink while *YesX* residues are in blue color. (B) Molecular envelop of *CtRGL*. Active site residues are shown in yellow (C) Molecular envelop of *YesX*. Active site residues are shown in yellowand the loop (⁴³⁹PPGNDGMSY⁴⁴⁷) is shown in red.

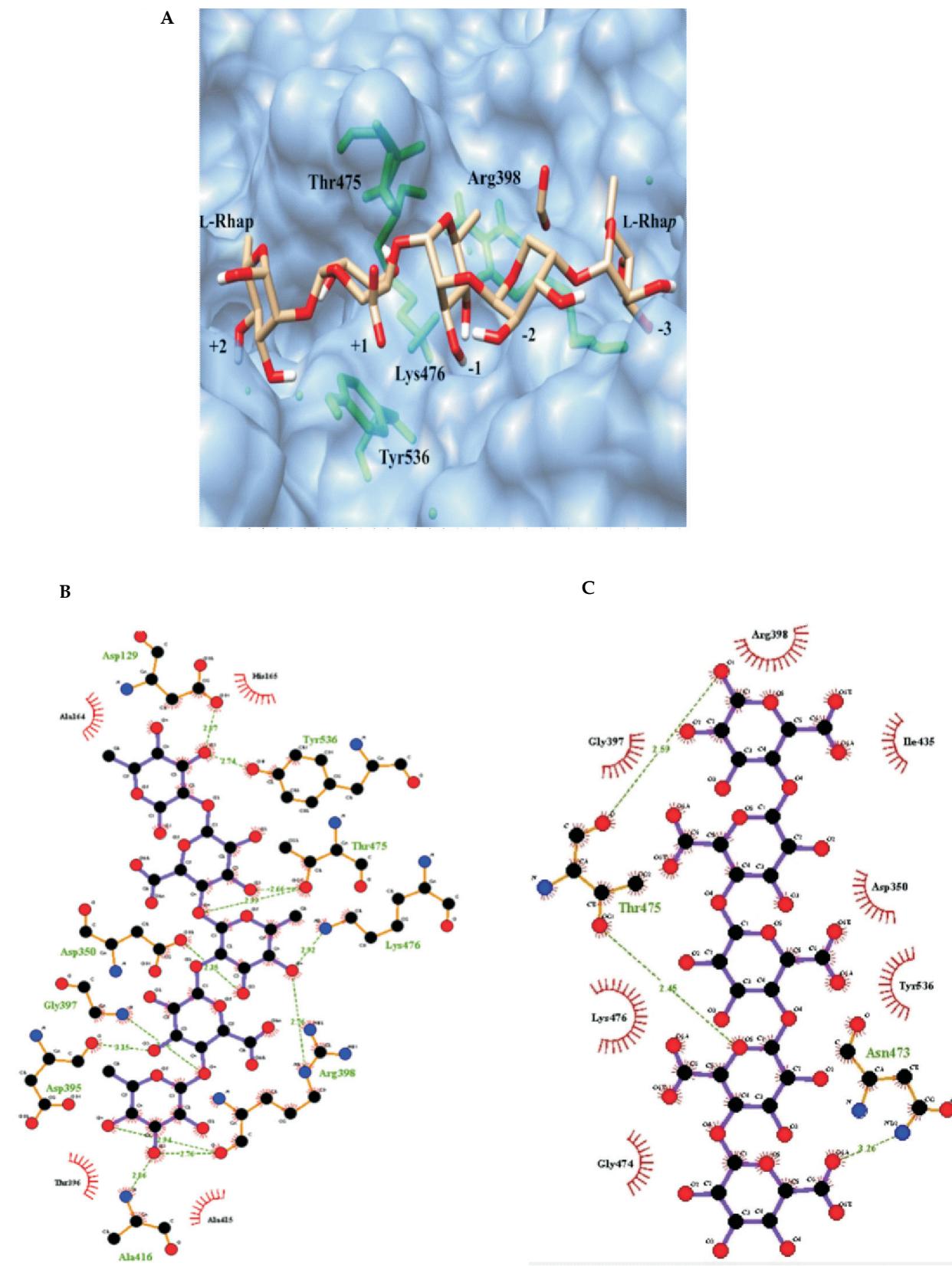


Figure 5: (A) Surface view of CtRGL showing docked RG I pentasaccharide at the active cleft. (B) Schematic representation of interaction between residues of CtRGL and RG I pentasaccharide. Ca^{2+} have been shown as green spheres.

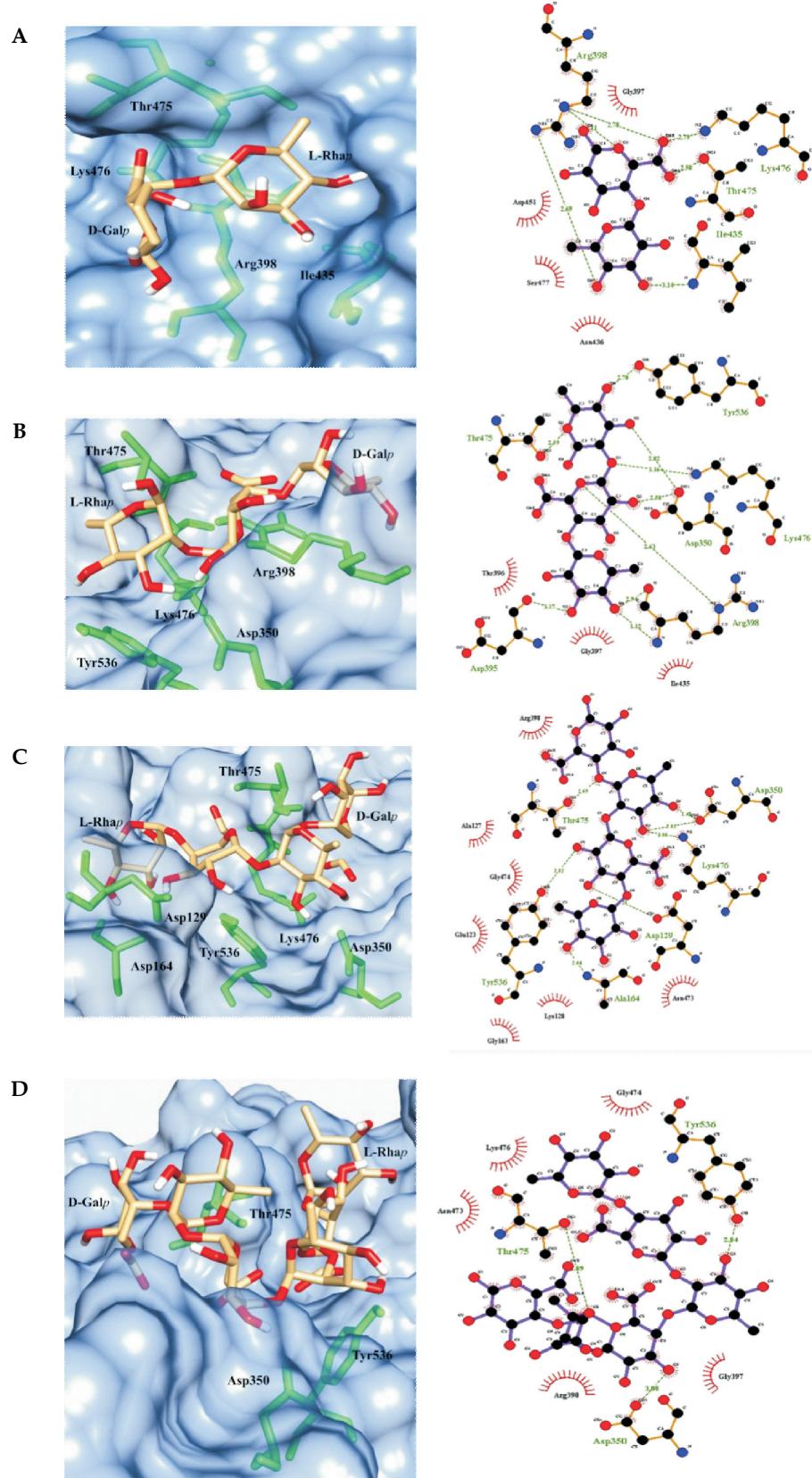


Figure 6: Comparison of poses of various ligands docked with CtRGL. Surface view and schematic representation of interaction between CtRGL and (A) RG I disaccharide, (B) RG I trisaccharide, (C) RG I tetrasaccharide and (D) RG I hexasaccharide.

the *CtRGL* active site by at least 5 Hydrogen bonds. However, RG I hexasaccharide could make only 3 hydrogen bonds. This may be due to its folded conformation at the active site (Fig. 6). Surface view of the active site after docking RG I hexasaccharide revealed that the ligand could not be accommodated at the active site and thus could not bind *CtRGL* along its length, like other ligands (Fig. 6).

The non-reducing end of a polysaccharide is referred to as $-n$ subsite and the reducing end is referred as $+n$ subsite. The cleavage is said to occur between -1 and $+1$ subsites (Davies *et al.*, 1997). The position of docked RG I pentasaccharide residues was determined based on the position of α -L-Rhap and α -D-GalpA residues in structures of YesW (2ZUX, 2Z8S) complexed with ligands (Fig. 5A). In a β -elimination reaction catalyzed by polysaccharide lyase, a catalytic base abstracts the proton from C-5 of the uronic-acid residue while a catalytic acid donates a proton to the glycoside bond undergoing cleavage. The docking study results indicated that Lys476 might act as a catalytic base and Thr475 might act as catalytic acid.

Conclusion

The present study explores the catalytic residues, secondary structure, tertiary structure fold and structural determinants of enzymatic reaction of a rhamnogalacturonan I lyase, *CtRGL* form *Clostridium thermocellum*. *CtRGL* possess conserved active site residues, Arg398, Thr475, Lys476 and Tyr536. The secondary structure composition analysis showed 2.75% of *CtRGL* structure is organized as α -helix, 30.1% as β -strands and 67.15% as loops. Modeled *CtRGL* structure showed an eight bladed β -propeller fold. Amino acid sequence analysis and comparison of *CtRGL* structure with structure of an exo-rhamnogalacturonan lyase, YesX from *Bacillus subtilis* showed that *CtRGL* lacks a nine amino acid long loop. This loop hinders YesX to access internal regions of polysaccharide chain. The absence of this loop indicated towards an endo-lytic mode of action for *CtRGL*. Docking studies revealed that active cleft of *CtRGL* can efficiently bind upto five residues of rhamnogalacturonan I chain. The recognition of α -L-Rhap residues was found to be crucial for substrate binding. Based on results of docking analysis it is predicted that Lys476 plays the role of catalytic base while Thr475 might act as catalytic acid.

Acknowledgement

The authors thank Joyeeta Mukherjee and Prof. M.N. Gupta of Department of Chemistry, Indian Institute of Technology Delhi for CD analysis.

Conflict of interest

The authors declare no conflict of interest.

Abbreviations

CBM, carbohydrate binding module; RG I, Rhamnogalacturonan I; RG Lyase, Rhamnogalacturonan lyase; CD, Circular Dichroism

References

- O'Neill, M. A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A. G., and Albersheim, P. (1996). Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by borate ester in vitro conditions for the formation and hydrolysis of the dimer. *J. Biol. Chem.* 271, 22923-22930.
- Naran, R., Pierce, M. L., and Mort, A. J. (2007). Detection and identification of rhamnogalacturonan lyase activity in intercellular spaces of expanding cotton cotyledons. *Plant J.* 50, 95-107.
- Laatu, M., and Condemeine, G. (2003). Rhamnogalacturonate lyase RhiE is secreted by the out system in *Erwinia chrysanthemi*. *J Bacteriol.* 185, 1642-1649.
- Silva, I. R., Jers, C., Meyer, A. S., and Mikkelsen, J. D. (2016). Rhamnogalacturonan I modifying enzymes: an update. *Nature Biotechnol.* 33, 41-54.
- Lombard V., Golaconda, R. H., Drula, E., Coutinho, P. M. and Henrissat, B. (2014). The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490-D495.
- Kofod, L.V., Kauppinen, S., Christgau, S., Andersen, L. N., Heldthansen, H. P. and Dorreich, K. (1994). Cloning and characterization of two structurally and functionally divergent rhamnogalacturonases from *Aspergillus aculeatus*. *J. Biol. Chem.* 269, 29182-29189.
- Bauer, S., Vasu, P., Persson, S., Mort, A. J. and Somerville, C. R. (2006). Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proc. Natl. Acad. Sci. USA*, 103, 11417–11422.
- Iwai, M., Yamada, H., Ikemoto, T., Matsumoto, S., Fujiwara, D., Takenaka, S., and Sakamoto, T. (2015). Biochemical characterization and overexpression of an endo-rhamnogalacturonan lyase from *Penicillium chrysogenum*. *Mol. Biotechnol.* 57, 539-548.
- Lin, X., Kaul, S., Rounseley, S., Shea, T. P., Benito, M. I and Town, C. D. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402, 761-768.
- McKie, V.A., Vincken, J.P., Voragen, A.G., van den Broek, L.A., Stimson, E. and Gilbert, H.J. (2001). A new family

- of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. *Biochem. J.* 355, 167–177.
- Pages, S., Valette, O., Abdou, L., Belaich, A., and Belaich, J. P. (2003). A rhamnogalacturonan lyase in the *Clostridium cellulolyticum* cellulosome. *J Bacteriol.* 185, 4727–4733.
- Silva, I. R., Jers, C., Otten, H., Nyffenegger, C., Larsen, D. M., Derkx, P. M., and Larsen, S. (2014). Design of thermostable rhamnogalacturonan lyase mutants from *Bacillus licheniformis* by combination of targeted single point mutations. *Appl Microbiol Biotechnol.* 98, 4521–4531.
- Dhillon, A., Fernandes, V. O., Dias, F. M., Prates, J. A., Ferreira, L. M., Fontes, C. M., and Goyal, A. (2016). A new member of family 11 polysaccharide lyase, rhamnogalacturonan lyase (CtRGLf) from *Clostridium thermocellum*. *Mol. Biotechnol.* 58, 232–240.
- Ochiai, A., Itoh, T., Kawamata, A., Hashimoto, W., and Murata, K. (2007). Plant cell wall degradation by saprophytic *Bacillus subtilis* strains: gene clusters responsible for rhamnogalacturonane polymerization. *Appl Environ Microbiol.* 73, 3803–3813.
- Ochiai, A., Itoh, T., Maruyama, Y., Kawamata, A., Mikami, B., Hashimoto, W., and Murata, K. (2007). A Novel Structural Fold in Polysaccharide Lyases *Bacillus subtilis* family 11 rhamnogalacturonan lyase YesW with an eight-bladed β -propeller. *J. Biol. Chem.* 282, 37134–37145.
- Ochiai, A., Itoh, T., Mikami, B., Hashimoto, W., and Murata, K. (2009). Structural determinants responsible for substrate recognition and mode of action in family 11 polysaccharide lyases. *J. Biol. Chem.* 284, 10181–10189.
- Fontes, C. M., and Gilbert, H. J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* 79, 655–681.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., and Thompson, J. D. (2011). Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202.
- Kelly, S. M., Jess, T. J., and Price, N. C. (2005). How to study proteins by Circular Dichroism. *Biochim. Biophys. Acta*, 1751, 119–139.
- Perez-Iratxeta, C., and Andrade-Navarro, M. A. (2008). K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Struct. Biol.* 8, 25.
- 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. *Curr Protoc Bioinformatics* 15, 5.6.1–5.6.30.
- Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., and Karplus, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins*, 77, 114–22.
- Laskowski, R. A. (2001). PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* 29, 221–222.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 16, 2785–2791.
- Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* 29, 622–655.
- Holm, L. and Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545–549.
- Verma, A. K., and Goyal, A. (2014). *In silico* structural characterization and molecular docking studies of first glucuronoxyran-xylanohydrolase (Xyn30A) from family 30 glycosyl hydrolase (GH30) from *Clostridium thermocellum*. *Mol. Biol.* 48, 278–286.
- Davies, G. J., Wilson, K. S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* 321 (Pt 2), 557.