Cloning, expression, purification, structure and functional Characterization of Rhamnogalacturonan Lyase (*Ct*RGL) from family 11 Polysaccharide Lyase (PL11) and associated Carbohydrate Binding Module 35 (CBM35) from *Clostridium thermocellum* ATCC 27405

PhD Thesis

by

Arun Dhillon

October 2017

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI – 781039, ASSAM, INDIA

Cloning, expression, purification, structure and functional Characterization of Rhamnogalacturonan Lyase (*Ct*RGL) from family 11 Polysaccharide Lyase (PL11) and associated Carbohydrate Binding Module 35 (CBM35) from *Clostridium thermocellum* ATCC 27405

A Thesis

Submitted in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy

by

Arun Dhillon

Under supervision of

Professor Arun Goyal

October 2017

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI – 781039, ASSAM, INDIA

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

STATEMENT

I do hereby declare that the content embodied in this thesis entitled as **"Cloning, expression, purification, structure and functional Characterization of Rhamnogalactouronan Lyase (*Ct*RGL) from family 11 Polysaccharide Lyase (PL11) and associated Carbohydrate Binding Module 35 (CBM35) from *Clostridium thermocellum* ATCC 27405"** is the result of investigations carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India under the guidance of Professor Arun Goyal.

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

October, 2017 *Arun Dhillon*

(11610610)

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

CERTIFICATE

It is certified that the work described in this thesis entitled **"Cloning, expression, purification, structure and functional Characterization of Rhamnogalactouronan Lyase (*Ct*RGL) from family 11 Polysaccharide Lyase (PL11) and associated Carbohydrate Binding Module 35 (CBM35) from *Clostridium thermocellum* ATCC 27405"** by **Aun Dhillon (Roll No. 11610610)** for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision at the Department of Biosciences & Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India and this work has not been submitted elsewhere for a degree.

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

ACKNOWLEDGEMENTS

It is a pleasant task for me to express my gratitude to all those who contributed in many ways and had profound impact on this study. It would not have been possible for me to write this thesis without the support and encouragement of many people including my supervisor, doctoral committee members, my family, friends and colleagues.

First, I would like to express my sincere gratitude to my family for my education and strength bestowed upon me to complete this work.

I wish to thank and acknowledge my thesis supervisor, Professor Arun Goyal, Department of Biosciences and Bioengineering, IIT Guwahati for his guidance, support, encouragement and providing me with the necessary instructions and research facilities. I thank him for reviewing my thesis progress and making corrections numerous times.

I would also like to express my sincere gratitude to all my doctoral committee members Dr. Debasish Das, Dr. A. B Kunnumakkara, Prof. T. Punniyamurthy, Dr. Ranjan Tamuli and Prof. Utpal Bora for their valuable suggestions and constructive criticism that has led to the successful completion of my thesis.

I am thankful to Department of Biosciences & Bioengineering and Central Instrumentation Facility (CIF), IITG for providing me instruments for my research work.

I would also like to express my sincere gratitude to Prof. S.S Ghosh, IIT Guwahati, his student Dr. Mohitosh, Prof. M.N. Gupta, Department of Chemistry, IIT Delhi and his PhD student Joyeeta Mukherjee for Circular Dichroism (CD) analysis. I sincerely thank Prof. Ashish, IMTECH, Chandigarh, for providing SAXS facility.

I would also like to thank the present and previous heads of the Department of Biosciences & Bioengineering, IIT Guwahati, Prof. K. Pakshirajan, and Prof. Venkata. V. Dasu and Prof. Arun Goyal for providing me with the necessary facilities.

I am also thankful to my seniors Dr. Shadab Ahamed, Dr. Rishikesh Shukla, Dr. Shraddha Shukla, Dr. T.J.M. Rao, Dr. Deeplina Das, Dr. Arabinda Ghosh, Dr. A.K. Verma, Dr. Damini Kothari, Dr. Soumyadeep Chakraborty and Dr. Suchi Singh for their help and suggestions. I am immensely thankful to my research group members Aruna, Rwivoo, Vikky, Sumitha, Kedar, Shweta, Priyanka, Abhijeet, Ajit. I am grateful to my batch mates at IIT Guwahati and all the people with whom I have worked at the Department of Biosciences and Bioengineering for their cooperation and support.

I would like to thank my friends Himanshu, Yogi, Mohan, Payel, Sid, Dimple, Sunanda, Sumit, Mansa, Anuradha and Varun for their support.

I wish to acknowledge the support received from other teaching and non-teaching staff of the Department of Biotechnology, IIT Guwahati.

I wish to acknowledge M.H.R.D, Govt. of India for providing financial assistance and also Department of Biotechnology, Govt. of India, New Delhi for providing me fellowship through its sponsored project.

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 and friends.

Arun Dhillon
October, 2017

SYNOPSIS

Introduction

Carbohydrates are defined as poly-hydroxy aldehydes (e.g. glucose), ketones (e.g. fructose), alcohols (e.g. xylitol), acids (e.g. gluconolactone) and their derivatives form. The sugars or saccharides are the typical carbohydrates. The plant cell wall (PCW) is principally a network of polysaccharides *viz.*, cellulose, pectin and hemicelluloses. The chains of cellulose are embedded in a matrix of pectin. Pectin refers to a family of structurally complex polysaccharides present in primary cell wall, middle lamella and cell corners. Pectin is present in walls of all higher plants, gymnosperms, pteridophytes and bryophytes. Approximately, 35% of primary cell walls in dicots and non-graminaceous monocots, 2–10% of grass and up to 5% of walls in woody tissues are composed of pectin. Pectin plays important role in cell–cell adhesion, signaling, cell expansion, leaf abscission and fruit development. There are three primary members of pectin family of polysaccharides, homogalacturonan, rhamnogalacturonan I (RG-I) and rhamnogalacturonan II. RG I main chain is made of alternating α-L-rhamnopyranosyl (α-L-Rha*p*) and α-D-galacturonic acid (α-D-Gal*p*A) residues. The monomeric unit of RG I main chain is a disaccharide, [→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→]. The α-L-Rha*p* residues in the RG-I backbone are decorated with side chains of individual, linear or branched α-L-arabinose (α-L-Ara*f*) and β-D-Galactopyranose (β-D-Gal*p*) residues. The side chains include α-1, 5-linked L-arabinan, β-1, 4-linked D-galactans and β-1, 3-linked D-galactan. The α-1, 5-linked L-arabinan may be substituted with α-1, 2- and α-1, 3-linked arabinose or arabinan

branching. The β-1, 4-linked D-galactans may be substituted with *O3*-linked L-arabinose or arabinan branching. The β-1, 3-linked D-galactan may be substituted with β-6-linked galactan or arabinogalactan branching. α-L-fucopyranose and β-D-Glc*p*A may also be present in the side chains. RG-I is referred to as the hairy region of pectin due to the presence of side chains.

Carbohydrate active enzymes catalyse the breakdown, biosynthesis or modification of carbohydrates and glycoconjugates (www.cazy.org). Carbohydrate active enzymes have been classified into different families based on protein sequence similarity. They have been classified in the CAZy database as: Glycosyl Transferases (GTs), Glycoside Hydrolases (GHs), Polysaccharide Lyases (PLs) and Carbohydrate Esterases (CEs). Polysaccharide lyases (PLs) cleave glycosidic bonds in uronic-acid containing polysaccharides. They use a β-elimination reaction mechanism to generate an unsaturated hexenuronic acid residue and a new reducing end at the cleavage site. The cleavage of polysaccharides by PLs proceeds through three stages. In the first stage carboxyl group of the substrate is neutralized by positively charged amino acids. Second, a basic amino acid abstracts proton from C-5 carbon. Next, the cleavage of the glycosidic bond is mediated by proton donation from a catalytic acid, to yield a hexenuronic acid moiety at the newly formed non-reducing end. Monosaccharide conformation at the PL active site determines whether the proton to be removed from C-5 and the departing oxygen on C-4 are syn or anti to each other. This determines whether the reaction is a syn or an anti β-elimination.

In the CAZy database the family 11 polysaccharide lyase (PL11) contains a total of 509 protein sequences as of May 2017 (http://www.cazy.org/PL11.html). Out

of 509 sequences 5 belong to archaea, 503 belong to bacteria, and 1 belongs to a eukaryote. The currently known members are endo-rhamnogalacturonan lyase (EC 4.2.2.23) or exo- rhamnogalacturonan lyase (EC 4.2.2.24). Six members of family PL11, one each from *Cellvibrio japonicus*, *Clostridium cellulolyticum*, *Bacillus licheniformis*, *Clostridium thermocellum* and two from *Bacillus subtilis* have been biochemically characterized so far. All of these are endo-rhamnogalacturonan lyase except YesX from *B. subtilis*, which is an exo- rhamnogalacturonan lyase.

Carbohydrate active enzymes often display a modular architecture. Their catalytic domains may be associated with one or more specialized substrate binding modules referred to as Carbohydrate Binding Modules (CBMs). CBMs have been to shown to enhance polysaccharide degradation by bringing catalytic modules in proximity of substrates. Role of CBMs in processivity of associated catalytic module has also been reported. CBMs have the potential to alter the specificity of catalytic modules. The evidence for potential of CBMs to direct their cognate catalytic modules to regions of plant cell wall undergoing degradation is also available in literature. A CBM has also been reported to facilitate the attachment of its catalytic module to the cell surface. Based on the sequence similarity, CBMs are grouped into 71 families as listed in CAZy database (http://www.cazy.org).

Clostridium thermocellum is an anaerobic, Gram positive, rod shaped, thermophillic bacterium. *C. thermocellum* displays on its cell surface a muti-enzyme complex called as cellulosome. The family 11 polysaccharide lyase (PL11) from *Clostridium thermocellum* under present study contains a family 35 CBM (Fig. 1).

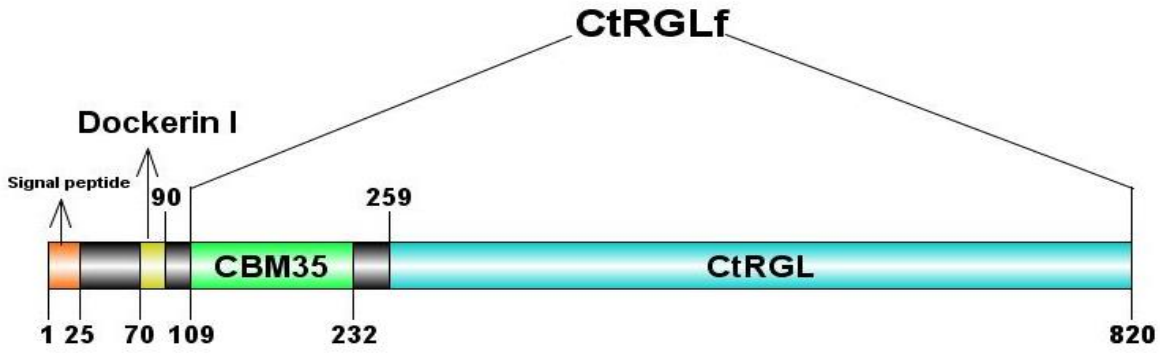


Fig. 1 Molecular architecture of modular protein ABN51485.1 from *C. thermocellum* consisting of a 25 aa signal peptide, N-terminal 124 aa family 35 carbohydrate binding module (CBM35), C-terminal 561 aa family 11 polysaccharide lyase catalytic (*Ct*RGL) module.

Present work

The present work entitled as "Cloning, expression, purification, structure and functional Characterization of rhamnogalactouronan lyase (*Ct*RGL) from family 11 Polysaccharide Lyase (PL11) and associated Carbohydrate Binding Module 35 (CBM35) from *Clostridium thermocellum* ATCC 27405" has been divided into 6 chapters.

Chapter 1 is the General Introduction, which describes the major plant cell wall carbohydrates. It presents an elaborate description of pectin and its components. This chapter describes different types of carbohydrate active enzymes and their sequence based classification. It presents a detailed description of polysaccharide lyases including their mechanism of action. It includes a description of family 11 polysaccharide lyase. Different types of pectin degrading enzymes and their applications have also been discussed in this chapter. This chapter reviews the carbohydrate binding modules, their function, classification and application. A

description of *Clostridium thermocellum* and its cellulosome has also been included in this chapter.

Chapter 2 describes amplification and cloning of the full length gene encoding family 11 polysaccharide lyase, *Ct*RGL*f* and its truncated derivatives, *Ct*RGL and *Rgl*-CBM35 from the genomic DNA of *Clostridium thermocellum* ATCC 27405 (GenBank Accession No: ABN51485.1). The molecular architecture revealed a type 1 dockerin domain towards N-terminal followed by a family 35 carbohydrate binding module (*Rgl*-CBM35, 16.4 kDa) and family 11 polysaccharide lyase catalytic module (*Ct*RGL, 64 kDa). The PCR amplified fragment of full length gene encoding *Ct*RGL*f* showed a band of ~2.1 kb, whereas, the gene encoding catalytic module (*Ct*RGL) and binding module (*Rgl*-CBM3), displayed the sizes of ~ 1.6 kb and ~ 0.3 kb, respectively. The restriction enzyme digested fragments of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 were ligated with linearized pET-28a(+) vector. The ligated mixture was transformed into *E. coli* DH5α competent cells. The positive clones containing recombinant plasmid DNA were screened by restriction enzyme digestion using enzymes, *NheI* and *XhoI*. The restriction enzyme digested products were electrophoresed and the band of ~ 5.4 kb was produced for pET-28a(+) vector and corresponding bands of ~2.1 kb, ~ 1.6 kb and ~ 0.3 kb were produced from the insert fragments for genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35, respectively. The recombinant plasmids containing genes encoding *Ct*RGL*f* and *Rgl*-CBM35 were transformed into *E. coli* BL-21 (DE3) cells. Recombinant plasmid containing gene encoding *Ct*RGL was transformed into *E. coli* BL-21 (DE3) pLysS cells. The recombinant proteins were expressed and purified. The purified recombinant proteins

displayed a band of approximately, 80 kDa for *Ct*RGL*f*, 70 kDa for *Ct*RGL and a band of approximately, 16 kDa for *Rgl*-CBM35 on SDS-PAGE gels. The amount recombinant *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 proteins obtained from 100 ml *E.coli* cultures after purification by Immobilized Metal Ion Affinity Chromatography (IMAC) were 0.4 mg, 0.3 mg and 4.0 mg, respectively.

Chapter 3 describes the biochemical and functional characterisation of *Ct*RGL*f* and *Ct*RGL. *Ct*RGL*f* and *Ct*RGL showed maximum enzyme activity at pH 8.5. *Ct*RGL*f* was active at pH range 7.0-10.0 while *Ct*RGL was active only at pH 7.0-9.0. The optimal temperature of *Ct*RGL*f* and *Ct*RGL was 70ºC and 60ºC respectively. *Ct*RGL*f* retained 90% of its activity after incubation at 60ºC for 30 min. *Ct*RGL retained 45% of its activity after incubation at 60ºC for 30 min. Both *Ct*RGL*f* and *Ct*RGL showed high activity towards rhamnogalacturonan I conatining polysaccharides. Maximum activity of both, *Ct*RGL*f* and *Ct*RGL was against rhamnogalacturonan from soybean (RGS) at 9.8 U/mg and 5.8 U/mg respectively. *Ct*RGL*f* and *Ct*RGL had similar values of K_m, 4.8 mg/ml and 5.1 mg/ml respectively, with RGS. The presence of 3 mM Ca^{2+} increased the activity of *Ct*RGL*f* and *Ct*RGL by 1.5 and 1.3 folds respectively. Treatment of *Ct*RGL*f* and *Ct*RGL with 10 mM EDTA completely abolished their activity. The TLC analysis of the products formed by the action of *Ct*RGL*f* and *Ct*RGL showed that they endolytically cleaved RGS.

Chapter 4 describes with the structural characterization of *Ct*RGL. Multiple sequence alignment revealed that *Ct*RGL has conserved active site as well as Ca^{2+} 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 *Ct*RGL based on homology modeling showed a β-propeller fold. The structure validation by Ramachandran plot revealed 99.6% amino acid residues in the allowed region. Comparison of *Ct*RGL structure with that of YesW and YesX proteins from *Bacillus subtilis* showed conserved catalytic cleft and endo-lytic mode of action. Docking analysis of *Ct*RGL 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 β-elimination. This study provides an insight into the structural determinants for mode of action and substrate recognition.

Chapter 5 describes the substrate binding analysis and structural characterization of *Rgl*-CBM35. *Rgl*-CBM35 displayed binding with β-D-glucuronic acid (β-D-Glc*p*A), Δ4,5-anhydro-D-galactopyranosyluronic acid (Δ4,5-Gal*p*A), rhamnogalacturonan I, arabinan, galactan, glucuronoxylans and arabinoxylans. *Rgl*-CBM35 contains a conserved ligand binding site in the loops known for binding β-D-Glc*p*A and Δ4,5-Gal*p*A moiety of unsaturated RG I and pectic-oligosaccharides. Mutagenesis revealed that Asn118 plays an important role in binding β-D-Glc*p*A, Δ4,5-Gal*p*A, sugarbeet arabinan and potato galactan at its conserved ligand binding site present in surface exposed loops. EDTA-treated *Rgl*-CBM35 showed no affinity towards β-D-Glc*p*A and Δ4,5-Gal*p*A underscoring Ca^{2+} mediated ligand recognition. Contrastingly, the EDTA-treated *Rgl*-CBM35 and its mutant N118A displayed affinity for sugarbeet arabinan and potato galactan. The curtailed affinity of Y37A/N118A and R69A/N118A double mutants towards sugarbeet arabinan emphasized the presence of a second ligand binding site. *Rgl*-CBM35 is the first CBM reported to primarily target

RG I and also is the first member of family 35 CBM possessing at least two ligand binding sites.

Chapter 6 describes the application of *Ct*RGL*f* in bioscouring of cotton fabric and degumming of jute fibers. The best concentration for both bio-scouring of cotton and degumming of jute was found to be 12 mg/ml (3.2 U/ml) of crude *Ct*RGL*f*. The cotton fabric treated with 12 mg/ml (3.2 U/ml) of crude *Ct*RGL*f* took 30 s to absorb a drop of water, while the *Ct*RGL*f* untreated fabric took 40 min. The crude *Ct*RGL*f* reduced the water absorption time of cotton fabric by 98% after 45 min of treatment, where as, in the industries, the usual scouring time is 1 h. The FESEM images showed that the crude *Ct*RGL*f* (12 mg/ml; 3.2 U/ml), effectively removed the waxy compounds from the surface of jute fibers after 1h of treatment. These results show that *Ct*RGL*f* has a potential application for textile industry. To best of our knowledge, this is the first study on the use of rhamnogalacturonan lyase for bio scouring of cotton fabric and degumming of jute fiber.

CONTENTS

Chapter 1

General Introduction

1. Carbohydrates

Carbohydrates are defined as poly-hydroxy aldehydes (e.g. glucose), ketones (e.g. fructose), alcohols (e.g. xylitol), acids (e.g. gluconolactone) and their derivatives form (Nelson *et al*., 2008). The sugars or saccharides are the typical carbohydrates. Carbohydrates consisting of a single polyhydroxy aldehyde or ketone molecule are called monosaccharides (Wade, 1999). D-Glucose is the most abundant monosaccharide in nature. Monosaccharides with four or more carbon atoms mostly exist in cyclic form (Nelson *et al*., 2008). More than one monosaccharide molecules may join to form a chain of saccharides called oligosaccharides. Di-saccaharides are the most prevalent oligosaccharides made up of two monosaccharide molecules. Sucrose consisting of D-Glucose and D-Fructose, is a common example of a disaccharide. 20 or more monosaccharide molecules may join to form polysaccharides e.g. cellulose (Nelson *et al*., 2008). Cellulose contains several thousand D-Glucose units (O'sullivan, 1997). Carbohydrates carry out important cellular functions. They are used as an energy source across all life forms, serve structural and protective role and act as signaling molecule among several other functions (Nelson *et al*., 2008).

1.1 Plant cell wall carbohydrates

The plant cell wall (PCW) is principally a network of polysaccharides *viz.*, cellulose, pectin and hemicelluloses (Mohnen *et al*., 2008). The chains of cellulose are embedded in a matrix of pectin (Fig. 1.1). The cellulose chains are interlinked by hemicellulose (Taiz *et al*., 2010). These carbohydrate components of PCW are linked by covalent and non-covalent interactions. Apart from these components, glycoproteins, minerals, enzymes and phenolic esters are the other components of PCW (Matsunaga *et al*., 2004). The cell wall is both structurally and functionally important for plant cells. The interconnected network of polysaccharide provides strength to plant cells. PCW selectively filters the molecules with diameter greater than 4 nm and forms channels for fluid transport (Alberts *et al*., 1997).

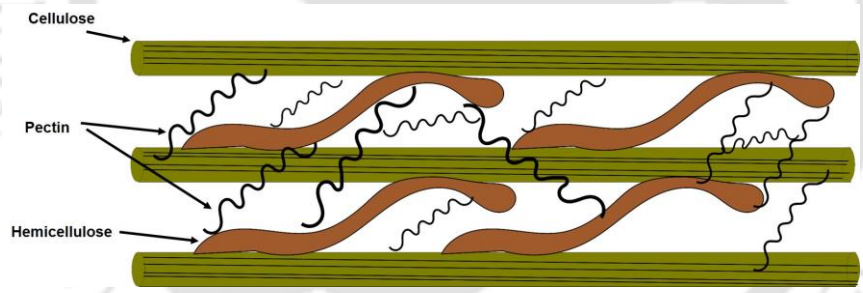


Fig. 1.1 Schematic representation of the plant cell wall showing the association between cellulose, hemicellulose and pectin.

Cellulose is linear un-branched homopolymer of β-1,4-linked D-glucopyransoe residues (Taiz *et al*., 2010). Cellulose chains make extensive non-covalent interactions with each other. The tensile strength cellulose is comparable to that of steel. Cellulose component of plant cell wall is organized into crystalline and amorphous domains (O'Sullivan, 1997). Hemicellulosic polysacchardies are hetropolymers. They are named according to the component of their main chain. The most common hemicellulosic polysaccharides are xylans and mannans. Xylan is made of β-1-4-

linked xylopyranose residues (Scheller *et al*., 2010). Xylan may also conatin α-L arabinofuranosyl (arabinoxylans), β-D-glucucopyranosyl uronic acid (β-D-Glc*p*A) or 4-*O*-methyl-D-glucucopyranosyl uronic acid (glucuroxylans) residues as side chains (Scheller *et al*., 2010 and Cosgrove, 1999).

Mannans are homo-polymers and contain β-1,4-linked D-mannopyranose backbone that is substituted with galactose in galactomanan or glucose in glucomanan (Cui *et al*., 2009). Mannans may be further classified into three subfamilies: (1) glucomannan, (2) galactomannan and (3) galactoglucomanan (Petkowicz *et al*., 2001). The glucomannan backbone of β-1, 4-linked D-mannopyranosyl residues contains randomly distributed β-D-glucose residues in a 3:1 ratio (mannose: glucose). The backbone of galactomannan consists of β-1, 4-linked D-mannopyranosyl residues that is substituted with side chains of single α-1,6-linked D-galactopyranosyl residues. Galactoglucomannan backbone is made of β-1, 4-linked D-mannopyranosyl residues and randomly distributed β-D-glucose residues linked to mannose residue by β-1, 4-linkage and the D-glucose residues are substituted with α-1,6-linked D-galactose residues (Moreira, 2008).

Apart from the polysaccharides, lignin is also an important component of PCW. Lignin refers to a group of aromatic polymers derived from oxidative polymerization of 4-hydroxyphenylpropanoids (Vanholme *et al*., 2010). Lignin provides strength to cell wall and imparts resistance to microbial attack. The complex architectute of PCW varies with species, tissue, developmental stage and environment among other factors (Taiz *et al*., 2010).

1.2 Pectin and its components

Pectin refers to a family of structurally complex polysaccharides present in primary cell wall, middle lamella and cell corners (Fig. 1.2) (O'Neill *et al*., 1990; Ridley *et al*., 2001). Pectin is present in walls of all higher plants, gymnosperms, pteridophytes and bryophytes (Mohnen *et al*., 2008). Approximately, 35% of primary cell walls in dicots and non-graminaceous monocots, 2–10% of grass and up to 5% of walls in woody tissues are composed of pectin (O'Neill *et al*., 1990; Ridley *et al*., 2001). Pectin plays important role in cell–cell adhesion, signaling, cell expansion, leaf abscission and fruit development (Willats *et al*., 2001).

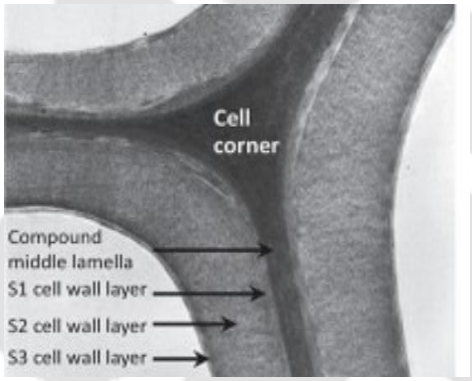


Fig. 1.2 Transmission electron microscopic image of a cross section of early wood xylem of Norway spruce (*Picea abies*). Compound middle lamella (middle lamella and primary cell wall) and various layers (S1, S2 ,S3) of secondary cell wall have been indicated (Johansen, 2016).

1.2.1 Types of Pectin

There are three primary members of pectin family of polysaccharides, Homogalacturonan (HG), rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II) (Fig. 1.2).

1.2.1.1 Homogalacturonan

Homogalacturonan is linear homopolymer of a α-1,4-linked D-galactopyranosyluronic acid (α-Gal*p*A) (Mohnen *et al*., 2008). The α-Gal*p*A residues

of homogalacturonan may be methylesterified at the C-6 carboxyl or O-acetylated at O-2/O-3 (O'Neill *et al*., 1990).

1.2.1.2 Rhamnogalacturonan I

RG I main chain contains alternating α-Gal*p*A and α-L-rhamnopyranosyl (α-L-Rha*p*) residues (Mohnen *et al*., 2008) (Fig. 1.3). The monomeric unit of RG I main chain is a disaccharide, [→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→]. The α-L-Rha*p* residues in the RG-I backbone are decorated with side chains of individual, linear or branched α-L-Ara*f* and β-D-Galactopyranose (β-D-Gal*p*) residues. The side chains include α-1, 5-linked L-arabinan, β-1, 4-linked D-galactans and β-1, 3-linked D-galactan (Nakamura *et al*., 2002). The α-1, 5-linked L-arabinan may be substituted with α-1, 2- and α-1, 3-linked arabinose or arabinan branching. The β-1, 4-linked D-galactans may be substituted with *O3*-linked L-arabinose or arabinan branching. The β-1, 3-linked D-galactan may be substituted with β-6-linked galactan or arabinogalactan branching. α-L-fucopyranose and β-D-Glc*p*A may also be present in the side chains. RG-I is referred to as the hairy region of pectin due to the presence of side chains (O'Neill *et al*., 2003).

1.2.1.3 Rhamnogalacturonan II

Rhamnogalacturonan II also consists of 1, 4-linked α-D-Gal*p*A residues which are decorated with side chains containing upto 12 different monosaccharides (Fig. 1.3) (O'Neill *et al*., 2004). The monosaccharides are linked by using more than 20 different linkages. β-D-apiosyl residues present in the side chains are involved in RG-II dimerization. Interactions of RG-II with HG and RG-I leads to an intricate macromolecular pectin network (Matsunaga *et al*., 2004). The other less abundant members of pectin family are xylogalacturonan (XGA) and apiogalacturonan (AP).

XGA is a homogalacturonan that conatins substitution of β-D-xylopyranose residues O-3 position. In apiogalacturonans the homogalacturonan main chain is substituted at O-2 or O-3 with D-apiofuranose residues (Zandleven *et al*., 2006).

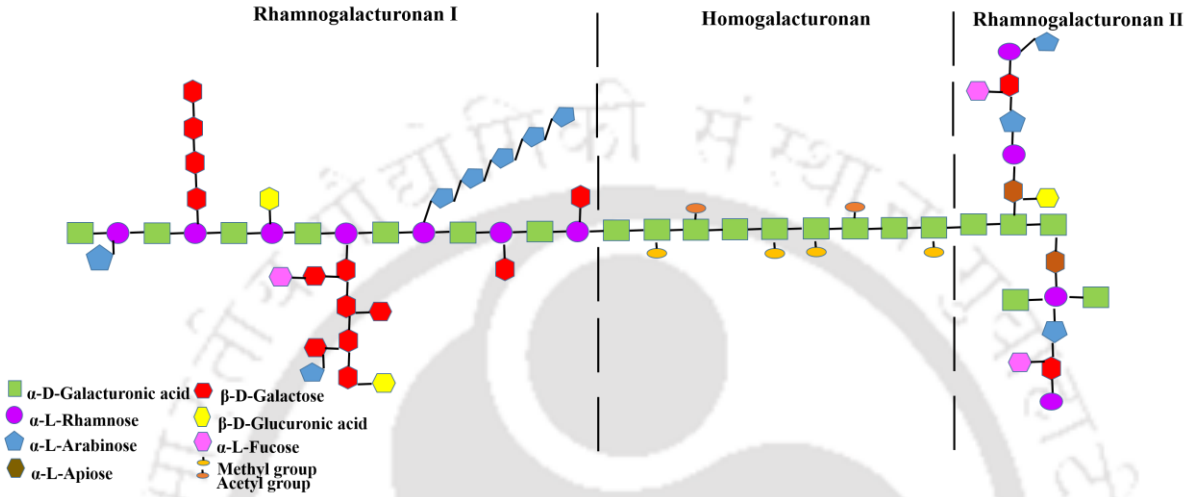


Fig. 1.3 Schematic representation of pectin.

1.3 Pectin structure and biosynthesis

Experimental results support a model for pectin structure that suggests that HG, RG-I, and RG-II are linked via their backbones (Nakamura *et al*., 2002). Evidence is also available that points towards strong interactions between pectin and other cell wall polysaccharides such as xylan and xyloglucan (Nakamura *et al*., 2002, Popper *et al*., 2007). Studies involving autoradiographic pulse chase experiments, immunocyto-chemical localization with anti-pectin-specific antibodies and pectin biosynthetic enzyme subcellullar fractionation have shown that pectin is synthesized in the Golgi lumen by membrane bound or associated glycosyltransferases.

1.4 Carbohydrate-active enzymes

Carbohydrate active enzymes catalyse the breakdown, biosynthesis or modification of carbohydrates and glycoconjugates (www.cazy.org). Carbohydrate

active enzymes have been classified into different families based on protein sequence similarity (Cantarel *et al.*, 2009). They have been classified in the CAZy database as: Glycosyl Transferases (GTs), Glycoside Hydrolases (GHs), Polysaccharide Lyases (PLs) and Carbohydrate Esterases (CEs). Among the four classes, enzymes of only glycosyl transferase class catalyse the formation of a bond, the glycosidic bond. Glycoside hydrolases catalyze the hydrolytic cleavage of the glycosidic bond whereas polysaccharide lyase enzymes catalyse non-hydrolytic cleavage of glycosidic bonds using a β-elimination mechanism. The carbohydrate esterase enzymes catalyze the de-O-acylation or de-N-acylation of substituted saccharides. As of September 2017, there are 104 families of GTs, 145 families of GHs, 27 families of PLs and 16 families of CEs. Carbohydrate active enzymes often display a modular architecture (Fig. 1.4). Their catalytic domains may be associated with one or more specialized substrate binding modules referred to as Carbohydrate Binding Modules (CBMs).

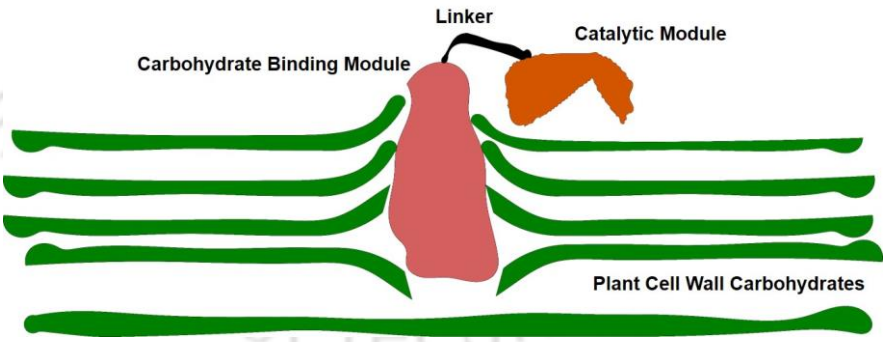


Fig. 1.4 Schematic representation of a modular carbohydrate active enzyme possessing a carbohydrate-binding molecule (CBM) connected to the catalytic domain (CD) via a linker peptide.

1.4.1 Polysaccharide lyases

Polysaccharide lyases (PLs) cleave glycosidic bonds in uronic-acid containing polysaccharides. They use a β-elimination reaction mechanism to generate an unsaturated hexenuronic acid residue and a new reducing end at the cleavage site (Yip *et al*., 2006). The enzymes of this class have been identified in bacteriophages, archaea, eubacteria, fungi, algae, plants and mammals (Sutherland, 1995). Involvement of PLs has been studied in diverse biochemical processes such as biomass degradation, tissue matrix recycling and pathogenesis (Lombard *et al*., 2010). The cleavage of polysaccharides by PLs proceeds through three stages (Gacesa, 1987). In the first stage carboxyl group of the substrate is neutralized by positively charged amino acids. Second, a basic amino acid abstracts proton from C-5 carbon. Next the cleavage of the glycosidic bond is mediated by proton donation from a catalytic acid, to yield a hexenuronic acid moiety at the newly formed non-reducing end. Monosaccharide conformation at the PL active site determines whether the proton to be removed from C-5 and the departing oxygen on C-4 are syn or anti to each other (Lombard *et al*., 2010). This determines whether the reaction is a syn or an anti β-elimination (Fig. 1.5).

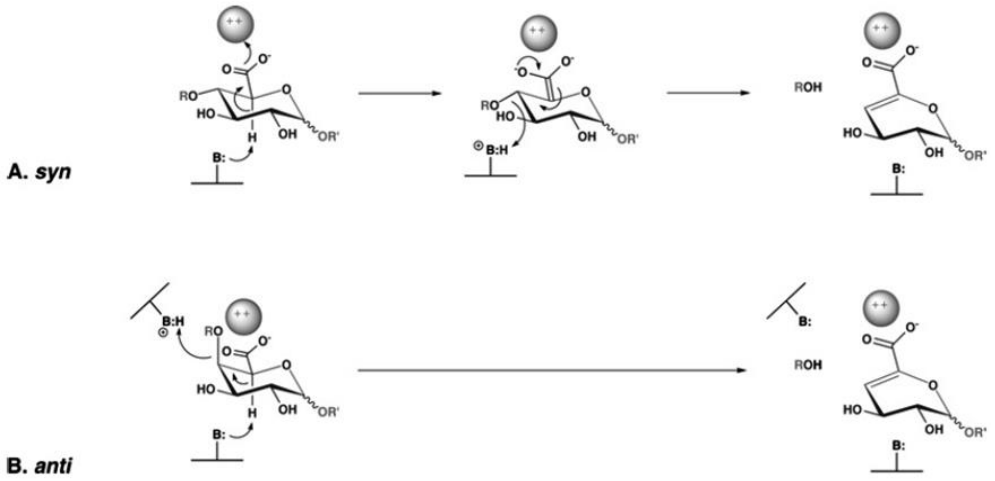


Fig. 1.5 Types of β- elimination mechanism. (A) syn-β-elimination and (B) anti-β-elimination. In both cases, after action of polysaccharide lyase a 4,5 unsaturated galacturonate moiety is formed at the new non-reducing end of the chain. The C-5 proton adjacent to the carbonyl group is abstracted by a basic amino acid at the catalytic centre (B:). Removal of the glycosidic oxygen is facilitated by proton donation from an acidic amino acid residue (B:H) (adapted from Lombard *et al*., 2010).

According to the nomenclature proposed by Davies et al. 1997, the monosaccharide residues are numbered +1, +2, … and so on from the cleavage site towards the reducing end and −1,−2, etc. towards the non-reducing end. The cleavage occurs between +1 and -1 site. Substrate recognition by PLs is often mediated by bivalent cations such as Ca^{2+}, or positively charged amino acid side chains which interact with uronic acid groups (Yip *et al*., 2004). The bivalent cation is thought to stabilise the transient anion generated during the reaction.

1.5 Family 11 polysaccharide lyase

In the CAZy database the family 11 polysaccharide lyase (PL11) contains a total of 509 protein sequences as of May 2017 (http://www.cazy.org/PL11.html). Out of 509 sequences 5 belong to archaea, 503 belong to bacteria, and 1 belongs to a

eukaryote. The currently known members are endo-rhamnogalacturonan lyase (EC 4.2.2.23) or exo- rhamnogalacturonan lyase (EC 4.2.2.24). Six members of family PL11, one each from *Cellvibrio japonicus*, *Clostridium cellulolyticum*, *Bacillus licheniformis*, *Clostridium thermocellum* and two from *Bacillus subtilis* have been biochemically characterized so far (Silva *et al*., 2016). All of these are endo-rhamnogalacturonan lyase except YesX from *B. subtilis* which is an exo-rhamnogalacturonan lyase. Rhamnogalacturonan lyases are also found in family PL4. Tertiary structure of YesW (endo-rhamnogalacturonan lyase) from *B. subtilis* , YesX (exo- rhamnogalacturonan lyase) from *B. subtilis* and YesW_Bl (endo-rhamnogalacturonan lyase) from *Bacillus licheniformis* have been determined. They take a common eight bladed β-propeller fold. The structure contains Ca^{2+} ions bound at the blades made up of β-sheets (Silva *et al*., 2016).

1.6 Different types of pectin degrading enzymes

1.6.1 Polygalacturonases

Endo-polygalacturonase (3.2.1.5) hydrolyses the α-(1,4)-glycosidic linkages of polygalacturonic acid randomly resulting in oligogalacturonides. Exo-polygalacturonase (3.2.1.67) cleave the substrate to produce mono-galacturonate (Rombouts and Pilnik, 1980). They belong to family 28 of glycoside-hydrolases (GH28) (http://www.cazy.org/GH28.html)

1.6.2 Pectin methyl esterases

Pectin methyl esterases (PME) (EC 3.1.1.11) cause the de-esterification of the methoxy group substitution of galacturonic acid moiety in pectin. Removal of the methoxy group by PME facilitates polygalacturonases and pectate lyases to cleave the non-esterified polygalacturonan chain (Yadav *et al*., 2009). PME have been classified

under family 8 carbohydrate esterase (CE8) in the CAZy database (http://www.cazy.org/PL8.html).

1.6.3 Pectin acetyl esterases

Pectin acetyl esterases (PAE) (EC 3.1.1.6) remove the acetyl group present at the C-6 carbon of the galacturonic acid moiety of pectin producing pectic acid and acetate (Shevchik *et al*., 1997). They have been classified under family 12 and 13 carbohydrate esterase (CE12 and CE13) in the CAZy database (http://www.cazy.org/html).

1.6.4 Pectate lyases

Endo-pectate lyases (EC 4.2.2.2) randomly cleave polygalacturonic acid while exo-pectate lyases (EC 4.2.2.9) catalyze the cleavage of polygalacturonic acid from non-reducing end (Jayani *et al*., 2005). As a result of their action α-Δ-4,5-unsaturated-Gal*p*A is produced at the non-reducing end (http://www.cazy.org/PL8.html). Pectate lyases have been classified under family 1, 2, 3, 9 and 10 polysaccharide lyase in the CAZy database (http://www.cazy.org/html).

1.6.5 Pectin lyases

Pectin lyases (4.2.2.10) catalyse the breakdown of the α-1,4-glycosidic bond in highly esterified polygalacturonic acid. They have an endo mode of action (Yadav *et al*., 2009). They have been classified under polysaccharide lyase 1 (http://www.cazy.org/PL1.html)

1.6.6 Rhamnogalacturonan I rhamnohydrolases

Rhamnogalacturonan I rhamnohydrolases (EC 3.2.1.174) cause the hydrolytic cleavage of the α-1,4-glycosidic bonds between L-Rha*p* and D-Gal*p*A at the non-reducing terminus releasing single L-Rha*p* residues (Fig. 1.6 A) (Silva *et al*.,

2016). They have been classified into glycoside hydrolase family 78 (GH78) (http://www.cazy.org/GH78.html).

1.6.7 Rhamnogalcturonan I galacturonohydrolases

Rhamnogalcturonan galacturonohydrolases (EC 3.2.1.173) catalyse the removal of terminal non-reducing galacturonosyl residue by breaking of α-1,2-glycosidic bonds between D-Gal*p*A and L-Rha*p* at the non-reducing terminus releasing single D-Gal*p*A residues (Fig. 1.6 B) (Silva *et al*., 2016). They have been classified into glycoside hydrolase family 28 (GH28) (http://www.cazy.org/GH28.html).

1.6.8 Rhamnogalacturonan I endo-hydrolases

Rhamnogalacturonan I endo-hydrolases (EC 3.2.1.171) catalyse the cleavage of α-1,2-glycosidic bonds between D-Gal*p*A and L-Rha*p* in a random fashion releasing oligogalacturonates (Fig. 1.6 C) (Silva *et al*., 2016). They have been classified into glycoside hydrolase family 28 (GH28) (http://www.cazy.org/GH28.html).

1.6.9 Rhamnogalacturonan lyases

Rhamnogalacturonan lyases catalyze the cleavage of α-1,4-glycosidic bond between Rha*p* and Gal*p*A residues of rhamnogalacturonan I main chain through β-elimination reactin mechanism (Fig. 1.6 D) (Silva *et al*., 2016). The oligosaccharides products formed have α-Δ-4,5-unsaturated-Gal*p*A residue at their non-reducing end (Fig 1.6 D). Endo-rhamnogalacturonan (EC 4.2.2.23) lyases cleave the main chain randomly while exo-rhamnogalacturonan lyases (EC 4.2.2.24) specifically cleave at the terminal glycosidic bond. Rhamnogalacturonan lyases are grouped in families 4 and 11 of polysaccharide lyases (PL4 and PL11) (http://www.cazy.org/PL4.html, http://www.cazy.org/PL11.html).

1.6.10 Unsaturated rhamnogalacturonyl hydrolases

Unsaturated rhamnogalacturonyl hydrolases (EC 3.2.1.172) are active only on RG I oligomers with Δ-4,5-unsaturated-Gal*p*A residue at the non-reducing end (Silva *et al*., 2016). They catalyse the cleavage of the α-1,2 glycosidic bond between the Δ-4,5-unsaturated-Gal*p*A and L-Rha*p* releasing single unsaturated D-Gal*p*A (Fig. 1.6 E). They have been classified into glycoside hydrolase family 105 (GH105) (http://www.cazy.org/GH105.html).

1.6.11 Rhamnogalacturonan acetylesterases

Rhamnogalacturonan acetylesterase (EC 3.1.1.86) provide RGI main chain cleaving enzymes access by to hydrolytically cleaving acetyl groups from rhamnogalacturonan I chain (Searle *et al*., 1992). They have been classified into carbohydrate esterase family 12 (http://www.cazy.org).

1.6.12 Arabinases

Endo-arabinases (EC 3.2.1.99) catalyse the hydrolysis of α-1,5-linked arabinan side chains of rhamnogalacturonan I (Silva *et al*., 2016). Terminal non-reducing arabinose residues are removed by a-L-arabinofuranosidases (EC 3.2.1.55). They have been classified into glycoside hydrolase family 43 (GH43) (http://www.cazy.org).

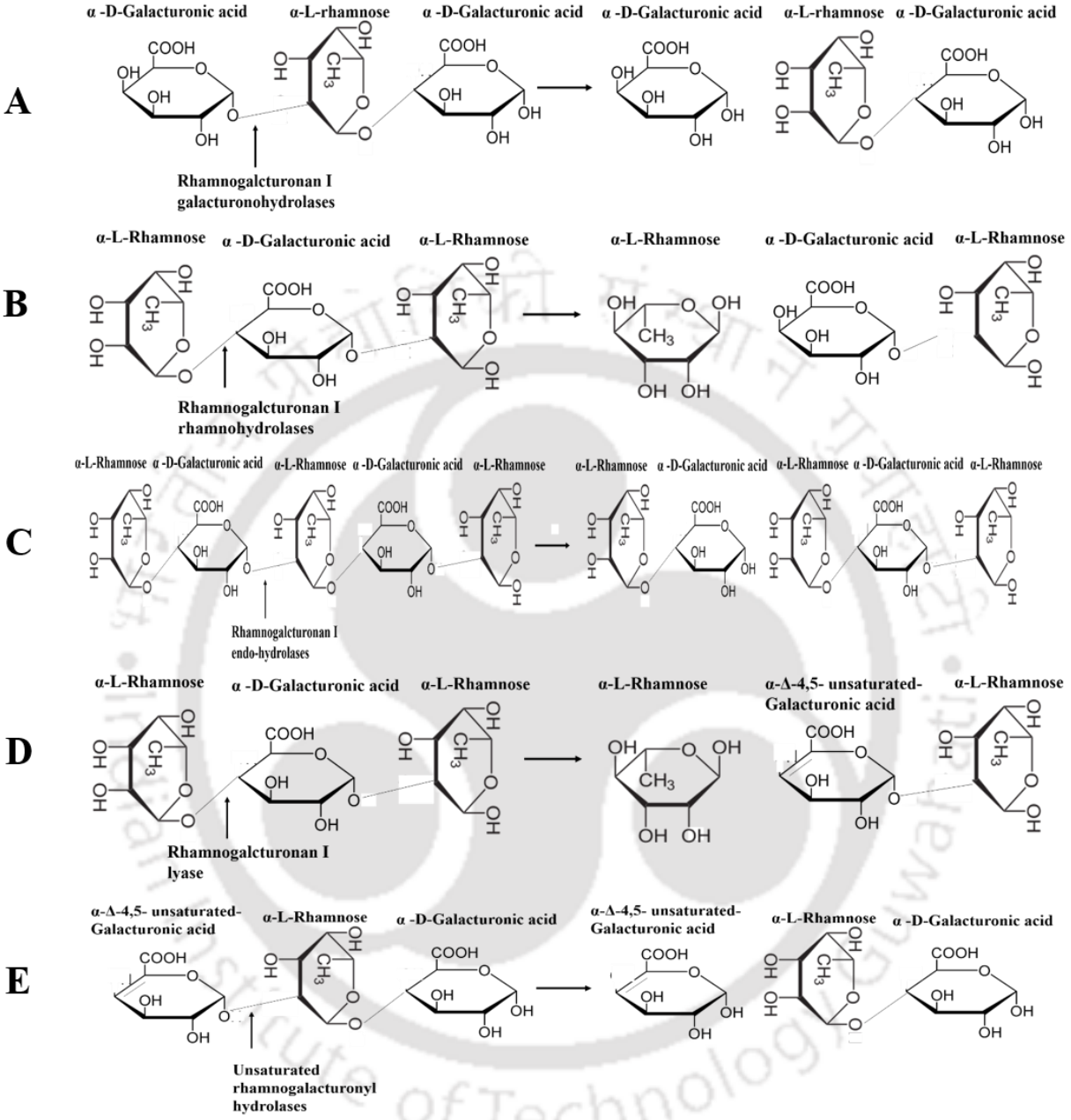


Fig. 1.6 Various types of rhamnogalacturonan I main chain degrading enzymes and their site of action. (A) Rhamnogalacturonan I rhamnohydrolases, (B) Rhamnogalacturonan galacturonohydrolases, (C) Rhamnogalacturonan I endohydrolase, (D) Rhamnogalacturonan I lyase and (E) Unsaturated rhamnogalacturonyl hydrolase. Arrow indicates the attack site on the polysaccharide where the enzymes cleave.

1.6.13 Galactanases

Endo-galactanases (EC 3.2.1.89) cleave the β-1,3-, β -1,4-, or β-1,6 linked galactose residues in the side chains They have been classified into glycoside hydrolase family 53 and 61. β-D-galactosidases (EC 3.2.1.23) remove the β-D-galactose residues from the terminal non-reducing end of galactan side chains (Silva *et al*., 2016). β-D-galactosidases have been classified into glycoside hydrolase family 1, 2 and 42 (GH1 and GH42) (http://www.cazy.org).

1.7 Applications of microbial pectin degrading enzymes

Production of pectinolytic enzymes (enzymes that degrade pectin component of plant cell wall) occupies around 10% of the overall manufacturing of enzyme preparations. These enzymes are widely used in the food industry in the production of juices, fruit drinks and wines (Kashyap *et al*., 2001).

1.7.1 Fruit processing industries

In production of clear apple, grape and pear juice pectinolytic enzymes are used for clarification (Alkorta *et al*., 1998 and Kashyap *et al*., 2001). Grapes have high pectin content, pectinolytic enzymes are used to reduce haze or gelling of grape juice. During cider production pectinolytic enzymes are used to generate coagulum. The coagulum entraps pectate-protein and tannins, which are eventually removed from juice (Grassin and Fauquembergue, 1996).

1.7.2 Oil extraction

Oil extraction from rape seeds, coconut germ, sunflower seed, palm, kernel and olives involves organic solvents such as hexane which is potentially toxic (West, 1996 and Kashyap *et al*., 2001). Olivex, an enzyme preparation containing

pectinolytic enzymes has been shown to increase the yield of extraction with increased content of polyphenols and vitamin E.

1.7.3 Maceration of plant tissue

The process of conversion of organize plant tissue into suspension of intact cells is maceration. The macerated plat tissue is used in pulpy juices, baby foods and yoghurts. Pectinolytic enzymes are used during maceration which allow the increased number of cells to stay intact than mechanical treatment (Kashyap *et al*., 2001).

1.7.4 Coffee and tea processing

During processing of coffee beans pectinolytic enzymes are sprayed over beans to removes the surface layer of pectic substances (Amorim and Amorim, 1977). The pectinolytic enzyme treatment of tea leaves improves quality of instant tea powders by degrading pectic substances (Carr, 1985).

1.7.5 Textile and fiber processing

Bio-scouring is an alternative to alkali assisted removal of pectin associated wax from cotton fibers. Degradation of pectin by pectinolytic enzymes causes removal of hydrophobic wax Chakraborty *et al*., 2015). This reduces the effluent discharge and demand for water (Gurucharanam and Deshpande, 1986). Plant fibres such as ramie, jute and flex need to be process to remove non-cellulosic pectin associated wax before the fibers could be spun into yarn. Pectinolytic enzymes have proven to be an efficient alternative to traditional water retting and alkaline degumming.

1.8 Carbohydrate binding modules

Carbohydrate Binding Modules (CBMs) were discovered as modules bound to cellulose and were referred as Cellulose Binding Modules (van Tilbeurgh *et al*., 1986). They are associated with catalytic modules by the means of short, flexible

linker peptides (Gilbert *et al*., 2013) After the discovery of modules capable of binding other carbohydrates the term Carbohydrate Binding Module was adopted (Boraston *et al*., 1999). CBMs fold independent of their cognate catalytic module and are non-catalytic in nature (Abbott *et al*., 2014). CBMs usually are made up of around 200 amino acids and may be present at C-terminal or N-terminal of a catalytic module (Shoseyov *et al*., 2006). The CBMs have been classified into amino acid sequence based families. Currently (May, 2017), there are 81 families of CBMs and new families are constantly being added to the classification upon discovery of novel CBM sequences (http://www.cazy.org/Carbohydrate-Binding-Modules.html).

CBMs are known to perform three different functions. CBMs have been to shown to enhance polysaccharide degradation by bringing catalytic modules in proximity of substrates (Boraston *et al*., 2004). Role of CBMs in processivity of associated catalytic module has also been reported. CBMs have the potential to alter the specificity of catalytic modules (Boraston *et al*., 2004). The evidence for potential of CBMs to direct their cognate catalytic modules to regions of plant cell wall undergoing degradation is also available in literature (Montanier *et al*., 2009). A CBM has also been reported to facilitate the attachment of its catalytic module to the cell surface (Montanier *et al*., 2009).

Seven different types of folds adopted by different CBMs are known: 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 (Cantarel *et al*. 2009). The β-sandwich fold is most prevalent (Boraston *et al*., 2004). Apart from the sequence based family classification CBMs have also been grouped as Type A, B and C (Fig. 1.7) (Gilbert *et al*., 2013). Type A CBMs recognize crystalline surface of

polysaccahrides, typically presenting a planar ligand binding surface. The Type B CBMs bind internal regions of polysaccharide chains while Type C CBMs bind the terminal residues of polysaccharide chains. For instance, members of family CBM35 could be classified as Type C CBMs as they use their variable loops to interact with monosaccharides and terminal residues of polysaccharides. The web page https://www.cazypedia.org/index.php/Carbohydrate_Binding_Module_Family_35 provids information about structure –function relation in family CBM35 members in a nutshell.

1.9 Applications of carbohydrate binding modules

CBMs have been the subject of various studies and patents aimed at utilization of CBMs for practical applications.

1.9.1 *CBMs for fiber modification*

CBMs from *Clostridium cellulovorans* were fused to produce a cellulose cross-linking protein. The fused CBMs when applied to Whatman filter paper enhanced its mechanical properties, such as tensile strength and ability to stretch. The filter paper was transformed into water-repellent material (Nussinovitch *et al*., 2002). A fusion protein containing a CBM and a cutinase have been used for degradation of polyethylene terephthalate (PET) fiber which is major land and water pollutant (Ribitsch *et al*., 2013).

1.9.2 *CBMs for bioremediation*

A fusion construct of CBM and an organophosphorus hydrolase has been used for degradation of nerve gas (Wang *et al*., 2002). Bacterial cells displaying the fusion enzyme were immobilized on cellulose matrix allowing organophosphorus hydrolase

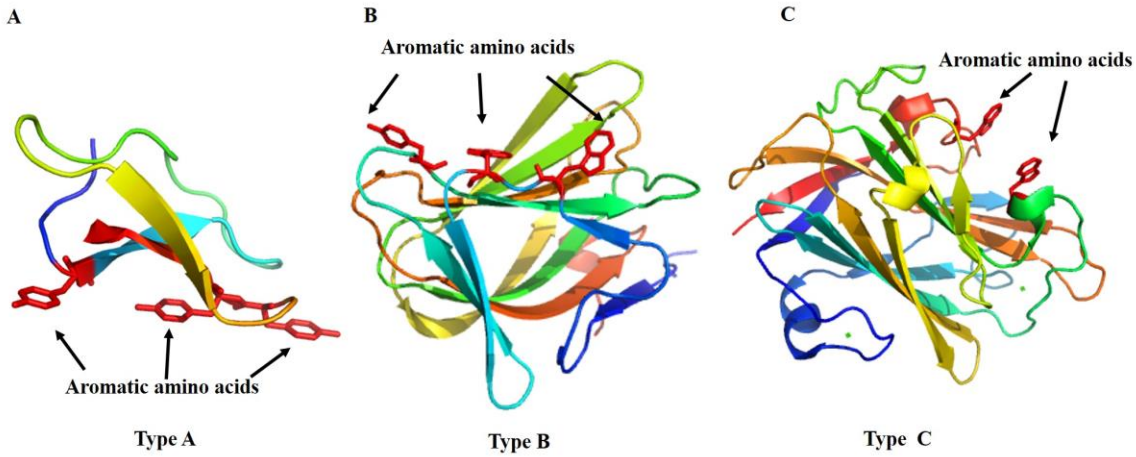


Fig. 1.7 Representation of the binding sites of three classes of CBMs based on the type of ligands they bind. (A) Type A CBMs bind crystalline surface of polysaccahrides (e.g. PDB code 1CBH), (B) Type B CBMs bind internal regions of polysaccharide chains (e.g. PDB code 1GWL) and (C) Type C CBMs bind terminal residues of polysaccharides (e.g. PDB code 1I82). Aromatic amino acids involved in ligand binding have been shown in red.

to degrade nerve gas. A fusion protein of atrazine chlorohydrolase (AtzA) a CBM immobilized on cellulose matrix was shown to dechlorinate atrazine (Kauffmann *et al*., 2000).

1.9.3 *CBMs as molecular probes*

Fluorescent protein (mOrange2) tagged CBM have been demonstrated to detect surface exposed xylan in kraft pulp thereby facilitating optimization of wood treatment process in paper industry (Khatri *et al*., 2016). The use of his-tagged CBMs for studying localization of various carbohydrate ligands of plant cell wall has been explored (McCartney *et al*., 2004).

1.9.4 *CBMs in detergent industry*

Patents have also described use of CBMs for targeting amylases, proteases present in detergents to cellulosic textile fibers for removal of starch and protein impurities (Osten *et al*., 2000). Addition of CBMs, carrying attached fragrant

particles, to detergents has shown to reduce cost and amount of fragrant particles in the detergent to be used as the CBM keep the pleasant smelling particle immobilized on the fabric (Berry *et al*., 2001).

1.9.5 *CBMs as affinity tags*

Proteins fused to CBMs have been purified where CBMs act as affinity tags (Boraston *et al*., 2001). A CBM from *Clostridium perfringens* has been developed as antibody mimetic for protein purification (Suderman *et al*., 2017)

1.10 *Clostridium thermocellum (renamed as Ruminiclostridium thermocellum)*

Clostridium thermocellum is an anaerobic, Gram positive, rod shaped, thermophillic bacterium (Fig. 1.8). It was isolated in 1926 by Viljoen et al. while identifying novel organisms capable of degrading cellulose. *C. thermocellum* is able to produce ethanol directly from cellulose (Bayer *et al*., 2000). It was later established that it could also grow on cellobiose, xylose and hemicelluloses. Ethanol is one of the main product of fermentation by *C. thermocellum* due to which it has attracted scientific attention for development into a potential organism for Consolidated Bioprocessing (Akinosho *et al*., 2014). *C. thermocellum* displays on its cell surface a muti-enzyme complex called as cellulosome (Lamed *et al*., 1983) (Fig. 1.9).

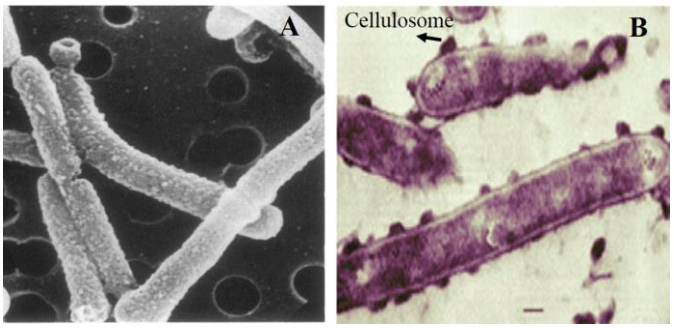


Fig. 1.8 (A) Scanning electron microscope (SEM) images rod shaped, Gram positive of *Clostridium thermocellum* ATCC 27405 (adapted from Lamed *et al*., 1987); (B) Cellulosome at the surface of *Clostridium thermocellum* cells (adapted from Fontes *et al*., 2010).

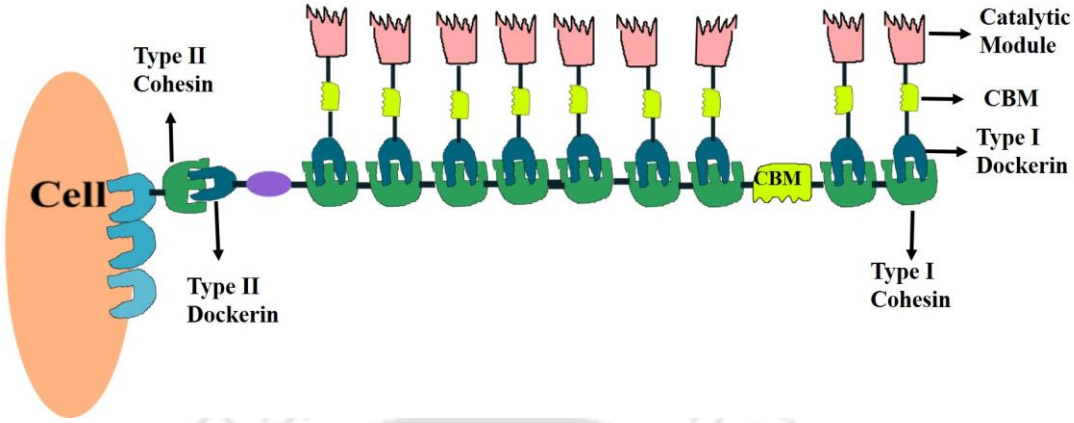


Fig. 1.9 Structure of Cellulosome from *Clostridium thermocellum*, where multiple enzymatic subunits are attached to a Scaffoldin subunit connected by cohesion-dockerin interaction. In the figure CBM is a carbohydrate binding module.

This cellulosome is known to be constituted by cellulases, hemicellulases, pectinases, chitinases, glycosidases, and esterases which assemble on a scaffold polypeptide (CipA). The cellusomal enzymes are modular in nature. The molecular weight of cellulsome has been estimated to be greater of than 2×10^6 Da (Akinosho *et al*., 2014). The assembly of such a big protein complex is made possible by two very specific interactions. The first type of interaction is between type I Dockerin module of cellusomal enzymes and type I Cohesin modules of the scaffold protein (CipA). The second type of interaction is between the type II Dockerin at the C-terminal of CipA and type I Cohesin module present on the cell surface (Fontes *et al*., 2010). The *C. thermocellum* genome encodes ~70 cellulsomal proteins (Hirano *et al*., 2016). It has been reported that composition of the cellulosome varies according to the carbon source on which the *C. thermocellum* cells are grown. The cellulosomal complex has also been reported from many microorganisms such as *C. cellulouorans, C. alkalicellulosi* and *C. clariflavum* (Bayer *et al*., 1994, Artzi *et al*., 2017).

1.11 Significance and objectives of the present study

1.11.1 Significance of this study

Pectin degrading enzymes find application in various industrial processes like bioscouring, juice extraction, vegetable puree production, fiber retting and waste water treatment and many others. Therefore, the search of better enzymes that degrade various components of pectin, is important. Several enzymes from the cellulosomal complex of *Clostridium thermocellum* have been studied but never an enzyme from PL11 family. *Clostridium thermocellum* genome contains a gene encoding a putative rhamnogalacturonan lyase which has been classified under family 11 polysaccharide lyase. Till date this type of enzyme has not been reported from *Clostridium thermocellum.* Cloning and purification of this enzyme would enable to characterize its biochemical and structural aspects. This will enable to make desirable changes at the DNA level, to obtain an enzyme which can easily fulfill the need of modern day food, beverage, textile and oil industries. Determination of the *in-silico* structure will help in understanding the key residues involved in catalysis and the mode of enzyme action. This PL11 enzyme (*Ct*RGL*f*) from *Clostridium thermocellum* contains putative family 35 Carbohydrate Binding Module (*Rgl*-CBM35). It has been proved through different studies that CBMs facilitate the binding of enzymes on the substrate resulting in an enhanced activity (Boraston *et al*., 2004). Characterization of *Rgl*-CBM35 could throw light on its possible novel substrate specificity. The proposed study involves the cloning, expression and purification of a modular family 11 polysaccharide lyase, *Ct*RGL*f* (GenBank Acession number ABN51485.1, Uniprot ID A3DC06) its associated carbohydrate binding module (*Rgl*-CBM35) and the catalytic module (*Ct*RGL) from *Clostridium thermocellum* ATCC 27405 genome. BLAST analysis

showed that *Ct*RGL*f* is a putative rhamnogalacturonan lyase. Biochemical and functional characterization of the full length *Ct*RGL*f* and its truncated derivative *Ct*RGL will be carried out to determine their activity against different pectic polysaccharides and to understand their cleavage pattern. Application of *Ct*RGL*f* in bioscouring of cotton fabric will be studied. Binding assays of the CBM, *Rgl*-CBM35 will be carried out to determine its affinity towards various plant cell wall carbohydrates. Site- directed mutagenesis of *Rgl*-CBM35 will be undertaken to determine the key ligand binding residues.

The reasons for selecting the family 11 polysaccharide lyase (PL11) and associated family 35 carbohydrate binding modules (CBM35) from *Clostridium thermocellum* are summarized:

1. *Ct*RGL*f* from *Clostridium thermocellum* is a putative cellulosomal enzyme which are known to be robust and efficient in degrading polysaccharides (Fontes *et al*., 2010).

2. Functional characterization of family 11 polysaccharide lyase (PL11) is important as all the enzymes belonging to family 11 PL may have the same β-elimination mechanism of catalysis but the degradation products will indicate its exo or endo mode during cleavage.

3. Family 11 PLs are generally rhamnogalacturonan lyases which are essential for complete degradation of pectin, hence they can be widely used in removal of pectic polysaccharide during various industrial processes.

4. It is important to determine the role of CBM35 in substrate binding and also if they influence the activity of the catalytic module.

1.11.2 Specific objectives

1. Cloning and expression of family 11 polysaccharide lyase full length module (*Ct*RGL*f*) and its truncated derivatives (*Ct*RGL and *Rgl*-CBM35) from *Clostridium thermocellum* genomic DNA.
2. Biochemical and functional characterization of catalytic modules, full length (*Ct*RGL*f*) and the truncated catalytic module (*Ct*RGL).
3. Structure analysis and substrate binding analysis of Carbohydrate Binding Module, *Rgl*-CBM35.
4. Structure modeling and ligand docking study of catalytic (*Ct*RGL) module.
5. Application of *Ct*RGL*f* in bioscouring of cotton fabric and degumming of jute.

References

Abbott, D. W., van Bueren, A.L. (2014) Using structure to inform carbohydrate binding module function. Current Opinion in Structural Biology, 28:32-40

Akinosho, H., Yee, K., Close, D., Ragauskas, A. (2014) The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. Frontiers in Chemistry, 2.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K.,Walter, P. (1997) Molecular Biology of the Cell Garland Science, New York, 2002.

Alkorta, I., Garbisu, C., Llama, M. J., Serra, J. L. (1998) Industrial applications of pectic enzymes: a review. Process Biochemistry, 33: 21-28.

Amorim, H .V., Amorim, V. L. (1977) Coffee enzyme and coffee quality. In Enzymes in Food and Beverage Processing. Ori, R., St. Angelo, J. Ed. ACS Symposium Series, 47:27-56.

Artzi, L., Bayer, E. A., Moraïs, S. (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nature Reviews Microbiology, 15: 83.

Bayer, E. A., Morag, E., Lamed, R. (1994) The cellulosome-a treasure-trove for biotechnology. Trends in Biotechnology, 12(9): 379-386.

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

Bayer, E.A., Shoham, Y., 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.

Berry, M. J., Davis, P. J., Gidley, M. J. (2001) *U.S. Patent No. 6,225,462*. Washington, DC: U.S. Patent and Trademark Office.

Brett, C. T., Waldron, K. W. (1996) Physiology and biochemistry of plant cell walls, 2nd ed. Chapman and Hall, London.

Boraston, A. B., McLean, B. W., Kormos, J. M., Alam, M., Gilkes, N. R., Haynes, C. A., Tomme, P., Kilburn, D. G., Warren, R. A. J. (1999) Carbohydrate-binding modules: diversity of structure and function. In recent advances in carbohydrate bioengineering (Gilbert, H. J., Davies, G. J., Henrissat, B. and Svensson, B., eds.), Royal Society of Chemistry, Cambridge, 202–211.

Boraston, A. B., McLean, B. W., Guarna, M. M., Amandaron-Akow, E., Kilburn, D. G. (2001) A family 2a carbohydrate-binding module suitable as an affinity tag for proteins produced in *Pichia pastoris*. Protein Expression and Purification, 21(3): 417-423.

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

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

Carr, J.G. (1985) Tea, coffee and cocoa. In: Wood BJB, editor. Microbiology of fermented foods, vol. 2. London: Elsevier Science Ltd., p. 133–154.

Chakraborty, S., Jagan Mohan Rao, T., Goyal, A. (2017) Immobilization of recombinant pectate lyase from *Clostridium thermocellum* ATCC-27405 on magnetic nanoparticles for bioscouring of cotton fabric. Biotechnology progress, 33(1): 236-244.

Cosgrove, D.J. (1999) Enzymes and other agents that enhance cell wall extensibility. Annual Review of Plant Physiology and Plant Molecular Biology, 50: 391–417.

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

Davies, G. J., Wilson, K. S., Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochemical Journal, 321(2):557-559.

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

Gacesa, P. (1987). Alginate-modifying enzymes: a proposed unified mechanism of action for the lyases and epimerases. FEBS Letters, 212(2):199-202.

Gilbert, H.J., Knox, J.P., Boraston, A.B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate binding modules. Current Opinion in Structural Biology, 23: 669–677.

Grassin, C., Fauquembergue, P. (1996) Application of pectinases in beverages. In Visser J., Voragen , A, Ed. Pectins and Pectinases, Elsevier, 14:453-462.

Gurucharanam, K., Deshpande, K. S. (1986) Polysaccharides of *Curbularia lunata*: Use in degumming of ramie fibers. Indian Journal of Phytopathology, 39(3): 385-389.

Hirano, K., Kurosaki, M., Nihei, S., Hasegawa, H., Shinoda, S., Haruki, M., Hirano, N. (2016) Enzymatic diversity of the *Clostridium thermocellum* cellulosome is crucial for the degradation of crystalline cellulose and plant biomass. Scientific Reports, 6.

Jayani, R.S., Saxena, S., Gupta, R. (2005) Microbial pectinolytic enzymes: A review. Process Biochemistry, 40: 2931–44.

Johansen, K. S. (2016). Lytic Polysaccharide Monooxygenases: The Microbial Power Tool for Lignocellulose Degradation. Trends in Plant Science, 21(11): 926-936.

Kashyap, D.R., Vohra, P.K., Tewari, R. (2001) Application of pectinases in the commercial sector: a review. Bioresource Technology, 77: 215–27.

Kauffmann, C., Shoseyov, O., Shpigel, E., Bayer, E. A., Lamed, R., Shoham, Y., Mandelbaum, R. T. (2000) Novel methodology for enzymatic removal of atrazine from water by CBD-fusion protein immobilized on cellulose. Environmental Science & Technology, 34(7): 1292-1296.

Khatri, V., Hébert-Ouellet, Y., Meddeb-Mouelhi, F., Beauregard, M. (2016) Specific tracking of xylan using fluorescent-tagged carbohydrate-binding module 15 as molecular probe. Biotechnology for Biofuels, 9(1):74.

Lamed, R., Setter, E., Bayer, E. A. (1983). Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. Journal of Bacteriology, 156(2), 828-836.

Lamed, R., Naimark, J., Morgenstern, E., Bayer, E.A. (1987) Specialized cell surface structures in cellulolytic bacteria. Journal of Bacteriology, 169: 3792–3800.

Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M., Henrissat, B. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. Biochemical Journal, 432(3): 437–444.

Matsunaga, T., Ishii, T., Matsumoto, S., Higuchi, M., Darvill, A., Albersheim, P., O'Neill, M.A. (2004) Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes, and bryophytes. Implication for the evolution of vascular plants. Plant Physiology, 134: 339–351.

McCartney, L., Gilbert, H. J., Bolam, D. N., Boraston, A. B., & Knox, J. P. (2004) Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. Analytical Biochemistry, 326(1): 49-54.

Mohnen, D., Bar-Peled, M., Somerville, C. (2008) Biosynthesis of Plant Cell Walls. Biomass Recalcitrance, Chapter 5 (eds. Himmel, M.), pp. 94–187. Blackwell Publishing, Oxford.

Moreira, L. R. S. (2008) An overview of mannan structure and mannan-degrading enzyme systems. Applied Microbiology and Biotechnology, 79(2): 165.

Montanier, C., Van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Gloster, T. M. (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proceedings of the National Academy of Sciences, 106(9): 3065-3070.

Nakamura, A., Furuta, H., Maeda, H., Takao, T., Nagamatsu, Y. (2002) Analysis of the molecular construction of xylogalacturonan isolated from soluble soybean

polysaccharides. Bioscience, Biotechnology and Biochemistry, 66(5): 1155-1158.

Nakamura, A., Furuta, H., Maeda, H., Takao, T., Nagamatsu, Y. (2002) Structural studies by stepwise enzymatic degradation of the main backbone of soybean soluble polysaccharides consisting of galacturonan and rhamnogalacturonan. Bioscience, Biotechnology and Biochemistry, 66(6): 1301-1313.

Nelson, D. L., Lehninger, A. L., Cox, M. M. (2008) Lehninger Principles of Biochemistry. Macmillan.

Nussinovitch, A., Shpigel, E., Shoseyov, O. (2002) Recombinant cellulose crosslinking protein: a novel paper-modification biomaterial. Cellulose, 9(1): 91-98.

O'Neill, M., Albersheim, P., Darvill, A. (1990) The pectic polysaccharides of primary cell walls. In: Methods in Plant Biochemistry, Vol. 2. (eds. Dey D.M.), pp: 415–441. Academic Press, London.

O'Neill, M. A., York , W. S (2003) The composition and structure of plant primary cell walls. In The Plant Cell Wall. Edited by Rose JKC. Ithaca, New York: Blackwell Publishing/CRC Press; 2003:1-54.

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

O'sullivan, A. C. (1997) Cellulose: the structure slowly unravels. Cellulose, 4(3): 173-207.

Osten, C., Bjornvad, M. E., Vind, J., & Rasmussen, M. D. (2000) U.S. Patent No. 6,017,751. Washington, DC: U.S. Patent and Trademark Office.

Petkowicz, C. D. O., Reicher, F., Chanzy, H., Taravel, F. R., Vuong, R. (2001) Linear mannan in the endosperm of *Schizolobium amazonicum*. Carbohydrate Polymers, 44(2): 107-112.

Popper, Z. A., Fry, S. C. (2008) Xyloglucan-pectin linkages are formed intra-protoplasmically, contribute to wall-assembly, and remain stable in the cell wall. *Planta*, *227*(4), 781-794.

Ribitsch, D., Yebra, A. O., Zitzenbacher, S., Wu, J., Nowitsch, S., Steinkellner, G., Gruber, K. (2013) Fusion of binding domains to *Thermobifida cellulosilytica* cutinase to tune sorption characteristics and enhancing PET hydrolysis. Biomacromolecules, 14(6): 1769-1776.

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

Rombouts, F.M., Pilnik, W. (1980) Pectic enzymes. In: Rose AH, Ed. Microbial Enzymes and Bioconversions. Academic Press London, 5: 227–72.

Scheller, H.V., Ulvskov, P. (2010) Hemicellulose. Annual Review Plant Biology, 61: 263–289.

Searle-Van Leeuwen, M.J.F, Van Den Broek, L.A.M, Schols H.A., Beldman, G., Voragen, A.G.J. (1992) Rhamnogalacturonan acetylesterase: a novel enzyme from *Aspergillus aculeatus*, specific for the deacetylation of hairy (ramified) regions of pectins. Applied Microbiology Biotechnology, 38: 347-9.

Shevchik, V.E., Hugouvieux-Cotte-Pattat, N. (1997) Identification of a bacterial pectin acetyl esterase in *Erwinia chysanthemi* 3937. Molecular Microbiology, 24(6): 1285–1301.

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

Silva, I. R., Jers, C., Meyer, A. S., Mikkelsen, J. D. (2016) Rhamnogalacturonan I modifying enzymes: an update. New Biotechnology, 33(1), 41-54.

Sinitsyna, O.A., Fedorova, E.A., Semenova, M.V., Gusakov, A.V., Sokolova, L.M., Bubnova, T.M., Okunev, O.N., Chulkin, A.M., Vavilova, E.A., Vinetsky, Y.P., Sinitsyn, A.P. (2007) Isolation and characterization of extracellular pectin lyase from *Penicillium canescens*. Biochemistry (Moscow), 72(5): 565–71.

Suderman, R. J., Rice, D. A., Gibson, S. D., Strick, E. J., Chao, D. M. (2017) Development of polyol-responsive antibody mimetics for single-step protein purification. Protein Expression and Purification, 134: 114-124.

Sutherland, I. W. (1995) Polysaccharide lyases. FEMS Microbiology Review, 16: 323–347.

Taiz, L., Zeiger, E. (2010) Cell walls: structure, biogenesis, and expansion. Plant Physiology. Sinauer Associates, Sunderland, 327.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., Boerjan, W. (2010) Lignin biosynthesis and structure. Plant Physiology, 153(3): 895-905.

Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R. and Pettersson, G. (1986) Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. FEBS Letters, 204: 223–227.

Viljoen, J. A., Fred, E. B., Peterson, W. H. (1926) The fermentation of cellulose by thermophilic bacteria. The Journal of Agricultural Science, 16(1): 1-17.

Wade Jr, L.G. (1999) Carbohydrates and Nucleic Acids. In Jaworski A, Ed. Organic Chemistry. Prentice-Hall Inc, 5:1101-1154.

Wang, A. A., Mulchandani, A., Chen, W. (2002) Specific adhesion to cellulose and hydrolysis of organophosphate nerve agents by a genetically engineered *Escherichia coli* strain with a surface-expressed cellulose-binding domain and organophosphorus hydrolase. Applied and Environmental Microbiology, 68(4): 1684-1689.

Willats, W.G.T, McCartney, L., Mackie, W., Knox, J.P. (2001) Pectin: cell biology and prospects for functional analysis. Plant Molecular Biology, 47: 9–27.

Yadav, S., Yadav, P. K., Yadav, D., Yadav, K. D. S. (2009) Pectin lyase: a review. Process Biochemistry, 44(1): 1-10.

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

Yip, V.L.Y., Varrot, A., Davies, G.J., Rajan, S.S., Yang, X.J., Thompson, J., Anderson, W.F., Withers, S.G. (2004) An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 β-glycosidase from *Thermotoga maritima*. Journal of American Chemical Society, 126: 8354–8355.

Zandleven, J., Beldman, G., Bosveld, M., Schols, H. A., Voragen, A. G. J. (2006) Enzymatic degradation studies of xylogalacturonans from apple and potato, using xylogalacturonan hydrolase. Carbohydrate Polymers, 65(4): 495-503.

Chapter 2

Cloning, expression and purification of family 11 polysaccharide lyase, full length module (*Ct*RGL*f*), its truncated catalytic module (*Ct*RGL) and carbohydrate binding module (*Rgl*-CBM35) from *Clostridium thermocellum* ATCC 27405

2.1 Introduction

The major components of plant cell wall (PCW) are the polysaccharides that assemble to form a network. These polysaccharides are cellulose, hemicellulose and pectin. The PCW is able to fulfil its structural role by providing strength and protection to the cell owing to the presence of cellulose and hemicellulose (Vincken *et al*., 2003). PCW has several functional roles such as cell-cell adhesion, cell signalling, wall porosity, pollen tube growth, leaf abscission. The pectin component of the cell wall has been credited to play these roles (Ridley *et al*., 2001). Structural role of pectin in promoting upright growth of plants has also been reported (Matsunaga *et al*., 2004). Pectin is predominantly localised in the primary cell wall of all higher plants, gymnosperms, pteridophytes and bryophytes (O'Neill *et al*., 1996**)**. The principal components of pectin are homogalacturonan (HG) and the substituted HG. The substituted HG are of two types: Rhamnogalacturonan I (RG I)

and Rhamnogalacturonan II (RG II). Xylogalacturonans and apiogalacturonans are other less abundant substituted HG (Oomen *et al*., 2002). Homogalacturonan is a homopolymer of α-(1→4) linked D-galactopyranosyluronic acid (D-Gal*p*A) residues which may be methylated at C-6. RG I is composed of a main chain of alternating L-rhamnopyranosyl (L-Rha*p*) and D-Gal*p*A residues. The repeating monomeric unit is a disaccharide [→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→]. 20-80% of L-Rha*p* residues are substituted with individual, linear or branched chains of L-arabinofuranosyl (L-Ara*f*) and D-Gal*p*. The side chains of RG I may be substituted with L-fucose or D-glucuronic acid (GlcA) (O'Neill *et al*., 1996). RG I has been demonstrated to be crucial for normal development of periderm in potato (Oomen *et al*., 2002). Transgenic potato plants that expressed RG I cleaving enzyme developed morphological abnormalities like swelling of periderm cells. RG II main chain is also composed of α-(1→4) linked D-Gal*p*A with L-Rha*p* residues as substitutions apart from 12 other different monosaccharide residues (O'Neill *et al*., 1996).

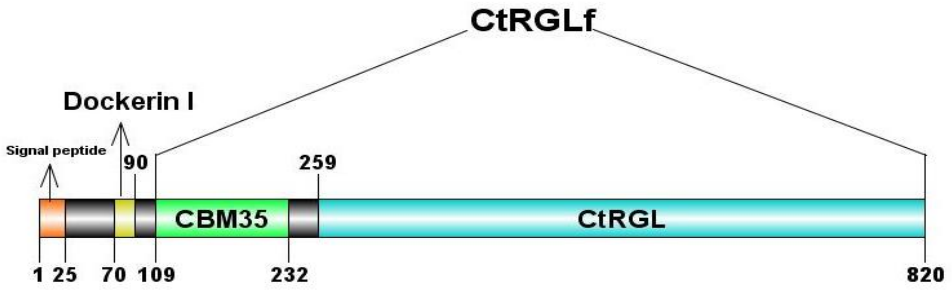
Degradation of plant cell wall makes a reservoir of nutrients available for recycling. Nature has bestowed a diverse group of microorganisms with enzymes to breakdown the plant cell wall polysaccharides (Ochiai *et al*., 2007**,** Pages *et al*., 2003 and McKie *et al*., 2001). When pectin comes under microbial attack both glycoside hydrolases and polysaccharide lyases are recruited. Glycoside hydrolases cleave the glycosidic bonds via an acid-base catalytic mechanism (Koshland *et al*., 1953). Polysaccharide lyases cleave their substrates *via* a β-elimination mechanism, generating a double bond between C-4 and C-5 in the residue at the non-reducing end (Moran *et al*., 1968). Glycoside hydrolases and polysaccharide lyases have been classified into different families based on sequence similarity (Lombard *et al*., 2014).

Many of the plant cell wall polysaccharide degrading enzymes are modular in nature and have one or more specialised substrate binding module(s) referred to as Carbohydrate Binding Module(s) (CBM) in addition to a catalytic module.

The modular protein, cthe_0246 (GenBank accession number ABN51485.1 and Uniprot ID A3DC06) of *Clostridium thermocellum* ATCC 27405 is a modular carbohydrate-active enzyme. Its nucleotide and protein sequences were retrieved from the CAZy database and analyzed by different computational approaches including the BLAST tool. BLAST analysis revealed that the protein, cthe_0246 contained a N-terminal Dockerin I followed by a family 35 carbohydrate binding module (*Rgl*-CBM35) and a putative rhamnogalacturonan lyase module (*Ct*RGL) belonging to polysaccharide lyase family 11. These modules are connected by linker sequences.

A schematic representation of the molecular architecture of *Ct*RGL*f* is presented in Fig. 2.1. The N-terminal cleavage site of the signal peptide was predicted between Ala25 and Gly26 by SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP-3.0/). Conserved Domains Database was referred for determining the expanse of domains (http://www.ncbi.nlm.nih.gov/cdd/). Towards the N-terminal, downstream of the signal peptide a stretch of 20 amino acids showed similarity to Dockerin I from *C. thermocellum* (Fig. 2.1). The second module spanning from amino acids residue 109 to 232 was identified to be a family 35 CBM. Towards C-terminal the catalytic module ranges from amino acid residue 259 to 820. The interaction between enzymes borne Dockerin modules and the Cohesin modules of a scaffold protein gives rise to the celluosomal complex (Fontes *et al*., 2010). The presence of a putative Dockerin I at N-terminal and the subcellular

localisation score predicted by PSORT server indicated that the protein encoded by sequence Cthe_0246 (*Ct*RGL*f* is a derivative of Cthe_0246 without the N-terminal Dockerin I) is an extracellular enzyme and probably integrates as a component of the *C. thermocellum* cellulosome.



MKKTLVFLTALSLIFTLFISYSLSAGPASTKYGDLNADGKINSTDYNLGKRLILRTISELPIS
NGSVAFDLNGDSKVDSTDLTALKRYLLGVIDKFPVGTDIPSQTQKTRYQAEDAMLYKAFEETI
HAGYDGRSYVNYDNEPGGYIEWNVNVSSSGTYKLIFRYANGSNNNRPMEIRVNSNLVAGSLDF
YPTSAWTVWNDQSIVVTLNAGNNVIRATGIASDGGPNVDYLEVIPTNEPPAPTPSPTPTVGPT
PAGARQMERLDRGLVAVKVNNGVFLSWRMFGTDPSNIAFNLYRNGTKINSTPITGATNYVDTG
GTTSSTYTVRAVINGQEQEASKPVSVWAQNYLQIPIQPPSSAYEANDCSAADLDGDGEYEIVL
KWEPNNAKDNSQSGYTDNVYLDAYKLNGTRLWRIDLGRNIRAGAHYTQFMVYDLDGDGKAEVA
CKTADGTRDGKGNVIGNPNADYRNSSGYILSGPEYLTVFDGQTGAAITTVDYDPPRGNVSSWG
DNYGNRVDRFLACIAYLDGQRPSLVMCRGYYTRSVLVAWDFRNGRLTKRWVFDGNNYSGYNGQ
GNHNLSVADVDGDGRDEIIYGACTIDDNGKGLYTSGLGHGDALHVGDLNPNRPGLEIWSCFES
SGGAALRDARTGEVLFRWHRSSDTGRACAADITASSPGAELWAAGSPLFSCTGQNIGTAPSQI
NFAIWWDGDELRELLDGITISKYGVGTLFTATGCASNNGTKSTPCLQADLLGDWREEVIFRTS
DNRYLRIYTTTATTNRRIYTLMHDPVYRLGIAWQNVAYNQPPHTSFFIGAGMAEPPKPNIYLV
P

Fig. 2.1 Molecular architecture and sequence of protein cthe_0246 showing boundaries and designation of different domains. Same color codes have been used in molecular architecture and protein sequence to denote domain boundaries. Linkers are shown in black colour.

In the present study the genes encoding *Ct*RGL*f* and its truncated derivatives *Ct*RGL and *Rgl*-CBM35 were cloned. All the proteins were expressed in *Escherichia coli* and purified by immobilized metal ion affinity chromatography (IMAC) for further biochemical, functional and structural characterization.

2.2 Materials and Methods

2.2.1 Chemicals, reagents and kits

The oligonucleotide primers for PCR amplification of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 were procured from Eurofins, India. BIOTAQ DNA polymerase was supplied by Biolone, UK. dNTPs and $MgCl_2$ were obtained from Sigma-Aldrich, India. PCR tubes (0.2 ml) were from Axygen, Germany. Restriction enzymes *Nhe*I, and *Xho*I were purchased from Promega, USA. The expression vector, pET28a(+) was purchased from Novagen, Germany. T_4 DNA ligase and 10x ligase buffers 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, nuclease free water (pH 8.0) and components of polyacrylamide gel electrophoresis were obtained from Sigma-Aldrich, India. The GenElute miniprep asmid isolation kit and GenElute gel-extraction kit was from Sigma-Aldrich, India. DNA was electrophoresed on agarose gels prepared using Agarose, with low EEO purchased from Sigma-Aldrich, India. DNA marker, Hyperladder I was purchased from Bioline, UK. Protein markers were procured from Fermentas, Canada and Bangalore GeNei, India. Disodium ethylenediamine tetra acetate salt (EDTA), glucose, sodium hydroxide, sodium dodecyl sulphate (SDS), LB medium and SOC medium components were supplied by Himedia Pvt. Ltd., India. The antibiotic, kanamycin was procured from Sigma-Aldrich, India. SDS-PAGE was performed using Mini-PROTEAN Tetra Cell purchased from Bio-Rad Laboratories (India) Private Limited. The protein staining dye Coomassie Brilliant Blue R250 was procured from Himedia Pvt. Ltd., India and methanol from Merck, India. The genomic DNA of *Clostridium thermocellum* ATCC 27405 was purchased

from Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures.

2.2.2 Microorganisms

Commercially available *E. coli* DH5α, *E. coli* BL-21 (DE3) and *E. coli* BL-21 (DE3) pLysS cells were obtained from Novagen, Germany.

2.2.3 PCR amplification of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35

The genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 were amplified with designed oligonucleotide primers using the *Clostridium thermocellum* ATCC 27405 genomic DNA as template. The primers contained the restriction enzyme sites of *Nhe*I and *Xho*I as mentioned in Table 2.1. Amplification of each module was done following the scheme presented in Fig. 2.2. Components of 50 µl PCR reaction mixture and the PCR cycles for amplification are given in Tables 2.2 and 2.3, respectively. PCR amplification was performed in a thermal cycler (Applied Biosystems, GeneAmp PCR System 9700). The PCR amplicons were electrophoresed on a 0.8 % (w/v) agarose gel along with a DNA marker (Hyperladder I) as mentioned in Section 2.2.4.

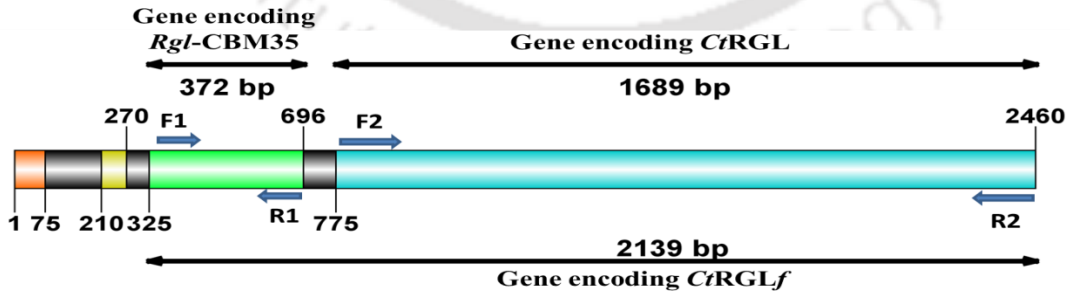


Fig. 2.2 Schematic presentation showing primers used for PCR amplification of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 from 2460 bp gene sequence of protein ABN51485.1 of *Clostridium thermocellum.*

Table 2.1 Oligonucleotide primers used for PCR amplification of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 from *Clostridium thermocellum*, nucleotides shown in bold represent the restriction enzyme sites.

Module	Primer name	Primer sequence
*Ct*RGL*f*	F1	5'- cg**gctagc**acgagatatcaggctgagg-3'
	R2	5'- cc**ctcgag**ttacggcacaaggtaaatatttgg-3'
*Ct*RGL	F2	5'- cg**gctagc**atggagagactggacagag-3'
	R2	5'- cc**ctcgag**ttacggcacaaggtaaatatttgg-3'
Rgl-CBM35	F1	5'- cg**gctagc**acgagatatcaggctgagg-3'
	R1	5'- cc**ctcgag**tta acttcaagataatccacattcgg-3'

Table 2.2 PCR reaction setup for amplification of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 from *Clostridium thermocellum.*

PCR components	Volume (µl)	Final concentration
10x reaction buffer	5.0	1x
$MgCl_2$ (50 mM)	1.5	1.5mM
dNTP mix (100 mM)	1.0	2 mM
Forward primer (15 µM)	1.5	0.45 µM
Reverse primer (15 µM)	1.5	0.45 µM
Sigma water, pH 8.0	36.5	--
Genomic DNA (15 µg/ml)	0.5	7.5 ng
Taq DNA polymerase (2.5 U/µl)	2.5	6.25 U
Total	50.0	--

Table 2.3 Conditions for PCR thermal cycles for amplification of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 from *Clostridium thermocellum.*

Steps		Time (min)
I. Denaturation at 94°C		4
II. 30 cycles of	i) Denaturation at 94°C	0.5
	ii) Annealing at: 49°C	1
	iii) Extension at 72°C	2
III. Final extension at 72°C		10

2.2.4 Agarose gel electrophoresis of PCR amplified products

The PCR amplified products were electrophoresed on 0.8% (w/v) agarose gel prepared in 1x TAE buffer. A stock solution of TAE buffer was prepared according to Sambrook and Russell (2001) keeping the concentrations of components to 10x

(400 mM Tris-acetate, 10 mM EDTA, pH 8.0). A gel was prepared by dissolving agarose (400 mg for 0.8% (w/v) and 500 mg for 1.0% (w/v) gel) in 50 ml of 1x TAE buffer by heating in a microwave oven 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 well and poured on the casting apparatus, comb was placed and the gel was allowed to solidify. 1x TAE (Tris-acetate-EDTA) buffer was used for preparation of agarose gels and also as an electrophoresis buffer (Sambrook and Russel, 2001). The DNA sample and DNA loading dye were mixed in 4:1 ratio and the gel was run at constant 40 Volt till the dye migrated over 70% of the gel length. The DNA bands were then visualized under UV illumination in a gel documentation system (BioRad XR).

2.2.4.1 DNA loading buffer

The DNA or sample loading buffer was prepared by mixing the components mentioned below in Table 2.4. A 5x stock solution of DNA loading buffer was prepared and mixed with 4 volumes of DNA to make it to 1x before loading on to agarose gels. The final pH of the DNA loading dye adjusted to pH 8.0.

Table 2.4 Composition of 5x DNA loading buffer.

Components	Final concentration (5x)
Tris-HCl (pH 8.0)	50 mM
Glycerol	25% (w/v)
EDTA	5.0 mM
Bromophenol blue	0.2% (w/v)
Xylene cyanol	0.2% (w/v)

2.2.5 Extraction of DNA from agarose gel

The PCR amplified DNA or other plasmids were purified from agarose gel using a kit (Sigma GenElute), following the protocol provided by the manufacturer as discussed in Section 2.2.5.1. The extracted DNA was eluted in 30 µl elution buffer supplied with the kit (Sigma-Aldrich, USA).

2.2.5.1 Protocol for extraction of DNA from agarose gel

1. 1.5 ml sterile microcentrifuge tube was weighed and the weight was noted.
2. The PCR amplified DNA or plasmid was excised from gel using sharp sterile scalpel and transferred to the micro-centrifuge tube. The tube was weighed again and the weight of excised gel was determined by subtracting the weight of the empty tube.
3. Now, 3 volumes of Gel solubilisation solution were added to every 1 volume of gel (100 mg ~ 100 µl).
4. The micro-centrifuge tube containing excised gel was incubated at 50°C for 10 min (or until the gel slice dissolved completely)
5. 1 gel volume of isopropanol was added to this solution.
6. GenElute binding column G was placed in a 2 ml collection tube provided with the kit. 500 µl of column preparation solution was added over column membrane and centrifuged at 16,000***g*** for 1 min. The flow through was discarded before the next step.
7. The solution containing PCR-amplified DNA or plasmid (~700 µl) were added to DNA binding columns and centrifuged at 16,000***g*** for 1 min at room temperature discarding the flow through. If the volume was more than 700 µl,

the remaining solution was centrifuged similarly and again the flow through was discarded.

8. 700 µl of Wash solution was added to each the DNA bound spin column and the mixture centrifuged at 16000g for 1 min at room temperature, discarding the flow through. The column was given an additional spin of 1 min at 16000***g*** to completely remove the residual ethanol.
9. Now the column containing bound DNA was placed on a fresh 1.5 ml sterile microcentrifuge tube. 30 µl of DNAse free water (Sigma-Aldrich, USA) or eluent solution (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 then centrifuged at 16000***g*** for 1 min. For efficient recovery of DNA, the elution solution was preheated to 65°C prior to adding it to the membrane. Eluting at 65°C improves the DNA recovery by 2 to 3-fold.
10. The PCR-amplified DNA or plasmid were eluted from GenElute spin columns and collected in 1.5 ml sterile microcentrifuge tube and stored at -20°C for further use.

2.2.6 Preparation of culture medium

The most commonly used LB medium for growing the *E. coli* cells containing recombinant plasmid was prepared by dissolving the ingredients (Table 2.5) in 800 ml of deionized water. The pH was adjusted to 7.2 and final volume was made up to 1 litre. 100 ml of LB medium was then transferred to 250 ml conical flask and autoclaved at 121°C at 15 psi for 20 min. The filter sterilized antibiotic (Kanamycin; 50 µg/ml) was added to autoclaved and cooled LB medium prior to inoculation.

Table 2.5 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

2.2.6.1 Preparation of LB-agar medium

LB agar medium was prepared by boiling 2% (w/v) Agar Agar type I in broth medium. The medium was autoclaved as described in Section 2.2.6 cooled to around 50-55°C and appropriate amount of antibiotics (kanamycin; 50 µg/ml) was added under laminar air flow. 34 µg/ml of chloramphenicol was used when *E. coli* BL21 (DE3) pLysS cells were cultured. 25 ml of medium supplemented with antibiotics were poured in sterile petri plates and allowed for 15- 20 min to solidify.

2.2.6.2 Preparation of Terrific Broth

Terrific Broth (TB) is a highly nutritious medium that allows extended growth phases for recombinant *E. coli* cells. It was prepared by dissolving the ingredients mentioned in Table 2.6 in 800 ml of deionized water. After addition and mixing of all the ingredients the volume was made up to 1000 ml. pH of the medium was adjusted to 7.2. 100 ml of TB medium was then transferred to 250 ml conical flask sand autoclaved at 121°C at 15 psi for 20 min. The filter sterilized chloramphenicol (34 µg/ml) and kanamycin (50 µg/ml) were added to autoclaved and cooled TB medium prior to inoculation.

Table 2.6 Composition of Terrific Broth (Sambrook *et al*., 1989).

Components	Final concentration (%, w/v)
Tryptone	1.2
Yeast extract powder	2.4
Potassium phosphate (dibasic)	0.94
Potassium phosphate (monobasic)	0.22
Glycerol	0.4

2.2.7 Preparation of SOC medium

The SOC (super optimal medium with catabolic repression) was prepared using ingredients mentioned in Table 2.7. It is a modified SOB (super optimal broth) with addition of glucose (Hanahan, 1983). Bactotryptone, yeast extract powder and NaCl was autoclaved. 1 M stock solutions of KCl, $MgCl_2$, $MgSO_4$ and glucose were filter-sterilized and required quantities added to above solution in the laminar hood to finally make the SOC medium.

Table 2.7 Composition of SOC medium (Sambrook *et al*., 1989)

Component	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

2.2.8 Preparation of *E. coli* DH5α competent cells

Day 1

1. 50 µl of culture of *E. coli* DH5α cells from glycerol stock was inoculated into 5.0 ml LB medium (Sambrook *et al*., 1989) contained in a test tube and grown overnight at 37ºC and 180 rpm.

2. 0.1 M $CaCl_2$ solution was filter-sterilized by passing through 0.22 µm filter in laminar air flow and kept in refrigerator.

Day 2

3. 1.0 ml culture from day 1 was inoculated into 100 ml LB medium kept in 250 ml conical flask and incubated at 37ºC with 180 rpm till cell Optical Density reached 0.4-0.6 at 550 nm.
4. Micro-centrifuge tubes, 50 ml centrifuge tubes (round bottom) and micro tips were autoclaved and kept on ice and placed in a laminar air flow hood.
5. 40 ml culture was transferred aseptically to round bottom centrifuge tubes.
6. The tubes were centrifuged at 4ºC with 4000***g*** for 10 min.
7. The step was repeated to centrifuge the entire 100 ml culture.
8. The cell pellet was re-suspended in 3-4 ml sterile, ice-chilled 0.1 M $CaCl_2$ solution followed by making up the final volume to 20 ml. The cell suspension in centrifuge tubes was kept on ice for 10 min.
9. The tube was centrifuged again at 4000***g*** at 4ºC for 10 min.
10. The supernatant was carefully removed and the pellet re-suspended 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.

2.2.9 Cloning of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 into pET28a(+) vector

The pET-28a(+) is a modified form of pBR322 plasmid. It is a frequently used vector for cloning and expression of recombinant proteins in *E. coli*. pET-28a(+) vector has a strong T7 promoter system originally developed by Studier and colleagues (Studier and Moffatt, 1986; Studier *et al.*, 1990). The expression of genes cloned in pET plasmids is under the control of T7 bacteriophage promoter. The cloned genes are transcribes by T7 RNA polymerase of the host cell. The genes cloned in pET vectors remain transcriptionally silent in the uninduced state. The proteins encoded by the cloned genes contain a His_6-Tag which single step purification using affinity chromatography. The pET-28a(+) vector allows for incorporation of expressed protein with an N-terminal His_6-Tag/thrombin/T7-Tag in addition to an optional C-terminal His_6-Tag sequence (Fig. 2.3). The location of sequence encoding His-Tag, T7 promoter, T7 terminator, kanamycin resistance and f1 origin are indicated in the Fig. 2.3.

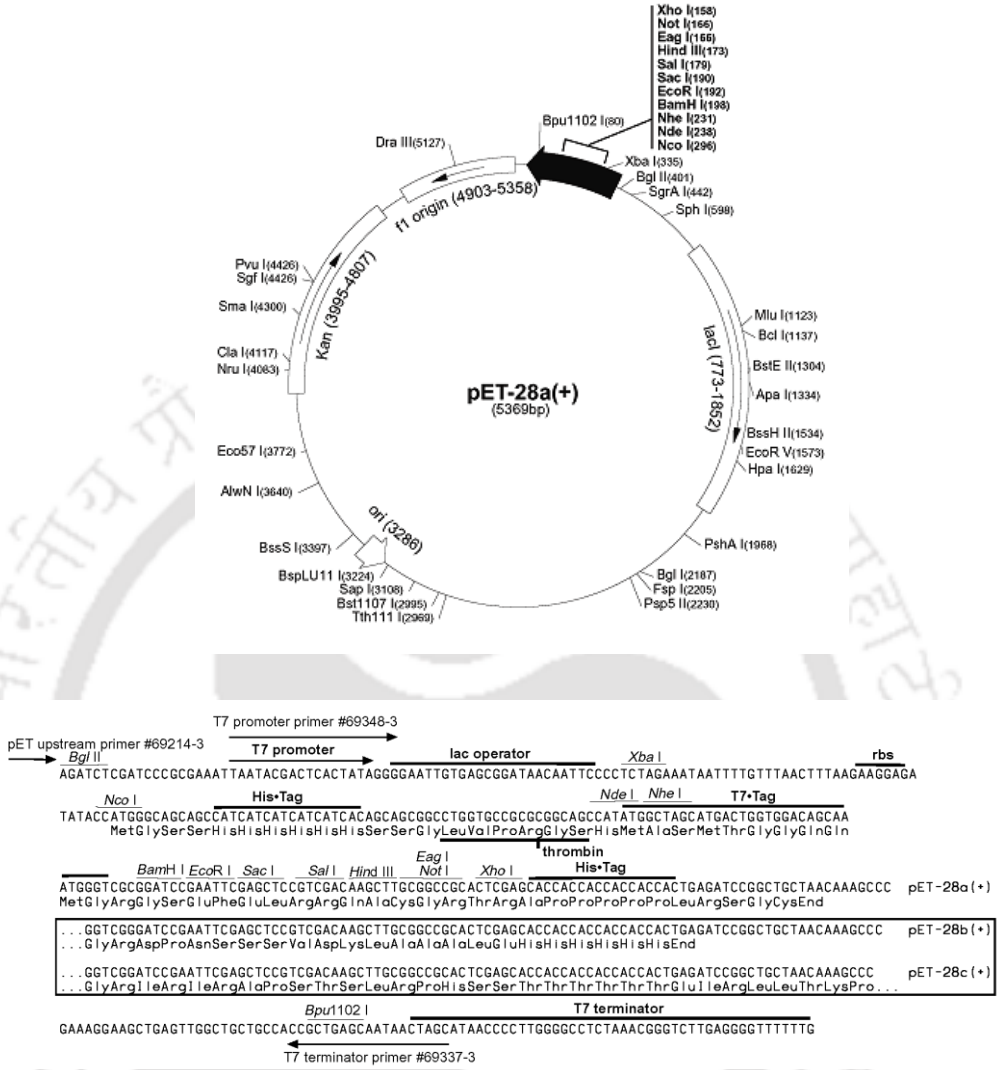


Fig. 2.3 Restriction map of the pET-28a(+) expression vector showing multiple cloning site (158-203 bp), restriction enzyme sites, N-terminal His_6-Tag coding sequence (270-287 bp), C-terminal His_6-Tag coding sequence (140-157 bp), T7 promoter (370-386), T7 terminator (26-72 bp), pBR322 origin (3286 bp), kanamycin marker (3995-4807 bp) and a f1 origin (4903-5358). *Nhe*I cuts at 231 and *Xho*I at 158.

2.2.9.1 Restriction digestion of PCR amplified genes encoding *Ct*RGL*f*, *Ct*RGL, *Rgl*-CBM35 and pET-28a(+) plasmid DNA

The pET-28a(+) vector was digested with *Nhe*I-*Xho*I restriction enzymes (Table 2.8). PCR amplified genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 were also digested with *Nhe*I-*Xho*I to prepare them for ligation with restriction digested

pET-28a(+) vector (Table 2.9). The digestion reactions were incubated in a water bath at 37°C for 90 min. The *Nhe*I-*Xho*I digested pET vector and PCR amplified genes were purified from agarose gel as described in Section 2.2.5.

Table 2.8 Restriction enzyme digestion of pET-28a (+) plasmid DNA.

Reaction components	**Volume (µl)**
10x buffer	3.0
Nuclease free water	4.5
Bovine serum albumin (10 mg/ml)	0.5
Plasmid DNA (approx. 13 ng/ µl)	20.0
*Nhe*I (10 U/µl)	1.0
*Xho*I (10 U/µl)	1.0
Total	30.0

Table 2.9 Restriction enzyme digestion of PCR amplified genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35.

Reaction component	Gene encoding *Ct*RGL*f* **(µl)**	Gene encoding *Ct*RGL **(µl)**	Gene encoding *Rgl*-CBM35 **(µl)**
10x buffer	3.0	4.0	4.0
Nuclease-free water	4.5	13.5	6.5
Bovine serum albumin (10 mg/ml)	0.5	0.5	0.5
PCR amplified product	20 (~125 ng)	20 (~100 ng)	20(~100 ng)
*Nhe*I (10 U/µl)	1 (118 ng)	1 (68.7 ng)	1 (15 ng)
*Xho*I (10 U/µl)	1	1	13.5
Total	30	40	30

2.2.9.2 Ligation of restriction digested genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 into pET-28a(+) vector

The *Nhe*I-*Xho*I digested genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 were ligated into pET-28a(+) vector, which was also digested with same restriction enzymes as described in Section 2.2.9.1. Three ligation reactions were setup using the reaction components mentioned in Table 2.10 and incubated at 16°C overnight to

get maximum number of transformants. The reactions were setup at an insert: vector molar ratio of 3:1, where the amount of insert is calculated as mentioned below:

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

$$\frac{72 \text{ (ng)} \times 2.1 \text{ (kb)}}{5.369 \text{ (kb)}} \times \frac{3}{1} = 118 \text{ ng } (Ct\text{RGL}f)$$

$$\frac{72 \text{ (ng)} \times 1.68 \text{ (kb)}}{5.369 \text{ (kb)}} \times \frac{3}{1} = 68.7 \ (Ct\text{RGL})$$

$$\frac{72 \text{ (ng)} \times 0.372 \text{ (kb)}}{5.369 \text{ (kb)}} \times \frac{3}{1} = 15.1 \text{ ng } (Rgl\text{-CBM35})$$

Table 2.10 Components reaction for ligating genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 into pET-28a (+) expression vector

Reaction component	Gene encoding *Ct*RGL*f* **(µl)**	Gene encoding *Ct*RGL **(µl)**	Gene encoding *Rgl*-CBM35 **(µl)**
10x Rapid Ligation Buffer	1.5	2	2
pET-28a (+) Vector	4 (100 ng)	9 (72 ng)	9 (72 ng)
Restriction digested product	6 (118 ng)	2 (68.7 ng)	3 (15 ng)
T4 DNA Ligase (3 Units/µl)	1	1	1
Nuclease-free water	2.5	6	5
Total	15	20	20

2.2.9.3 Transformation of ligated recombinant DNA into *E. coli* DH5α cells

The *E. coli* DH5α competent cells were transformed with ligation reactions, after overnight ligation. Preparation of *E. coli* competent cell preparation has been described in Section 2.2.8. The step-wise transformation protocol is described below:

1. The micro-centrifuge tube containing competent cell (200 µl) was taken out from -80°C and kept on ice for 5 min.

2. 10 µl of ligation mixture was added to cells and the tube was gently tapped 4-5 times and kept on ice for 30 min. The cells were occasionally tapped gently during 30 min incubation.
3. The cells were given a heat shock at 42°C for 40s.
4. The cells were immediately transferred back to the ice and kept for 5 min.
5. 800 µl of super optimal medium with catabolite repression (SOC) (Hanahan, 1983; Sambrook *et al*., 1989; given in Section 2.2.7) (previously incubated at 37°C) was added to the transformed cells.
6. The transformed cells were incubated at 37°C in a shaking incubator at 220 rpm for 1h.
7. The cells were centrifuged at 2000***g***, 25°C for 5 min.
8. 800 µl supernatant was discarded and the cell pellet was re-suspended in remaining 200 µl supernatant.
9. The 200 µl cells were spread plated on LB agar plates as described in Section 2.2.6.1 supplemented with antibiotics. The LB agar plates were incubated overnight at 37°C.
10. The transformation efficiency was calculated using the following formula,

$$\text{Transformation efficiency} = \frac{\text{No. of colonies on LB plate}}{\text{Amount of insert (µg)}} = \text{cfu/µg}$$

The 15-20 µl of ligation mixture was added to 200 µl *E. coli* DH5α competent cells, following the above transformation protocol. The transformed DH5α cells were plated on LB plates supplemented with kanamycin (50 µg/ml) and grown overnight at 37°C, 180 rpm.

2.2.9.4 Isolation of plasmid DNA from transformed colonies by miniprep kit

Overnight incubated plates were observed for colonies. Colonies preferably from the centre of the plate were randomly picked in a laminar air flow and grown overnight in 5 ml LB medium supplemented with kanamycin (50 µg/ml). The plasmid DNA from this 5 ml culture was isolated by miniprep kit (Sigma-Aldrich) following the protocol mentioned in Section 2.2.9.4.1.

2.2.9.4.1 Plasmid isolation protocol by miniprep kit

1. 9 ml from each of the grown culture containing recombinant plasmid were pelleted in 1.5 ml microcentrifuge tube aseptically.
2. The cells were then centrifuged at 14000***g*** for 1 min and the process was repeated six times with 1.5 ml culture (Total 9ml culture).
3. The resulting cell pellet of each recombinant derivative was re-suspended in 200 µl resuspension solution and vortexed. RNAse at final concentration of 0.3 mg/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 to ensure mixing and 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 centrifuged at 16,000***g*** for 10 min.
7. The DNA binding columns were prepared and activated by adding 500 µl of column preparation solution to binding column and centrifuging at 14,000***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, centrifuged at 14,000***g*** for 1 min and the flow through in the collection tube was discarded again.
9. The plasmid DNA bound to the column was washed with wash solution and spun at 14,000***g*** for 1 min. The flow through was discarded and the column was given another 1 min spin at 14,000***g*** for removing the wash solution completely.
10. The DNA binding column was transferred to a fresh sterile microcentrifuge tube and 30 µl of TE buffer solution or DNAse free water was added at the centre of binding column. The microcentrifuge tube was allowed to stand for 10 min at room temperature and then plasmid DNA was eluted by centrifugation at 14,000***g*** for 1 min.
11. The eluted plasmid DNA in sterile microcentrifuge tube was stored at -20°C.

2.2.9.5 Screening of recombinant plasmid DNAs for positive clones by restriction digestion

15 µl of recombinant plasmids from pET-28a(+) clones of *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 that were isolated in the last step, were taken in separate fresh sterile micro-centrifuge tubes for restriction enzyme digestion analysis. The recombinant plasmid DNA of each of the above mentioned derivatives was digested with restriction enzymes, *Nhe*I and *Xho*I, to check for positive clones following a 30 µl reaction mixture set up as mentioned in Table 2.8.

2.2.10 Preparation of competent *E. coli* BL-21 (DE3) and *E. coli* BL-21 (DE3) pLysS cells

The competent *E.coli* BL-21 (DE3) and *E. coli* BL-21 (DE3) pLysS cells were prepared by calcium chloride method following the protocol as discussed in Section 2.2.8. 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.

2.2.11 Transformation of recombinant plasmids containing genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 into *E. coli* BL21 (DE3) or *E. coli* BL-21 (DE3) pLysS cells

2 µl of from each of the recombinant plasmid of positive pET-28a(+) clone isolated in Section 2.2.9.5 was used for transformation of 200 µl *E. coli* BL-21 or *E. coli* BL-21 pLysS cells for protein expression following the transformation protocol described in Section 2.2.9.3. Recombinant plasmids containing genes encoding *Ct*RGL*f* and *Rgl*-CBM35 were transformed into *E. coli* BL-21 cells and plated on LB agar plates supplemented with kanamycin (50 µg/ml) and grown overnight at 37°C. While recombinant plasmid containing gene encoding *Ct*RGL was transformed into *E. coli* BL-21 pLysS cells and plated on LB agar plates supplemented with kanamycin (50 µg/ml) and chloramphenicol (34 µg/ml) followed by overnight incubation at 37°C.

2.2.12 Expression of recombinant *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35

E. coli BL21(DE3) cells used as host for expression of *Ct*RGL*f* and *Rgl*-CBM35 were cultured in 100 ml of LB medium supplemented with kanamycin (50 µg/ml) at 37°C, 180 rpm. *Ct*RGL was expressed in *E. coli* BL21 (DE3) pLysS cells, which were cultured in 100 ml Terrific Broth supplemented with chloramphenicol

(34 µg/ml) and kanamycin (50 µg/ml) incubated at 37°C and 180 rpm. After the cell growth reached mid exponential phase (A_{600} = 0.6). The cells expressing *Ct*RGL*f* and *Ct*RGL were cooled and after IPTG induction further cultured at 16°C. While the cells expressing *Rgl*-CBM35, were cooled and after IPTG induction further cultured at 24°C. After cooling the cultures to appropriate temperatures isopropyl-β-D-thiogalactopyranoside (IPTG) was added at a final concentration of 1 mM and further incubated at 180 rpm for 20 h.

2.2.13 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 was used to separate components of a protein mixture based on their size (Laemmli, 1970; Sambrook *et al*., 1989). Analysis of expression and purification of *Ct*RGL*f* and *Ct*RGL was done on 12% SDS-PAGE gels, whereas *Rgl*-CBM35 was analysed on 14% SDS-PAGE gel using ingredients mentioned in Tables 2.11 and 2.12.

Table 2.11 Composition of SDS-PAGE components for preparation of resolving gel.

Component	**12% gel volume (ml)**	**14% gel volume (ml)**
Acrylamide solution *(30%,w/v)	4.0	4.67
Deionized water	0.7	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
Total	10ml	10ml

**mixture of 29.2% acrylamide and 0.8% N,N'-Methylenebisacrylamide*

Table 2.12 Composition of SDS-PAGE components for preparation of stacking gel.

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

**mixture of 29.2% acrylamide and 0.8% N,N'-Methylenebisacrylamide*

SDS-PAGE was run using 1x Tris-Glycine (pH 8.3-8.5) running buffer at constant 40mA current. The expressed and purified protein samples were visualised after staining the gel with staining soultion containing (0.25% w/v) Coomassie Brilliant Blue (CBB) R-250 dye 100 ml solution of deionized water, methanol and glacial acetic acid in 5:4:1 ratio. The gels were de-stained by immersing the gel in de-staining solution containing deionized water, methanol and glacial acetic acid in 5:4:1 ratio. The gels were subjected to gentle rocking with periodic change of de-staining solution was done, until the protein bands were clearly visible.

2.2.14 Purification of recombinant *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 proteins

The recombinant *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 proteins containing a His_6-tag at the N-terminal were purified through a single step purification method based on immobilized metal-ion affinity chromatography (IMAC) as described in Section 2.2.14.1. The purification of these recombinant proteins was carried out by using 1.0 ml sepharose columns (GE Healthcare, HiTrap chelating HP). The composition of binding as well as elution buffers used for affinity column purification is mentioned in Table 2.13.

Table 2.13 Composition of buffers required for purification of recombinant proteins by affinity purification (IMAC).

Buffers	Composition
Equilibration buffer	50 mM Tris-HCl, pH 8.5 100 mM NaCl, 50 mM Imidazole
Elution buffer	50 mM Tris-HCl, pH 8.5 100 mM NaCl, 500 mM Imidazole
Cleaning buffer	50 mM Tris-HCl, pH 8.0 500 mM NaCl, 50 mM EDTA

2.2.14.1 IMAC purification protocol for recombinant *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35

1. The bacterial cells (100 ml culture) were harvested by centrifugation at 10,000***g*** at 4°C. The cell pellet was re-suspended in 5 ml of 50 mM Tris-HCl buffer, pH 8.5.
2. The cells were sonicated on ice for 15 min (10s on and 15s off pulse; with 33% amplitude) and centrifuged at 12,000***g*** at 4ºC for 30 min to get the crude cell free extract.
3. The cell free extract was passed through a 0.45 µm filter membrane before loading onto 1ml HiTrap chelating HP column. The column was pre-washed with 5 volumes of filtered and degassed water to remove the alcohol.
4. Column was charged using 2.0 ml of 0.1 M $NiSO_4$ solution and the unbound Ni^{2+} ions were washed away with 2-5 volumes of water.
5. Then the column was equilibrated with 10 volumes of equilibration buffer (Table 2.13).
6. The filtered cell free extract of recombinant protein was loaded on to the column at a flow rate of 0.5 ml/min.

7. The column was then washed with 50 column volumes of equilibration buffer to remove the unbound proteins.
8. The retained protein of interest was then eluted with elution buffer under a gradient of 0-100% imidazole concentration and 1 ml fractions were collected (Carvalho *et al.*, 2004).
9. The column was cleaned using cleaning buffer as mentioned in Table 2.13, washed with 5 volumes of water and incubated in 1N NaOH at 4°C for 2h. The column was then washed with 20 volumes of water to remove NaOH, and finally stored in 20% (v/v) ethanol at 4°C.

The purified recombinant proteins *Ct*RGL*f* and *Ct*RGL were dialyzed against 50 mM Tris-HCl buffer, pH 8.5 containing 100 mM NaCl and *Rgl*-CBM35 was dialyzed against 50 mM Tris-HCl buffer, pH 8.5. The purity and molecular mass of recombinant proteins were verified by SDS-PAGE as described in Section 2.2.13.

2.2.15 Protein concentration determination of purified recombinant proteins

The concentration of purified protein was determined from their corresponding absorbance at 280 nm using the equation below (Layne, 1957; Stoscheck, 1990). Absorbance was measured after proper dilution of the protein using a spectrophotometer (Gene Quant, GE) having a path length of 1 cm. The molar extinction co-efficient 153725 $M^{-1}cm^{-1}$ for *Ct*RGL*f*, 121400 $M^{-1}cm^{-1}$ for *Ct*RGL and 31400 $M^{-1}cm^{-1}$ for *Rgl*-CBM35, respectively were used.

$$\text{Concentration of protein (mg/ml)} = \frac{\text{Absorbance at 280 nm x Mol. weight (Da)}}{\text{Extinction coefficient } (M^{-1}cm^{-1}) \text{ x Path length (1 cm)}}$$

2.3 Results and Discussion

2.3.1 PCR amplification of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35

The genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 were amplified from genomic DNA of *Clostridium thermocellum* ATCC 27405 using the conditions mentioned in Section 2.2.3 and detected on 0.8% (w/v) agarose gel and are displayed in Fig. 2.4 below. The PCR products were purified from gel using gel extraction kit as mentioned in Section 2.2.5 and stored at -20°C.

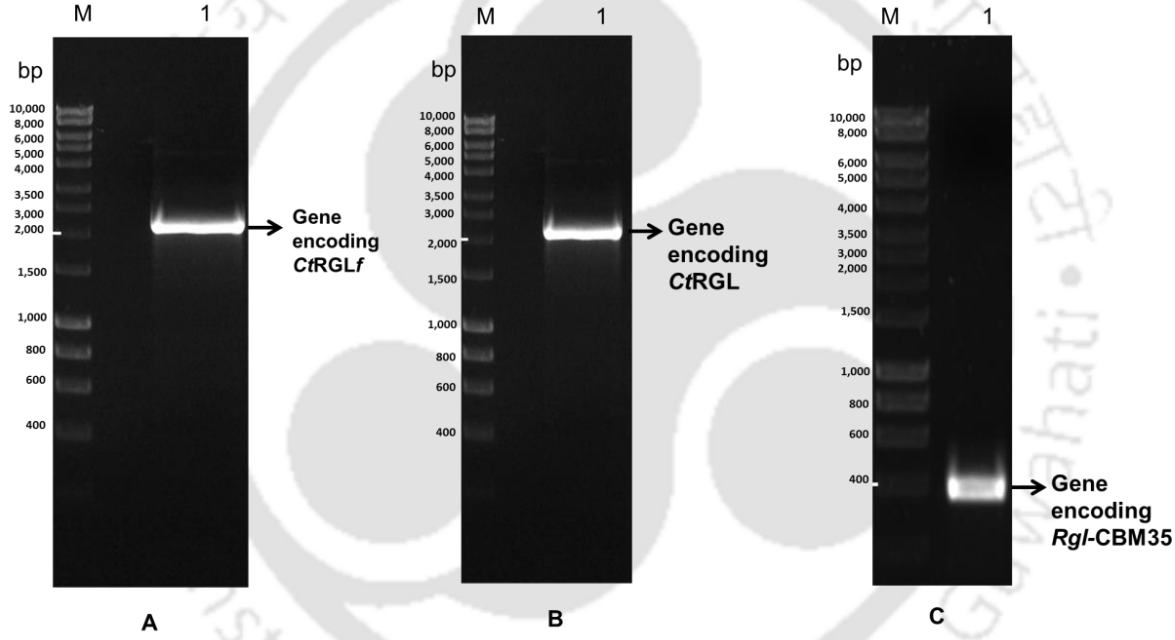


Fig. 2.4 Agarose gels (0.8%, w/v) showing PCR amplified genes encoding: **(A) *Ct*RGL*f*** (lane 1), **(B) *Ct*RGL** (lane 1) and **(C) *Rgl*-CBM35** (lane 1) around 2.1 kb, 1.6 kp and 0.3 kb, respectively. Lane M- DNA marker (Hyperladder I, Bioline)

2.3.2 Cloning of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 into pET-28a (+) vector

The restriction enzyme digested genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 were ligated with the linearized fragments of pET-28a(+) vector following the protocol mentioned in Section 2.2.10.1. The ligated product was transformed into *E. coli* DH5α competent cells and grown overnight on LB agar plates grown at 37°C

under stationary condition. The transformation efficiency of *E. coli* DH5α competent cells was 1.5 x10^6.

2.3.2.1 Isolation of recombinant plasmid DNA

Plasmid DNA from grown colonies after cloning into pET-28a(+) was isolated using Plasmid miniprep kit following the protocol mentioned in Section 2.2.13.3.1. The isolated plasmids were visualized after electrophoresis on 0.8% (w/v) agarose gel. ositive clones were confirmed by restriction digestion of this isolated plasmid DNA.

2.3.2.2 Restriction digestion of isolated plasmid DNA for confirmation of positive clone

The isolated plasmids were digested with *Nhe*I and *Xho*I restriction enzymes for confirming the positive clones. The plasmids after restriction digestion were electrophoresed on 0.8% (w/v) agarose gels. *NheI* and *XhoI* digested fragments of gene encoding *Ct*RGL*f*, (Fig. 2.5A; Lane 1), *Ct*RGL*f* (Fig. 2.5B; Lane 1) and *Rgl*-CBM35 (Fig. 2.5 C; Lane 1) were visualized on agarose gels around 2.1 kb, 1.6 kb and 0.3 kb respectively. Linearized pET-28a (+) vector was visualized at around 5.4 kb after restriction digestion. The positives clones were sequenced (Scigenom Labs Pvt. Ltd, India) and no undesired mutations were detected. (Fig. 2.6, 2.7, 2.8, 2.9, 2.10).

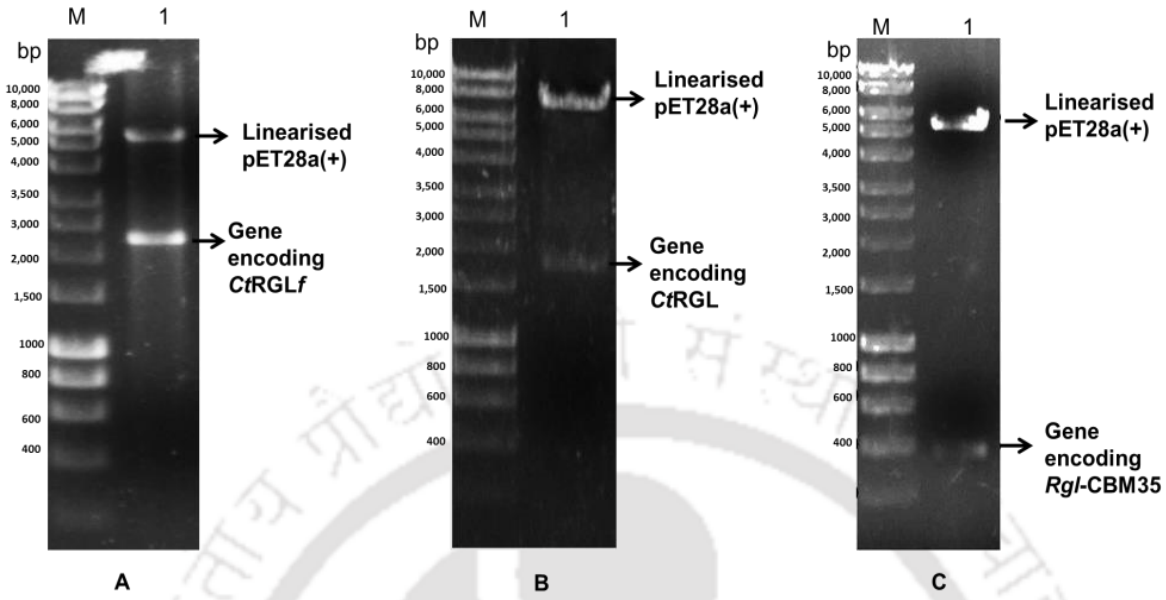
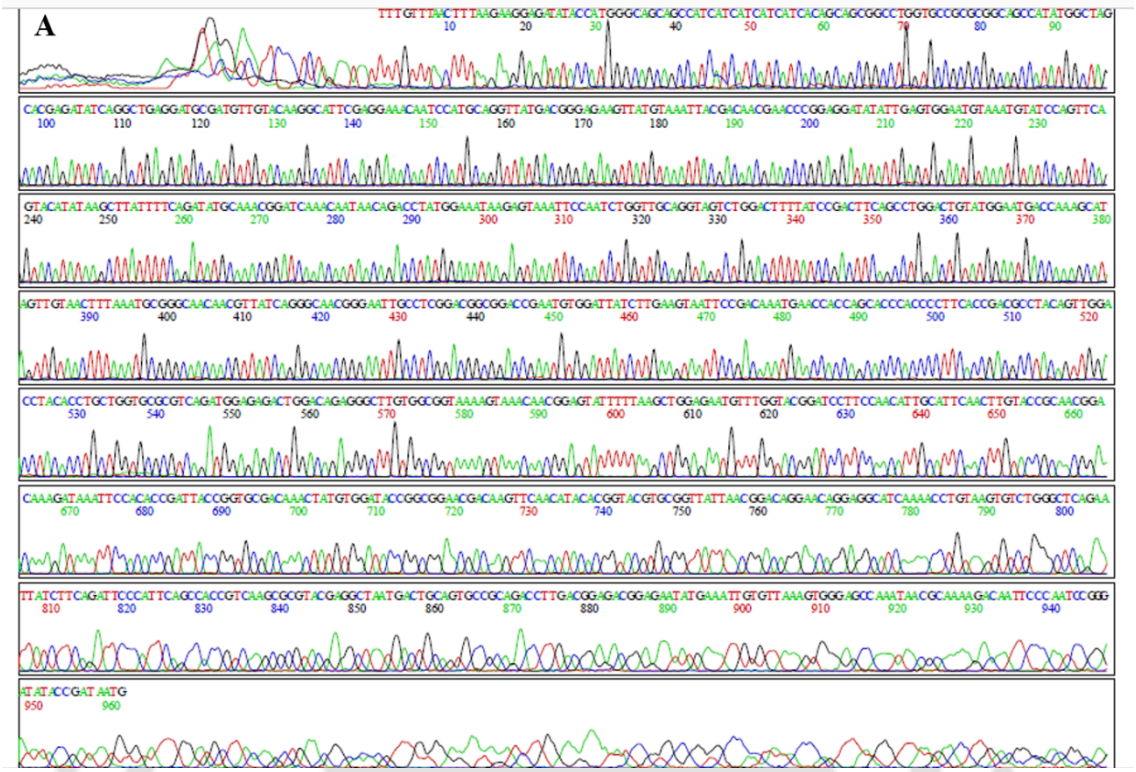


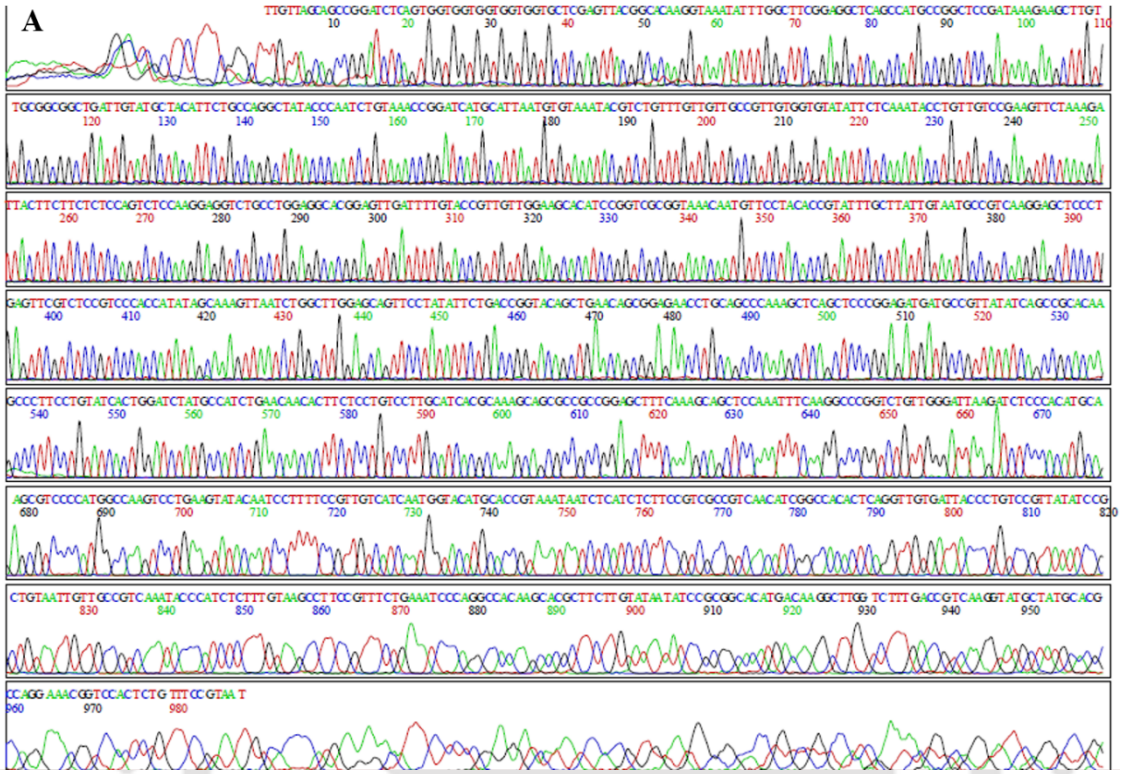
Fig. 2.5 Agarose gel (0.8%, w/v) showing *Nhe*I-*Xho*I digested recombinant plasmid containing genes encoding (A) *Ct*RGL*f*, (Lane 1: 2.1 kb), (B) *Ct*RGL, (Lane 2: 1.6 kb) and (C) *Rgl*-CBM35 (Lane 1: 0.3 kb). Lane M: DNA marker (Hyper ladder I, Bioline).



B

TTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCACA
GCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCACGAGATATCAGGCTGAGGATG
CGATGTTGTACAAGGCATTCGAGGAAACAATCCATGCAGGTTATGACGGGAGAAGTTATG
TAAATTACGACAACGAACCCGGAGGATATATTGAGTGGAATGTAAATGTATCCAGTTCAG
GTACATATAAGCTTATTTTCAGATATGCAAACGGATCAAACAATAACAGACCTATGGAAA
TAAGAGTAAATTCCAATCTGGTTGCAGGTAGTCTGGACTTTTATCCGACTTCAGCCTGGA
CTGTATGGAATGACCAAAGCATAGTTGTAACTTTAAATGCGGGCAACAACGTTATCAGGG
CAACGGGAATTGCCTCGGACGGCGGACCGAATGTGGATTATCTTGAAGTAATTCCGACAA
ATGAACCACCAGCACCCACCCCTTCACCGACGCCTACAGTTGGACCTACACCTGCTGGTG
CGCGTCAGATGGAGAGACTGGACAGAGGGCTTGTGGCGGTAAAAGTAAACAACGGAGTAT
TTTTAAGCTGGAGAATGTTTGGTACGGATCCTTCCAACATTGCATTCAACTTGTACCGCA
ACGGAACAAAGATAAATTCCACACCGATTACCGGTGCGACAAACTATGTGGATACCGGCG
GAACGACAAGTTCAACATACACGGTACGTGCGGTTATTAACGGACAGGAACAGGAGGCAT
CAAAACCTGTAAGTGTCTGGGCTCAGAATTATCTTCAGATTCCCATTCAGCCACCGTCAA
GCGCGTACGAGGCTAATGACTGCAGTGCCGCAGACCTTGACGGAGACGGAGAATATGAAA
TTGTGTTAAAGTGGGAGCCAAATAACGCAAAAGACAATTCCCAATCCGGGATATACCGAT
AATG

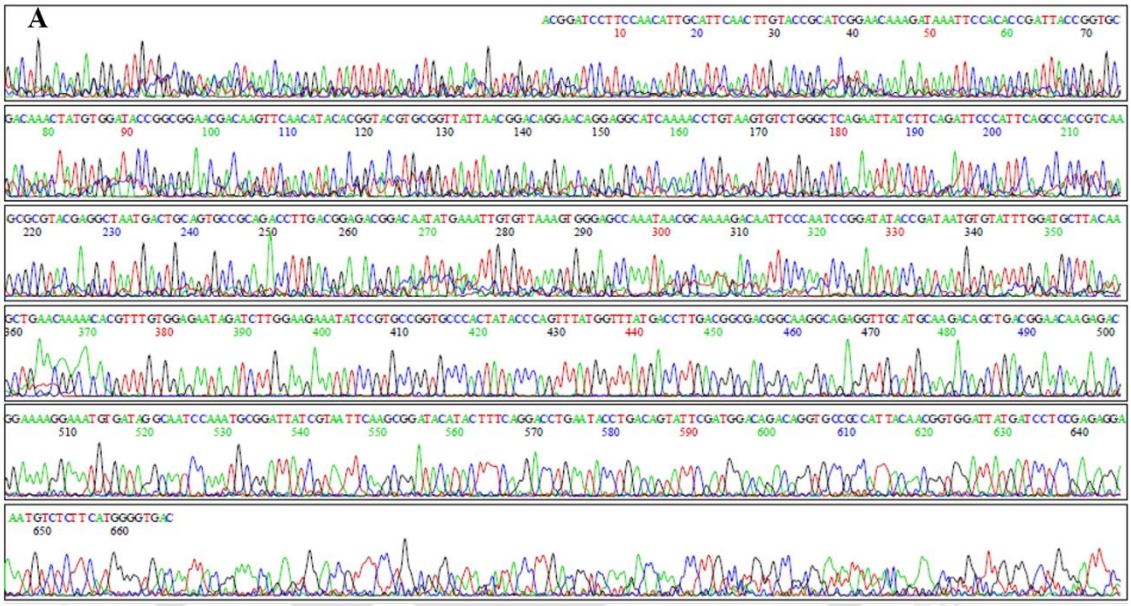
Fig. 2.6 DNA Sequencing of *Ct*RGL*f*. (A**)** Electropherogram showing the DNA sequencing result for cloned gene encoding *Ct*RGL*f*. Sequencing was done using T7 forward primer. (B) the deduced sequence.



B

TTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTACGGCACAAGGTAA
ATATTTGGCTTCGGAGGCTCAGCCATGCCGGCTCCGATAAAGAAGCTTGTGTGCGGCGGC
TGATTGTATGCTACATTCTGCCAGGCTATACCCAATCTGTAAACCGGATCATGCATTAAT
GTGTAAATACGTCTGTTTGTTGTTGCCGTTGTGGTGTATATTCTCAAATACCTGTTGTCC
GAAGTTCTAAAGATTACTTCTTCTCTCCAGTCTCCAAGGAGGTCTGCCTGGAGGCACGGA
GTTGATTTTGTACCGTTGTTGGAAGCACATCCGGTCGCGGTAAACAATGTTCCTACACCG
TATTTGCTTATTGTAATGCCGTCAAGGAGCTCCCTGAGTTCGTCTCCGTCCCACCATATA
GCAAAGTTAATCTGGCTTGGAGCAGTTCCTATATTCTGACCGGTACAGCTGAACAGCGGA
GAACCTGCAGCCCAAAGCTCAGCTCCCGGAGATGATGCCGTTATATCAGCCGCACAAGCC
CTTCCTGTATCACTGGATCTATGCCATCTGAACAACACTTCTCCTGTCCTTGCATCACGC
AAAGCAGCGCCGCCGGAGCTTTCAAAGCAGCTCCAAATTTCAAGGCCCGGTCTGTTGGGA
TTAAGATCTCCCACATGCAGAGCGTCCCCATGGCCAAGTCCTGAAGTATACAATCCTTTT
CCGTTGTCATCAATGGTACATGCACCGTAAATAATCTCATCTCTTCCGTCGCCGTCAACA
TCGGCCACACTCAGGTTGTGATTACCCTGTCCGTTATATCCGCTGTAATTGTTGCCGTCA
AATACCCATCTCTTTGTAAGCCTTCCGTTTCTGAAATCCCAGGCCACAAGCACGCTTCTT
GTATAATATCCGCGGCACATGACAAGGCTTGGTCTTTGACCGTCAAGGTATGCTATGCAC
GCCAGGAAACGGTCCACTCTGTTTCCGTAAT

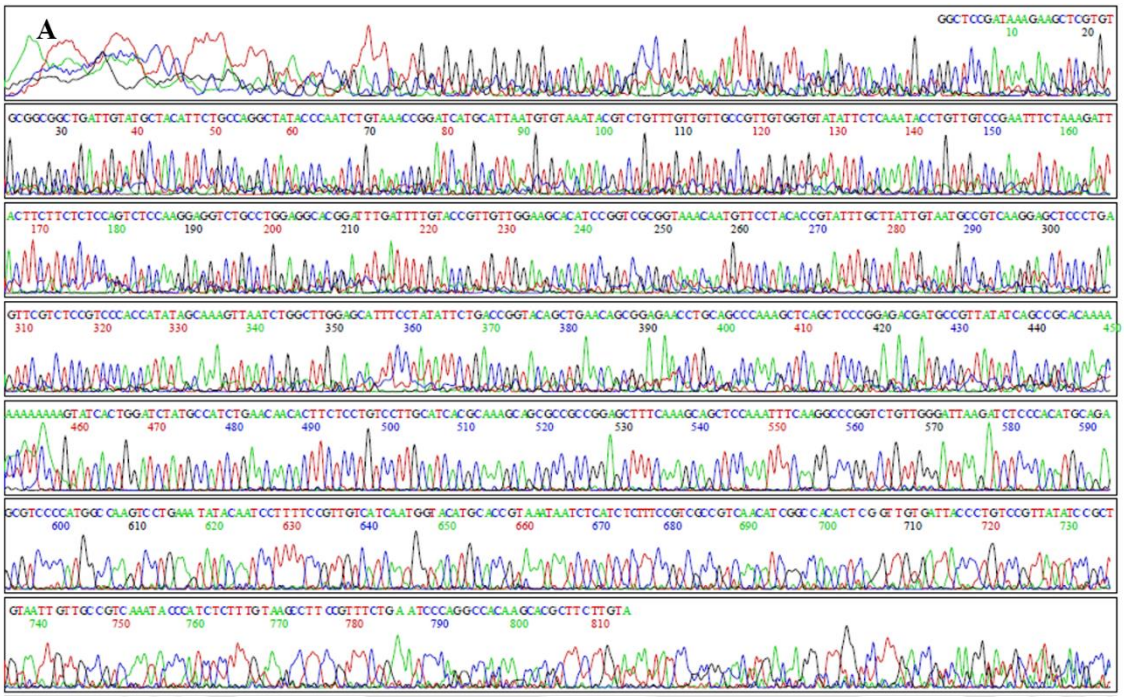
Fig. 2.7 DNA Sequencing of *Ct*RGL*f*, (A**)** Electropherogram showing the DNA sequencing result for cloned gene encoding *Ct*RGL*f*, Sequencing was done using T7 reverse primer. (B) the deduced sequence.



B

ACGGATCCTTCCAACATTGCATTCAACTTGTACCGCATCGGAACAAAGATAAATTCCACA
CCGATTACCGGTGCGACAAACTATGTGGATACCGGCGGAACGACAAGTTCAACATACACG
GTACGTGCGGTTATTAACGGACAGGAACAGGAGGCATCAAAACCTGTAAGTGTCTGGGCT
CAGAATTATCTTCAGATTCCCATTCAGCCACCGTCAAGCGCGTACGAGGCTAATGACTGC
AGTGCCGCAGACCTTGACGGAGACGGACAATATGAAATTGTGTTAAAGTGGGAGCCAAAT
AACGCAAAAGACAATTCCCAATCCGGATATACCGATAATGTGTATTTGGATGCTTACAAG
CTGAACAAAAACACGTTTGTGGAGAATAGATCTTGGAAGAAATATCCGTGCCGGTGCCCA
CTATACCCAGTTTATGGTTTATGACCTTGACGGCGACGGCAAGGCAGAGGTTGCATGCAA
GACAGCTGACGGAACAAGAGACGGAAAAGGAAATGTGATAGGCAATCCAAATGCGGATTA
TCGTAATTCAAGCGGATACATACTTTCAGGACCTGAATACCTGACAGTATTCGATGGACA
GACAGGTGCCGCCATTACAACGGTGGATTATGATCCTCCGAGAGGAAATGTCTCTTCATG
GGGTGAC

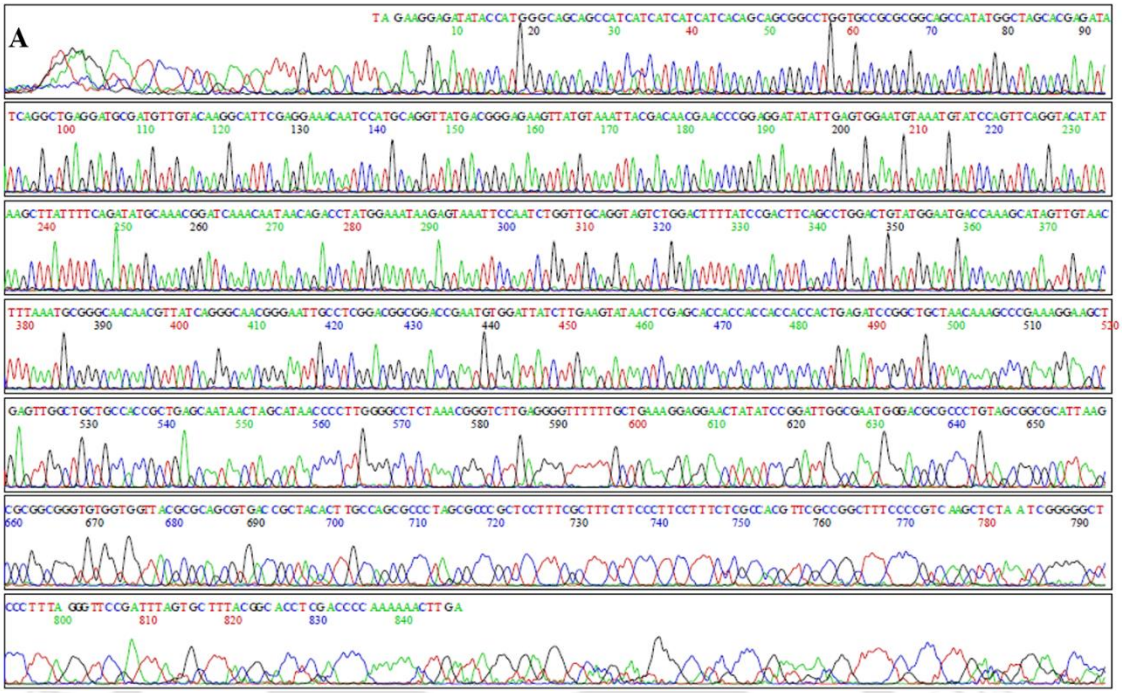
Fig. 2.8 DNA Sequencing of *Ct*RGL (A**)** Electropherogram showing the DNA sequencing result for cloned gene encoding *Ct*RGL. Sequencing was done using T7 forward primer. (B) the deduced sequence.



B

GGCTCCGATAAAGAAGCTCGTGTGCGGCGGCTGATTGTATGCTACATTCTGCCAGGCTAT
ACCCAATCTGTAAACCGGATCATGCATTAATGTGTAAATACGTCTGTTTGTTGTTGCCGT
TGTGGTGTATATTCTCAAATACCTGTTGTCCGAATTTCTAAAGATTACTTCTTCTCTCCA
GTCTCCAAGGAGGTCTGCCTGGAGGCACGGATTTGATTTTGTACCGTTGTTGGAAGCACA
TCCGGTCGCGGTAAACAATGTTCCTACACCGTATTTGCTTATTGTAATGCCGTCAAGGAG
CTCCCTGAGTTCGTCTCCGTCCCACCATATAGCAAAGTTAATCTGGCTTGGAGCATTTCC
TATATTCTGACCGGTACAGCTGAACAGCGGAGAACCTGCAGCCCAAAGCTCAGCTCCCGG
AGACGATGCCGTTATATCAGCCGCACAAAAAAAAAAAAGTATCACTGGATCTATGCCATC
TGAACAACACTTCTCCTGTCCTTGCATCACGCAAAGCAGCGCCGCCGGAGCTTTCAAAGC
AGCTCCAAATTTCAAGGCCCGGTCTGTTGGGATTAAGATCTCCCACATGCAGAGCGTCCC
CATGGCCAAGTCCTGAAATATACAATCCTTTTCCGTTGTCATCAATGGTACATGCACCGT
AAATAATCTCATCTCTTTCCGTCGCCGTCAACATCGGCCACACTCGGTTGTGATTACCCT
GTCCGTTATATCCGCTGTAATTGTTGCCGTCAAATACCCATCTCTTTGTAAGCCTTCCGT
TTCTGAATCCCAGGCCACAAGCACGCTTCTTGTA

Fig. 2.9 DNA Sequencing of *Ct*RGL (A**)** Electropherogram showing the DNA sequencing result for cloned gene encoding *Ct*RGL. Sequencing was done using T7 reverse primer. (B) the deduced sequence.



B

TAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGT
GCCGCGCGGCAGCCATATGGCTAGCACGAGATATCAGGCTGAGGATGCGATGTTGTACAA
GGCATTCGAGGAAACAATCCATGCAGGTTATGACGGGAGAAGTTATGTAAATTACGACAA
CGAACCCGGAGGATATATTGAGTGGAATGTAAATGTATCCAGTTCAGGTACATATAAGCT
TATTTTCAGATATGCAAACGGATCAAACAATAACAGACCTATGGAAATAAGAGTAAATTC
CAATCTGGTTGCAGGTAGTCTGGACTTTTATCCGACTTCAGCCTGGACTGTATGGAATGA
CCAAAGCATAGTTGTAACTTTAAATGCGGGCAACAACGTTATCAGGGCAACGGGAATTGC
CTCGGACGGCGGACCGAATGTGGATTATCTTGAAGTATAACTCGAGCACCACCACCACCA
CCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGC
TGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCT
GAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGC
GCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCC
GCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCT
CTAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAA
AACTTGA

Fig. 2.10 DNA Sequencing of *Rgl*-CBM35 (A**)** Electropherogram showing the DNA sequencing result for cloned gene encoding *Rgl*-CBM35. Sequencing was done using T7 forward primer. (B) the deduced sequence.

2.3.3 Expression and purification of recombinant proteins

The *E. coli* BL21 (DE3) competent cells were transformed with recombinant pET-28a (+) plasmids containing genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35. The transformation efficiencies of *E. coli* BL21 (DE3) and *E. coli* BL21 (DE3) pLysS cells were 1.8×10^6 and 5×10^5. The colonies were picked randomly and grown in 5 ml LB medium supplemented with kanamycin (50 µg/ml) as described in Section 2.2.12. The cells were induced for expression of protein at mid exponential stage as described in Section 2.2.12. Protein expression was analysed using SDS-PAGE gels by loading uninduced as well as the induced cells in adjacent wells, as depicted in Fig. 2.11, 2.12 and 2.13.

The recombinant proteins were purified by immobilized metal ion affinity chromatography as described in Section 2.2.14 and then dialysed for removal of imidazole and sodium chloride. The recombinant *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 expressed as soluble proteins and after purification displayed homogeneous single bands on SDS-PAGE gels (Fig. 2.11, 2.12 and 2.13). The molecular masses of the recombinant *Ct*RGL*f* , *Ct*RGL and *Rgl*-CBM35 including the N-terminal histidine tag were calculated to be 80.2 kDa, 63.9 kDa and 16.4 kDa respectively, which are in close agreement with those observed on SDS-PAGE gels.

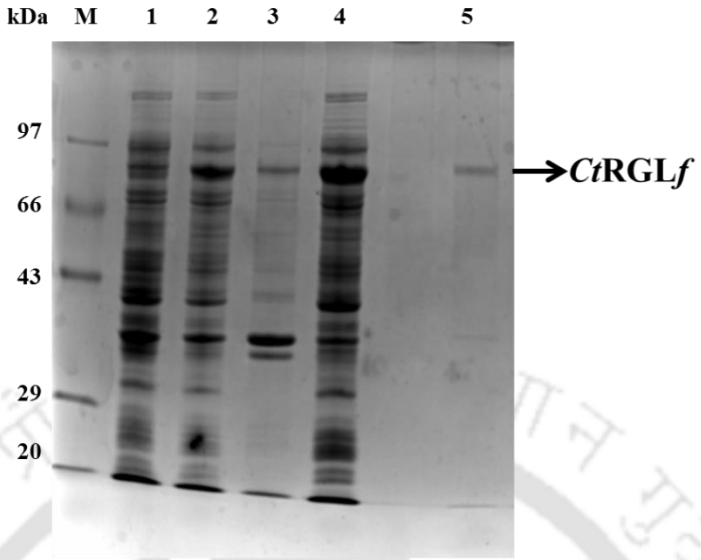


Fig.2.11 SDS-PAGE (12%) gel showing expression and purification of recombinant *Ct*RGL*f* in *E. coli* BL-21 cells, Lane M: Protein marker (Bangalore GeNei, India), Lane 1: Uninduced cells, Lane 2: Induced cells, Lane 3: Cell pellet (cell debris after sonication), Lane 4: Cell free extract, Lane 5: Purified and dialyzed *Ct*RGL*f* (80 kDa approx.).

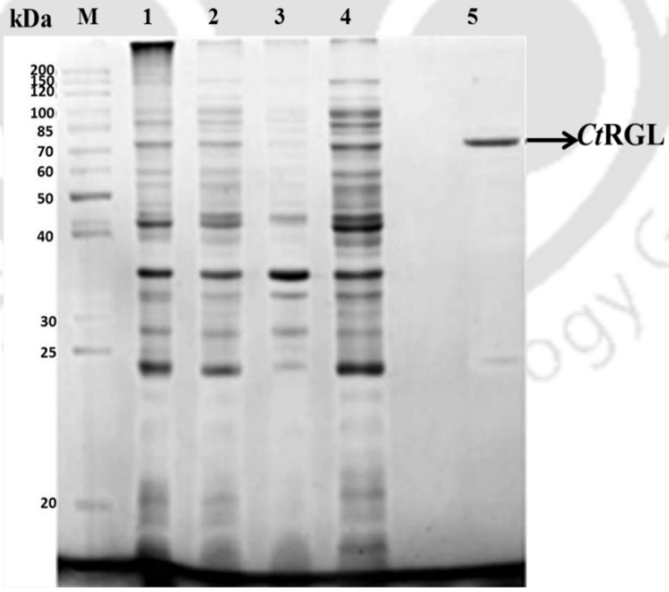


Fig. 2.12 SDS-PAGE (12%) gel showing expression and purification of *Ct*RGL in *E. coli* BL-21 (DE3) pLysS cells, Lane M: Page Ruler marker (Fermentas), Lane 1: Uninduced cells, Lane 2: Induced cells, Lane 3: Cell pellet (cell debris after sonication), Lane 4: Cell free extract, Lane 5: Purified and dialyzed *Ct*RGL (64 kDa approx.).

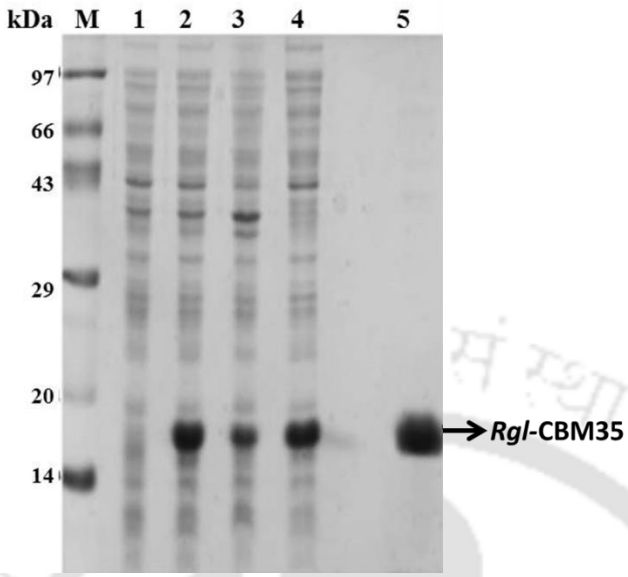


Fig. 2.13 SDS-PAGE (14%) gel showing expression and purification of recombinant protein *Rgl*-CBM35 from *E. coli* BL-21 (DE3) cells, Lane M: Protein marker (Bangalore GeNei, India), Lane 1: Uninduced cells, Lane 2: Induced cells, Lane 3: Cell pellet (cell debris after onication), Lane 4: Cell free extract, Lane 5: Last column wash, Lane 6: Purified and dialyzed *Rgl*-CBM35 (16 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 using the formula mentioned in Section 2.2.15 and listed in Table 2.14. The total amount of recombinant enzymes present in eluted 2 ml of was 0.4 mg (*Ct*RGL*f*), 0.3 mg (*Ct*RGL), and 4 mg (*Rgl*-CBM35) as displayed in Table 2.14.

Table 2.14 Purified recombinant proteins obtained from 100 ml cultures.

Recombinant protein	Protein concentration (mg/ml)	Volume of purified protein (ml)	Total amount of purified protein (mg)
*Ct*RGL*f*	0.20	2.0	0.4
*Ct*RGL	0.15	2.0	0.3
Rgl-CBM35	2.0	2.0	4.0

2.4 Conclusions

Family 11 polysaccharide lyase (PL11) *Ct*RGL*f* and its truncated derivatives, *Ct*RGL and *Rgl*-CBM35 were cloned from the genomic DNA of *Clostridium thermocellum* ATCC 27405 (GenBank Accession No: ABN51485.1). The molecular architecture revealed a type 1 dockerin domain towards N-terminal followed by a family 35 carbohydrate binding module (*Rgl*-CBM35, 16.4 kDa) and family 11 polysaccharide lyase catalytic module (*Ct*RGL, 64 kDa). The PCR amplified fragment of full length gene encoding *Ct*RGL*f* showed a band of ~2.1 kb, whereas, the gene encoding catalytic module (*Ct*RGL) and binding module (*Rgl*-CBM3), displayed the sizes of ~ 1.6 kb and ~ 0.3 kb, respectively. The restriction enzyme digested fragments of genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35 were ligated with linearized pET-28a(+) vector. The ligated mixture was transformed into *E. coli* DH5α competent cells. The positive clones containing recombinant plasmid DNA were screened by restriction enzyme digestion using enzymes, *NheI* and *XhoI*. The restriction enzyme digested products were electrophoresed and the band of ~ 5.4 kb was produced for pET-28a(+) vector and corresponding bands of ~2.1 kb, ~ 1.6 kb and ~ 0.3 kb were produced from the insert fragments for genes encoding *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35, respectively. The recombinant plasmids containing genes encoding *Ct*RGL*f* and *Rgl*-CBM35 were transformed into *E. coli* BL-21 (DE3) cells. Recombinant plasmid containing gene encoding *Ct*RGL was transformed into *E. coli* BL-21 (DE3) pLysS cells. The recombinant proteins were expressed and purified. The purified recombinant proteins displayed a band of approximately, 80 kDa for *Ct*RGL*f*, 70 kDa for *Ct*RGL and a band of approximately, 16 kDa for *Rgl*-CBM35 on SDS-PAGE gels. The amount recombinant *Ct*RGL*f*, *Ct*RGL and *Rgl*-CBM35

proteins obtained from 100 ml *E.coli* cultures after purification by Immobilized Metal Ion Affinity Chromatography (IMAC) were 0.4 mg, 0.3 mg and 4.0 mg, respectively.

References

Abbott, D. W., van Bueren, A. L. (2014) Using structure to inform carbohydrate binding module function. Current Opinion in Structural Biology, 28: 32-40.

Correia, M. A., Abbott, D. W., Gloster, T. M., Fernandes, V. O., Prates, J. A., Montanier, C., Dumon, C., Williamson, M. P., Tunnicliffe, B. B., Ziyuan, L., Flint, J. E., Gideon, D., Henrissat, B., Cutinho, P., Fontes, M. G. A., Gilbert, H. (2010) Signature active site architectures illuminate the molecular basis for ligand specificity in family 35 carbohydrate binding module. Biochemistry, 49(29): 6193-6205.

Duan, C. J., Feng, Y. L., Cao, Q. L., Huang, M. Y., Feng, J. X. (2016) Identification of a novel family of carbohydrate-binding modules with broad ligand specificity. Scientific Reports, 6: 9392.

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

Gilbert, H. J., Knox, J. P., Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Current Opinion in Structural Biology, 23(5): 669-677.

Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., Knox, J. P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. Proceedings of the National Academy of Sciences, 107(34): 15293-15298.

Koshland, D. E. (1953) Stereochemistry and the mechanism of enzymatic reactions. Biological Reviews, 28: 416-436.

Layne, E. (1957) Spectrophotometric and Turbidimetric Methods for Measuring Proteins. Methods in Enzymology 3: 447-455.

Lombard, V., Golaconda, R. H., Drula, E. (2014) The Carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research, 42: D490-D495.

Matsunaga, T., Ishii, T., Matsumoto, S. (2004) Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes and bryophytes: Implications for the evolution of vascular plants. Plant Physiology, 134: 339-351.

McKie, V. A., Vincken, J. P., Voragen, A. G. (2001) A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. Biochemical Journal, 355: 167–177.

Miyanaga, A., Koseki, T., Miwa, Y., Mese, Y., Nakamura, S., Kuno, A., Fushinobu, S. (2006) The family 42 carbohydrate-binding module of family54 α-L-arabinofuranosidase specifically binds the arabinofuranose side chain of hemicellulose. Biochemical Journal, 399(3): 503-511.

Montanier, C., Van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis., R. J., Grondin, G. G., Ghinet, M. G., Gloster, T. M., Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., Kerovuo, J., Brzezinski, R., Fontes, C. M., Davies, G. J., Boraston, A. B., Gilbert, H. J. (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proceedings of the National Academy of Sciences, 106(9): 3065-3070.

Moran, F. S, Nasuno, S., Starr, M. P. (1968) Extracellular and intracellular polygalacturonic acid trans-eliminase of *Erwinia carotovora*. Archives of Biochemistry and Biophysics, 123: 298–306.

Moreira, S., Castanheira, P., Casal, M., Faro, C., Gama, M. (2010) Expression of the functional carbohydrate-binding module (CBM) of human laforin. Protein Expression and Purification, 74(2): 169-174.

Ochiai, A., Itoh, T., Kawamata, A. (2007) Plant cell wall degradation by saprophytic Bacillus subtilis strains: gene clusters responsible for rhamnogalacturonande polymerization. Applied and Environmental Microbiology, 73: 3803-3813.

O'Neill, M. A., Warrenfeltz, D., Kates, K. (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. Journal of Biological Chemistry, 271: 22923-22930.

Oomen, R. J., Doeswijk, V. H., Bush, M. S. (2002) *In muro* fragmentation of the rhamnogalacturonan I backbone in potato (*Solanum tuberosum* L.) results in a reduction and altered location of the galactan and arabinan side-chains and abnormal periderm development. Plant Journal, 30: 403-413.

Pages, S., Valette, O., Abdou, L. (2003) A rhamnogalacturonan lyase in the *Clostridium cellulolyticum* cellulosome. Journal of Bacteriology, 185: 4727-4733.

Ribeiro, T., Santos-Silva, T., Alves, V. D., Dias, F. M., Luís, A. S., Prates, J. A., Ferreira, L., Romão, M., Fontes, C. M. (2010) Family 42 carbohydrate-binding modules display multiple arabinoxylan-binding interfaces presenting

different ligand affinities. Biochimica et Biophysica Acta, 1804(10): 2054-2062.

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

Stoscheck, C.M. (1990) Quantitation of Protein. Methods in Enzymology, 182: 50–69.

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

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

Vincken, J. P., Schols, H. A., Oomen, R. J. (2003) If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. Plant Physiology, 132: 1781-1789.

Chapter 3

Biochemical and functional characterization of *Ct*RGL*f* and *Ct*RGL

3.1 Introduction

The major components of plant cell wall are the polysaccharides that assemble to form a network. These polysaccharides are mainly cellulose, hemicellulose and pectin. Cellulose and hemicellulose in cell wall afford rigidity and protection to the cell (Vincken *et al*., 2003). The pectin content of the cell wall facilitates cell-cell adhesion, cell signalling, wall porosity, pollen tube growth and leaf abscission (Ridley *et al*., 2001). Structural role of pectin in promoting upright growth of plants has also been reported (Matsunaga*et al*., 2004). Pectin is predominantly localised in the primary cell wall of all higher plants, gymnosperms, pteridophytes and bryophytes (O'Neill *et al*., 1996). The principal components of pectin are homogalacturonan (HG) and the substituted HG: Rhamnogalacturonan I (RG I) and Rhamnogalacturonan II (RG II). Xylogalacturonans and Apiogalacturonans are other less abundant substituted HG (Ridley *et al*., 2001). Homogalacturonan is a homopolymer of α-(1→4) linked D-galactopyranosyluronic acid (D-Gal*p*A) residues which may be methylated at C-6. RG

I is composed of a main chain of alternating L-rhamnopyranosyl (L-Rha*p*) and D-Gal*p*A residues. The repeating monomeric unit is a disaccharide [→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→]. 20-80% of L-Rha*p* residues are substituted with individual, linear or branched chains of L-arabinofuranosyl (L-Ara*f*) and D-Gal*p*A. The side chains of RG I may be substituted with L-fucose or D-glucuronic acid (GlcA) (O'Neill (O'Neill *et al*., 1996). RG I has been held crucial for normal development of periderm in potato (Oomen *et al*., 2002). Transgenic potato plants that expressed RG I cleaving enzyme, developed morphological abnormalities like swelling of periderm cells. RG II main chain contrary to the name, is also composed of α-(1→4) linked D-Gal*p*A with L-Rha*p* residues as substitutions apart from 12 other different monosaccharide residues (O'Neill *et al*., 1996).

Degradation of plant cell wall makes a reservoir of nutrients available for recycling. Diverse groups of microorganisms possess enzymes to breakdown the cell wall polysaccharides (Ochiai *et al*., 2007; Pages *et al*., 2003; McKie *et al*., 2001). Microboal hydrolysis of pectin is mediated by glycoside hydrolases and polysaccharide lyases. Glycoside hydrolases cleave the glycosidic bonds *via* acid-base catalysis mechanism (Koshland *et al*., 1953). Polysaccharide lyases cleave their substrates *via* β-elimination mechanism, creating a double bond between C-4 and C-5 in the residue at the non-reducing end of the product (Moran *et al*., 1968). Glycoside hydrolases and polysaccharide lyases have been classified into different families based on sequence similarity (Lombard *et al*., 2014). Many of the plant cell wall polysaccharide degrading enzymes are modular in nature and have one or more specialised substrate binding module(s) referred to as Carbohydrate Binding Module(s) (CBM) in addition to a catalytic module.

Enzymes that can break down the cell wall polysaccharides are of key importance in conversion of lignocellulosic biomass into bio-ethanol and production of prebiotics (Koukiekolo *et al*., 2005; Aachary *et al*., 2011). The pectinases are one such class of industrial enzymes that are used in paper and textile industries, coffee and tea fermentation, treatment of feedstock for biofuel production and recovery of valuable products of plant origin like essential oils (Kashyap *et al*., 2001). Some anaerobic bacteria are better equipped at cleaving the recalcitrant structural polysaccharides, an attribute probably acquired along the evolutionary ladder. (Fontes *et al*., 2010). *C. thermocellum* is an anaerobic, thermophilic bacterium possessing cellulosome, a multi-enzyme complex which afforded a new meaning of enzyme modularity and concerted enzyme action (Lamed *et al*., 1983). The cellulosomal complex has also been reported from other microorganisms (Bayer *et al*., 1994). Several studies involving cellulases and hemicellulases from this bacterium in bio-ethanol production found them to be better than those of fungal origin. (Das *et al*., 2012; Anbar *et al*., 2012). However, studies on pectin degrading enzymes from *C. thermocellum* have been sporadic. The members of family 11 polysaccharide lyase (PL11) characterised earlier are rhamnogalacturonan lyases (RG lyases) from *Bacillus subtilis, Clostridium cellulolyticum, Cellvibrio japonicus* and *Bacillus licheniformis* (Ochiai *et al*., 2007; Pages *et al*., 2003; McKie *et al*., 2001; Silva *et al*., 2014). In the present study cloning, expression and biochemical characterisation of a putative cellulosomal, family 11 PL enzyme (*Ct*RGL*f*) was attempted. *Ct*RGL*f* comprises a RG Lyase catalytic module, designated as *Ct*RGL and an associated putative family 35 CBM. *Ct*RGL*f* endolytically cleaves RG I component of pectin. To our knowledge

this is the first modular RG lyase reported from *C. thermocellum* and is a new family 11 polysaccharide lyase member.

Biochemical and functional characterization of *Ct*RGL*f* and *Ct*RGL is essential to understand the mechanism of catalysis and to determine their substrate specificity. In the present study the full length module *Ct*RGL*f* and truncated catalytic module *Ct*RGL*f* were biochemically and functionally characterized. The enzyme activities of *Ct*RGL*f* and *Ct*RGL against various pectic polysaccharides and other carbohydrates were determined. The specific activity and kinetic parameters of both these enzymes were compared to explore the influence of CBMs on the enzyme activity.

3.2 Materials and Methods

3.2.1 Substrates and reagents

Rhamnogalacturonan from soybean (RGS), and potato (RGP), pectic galactan from lupin (PGL) and galactan from potato (GP) were purchased from Megazyme, Ireland and polygalacturonic acid (PGA), pectin (25% methyl esterified) were purchased from Sigma-Aldrich, India. Trizma (Tris base), glycine and sodium hydroxide were procured from Sigma-Aldrich, India. Disodium 2-[2-carboxylatomethyl-(carboxymethyl)-amino]-ethyl-(carboxymethyl)-amino] acetate (disodium EDTA) and salts of Ca^{2+}, Mg^{2+}, Mn^{2+}, Co^{2+} were procured from Himedia Laboratories Pvt. Ltd., India. Acetic acid, aniline, butanol, ethanol, ethyl acetate, phosphoric acid and diphenylamine were sourced from Fisher Scientific, India and readymade silica coated aluminium TLC plates were obtained from Merck, Germany.

3.2.2 Enzyme activity assay

Enzyme activity of *Ct*RGL*f* and *Ct*RGL was determined against different substrates in a reaction mixture containing 0.15% (w/v) of substrate dissolved in 50 mM Tris-HCl buffer (pH 8.5) containing 3 mM $CaCl_2$ incubated at 60°C for 5 min. After incubation the reaction mixture was kept on ice for 5 min to stop the reaction and then centrifuged at 13,000*g*. The supernatant was used to measure the absorbance. The enzyme activity was assayed by measuring the absorbance of Δ4,5 unsaturated D-Gal*p*A residue at the non-reducing end of the rhamnogalacturonates products at 235 nm (A_{235}) on a UV-Visible spectrophotometer (Gene-Quant, GE Healthcare, USA).

3.2.2.1 Calculation of enzyme activity

Enzyme activity was expressed as U/ml and specific activity as U/mg of protein. One unit of enzyme activity was defined as the amount of enzyme required to produce 1 µmol of 4,5-unsaturated rhamnogalacturonates (product) per minute. The enzyme activities of *Ct*RGL*f* and *Ct*RGL were calculated as described below,

$$\text{Enzyme activity (U/ml)} = \frac{\Delta A_{235} \times 1000}{\varepsilon \times t \times v}$$

Where,

ΔA_{235} = Absorbance of unsaturated rhamnogalacturonates

ε = Molar extinction of 4600 M^{-1} for Δ4,5 unsaturated D-Gal*p*A

t = Time of reaction in min

v = Total volume of reaction in ml

$$\text{Specific activity (U/mg)} = \frac{\text{Enzyme activity (U/ml)}}{\text{Concentration of protein used (mg/ml)}}$$

3.2.3 Substrate specificity of *Ct*RGL*f* and *Ct*RGL

The substrate specificity of *Ct*RGL*f* and *Ct*RGL was investigated using various pectic substrates (1%, w/v) namely rhamnogalacturonan from soybean (RGS), rhamnogalacturonan from potato (RGP), pectic galactan from lupin (PGL), galactan from potato (GP), polygalacturonic acid (PGA) and pectin (25% methyl esterified) under optimized conditions. *Ct*RGL*f* or *Ct*RGL was incubated with substrate in 50 mM Tris-HCl buffer (pH 8.5) containing 3 mM $CaCl_2$ for 5 min at optimum temperature and the A_{235} was recorded as mentioned in previous section.

3.2.4 Determination of optimum pH of *Ct*RGL*f* and *Ct*RGL

The optimum pH for maximum activity of *Ct*RGL*f* and *Ct*RGL was determined by incubating the enzymes with 0.15% (w/v) RGS at 60°C for 5 min in different buffers containing 3 mM $CaCl_2$: 50 mM MES (pH 6.0), 50 mM Tris-HCl (pH 7.0-9.0), 50 mM Glycine-NaOH (pH 10.0), 50 mM CAPS (pH 11.0). The A_{235} was measured and the enzyme activity was calculated as mentioned in Section 3.2.2.1.

3.2.5 Determination of optimum temperature of *Ct*RGL*f* and *Ct*RGL

To determine the optimum temperature for the activity of *Ct*RGL*f* and *Ct*RGL, the enzymes were incubated at different temperatures ranging from 30°C to 100°C in 50 mM Tris-HCl buffer (pH 8.5) containing 0.15% (w/v) RGS and 3 mM $CaCl_2$. The A_{235} was measured and the enzyme activity was calculated as described in Section 3.2.2.1.

3.2.6 Determination of temperature stability of *Ct*RGL*f* and *Ct*RGL

The effect of temperature on stability of *Ct*RGL*f* and *Ct*RGL was determined by incubating the enzymes in Tris-HCl buffer (pH 8.5) at different temperatures ranging from 30°C to 100°C for 30 min.

3.2.7 Determination of kinetic parameters of *Ct*RGL*f* and *Ct*RGL

Kinetic parameters of *Ct*RGL*f* and *Ct*RGL were determined by assaying their activity against different concentrations of substrates under optimized conditions of temperature and pH. The specific activity was calculated as mentioned in Section 3.2.2.1 and k_{cat} and K_m were determined using the Michaelis-Menten plot and Lineweaver-Burk plot.

3.2.8 Effect of metal ions on activity of *Ct*RGL*f* and *Ct*RGL

The effect of various metal ions on the activity of *Ct*RGL*f* or *Ct*RGL was determined by pre-treating them with 10 mM EDTA at 25°C for 60 min to remove any bound divalent metal ion, followed by buffer exchange to 50 mM Tris-HCl (pH 8.5). The activity was measured in presence of $CaCl_2$, $MgCl_2$, $MnCl_2$ or $CoCl_2$ at a final concentration of 3 mM. The effect of Ca^{2+} ions on the activity of *Ct*RGL*f* and *Ct*RGL was determined by varying the $CaCl_2$ concentration from 0 to 8 mM in the reaction mixture.

3.2.9 Thin-layer chromatography analysis of *Ct*RGL*f* and *Ct*RGL treated RGS

The mode of action of *Ct*RGL*f* and *Ct*RGL on RGS was determined by analysing the degradation products formed at different time intervals using thin layer chromatography. Separate reaction mixtures (1 ml) of *Ct*RGL*f* and *Ct*RGL were set that contained 0.15% (w/v) RGS and 3 mM $CaCl_2$ in 50 mM Tris-HCl buffer (pH 8.5). A separate reaction was set for each time interval (0, 5, 15 and 30 min and 1 h, 12 h and 24 h). Left over large undigested polysaccharide molecules were precipitated by adding 2 volumes of absolute ethanol followed by centrifugation at 13,000*g* for 10 min. The supernatant (1 ml) containing enzyme degraded products of RGS were collected in separate microcentrifuge tubes and concentrated to 20 µl by incubating at 60°C for 16 h. The concentrated sample (1 µl) was applied as a spot on a TLC plate (Merck, Germany) that was later developed in a solution containing 1-butanol, acetic acid and water in 5:2:3 ratio. The TLC plates were stained by diphenylamine-aniline-phosphoric acid reagent (1 ml 37.5% HCl, 2 ml aniline, 10 ml 85% H_3PO_4, 100 ml ethyl acetate and 2 g diphenylamine) and visualized after incubating at 80°C for 30 min (Zhang *et al*., 2009).

3.3 Results and Discussion

3.3.1 Substrate specificity of *Ct*RGL*f* and *Ct*RGL

The β-elimination reaction mechanism of *Ct*RGL*f* and *Ct*RGL was confirmed by increase in the absorbance at 235 nm, when the enzymes were incubated with their substrate RGS. Increase in absorbance was caused by the formation of double bond in the D-Gal*p*A residue at the newly formed non-reducing end. Both *Ct*RGL*f* and *Ct*RGL displayed high activity against two RGs (RGS and RGP) and moderate activity against pectic galactan (PGL). Both the enzymes showed low activity against polygalacturonic acid (PGA) and 25% methyl esterified pectin (Table 3.3). *Ct*RGL*f* and *Ct*RGL were primarily active against rhamnogalacturonans but could also cleave the homogalacturoanan component of pectin. It may be easy to comprehend that the low activity of *Ct*RGL*f* and *Ct*RGL against PGL may be due to the low content of L-Rha*p* and D-Gal*p*A residues in this substrate as compared to both RGs (RGS and RGP). It is noteworthy that this catalytic property of *Ct*RGL*f* and *Ct*RGL is in stark contrast to that of some other RG lyases, Rgl11Y and Rgl11A from *C. cellulolyticum* and *P. cellulosa,* respectively (Pages *et al*., 2003; McKie *et al*., 2001). Rgl11A and Rgl11Y displayed higher activity on PGP despites its low L-Rha*p* and D-Gal*p*A content. This highlights the role of β-D-galactopyranose (D-Gal*p*) substitutions for their enhanced activity on PGP as it is highly substituted with D-Gal*p* residues (Pages *et al*., 2003; McKie *et al*., 2001). All these substrates contain RG I main chain, but differ in the percentage of D-Gal*p* residues in their side chains. PGL and PGP contain around 74% and 87% D-Gal*p* respectively, while RGP and RGS contain only 20% D-Gal*p*. Very recently another RG lyase belonging to PL family 4, PcRGL4A from

Penicillium chrysogenum has been reported which too does not preferentially degrade substrate with higher D-Gal*p* substitution (Iwai *et al*., 2015).

Table 3.3 Activity of *Ct*RGL*f* and *Ct*RGL towards different polysaccharides.

Substrate (1%, w/v)	**Specific Activity *Ct*RGL*f* (U/mg)**	**Specific Activity *Ct*RGL (U/mg)**
Rhamnogalactouronan (Soyabean)	9.8	5.8
Rhamnogalactouronan (Potato)	9.1	5.4
Pectic galactan (Lupin)	4.7	3.0
Pectic galactan (Potato)	3.8	1.2
Pectin (25% methyl-esterified)	4.3	1.3
Polygalacturonic Acid	1.2	1.2

*The reaction of Ct*RGL*f and Ct*RGL *was carried out at 70°C and 60°C, respectively, for 5 min using pectic substrates (1%, w/v) dissolved in 50 mM Tris-HCl buffer (pH 8.5) containing 3 mM* $CaCl_2$.

3.3.2 Optimum pH for activity of *Ct*RGL*f* and *Ct*RGL

Optimum pH for the activity of both *Ct*RGL*f* and *Ct*RGL was 8.5 (Fig. 3.1), in line with other RG lyases from *B. subtilis, C. cellulolyticum, C. japonicus, and B. licheniformis* (Ochiai *et al*., 2007; Pages *et al*., 2003; McKie *et al*., 2001; Silva *et al*., 2014). The results showed that both these enzymes were active under alkaline conditions.

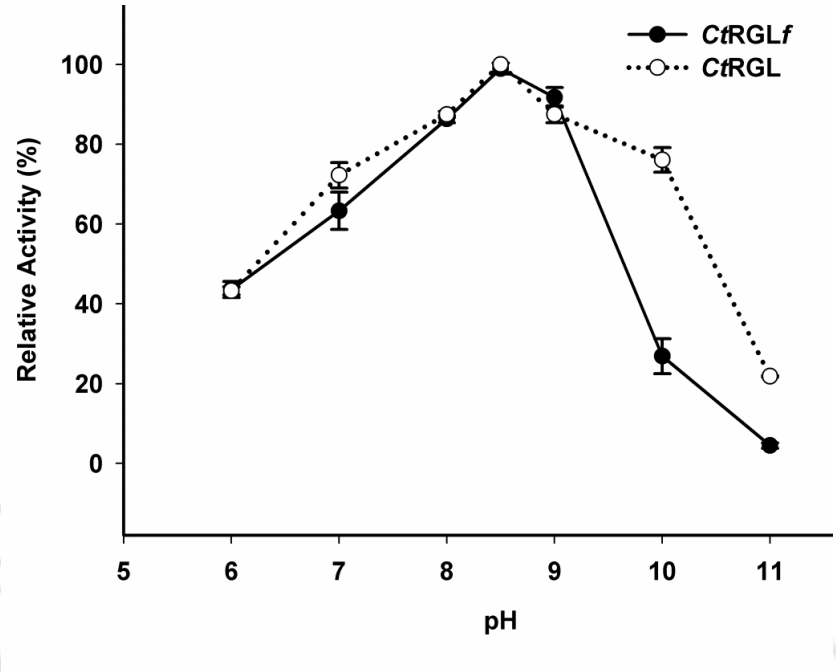


Fig. 3.1 Optimum pH for the activity of *Ct*RGL*f* and *Ct*RGL**.**

3.3.3 Optimum temperature for activity of *Ct*RGL*f* and *Ct*RGL

As expected, the optimum temperature for the activity of *Ct*RGL*f* was 70°C, whereas for *Ct*RGL it was 60°C (Fig. 3.2), since these enzymes were from a thermophilic bacterium. A dramatic decrease in activity was seen above 70°C for both *Ct*RGL*f* and *Ct*RGL. Thermophillic rhamnoogalacturonan lyases reported earlier showed optimum temperature of 66°C for YesW from *Bacillus subtilis* (Pages *et al*., 2003).

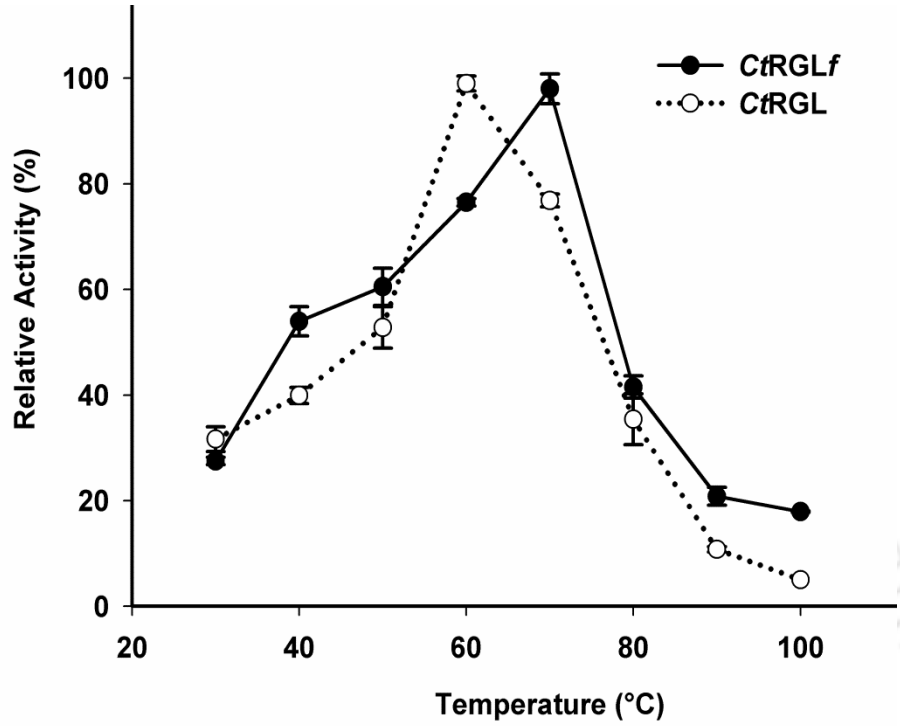


Fig. 3.2 Optimum temperature for activity of *Ct*RGL*f* and *Ct*RGL**.**

3.3.4 Temperature stability of *Ct*RGL*f* and *Ct*RGL

*Ct*RGL*f* was more thermostable than *Ct*RGL as it retained 90% of its activity after incubation at 60°C for 30 min, whereas, *Ct*RGL showed only 45% of its activity obtained under optimised assay conditions (Fig. 3.3). Higher thermal stability of *Ct*RGL*f* may be attributed to the presence of CBM35. Recently a Family 13 CBM has also been reported to provide thermostability to its associate catalytic module (Li *et al*., 2015). The exposure to high temperature disrupts the protein fold causing decreased stability and loss of enzyme activity (Branden *et al*., 1991; Creighton, 1992).

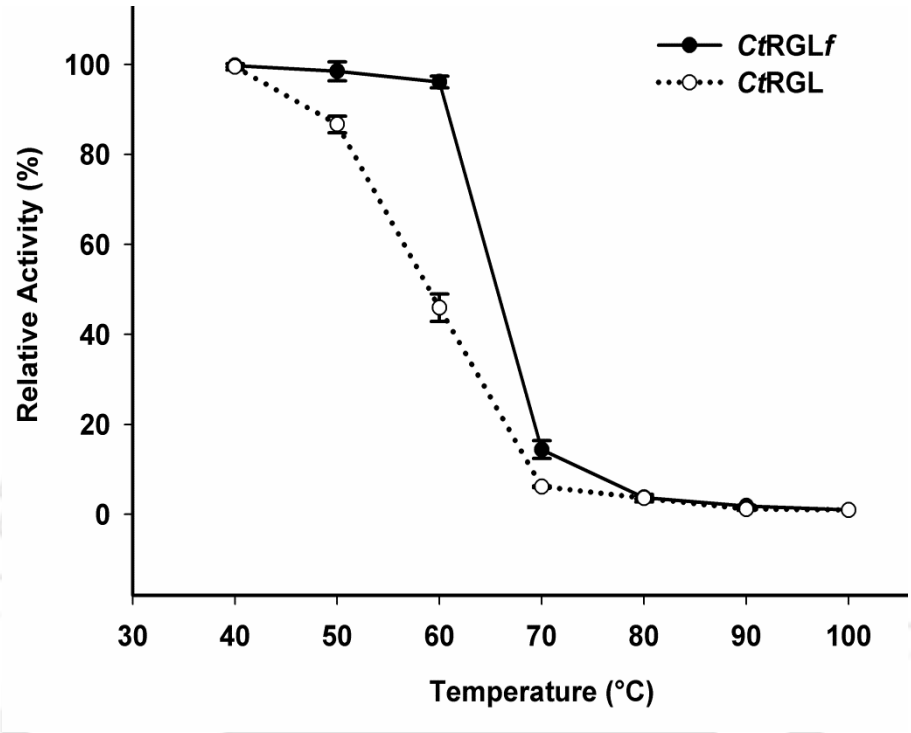


Fig 3.3 Thermal stability of *Ct*RGL*f* and *Ct*RGL. The activity of the enzymes incubated at 4°C was taken as 100%.

3.3.5 Kinetic parameters of *Ct*RGL*f* and *Ct*RGL

The K_m values of *Ct*RGL*f* and *Ct*RGL with RGS were identical at 4.8 and 5.1 mg/ml, respectively (Table 3.4). This indicated that *Ct*RGL, the truncated form of *Ct*RGL*f* can cleave RGS, independently even in the absence of CBM. The k_{cat} values of *Ct*RGL*f* and *Ct*RGL against RGS were 21 s^{-1} and 7 s^{-1}, respectively (Table 3.4). The lower k_{cat} value of *Ct*RGL may be attributed to the low quantitity of appropriately folded protein molecules.

Table 3.4 Kinetic parameters of *Ct*RGL*f* and *Ct*RGL. One unit of enzymatic activity was defined as the amount of enzyme in mg that produced 1 mmol/L of unsaturated product per minute**.**

Substrate (1%, w/v)	*Ct*RGL*f*			*Ct*RGL		
	K_m (mg/ml)	k_{cat} (s^{-1})	k_{cat}/K_m (ml/mg/s)	K_m (mg/ml)	k_{cat} (s^{-1})	k_{cat}/K_m (ml/mg/s)
Rhamnogalacturonan (Soyabean)	4.8	21.2	4.41	5.1	6.9	1.35
Rhamnogalacturonan (Potato)	4.9	21.0	4.28	5.3	6.0	1.13
Pectic galactan (Lupin)	10	7.7	0.77	11.2	4.8	0.43
Galactan (Potato)	4.7	6.4	1.36	6.4	2.1	0.32
Pectin (25% methyl-esterified)	8.2	9.8	1.19	7.6	2.3	0.3
Polygalacturonic acid	ND	ND	ND	ND	ND	ND

3.3.6 Effect of metal ions on the activity of *Ct*RGL*f* and *Ct*RGL

Sequence analysis of *Ct*RGL*f* revealed the presence of conserved Ca^{2+} ions binding amino acid residues. Therefore, the effect of divalent metal ions on the activity of *Ct*RGL*f* and *Ct*RGL was analysed after removing the metal ions by treating with 10 mM EDTA. This completely abolished the enzyme activity (Table 3.5). These results substantiated the important role of divalent metal ions on the activity of *Ct*RGL*f* and *Ct*RGL. In the absence of any of the metal ions, *Ct*RGL*f* and *Ct*RGL showed 6.5 and 4.4 U/mg enzyme activity, which was 33% and 23% less than their activity in presence of 3 mM Ca^{2+} ions. Among the various metal ions tested, Ca^{2+} ions (3 mM) restored the enzyme activity of EDTA-treated *Ct*RGL*f* and *Ct*RGL up to 45% and 35%, respectively Ca^{2+} (Table 3.5). However, Mn^{2+} ions (3 mM) resulted in higher reactivation than Ca^{2+} ions (3 mM), restoring the enzyme activity of EDTA treated *Ct*RGL*f* and *Ct*RGL by 96% and 77%, respectively. Ochiai *et al* (2007) have also reported that Mn^{2+} ions were more potent than Ca^{2+} ions in restoring the activity of EDTA-treated two RG lyases from *B. subtilis*.

Table 3.5 Effect of metal ions on the activity of *Ct*RGL*f* and *Ct*RGL.

Pre-treatment	Metal ion (3 mM)	Activity (%)	
		*Ct*RGL*f*	*Ct*RGL
No EDTA	-	67	77
	*Ca^{2+}	100	100
	Mn^{2+}	100	70
	Mg^{2+}	81	76
	Co^{2+}	14	14
EDTA (10 mM)	-	0	0
	Ca^{2+}	45	35
	Mn^{2+}	96	77
	Mg^{2+}	32	25
	Co^{2+}	48	29

**Specific activities of CtRGLf and CtRGL in presence of 3mM Ca^{2+} ions were 9.8 U/mg and 5.8 U/mg respectively*

Two different types of RG lyases have been reported, one which do not have an absolute requirement of calcium ions and show optimum activity at acidic pH and the other which have an obligatory requirement of calcium ions and show optimum activity at basic pH. RG lyase from *Erwinia chrysenthemi* is active at acidic pH and does not depend on Ca^{2+} ions for its activity (Laatu *et al*., 2003). Conversely the RG lyases from *B. subtilis*, *C. cellulolyticum*, *C. japonicas* and *B. licheniformis* are active at alkaline pH and require Ca^{2+} ions (Ochiai *et al*., 2007; Pages *et al*., 2003; Silva *et al*., 2014). *Ct*RGL*f* and *Ct*RGL were optimally active at alkaline pH and required Ca^{2+} ions for their activity (Fig. 3.4). The activities of *Ct*RGL*f* and *Ct*RGL were enhanced by 1.5 and 1.3 fold, respectively, by 3 mM Ca^{2+} ions.

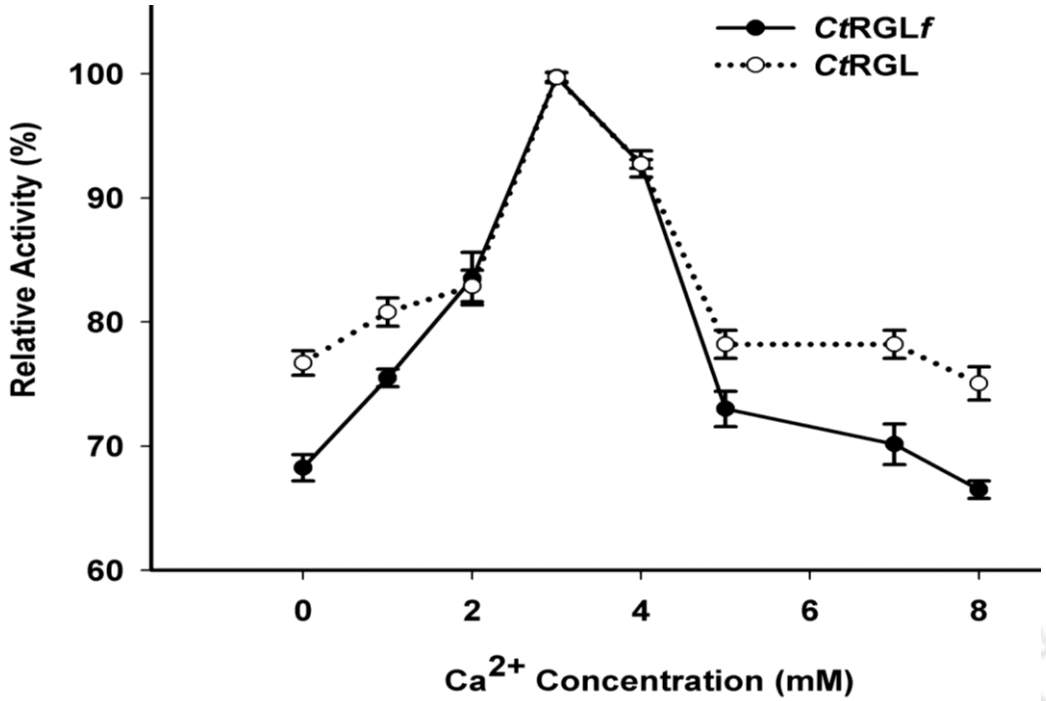


Fig. 3.4 Effect of Ca^{2+} ions ($CaCl_2$) on the activities of *Ct*RGL*f* and *Ct*RGL. The activity of both enzymes at 3 mM Ca^{2+} ions concentration was taken as 100%.

3.3.7 Effect of various compounds on activity of *Ct*RGL*f* and *Ct*RGL

The effect of various compounds on the activity of *Ct*RGL*f* was investigated (Table 3.6). In presence of the detergent SDS (5 mM), *Ct*RGL*f* retained more than 50% of its activity. Even in presence of the chaotropic agent Urea (5 M), a *Ct*RGL*f* retained 61% activity. In presence of reducing agent dithiothreitol (5 mM) *Ct*RGL*f* retained 75% activity while in presence of oxidizing agent hydrogen peroxide (5 mM) *Ct*RGL*f* retained 66% of its activity. D-Galacturonic acid (5 mM) had no effect on the activity of *Ct*RGL*f*. L-Rhamnose (5 mM) only marginally reduced the activity of *Ct*RGL*f* by 16%.

Table 3.6 Effect of various compounds on activity of *Ct*RGL*f*

Compound	Relative Activity *Ct*RGL*f* (%)
Control*	100
SDS (5 mM)	55
Urea (5 M)	61
Hydrogen peroxide (5 mM)**	66
Dithiothreitol (5 mM)	75
L-Rhamnose (5 mM)	84
D-Galacturonic acid (5 mM)	100

**Control reaction was carried out in presence of 3mM Ca^{2+} ions under optimised conditions.*
***Reaction mixture contaning hydrogen peroxide lacked Ca^{2+} ions as hydrogen peroxide decomposes in presence of calcium. Activity of Ct*RGL*f in presence of hydrogen peroxide was compared to activity of Ct*RGL*f in absence of Ca^{2+} ions and hydrogen peroxide.*

3.3.8 Thin-layer chromatography analysis of *Ct*RGL*f* and *Ct*RGL treated RGS

The profiles of time dependent enzymatic degradation of RGS by *Ct*RGL*f* and *Ct*RGL were analysed by TLC (Fig. 3.5). It has been reported earlier that the smallest possible reaction product generated by the activity of a RG lyase on RGS is unsaturated RG disaccharide whose mobility is similar to that of D-Gal*p*A . All the samples of both *Ct*RGL*f* and *Ct*RGL collected at different time intervals over 24 h displayed the products of variable sizes, appearing as discrete spots with much less mobility than D-Gal*p*A on TLC plate. This suggested that both *Ct*RGL*f* and *Ct*RGL cleave the substrate endolytically and produce oligosaccharides larger than the unsaturated RG disaccharide as the major products. This type of substrate cleavage pattern was also displayed by an endo- RG lyase (YesW) from *B. Subtilis* (Ochiai *et*

al., 2007). The pectin and pectic oligosaccharides have been reported to show anti-proliferative effects on the human colonic adenocarcinoma cell line HT29 (Olano *et al*., 2002). The RGI component of potato was reported to inhibit the proliferation of HT29 colon cancer cells (Cheng *et al*., 2013). These observations suggest that the RG oligosaccharides produced by action of *Ct*RGL*f* may be examined for therapeutic applications.

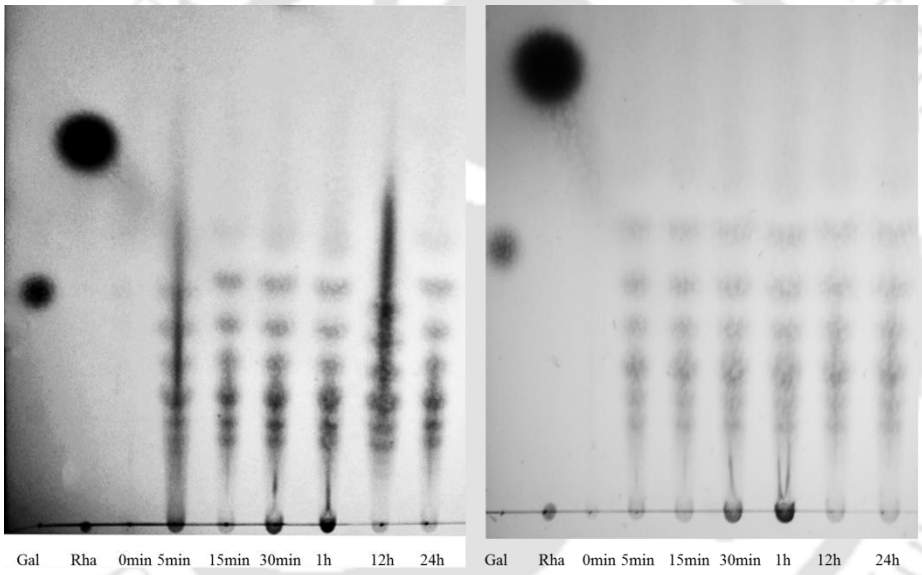


Fig. 3.5 TLC analysis of degradation products of (A) *Ct*RGL*f* and (B) *Ct*RGL. Gal- β-Gal*p*A and Rha- α-Rha*p* were used as standards.

3.4 Conclusions

*Ct*RGL*f* and *Ct*RGL showed maximum enzyme activity at pH 8.5. *Ct*RGL*f* was active at pH range 7.0-10.0 while *Ct*RGL was active only at pH 7.0-9.0. The optimal temperature of *Ct*RGL*f* and *Ct*RGL was 70ºC and 60ºC respectively. *Ct*RGL*f* retained 90% of its activity after incubation at 60ºC for 30 min. *Ct*RGL retained 45% of its activity after incubation at 60ºC for 30 min. Both *Ct*RGL*f* and *Ct*RGL showed high activity towards rhamnogalacturonan I conatining polysaccharides. Maximum activity of both, *Ct*RGL*f* and *Ct*RGL was against rhamnogalacturonan from soybean (RGS) at 9.8 U/mg and 5.8 U/mg respectively. *Ct*RGL*f* and *Ct*RGL had similar values of K_m, 4.8 mg/ml and 5.1 mg/ml respectively, with RGS. The presence of 3 mM Ca^{2+} increased the activity of *Ct*RGL*f* and *Ct*RGL by 1.5 and 1.3 folds respectively. Treatment of *Ct*RGL*f* and *Ct*RGL with 10 mM EDTA completely abolished their activity. The TLC analysis of the products formed by the action of *Ct*RGL*f* and *Ct*RGL showed that they endolytically cleaved RGS.

References

Aachary, A. A., Prapulla, S. G. (2011) Xylooligosaccharides (XOS) as an emerging prebiotic: microbial synthesis, utilization, structural characterization, bioactive properties, and applications. Comprehensive Reviews in Food Science and Food Safety, 10:2-16.

Anbar, M., Gul, O., Lamed, R. (2012) Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis. Applied and Environmental Microbiology, 78:3458-3464.

Bayer, E. A., Morag, E., Lamed, R. (1994) The cellulosome- a treasure-trove for biotechnology. Trends in Biotechnology, 12:379-386.

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

Cheng, H., Zhang, Z., Leng, J. (2013) The inhibitory effects and mechanisms of rhamnogalacturonan I pectin from potato on HT-29 colon cancer cell proliferation and cell cycle progression. International Journal of Food Science and Nutrition, 64:36-43.

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

Das, S. P., Ravindran, R., Ahmed, S. (2012) Bioethanol production involving recombinant *C. thermocellum* hydrolytic hemicellulase and fermentative microbes. Applied Biochemistry and Biotechnology, 167:1475-1488.

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

Iwai, M., Yamada, H., Ikemoto, T., Matsumoto, S., Fujiwara, D., Takenaka, S., Sakamoto, T. (2015) Biochemical characterization and overexpression of an endo-rhamnogalacturonan lyase from *Penicillium chrysogenum*. Molecular Biotechnology, 57(6):539-548.

Kashyap, D. R., Vohra, P. K., Tewari, R. (2001) Application of pectinases in the commercial sector: a review. Bioresource Technology, 77:215-227.

Koshland, D. E. (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews*, 28:416-436.

Koukiekolo, R., Cho, H. Y., Kosugi, A. (2005) Degradation of corn fibre by *Clostridium cellulovorans* cellulases and hemicellulases and contribution of scaffolding protein CbpA. Applied and Environmental Microbiology, 71:3504-3511.

Lamed, R., Setter, E., Bayer, E. A. (1983) Characterization of a *cellulose*-binding, cellulase-containing complex in *Clostridium thermocellum*. Journal of Bacteriology, 156:828-836.

Laatu, M., Condemine, G. (2003) Rhamnogalacturonate lyase RhiE is secreted by the out system in *Erwinia chrysanthemi*. Journal of Bacteriology, 185:1642-1649.

Li, S., Yang, X., Bao, M., Wu, Y., Yu, W., Han, F. (2015) Family 13 carbohydrate-binding module of alginate lyase from *Agarivorans* sp. L11 enhances its catalytic efficiency and thermostability, and alters its substrate preference and product distribution. FEMS Microbiology Letters, 362(10):fnv054.

Lombard, V., Golaconda, R. H., Drula, E. (2014) The Carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research, 42:D490-D495.

Matsunaga, T., Ishii, T., Matsumoto, S. (2004) Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes and bryophytes. Implications for the evolution of vascular plants. *Plant Physiology*, 134:339-351.

McKie, V. A., Vincken, J. P., Voragen, A. G. (2001) A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. Biochemcal Journal, 355:167–177.

Moran, F. S, Nasuno, S., Starr, M. P. (1968) Extracellular and intracellular polygalacturonic acid trans-eliminase of *Erwinia carotovora*. Archieves of Biochemistry and Biophysics, 123:298–306.

Ochiai, A., Itoh, T., Kawamata, A. (2007) Plant cell wall degradation by saprophytic *Bacillus subtilis* strains: gene clusters responsible for rhamnogalacturonande polymerization. Applied Environmental Microbiology, 73:3803-3813.

.Oomen, R. J., Doeswijk, V. H., Bush, M. S. (2002) *In muro* fragmentation of the rhamnogalacturonan I backbone in potato (*Solanum tuberosum* L.) results in a reduction and altered location of the galactan and arabinan side-chains and abnormal periderm development. *Plant Journal*, 30:403-413.

Olano, E., Rimbach, G. H., Gibson, G. R., Rastall, R. A. (2002) Pectin and pectic-oligosaccharides induce apoptosis in in vitro human colonic adenocarcinoma cells. Anticancer Research, 23(1A):341-346.

O'Neill, M. A., Warrenfeltz, D., Kates, K. (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. Journal of Biological Chemistry, 271:22923-22930.

Pages, S., Valette, O., Abdou, L. (2003) A rhamnogalacturonan lyase in the *Clostridium cellulolyticum* cellulosome. Journal of Bacteriology, 185:4727-4733.

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

Silva, I. R., Jers, C., Otten, H. (2014) Design of thermostable rhamnogalacturonan lyase mutants from *Bacillus licheniformis* by combination of targeted single point mutations. Applied Microbiology and Biotechnology, 98:4521-4531.

Vincken, J. P, Schols, H. A, Oomen, R. J. (2003) If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. Plant Physiology, 132:1781-1789

Zhang, Z., Xiao, Z., Linhardt, R. J. (2009) Thin layer chromatography for the separation and analysis of acidic carbohydrates. Journal of Liquid Chromatography and Related Technologies, 32:1711-1732.

Chapter 4

Structure model and substrate binding analysis of family 11 polysaccharide lyase (*Ct*RGL) catalytic module from *Clostridium thermocellum*

4.1 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 form α-(1,4) linked D-galactopyranosyl uronic acid (Gal*p*A) residues. RG I main chain contains alternating Gal*p*A and L-rhamnopyranosyl (L-Rha*p*) residues. The monomeric unit of RG I main chain is a disaccharide, [→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→]. The RG I main chain is decorated with linear or branched chains of α-L-arabinofuranosyl (α-L-Ara*f)* and β-D-galactopyranosyl residues (β-D-Gal*p*) 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 Gal*p*A residues substituted with various monosaccharide residues including L-Rha*p* (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 Condemine, 2003). RG I degrading enzyme may cleave RG I main chain by hydrolysis (glycoside hydrolases) or β-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 *Erwiniachrys anthemi* (renamed as *Dickeya dadantii*), *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 *Ct*RGL*f* (*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_Bl (*Bacillus licheniformis*) have been solved (Ochiai *et a*l., 2007; Ochiai *et al*., 2009; Silva *et al*., 2014). PL11 RG lyases take an 8-bladed β-propeller fold.

Clostridium thermocellum is a thermophillic, anaerobic bacterium that secretes a wide repertoire of PCW polysaccharide degrading enzymes. These enzymes assemble by means of protein-protein (Cohesin-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 pectate lyases (Fontes and Gilbert, 2010). However, only one RG I degrading enzyme, *Ct*RGL*f*, an RG lyase has been characterized (Dhillon *et al*., 2016). This study reports the structural characterization of the catalytic module, *Ct*RGL, present at the N-terminal of multi-modular RG lyase, *Ct*RGL*f* by homology modeling, circular dichroism and its docking analysis.

4.2. Materials and Methods

4.2.1 Amino acid sequence analysis of *Ct*RGL

The amino acid sequence of *Ct*RGL, the catalytic module of *Ct*RGL*f* (GenBank Acession 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).

4.2.2 Secondary structure analysis of *Ct*RGL

The amino acid sequence of *Ct*RGL (GenBank Acession no. ABN51485.1) was used to predict the secondary structure of *Ct*RGL using PsiPred tool (Jones, 1999). The secondary structure composition of *Ct*RGL was determined by circular dichroism (CD). The gene encoding *Ct*RGL from *Clostridium thermocellum* cloned, expressed and purified earlier (Dhillon *et al*., 2016) was used in the present study. 15 µM of purified recombinant *Ct*RGL dissolved in 1 ml of 50 mM TrisHCl buffer (pH 8.5) was used for the CD analysis in the Far UV range (190-250 nm). The CD spectrum was recorded on a spectro-polarimeter (J-815 Jasco Corporation, Tokyo) at 25°C. The CD spectrum of *Ct*RGL was expressed in terms of molar residual ellipticity (MRE deg cm^2dmol^{-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).

4.2.3 Homology modeling of *Ct*RGL

The Modeller 9.14 program was used to build the 3-dimensional structure of *Ct*RGL (Eswar *et al*., 2006). Multiple templates were used to model the 3D-structure of *Ct*RGL. 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 *Ct*RGL was then aligned to the sequence of template structures. The 'automodel' class was used for building *Ct*RGL 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.

4.2.4 Docking analysis of modeled *Ct*RGL

Molecular docking study was performed using AutoDock (Morris et al., 2007). The modeled *Ct*RGL structure (containing Ca^{2+} ions) after energy minimization was used. Autodock did not assign any charge to the Ca^{2+} ions while preparation of *Ct*RGL structure for docking. Rhamnogalacturonan I (RG I) ligands used were: RG I disaccharide- α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2), RG I trisaccharide- α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*, RG I tetrasaccharide- α-L-Rha*p*-(1→4)-α-D-

Gal*p*A-(1→2)-α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2), RG I pentasaccharide- α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p* and RG I hexasaccharide- α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2).

Homogalacturonan ligands used were, digalacturonic acid- α-D-Gal*p*A-(1→4)-α-D-Gal*p*A, trigalacturonic acid- α-D-Gal*p*A-(1→4)-α-D-Gal*p*A-(1→4)-α-D-Gal*p*A, tetragalacturonic acid- α-D-Gal*p*A-(1→4)-α-D-Gal*p*A-(1→4)-α-D-Gal*p*A-(1→4)-α-D-Gal*p*A and pentagalacturonic acid- α-D-Gal*p*A-(1→4)-α-D-Gal*p*A-(1→4)-α-D-Gal*p*A-(1→4)-α-D-Gal*p*A-(1→4)-α-D-Gal*p*A. PDB files for the ligands were prepared using the GLYCAM server (Kirschner et al., 2008). The ligands and enzyme (*Ct*RGL) 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 0.375Å. 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.

4.3 Results and Discussion

4.3.1 Sequence analysis of *Ct*RGL

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 (*Ct*RGL) 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 4.1).

Table 4.1 BLAST analysis of *Ct*RGL to identify its homologues.

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

Multiple sequence alignment of *Ct*RGL with these sequences provided information about the residues involved in catalysis (Fig. 4.1). 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 *Ct*RGL.

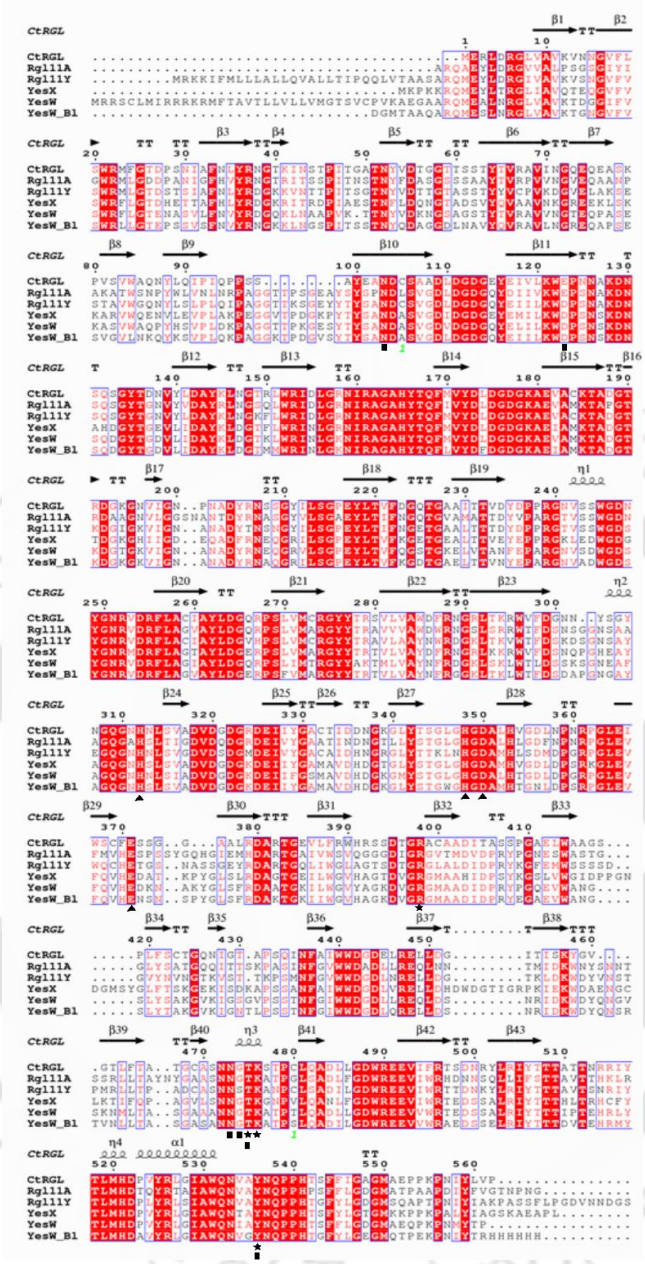


Fig. 4.1 Multiple sequence analysis of *Ct*RGL from *Clostridium thermocellum* with rhamnogalacturonan lyases from *Cellvibrio japonicus* (Rgl11A), *Bacillus subtilis* (YesW & YesX), *Bacillus licheniformis* (YesW_Bl) and *Clostridium cellulolyticum* (Rgl11Y). The conserved residues are shown against 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 Ca^{2+} ion at the active site have been indicated by symbol (▲). The secondary structure elements of *Ct*RGL have been shown above the sequences. ('α'- alpha helix, 'β'- beta strand, 'T'-turn and 'η'- 3_{10}helix).

The residues, Thr475 and Tyr536 make hydrogen bonds with rhamnose present at -1 subsite and are conserved in *Ct*RGL. The residues, Asn103, Glu123, Asn473 and Gly474 making hydrogen bonds with rhamnose residue present at +2 subsite are also conserved. *Ct*RGL also contains conserved rhamnose binding residues (Asn138, Arg158, Gly189 and Arg206) in addition to the residues present at the active site. The amino acid residues of YesW forming coordinate bonds with the Ca^{2+} ion involved in catalysis are conserved in *Ct*RGL as His312, His348, Asp350 and Glu371 (Fig. 4.1).

4.3.2 Secondary structure of *Ct*RGL

PsiPred server predicted that 1.7% of *Ct*RGL residues form α-helix, 31.7 % residues are arranged as β-strands and 66.6% of the residues occur as loops (Table 4.2, Fig. 4.2A). The secondary structure composition of *Ct*RGL was confirmed by CD analysis. The CD analysis results 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 4.3.2, Fig. 4.2B). The CD results corroborated with those predicted by PsiPred.

Table 4.2 Secondary structure composition of *Ct*RGL.

Secondary structure element	PsiPred (%)	CD analysis (%)
α-helix	1.7	2.75
β-sheet	31.7	30.1
Random coil	66.6	67.15

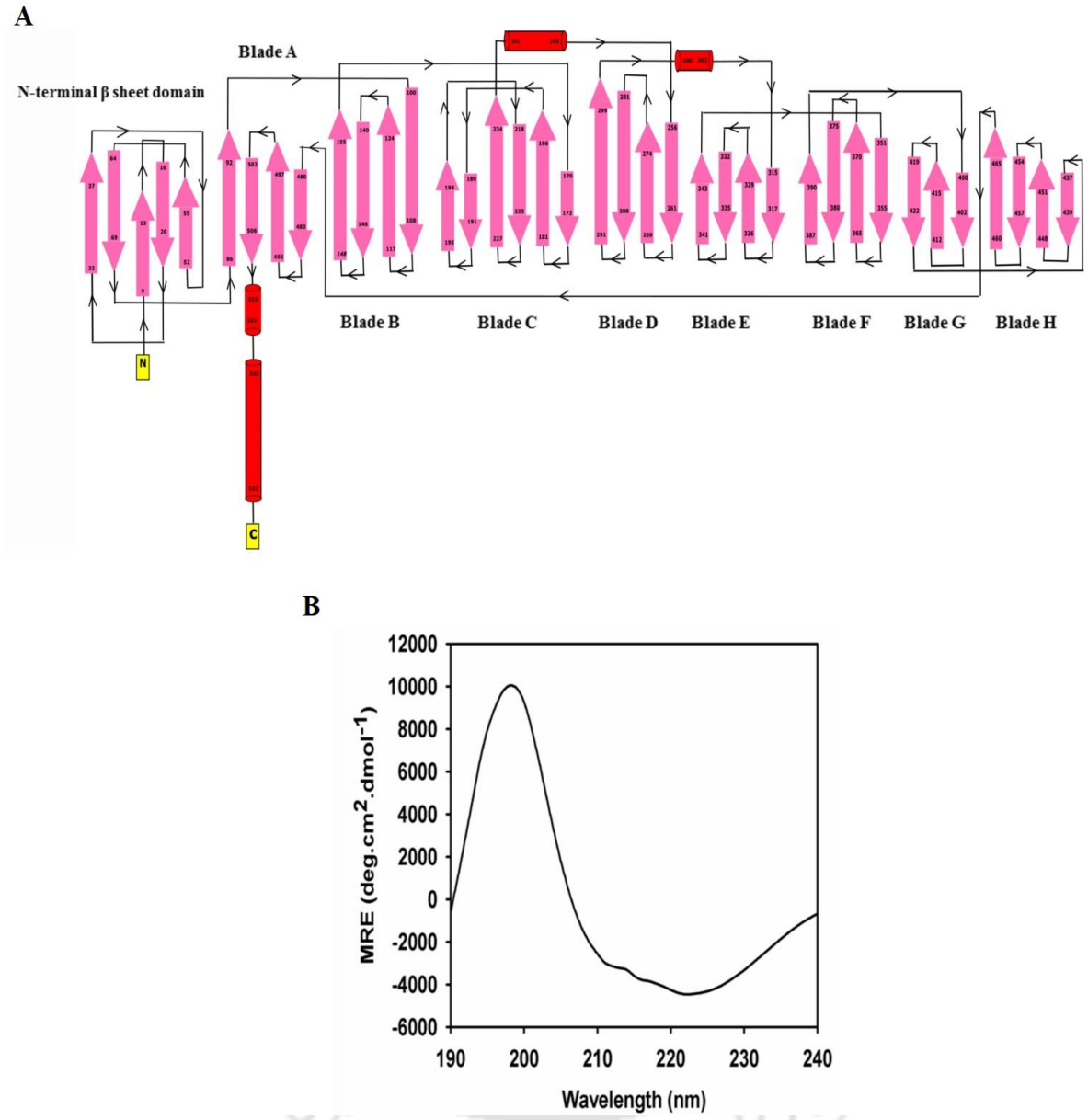


Fig. 4.2 (A) Topology diagram of modelled *Ct*RGL where β-sheets are shown in pink arrows and α-helices are shown as red cylinders. (B) Far UV-CD spectrum of *Ct*RGL.

4.3.3 3D-structure model of *Ct*RGL

Multiple sequence alignment showed stretches of semi-conserved residues among the aligned sequences (Fig. 4.1). Some stretches of amino acid sequence of *Ct*RGL 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 *Ct*RGL structure was organized into an N-terminal β-sheet domain (Val9-Val69) and a β-propeller domain (Tyr88-Pro562) (Fig. 4.3A). The Ramachandran plot developed for *Ct*RGL 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. 4.3B). Verify3D results showed that 100% of the residues had an average 3D/1Dscore≥ 0.2. The DALI server was used to find structural homologues of *Ct*RGL. The results showed that modeled *Ct*RGL structure is most similar to that of RG lyase, YesW with an rmsd of 0.5Å over 559 C_α atoms (Holm and Rosenström, 2010).

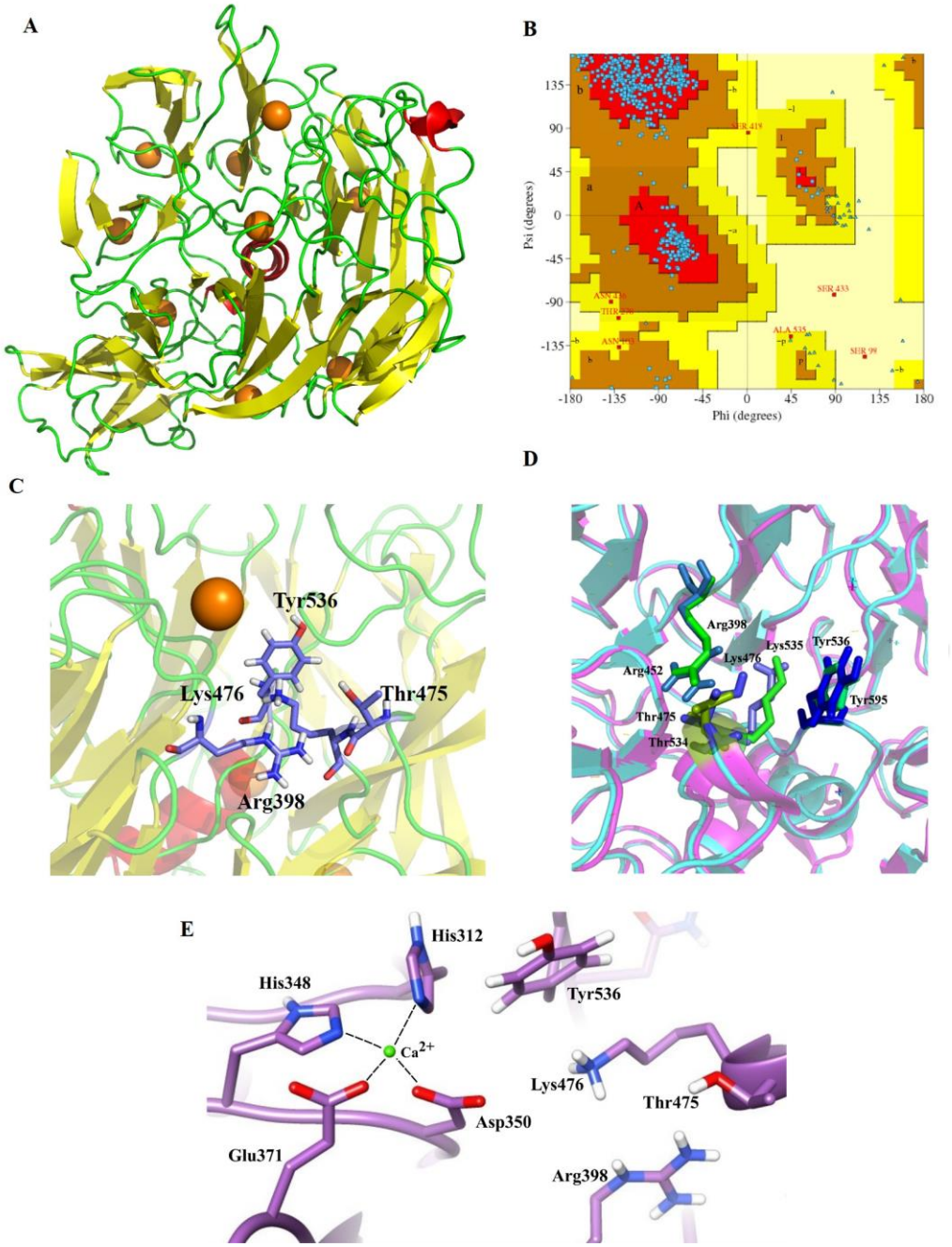


Fig. 4.3 (A) Modelled *Ct*RGL structure showing eight bladed β-propeller fold. Ca^{2+} ions are shown as orange spheres, (B) Ramachandran plot for modeled *Ct*RGL, (C) active site residues of *Ct*RGL where Ca^{2+} ion is shown as a sphere, (D) superposition of modeled *Ct*RGL structure (cyan) over structure of YesW (magenta) (PDB id: 2Z8S) shows similar orientation of active residues. *Ct*RGL residues are shown in blue while YesW residues are in light green and (E) Coordination of Ca^{2+} ion present at the active site by *Ct*RGL residues.

The N-terminal β-sheet domain of *Ct*RGL is composed of five β-strands arranged as two anti-parallel β-sheets (Fig. 4.2A). 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 *Ct*RGL (Ochiai *et al.*, 2009). The β-propeller domain consists of anti-parallel β-sheets arranged as eight blades (A-H) (Fig. 4.2A) of a propeller around a central α-helix (Fig. 4.3A). 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 Ca^{2+}ion binding sites and binds one Ca^{2+}ion except blade D and E. Blade D does not bind any Ca^{2+}ion while blade E binds two Ca^{2+}ions. Two Ca^{2+}ions are present in central cavity of the β-propeller. The modelled *Ct*RGL structure was superposed with its closest homologue, YesW. It was observed that a Ca^{2+}ion is present at the active site (Fig. 4.3C). The active site residues, Arg398, Thr475, Lys476 and Tyr536 of *Ct*RGL were oriented in the same manner as YesW active site residues (Fig. 4.3D). The Ca^{2+}ion present in the active cleft of YesW was shown to interact with the substrate (Ochiai *et al.*, 2009). The Ca^{2+} ion present at the active site of *Ct*RGL is therefore crucial and it is coordinated by conserved residues, His312, His348, Asp350 and Glu371 (Fig. 4.3E).

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. 4.1 and Fig. 4.4). 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 *Ct*RGL (Fig. 4.1). This indicated that *Ct*RGL is also an endo-RG lyase. Moreover, the endo-lytic mode of *Ct*RGL activity was also demonstrated earlier (Dhillon *et al*., 2016).

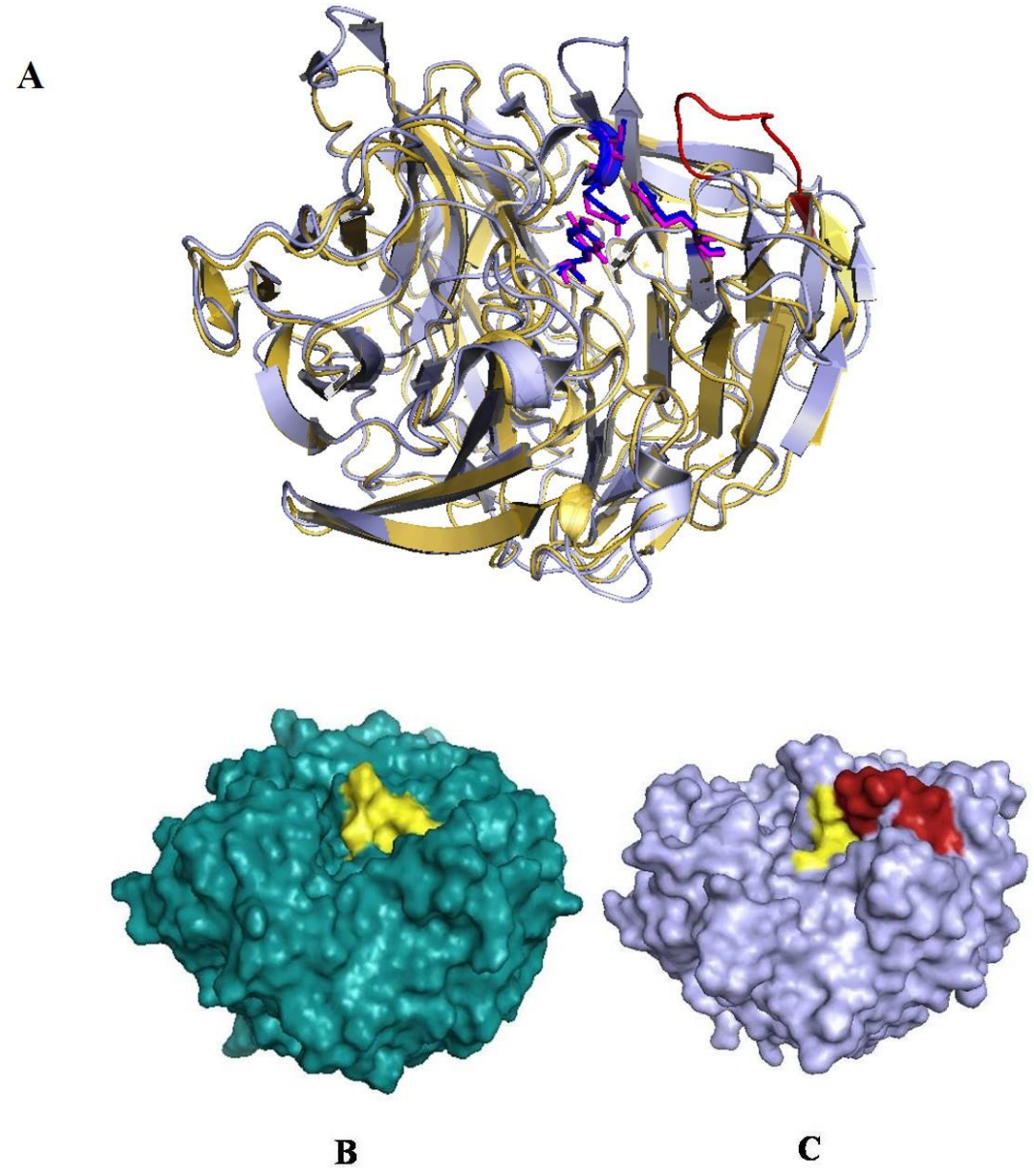


Fig. 4.4 (A) Superposition of modeled *Ct*RGL structure (yellow) over structure of exo-RG lyase, YesX (blue) (PDB id: 2ZUY). The loop, (^{439}PPGNDGMSY^{447}), responsible for exo-type activity of Yes X is shown in red. *Ct*RGL residues are shown in pink while YesX residues are in blue color. (B) Molecular envelop of *Ct*RGL. Active site residues are shown in yellow (C) Molecular envelop of YesX. Active site residues are shown in yellow and the loop (^{439}PPGNDGMSY^{447}) is shown in red.envelop of YesX. Active site residues are shown in yellow and the loop (^{439}PPGNDGMSY^{447}) is shown in red.

4.3.4 Docking analysis and ligand binding by *Ct*RGL

The interaction of *Ct*RGL with its ligands was analyzed by docking studies. The binding energies for these ligands are reported in Table 4.3. The binding energy of rhamnogalacturonan I oligosaccharides of up to 5 residues in length was -5 kcal/mol, but the binding of RG I hexa-saccharide [α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→2)] was comparatively weaker (-1 kcal/mol). This indicated that the active site of *Ct*RGL could only accommodate five residues.

Table 4.3 Docking analysis of interaction between *Ct*RGL 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, Lys476, Gly397

The RG I pentasaccharide docked on the surface of *Ct*RGL is shown in Fig. 4.5. 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. 4.5B). *Ct*RGL made eight H-bonds with α-L-Rha*p* residues of RG I pentasaccharide

and only 2 H-bonds with α-D-Gal*p*A residues. This indicated the importance of α-L-Rha*p* residues in ligand recognition. The pentagalacturonic acid oligosaccharide is held at the active site of *Ct*RGL by only 3 hydrogen bonds, whereas, RG I pentasaccharide is held by 12 hydrogen bonds (Fig. 4.5C). The weak binding of pentagalacturonic acid showed that *Ct*RGL has greater specificity for α-L-Rha*p* residues than α-D-Gal*p*A residues and that α-L-Rha*p* are crucial for the enzyme (Table 4.3).

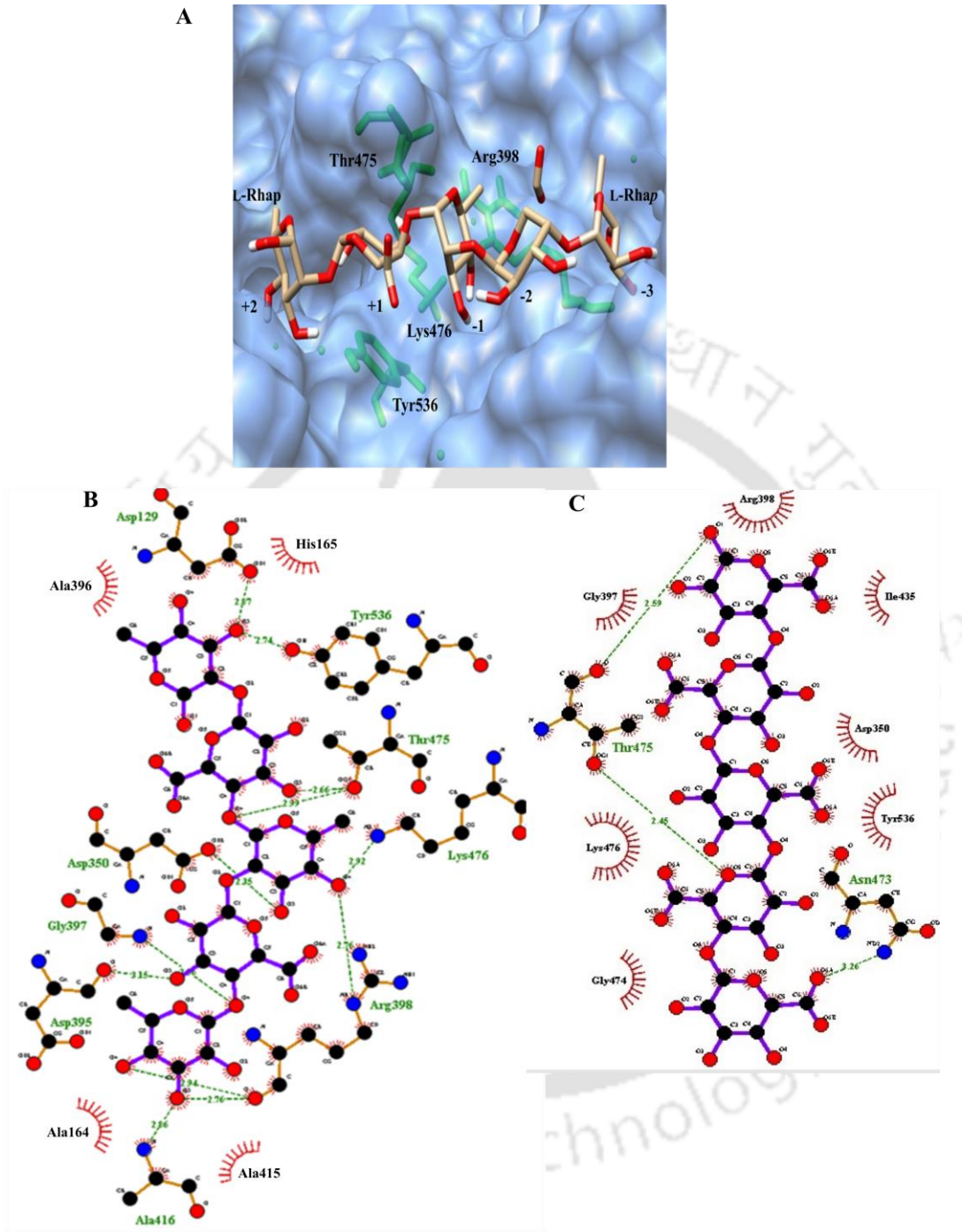


Fig. 4.5. (A) Surface view of *Ct*RGL showing docked RG I pentasaccharide at the active cleft. (B) Schematic representation of interaction between residues of *Ct*RGL and RG I pentasaccharide. Ca^{2+} have been shown as green spheres.

A comparison of docked poses of ligands was made (Fig. 4.6). It was found that RG I di-, tri-, tetra- and penta-saccharide were held at the *Ct*RGL 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. 4.6). 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 *Ct*RGL along its length, like other ligands (Fig. 4.6).

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-Rha*p* and α-D-Gal*p*A residues in structures of YesW (2ZUX, 2Z8S) complexed with ligands (Fig. 4.5A). 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.

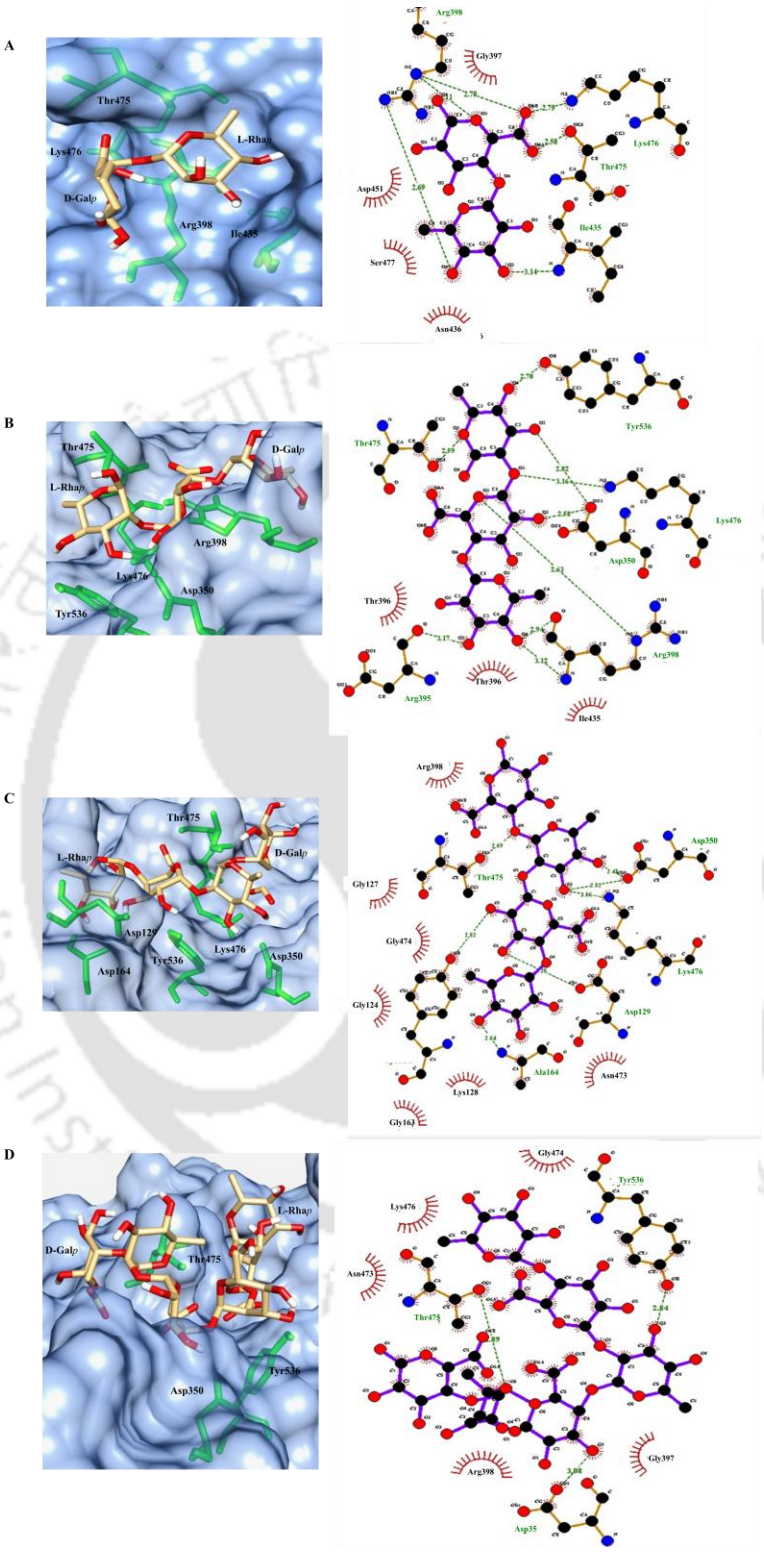


Fig. 4.6. Comparison of poses of various ligands docked with *Ct*RGL. Surface view and schematic presentation of interaction between *Ct*RGL and (A) RG I disaccharide, (B) RG I trisaccharide, (C) RG I tetrasaccharide and (D) RG I hexasaccharide.

4.4 Conclusions

The truncated N-terminal catalytic module (*Ct*RGL) of molecular size ~ 64 kDa of cellulosomal rhamnogalacturonan lyase, (*Ct*RGL*f*) from *Clostridium thermocellum* belonging to family 11 polysaccharide lyase (PL11) was structurally characterized by molecular modelling. Multiple sequence alignment revealed that *Ct*RGL has conserved active site as well as Ca^{2+} 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 *Ct*RGL based on homology modeling showed a β-propeller fold. The structure validation by Ramachandran plot revealed 99.6% amino acid residues in the allowed region. Comparison of *Ct*RGL structure with that of YesW and YesX proteins from *Bacillus subtilis* showed conserved catalytic cleft and endo-lytic mode of action. Docking analysis of *Ct*RGL 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 β-elimination. This study provides an insight into the structural determinants for mode of action and substrate recognition.

References

Altschul, S.F., Gish, W., Miller. W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. Journal of Molecular Biology, 215(3): 403-410.

Bauer, S., Vasu, P., Persson, S., Mort, A. J., Somerville, C. R. (2006) Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. Proceedings of the National Academy of Sciences (USA), 103: 11417–22.

Davies, G. J., Wilson, K. S., Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochemical Journal, 321 (Pt 2): 557.

Dhillon, A., Fernandes, V. O., Dias, F. M., Prates, J. A., Ferreira, L. M., Fontes, C. M., Goyal, A. (2016) A new member of family 11 polysaccharide lyase, rhamnogalacturonan lyase (*Ct*RGL*f*) from *Clostridium thermocellum*. Molecular Biotechnology, 58(4): 232-240.

Eswar, N., Marti-Renom, M. A., Webb, B., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., Sali. A. (2006) Comparative protein structure modeling with MODELLER. Current Protocols in Bioinformatics, 15: 5.6.1-5.6.30.

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

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

Iwai, M., Yamada, H., Ikemoto, T., Matsumoto, S., Fujiwara, D., Takenaka, S., Sakamoto, T. (2015) Biochemical characterization and overexpression of an

endo-rhamnogalacturonan lyase from *Penicillium chrysogenum*. Molecular Biotechnology, 57(6): 539-548.

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

Kelly, S. M., Jess, T. J., Price, N.C. (2005) How to study proteins by Circular Dichroism. Biochimica et Biophysica Acta, 1751: 119-139.

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. Journal of Computational Chemistry, 29: 622− 655.

Kofod, L.V., Kauppinen, S., Christgau, S., Andersen, L. N., Heldthansen, H. P, Dorreich, K. (1994) Cloning and characterization of two structurally and functionally divergent rhamnogalacturonases from *Aspergillus aculeatus*. Journal of Biological Chemistry, 269: 29182–89.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., 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.

Laatu, M., Condemine, G. (2003) Rhamnogalacturonate lyase RhiE is secreted by the out system in *Erwinia chrysanthemi*. Journal of Bacteriology, 185: 1642-1649.

Laskowski, R. A. (2001) PDBsum: summaries and analyses of PDB structures. Nucleic Acids Research, 29(1): 221-222.

Lin, X., Kaul, S., Rounsley, S., Shea, T. P, Benito, M. I, Town, C. D. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature, 402: 761–8.

Lombard V., Golaconda, R. H., Drula, E., Coutinho, P. M., Henrissat, B. (2014) The Carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research, 42: D490–D495.

McKie, V.A., Vincken, J.P., Voragen, A.G., van den Broek, L.A., Stimson, E., Gilbert, H.J. (2001) A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. Biochemical Journal, 355: 167–177.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., Olson, A.J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. Journal of Computational Chemistry, 16: 2785-91.

Naran, R., Pierce, M. L., Mort, A. J. (2007) Detection and identification of rhamnogalacturonan lyase activity in intercellular spaces of expanding cotton cotyledons. The Plant Journal, 50(1): 95-107.

Ochiai, A., Itoh, T., Kawamata, A., Hashimoto, W., Murata. K. (2007) Plant cell wall degradation by saprophytic *Bacillus subtilis* strains: gene clusters responsible for rhamnogalacturonande polymerization. Applied and Environmental Microbiology, 73: 3803-3813.

Ochiai, A., Itoh, T., Maruyama, Y., Kawamata, A., Mikami, B., Hashimoto, W., Murata, K. (2007) A Novel Structural Fold in Polysaccharide Lyases *Bacillus subtilis* family 11 rhamnogalacturonan lyase YesW with an eight-bladed β-propeller. Journal of Biological Chemistry, 282(51): 37134-37145.

Ochiai, A., Itoh, T., Mikami, B., Hashimoto, W., Murata, K. (2009) Structural determinants responsible for substrate recognition and mode of action in family 11 polysaccharide lyases. Journal of Biological Chemistry, 284(15): 10181-10189.

O'Neill, M. A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A. G., 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. Journal of Biological Chemistry, 271: 22923-22930.

Pages, S., Valette, O., Abdou, L., Belaich, A., Belaich, J. P. (2003) A rhamnogalacturonan lyase in the *Clostridium cellulolyticum* cellulosome. Journal of Bacteriology, 185: 4727-4733.

Perez-Iratxeta, C., Andrade-Navarro, M. A. (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. BMC Structural Biology, 8(1): 25.

Silva, I. R., Jers, C., Otten, H., Nyffenegger, C., Larsen, D. M., Derkx, P. M., Larsen, S. (2014) Design of thermostable rhamnogalacturonan lyase mutants from *Bacillus licheniformis* by combination of targeted single point mutations. Applied Microbiology and Biotechnology, 98: 4521-4531.

Silva, I. R., Jers, C., Meyer, A. S., Mikkelsen, J. D. (2016) Rhamnogalacturonan I modifying enzymes: an update. New Biotechnology, 33(1): 41-54.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Thompson, J. D. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7(1): 539.

Verma, A. K., Goyal, A. (2014) *In silico* structural characterization and molecular docking studies of first glucuronoxylan-xylanohydrolase (Xyn30A) from family 30 glycosyl hydrolase (GH30) from *Clostridium thermocellum*. Molecular Biology, 48(2): 278-286.

Chapter 5

Substrate binding analysis and structural characterization of family 35 carbohydrate binding module, *Rgl*-CBM35 associated with *Ct*RGL, a family 11 polysaccharide lyase

5.1 Introduction

The plant cell wall (PCW) is a highly networked structure principally composed of cellulose, hemicellulose, pectin and their derived carbohydrate polymers (Hervé *et al*., 2010; Gilbert *et al*., 2013). These polysaccharides are an abundant source of carbon fixed by photosynthesis (Hervé *et al*., 2010). Enzymes that modify and degrade these polysaccharides are critical, physiologically for plant growth and ecologically in plant pathogenesis as well as recycling of carbon as a nutrient. Paralleling the biochemical specialisation and structural complexity of PCW, most of such enzymes have evolved highly modular structural organisation such that they contain non-catalytic carbohydrate binding module(s) (CBM) connected via a linker to a catalytic module(s) (Gilbert *et al*., 2013). CBMs fold independently of the associated catalytic module and bind to carbohydrate ligands (Abbott *et al*., 2013). CBMs have been classified under different primary sequence based families in the CAZy database (http://www.cazy.org/). CBMs have been reported from a variety of organisms

ranging from bacteria, fungi to humans (Miyanaga *et al*., 2006; Ribeiro *et al*., 2010; Moreira *et al*., 2010). Recently a new family of CBM (CBM72) was established following the report of a CBM identified by metagenomic analysis which displayed wide ligand specificity (Duan *et al*., 2016). β-sandwich has emerged, from plethora of structural studies as the most abundant fold assumed by a variety of CBMs from diverse organisms (Abbott *et al*., 2013). Based on several functional studies on ligand specificity of CBMs, a generalisation had emerged that CBMs bind to substrates cleaved by the cognate catalytic module (Correia *et al*., 2010). However, the report of family 2 CBMs (CBM2) (Abbott *et al*., 2013) and the report of four CBM35 appended to divergent catalytic modules present an exception to this generalization (Montanier *et al*., 2013). In the process of enzymatic deconstruction of PCW, CBMs partake diverse roles of increasing the enzyme-substrate proximity (Abbott *et al*., 2013), targeting enzymes to regions of cell wall accessible and susceptible to degradation (Montanier *et al*., 2013). CBMs have also been shown to influence substrate specificity, mechanism of enzyme action, thermo-stability and assist the catalytic module in bacterial cell wall attachment (Abbott *et al*., 2013; Montanier *et al*., 2013).

Clostridium thermocellum is an anaerobic, thermophilic and Gram positive bacterium. It produces an array of PCW polysaccharide degrading enzymes which are assembled into a complex called cellulosome. Cellulosome has been also reported in *Clostridium cellulolyticum*, *Ruminococcus flavefaciens* and some other anaerobic microorganisms (Fontes *et al*., 2013). A rhamnogalacturonan lyase (RG lyase), *Ct*RGL associated with a family 35 CBM from *Clostridium thermocellum* was characterized (Dhillon *et al*., 2016). In this study CBM, the *Rgl*-CBM35 associated with *Ct*RGL was cloned, expressed and analysed for binding against various ligands.

5.2 Materials and Methods

5.2.1 Substrates and reagents

β-D-Glucuronic acid (β-D-Glc*p*A) and polygalacturonic acid were purchased from Hi-Media Laboratories, India. Avicel, barley β-glucan, carob galactomannan, ivory nut mannan, potato galactan, soybean rhamnogalacturonan I, rye arabino xylan, sugarbeet arabinan and wheat arabinoxylan were procured from Megazyme, Ireland. Oat spelt xylan, beechwood xylan, birchwood xylan, carboxymethyl cellulose, laminarin were purchased form Sigma-Aldrich Corporation, USA. Trizma (Tris base) and glycine were procured from Sigma-Aldrich Pvt. Ltd., USA. Disodium 2-[2-carboxylatomethyl-(carboxymethyl)-amino]-ethyl-(carboxymethyl)-amino] acetate (disodium EDTA) and calcium chloride were procured from Himedia Laboratories Pvt. Ltd., India.

5.2.2 Binding assays with soluble polysaccharides

5.2.2.1 Preparation of native-PAGE with soluble substrates

Binding of *Rgl*-CBM35 to 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 N,N,N′,N′-Tetramethylethylenediamine (TEMED) as described for SDS-PAGE in Chapter 2, Section 2.2.13. The only difference in native-PAGE was the absence of SDS in resolving gel and in electrophoresis buffer. Therefore, 7.5% resolving (native) gel was prepared following the protocol described in Chapter 2, Section 2.2.13 without SDS. 1.0% (w/v) stock solution of polysaccharides were made

by dissolving in sterile deionized water until a clear solution was obtained. Thereafter, necessary amount of substrate solution (from the stock) was added to the resolving gel prior to the polymerization. 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. Around 15 µg of *Rgl*-CBM35 was loaded in a well. Electrophoresis was carried out at 25°C in a Mini PROTEAN Tetra pack (Bio-Rad, USA) at a constant current of 15 mA/gel. The gels were stained with Coomassie Brilliant Blue R250 (Sigma) for identification of qualitative binding of the CBM.

5.2.2.2 Preparation of native-PAGE running buffer

The native-PAGE buffer was prepared using components as described in Table 5.1. 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 5.1).

Table 5.1 Composition of 5x native-PAGE running buffer.

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

5.2.2.3 Preparation of sample buffer

A 5x sample loading buffer was prepared by dissolving the components maintaining the concentration of components as described in Table 5.2 and pH of the buffer was adjusted to 6.8. The components were dissolved in the order as mentioned in Table 5.2 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 5.2 Composition of 5x sample loading buffer.

Components	Final concentration (5x buffer)
Tris-HCl (pH 6.8)	0.31M
Glycerol	40.0 (%, v/v)
Bromophenol Blue	0.025 (%, w/v)

5.2.3 Binding assays with insoluble polysaccharides

Qualitative binding analysis of *Rgl*-CBM35 against insoluble polysaccharides such as avicel, ivory nut mannan, and wheat arabinoxylan was carried out as explained by Bolam *et al.*, (2004). Around 20 µg of *Rgl*-CBM35 was mixed with 5 mg of insoluble polysaccharide in 200 µl of 50 mM Tris-HCl (pH 8.5) contained in a 1.5 ml micro-centrifuge tube. The tube was incubated at 25°C with gentle orbital shaking. After 2h of incubation, the tubes were centrifuged at 13,000*g* for 20 min at 4°C. The supernatant containing unbound protein was collected and the pellet was washed twice with 200 µl of 50 mM Tris-HCl (pH 8.5). The bound protein was recovered from the polysaccharide pellet by adding 200 µl of 10% (w/v) sodium dodecyl sulphate (SDS) followed by boiling in a water bath for 20 min. Control reactions containing BSA incubated with polysaccharides instead of *Rgl*-CBM35 were set up in order to check for non-specific binding. Another control reaction containing *Rgl*-CBM35 in 200 µl of 50 mM Tris-HCl (pH 8.5) was also set up in order to check for the protein precipitation during the experiment. The polysaccharide bound and un-bound *Rgl*-CBM35 samples were analysed by using 14% (w/v) SDS-PAGE gels.

5.2.4 Isothermal titration calorimetry

The protein, *Rgl*-CBM35 and ligands dissolved in 50 mM Tris-HCl buffer (pH 8.5) were titrated at 25°C using an isothermal titration calorimeter (MicroCal iTC200, GE Healthcare, USA). During titration 200 µl of 0.02-0.2 mM protein present in cell was stirred at 450 rpm and 25-32 successive injections of 1.0-1.5 µl of ligand were made at regular intervals of 120-150 s. Integrated heat effects were analysed by non-linear regression and fitted to a single binding site model using MicroCal Origin software. Association constant (K_a), enthalpy of binding (ΔH) and number of binding sites (*N*)values for each titration were obtained from the fitted data. While rest of the thermodynamic parameters were derived from the equation $-RT\ln K_a = \Delta G = \Delta H - T\Delta S$ (Bolam *et al.*, 2004). Ligands for which concentration of binding sites was not accurately known (RGI oligosaccharides and pectin oligosaccharides), the N value was fitted iteratively to be close to 1 by varying the molar concentration of ligand as described previously (Montanier *et al.*, 2011). Rhamnogalacturonan I (RGI) oligosaccharides were produced by the action of *Ct*RGL*f* on soybean rhamnogalacturonan I (Dhillon *et al.*, 2016) and pectin oligosaccharides by the action of PL1B on polygalacturonic acid (Chakraborty *et al.*, 2015). Both contain unsaturated galacturonic acid residue (Δ4,5-Gal*p*A) at the non-reducng end. In case of polysaccharides, the binding data was fitted by considering *Rgl*-CBM35 present in the sample cell as ligand. The K_a and ΔH values obtained by this approach are accurate but the N value may not be precise (Bolam *et al.*, 2004). To study the role of Ca^{2+} ions in ligand binding, inherent Ca^{2+} ions were removed from *Rgl*-CBM35. To remove any inherently bound Ca^{2+} ions, *Rgl*-CBM35 was treated with 10 mM EDTA in 50 mM Tris-HCl (pH 8.5) for 90 min at 25°C followed by removal of excess EDTA by

repeated washing by 50 mM Tris-HCl buffer (pH 8.5) on protein concentrator (Amicon Ultra-15 Centrifugal Filter Units, Merck Millipore, Germany).

5.2.5 Protein-melting study of *Rgl*-CBM35

The protein melting studies were carried using a UV-visible spectrophotometer (Varian Cary-100) equipped with a peltier system. 30 µg of *Rgl*-CBM35 dissolved in 1 ml 50 mM Tris-HCl buffer, pH 8.5 contained in a 1 ml quartz cuvette was heated from 40°C- 100°C at a rate, 3°C/min and the absorbance at 280 nm (A_{280}) was recorded following the method reported earlier (Dvortsov *et al*., 2009). The melting curve analysis of *Rgl*-CBM35 was carried out in presence of 10 mM $CaCl_2$, 10 mM EDTA (to remove inherent Ca^{2+} ions) and then in presence of both 10 mM $CaCl_2$ followed by 10 mM EDTA.

5.2.6 Amino acid sequence analysis

The amino acid sequence analysis of rhamnogalacturonan lyase (RG lyase), *Ct*RGL*f* reported in chapter 2, revealed the presence of a CBM (109-232 amino acid residues), *Rgl*-CBM35 towards the N-terminal of *Ct*RGL*f* (Fig. 2. 1). *Rgl*-CBM35 has been classified as a family 35 CBM as per the classification proposed by Lombard *et al.,* (2014). *Rgl*-CBM35 amino acid sequence was analysed by BLAST (http://www.rcsb.org/pdb/home/home.do), Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and ESPRIPT 3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) (Gouet *et al*., 1999).

5.2.7 Homology modelling and model validation

The 3D structure of *Rgl*-CBM35 was modelled using Modeller 9.14. Multiple templates from the PDB database with amino acid sequence identical to *Rgl*-CBM35 were searched using 'build_profile.py' script of Modeller. Family 35 CBMs with PDB

id 2VZP (Chi-CBM35 from *Amycolatopsis orientalis*), 2WIW (Rhe-CBM35 from *Clostridium thermocellum*), 2W3J (Pel-CBM35 from an environmental isolate) and 2W46 (Xyl-CBM35 from *Cellvibrio japonicus*) were chosen as templates for building a model. The generated models were ranked according to Discrete Optimized Protein Energy (DOPE) score. The model with lowest score was assigned the highest rank and selected for loop refinement step. Among the resulting models the one with least DOPE score was chosen for validation and further computational studies. The model was validated using freely accessible servers/databases over the World Wide Web such as Verify3D (http://services.mbi.ucla.edu/Verify_3D/) (Liithy *et al*., 1992) and ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) (Perez-Iratxeta *et al*., 2008). Ramachandran plot for the modelled structure was developed using PDBSum database (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html).

5.2.8 Circular Dichroism analysis of *Rgl*-CBM35

Circular Dichroism (CD) spectrum of *Rgl*-CBM35 was recorded in the Far UV range (190-250 nm) on a spectro-polarimeter (J-815, Jasco Corporation, Tokyo) at 25°C. 2.5 µM of *Rgl*-CBM35 in 50 mM Tris-HCl buffer (pH 8.5) was used for the analysis. The CD spectra were corrected for buffer contributions and the secondary structures were predicted using K2D3 server (Kelly *et al*., 2005). The CD spectrum of *Rgl*-CBM35 is expressed in terms of molar residual ellipticity (MRE deg cm^2dmol^{-1}) and plotted as a function of wavelength (Kelly *et al*., 2005).

5.2.9 Small Angle X-ray Scattering analysis of *Rgl*-CBM35

Scattering data were collected on a small-angle X-ray scattering system (SAXSpace, Anton Paar GmbH, Graz, Austria). 40µl each of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 (*Rgl*-CBM35 incubated with 10 mM EDTA for 90 min at 25°C)

and Ca^{2+}-EDTA-treated *Rgl*-CBM35 (*Rgl*-CBM35 incubated with 10 mM EDTA for 90 min at 25°C followed by addition of 10 mM $CaCl_2$ and further incubation for 30 min) was used for collection of SAXS data using a one-dimensional CMOS Mythen detector (Dectris, Baden, Switzerland) and a radiation wavelength of 1.5 Å. 20 mg/ml protein was used in each experiment. The sample-to-detector distance was 317 mm. Two individual 30 min exposures at 10°C were collected using thermostated quartz capillary of 1 mm diameter. These frames were reduced in SAXStreat software to calibrate the position of the primary beam. Buffer solution (50 mM Tris-HCl, pH 8.5) was used to measure background scattering. SAXSquant software was used to subtract the buffer contribution to obtain the scattering intensity, I as a function of momentum-transfer vector q ($q= 4\sin\theta/\lambda$ where λ and θ represent the wavelength of the X-rays and the scattering angle, respectively. Scattering pattern was visually inspected in Primus to confirm the absence of radiation damage and aggregation. Guinier analysis and radius of gyration (R_g) estimations were performed in Primus and confirmed by automatic analysis using AutoRG (Konarev *et al*., 2003). The distance distribution function P(r), maximum particle diameter (D_{max}) and forward scattering I(0) were computed automatically with AutoGNOM (Konarev *et al*., 2003) and compared with those determined in Primusqt (Konarev *et al*., 2003). The molecular weight estimation was done by DatMow program (Petoukhov *et al*., 2007). The *ab initio* method was used to determine the low resolution shapes of the *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 from the scattering curve by using Gasbor program (Svergun *et al*., 2001). The Dummy Residue Model (DRM) reconstruction was performed using 50 independent runs of Gasbor. These solutions were subsequently clustered by Damclust to create the final *ab initio* shape

(Petoukhov *et al*., 2012). The modelled structure of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 were positioned in the DRM using PyMOL.

5.2.10 Site-directed mutagenesis of *Rgl*-CBM35

Site-directed mutagenesis of Asn118 residue of *Rgl*-CBM35 to Ala was carried out by introducing mutation in the reverse primer (Table 5.3) during conventional PCR amplification. pET28a(+) vector containing gene encoding *Rgl*-CBM35 (pET28a(+)-*Rgl*-CBM35) was used as template. The forward and reverse primers were designed to contain *Nhe* I and *XhoI* restriction enzyme sites. Y37A and R69A mutations were introduced into the N118A mutant of *Rgl*-CBM35 by site-directed mutagenesis by following the method described earlier (Ke *et al*., 1997). Y37A and R69A mutations were introduced by PCR in two steps. To introduce Y37A mutation, in the first PCR step a mega-primer was generated using primers listed in Table 5.3. The product of the first PCR step was as forward primer in the second PCR step to get the amplicon (372 bp) containing Y37A/N118A mutations. Similarly, to introduce R69A mutation, in the first PCR step a mega-primer was generated using the primers listed in Table 5.3. In this case, product of the first PCR step was as reverse primer in the second PCR step to get the amplicon (372 bp) containing R69A/N118A mutations. pET28a(+) vector containing gene encoding *Rgl*-CBM35 N118A (pET28a(+)-*Rgl*-CBM35 N118A) mutant was used as template for generation of double mutants. Each, 50 µl reaction contained 1 µM each of forward and reverse primers, 0.25 mM dNTPs and 0.07 U/µl of *Pfu* Ultra DNA polymerase (Stratagene, USA). The conditions for PCR were: initial denaturation at 94°C for 4 min and 94°C for 30 s, 70°C for 45 s, 72°C for 60 s for a total of 30 cycles followed by final extension at 72°C for 10 min

on thermal cycler (Veriti, Applied Biosystems, U.S.A). The PCR condition and reaction set up are given in Table 5.4 and Table 5.5, respectively. The PCR amplified genes encoding *Rgl*-CBM35 N118A, *Rgl*-CBM35 Y37A/N118A and *Rgl*-CBM35 R69A/N118A mutants were cloned into pET28a(+) vector, sequenced and expressed in *E. coli* BL21 under same conditions optimised for *Rgl*-CBM35 as explained in Chapter 2, Section 2.2.9.

Table 5.3 Primers for construction of mutants by site-directed mutagenesis.

PCR Step	Primer	*Rgl*-CBM35 N118A	*Rgl*-CBM35 Y37A/N118A	*Rgl*-CBM35 R69A/N118A
I	Forward	CGGCTAGCACGAGATATCAG GCTGAGGATGCGATGTTGTAC AAGGC	CGGCTAGACCGAGATATCA GGCTGAGGATGCGATGTTGT ACAAGGC	CGGCTAGACCGAGATATCA GGCTGAGGATGCGATGTTGT ACAAGGC
I	Reverse	CCCTCGAGTTATACTTCAAGA TAATCCACAGCCGGTCCGCCG TCC	CATTTACATTCCACTCAATA GCTCCTCCGGGTTCGTTGTC GT	CAACCAGATTGGAATTTACT GCTATTTCCATAGGTCTGTT ATTGTTTGAT
II	Forward	-	Product of I step of PCR	Product of I step of PCR
II	Reverse	-	CCCTCGAGTTATACTTCAAG ATAATCCACAGCCGGTCCGC CGTCC	CCCTCGAGTTATACTTCAAG ATAATCCACAGCCGGTCCGC CGTCC

Table. 5.4 PCR conditions for generation of mutants by site-directed mutagenesis.

Mutant	Template	I PCR annealing temperature	I PCR amplicon	II PCR annealing temperature	II PCR amplicon
N118A	pET28a(+)-*Rgl*-CBM35	70 ºC	372 bp	-	-
Y37A/N118A	pET28a(+)-*Rgl*-CBM35 N118A	70 ºC	111 bp	70 ºC	372 bp
R69A/N118A	pET28a(+)-*Rgl*-CBM35 N118A	70 ºC	207 bp	70 ºC	372 bp

Table. 5.5 PCR reaction setup for generation of mutants by site-directed mutagenesis.

Reaction Component	Final Concentration	Volume Used (µl)
10X Reaction Buffer	1x	5.0
dNTPs (2.5 mM)	0.25 mM	5.0
Template (15 ng/)µl	0.3 ng/µl	1.0
Forward Primer	1 µM	0.5
Reverse Primer	1 µM	0.5
Pfu Ultra DNA polymerase (2.5 U/ µl)	0.07 U/ µl	1.4
Nuclease Free Water	-	36.6
Total Reaction Volume		50 µl

5.3. Results and Discussion

5.3.1 Amino acid sequence analysis of *Rgl*-CBM35

The modular protein encoded by Cthe_0246 gene of *Clostridium thermocellum* consists of an N-terminal Dockerin I followed by a CBM (*Rgl*-CBM35) and catalytic RG Lyase module (*Ct*RGL). The BLAST analysis of *Rgl*-CBM35 amino acid sequence showed that it is homologous to CBM35 associated with a variety of enzymes from *Clostridium thermocellum* (Rhe-CBM35; 69% identity), *Amycolatopsis orientalis* (Chi-CBM35; 44% identity), *Paenibacillus barcenonensis* (Xyn30D-CBM35; 37% identity) and *Cellvibrio japonicus* (Xyl-CBM35, 34% identity) (Montanier *et al*., 2009; Valenzuela *et al*., 2012).

Multiple sequence alignment of *Rgl*-CBM35 amino acid sequence with its closest homologues *viz.* Rhe-CBM, Chi-CBM35, Xyn30D-CBM35, Xyl-CBM35 showed conserved amino acid residues involved in Ca^{2+} ion and ligand binding (Fig. 5.1). All the above mentioned homologues of *Rgl*-CBM35 except Rhe-CBM35 were found to bind two Ca^{2+} ions. One of the Ca^{2+} ions is conserved across several CBM families (Correia *et al*., 2010) and has been shown to impart thermostability (Hachem *et al*., 2000). Based on the information gathered from crystal structures of Chi-CBM35, Xyn30D-CBM35 and Xyl-CBM35, the residues, Gln4, Glu6 and Asp120 of *Rgl*-CBM35 were elucidated to be involved in binding of the structural Ca^{2+} ion. The second Ca^{2+} ion is not structurally conserved and has been credited to play a pivotal role in uronic acid recognition (Montanier *et al*., 2009). The consensus sequence, Asn/Asp-Tyr/Thr-X-Asn is responsible for binding the second, functional Ca^{2+} ion in family 35 CBMs. In *Ct*CBM35-Gal, a member of family CBM35, the Asn/Asp residue is replaced with Gly and the structural consequence of this replacement is that

it disables *Ct*CBM35-Gal to employ Ca^{2+} ion for ligand binding (Correia *et al*., 2010). Structural plasticity of β-sandwich fold allows presence of two potential ligand binding sites on a single CBM molecule.

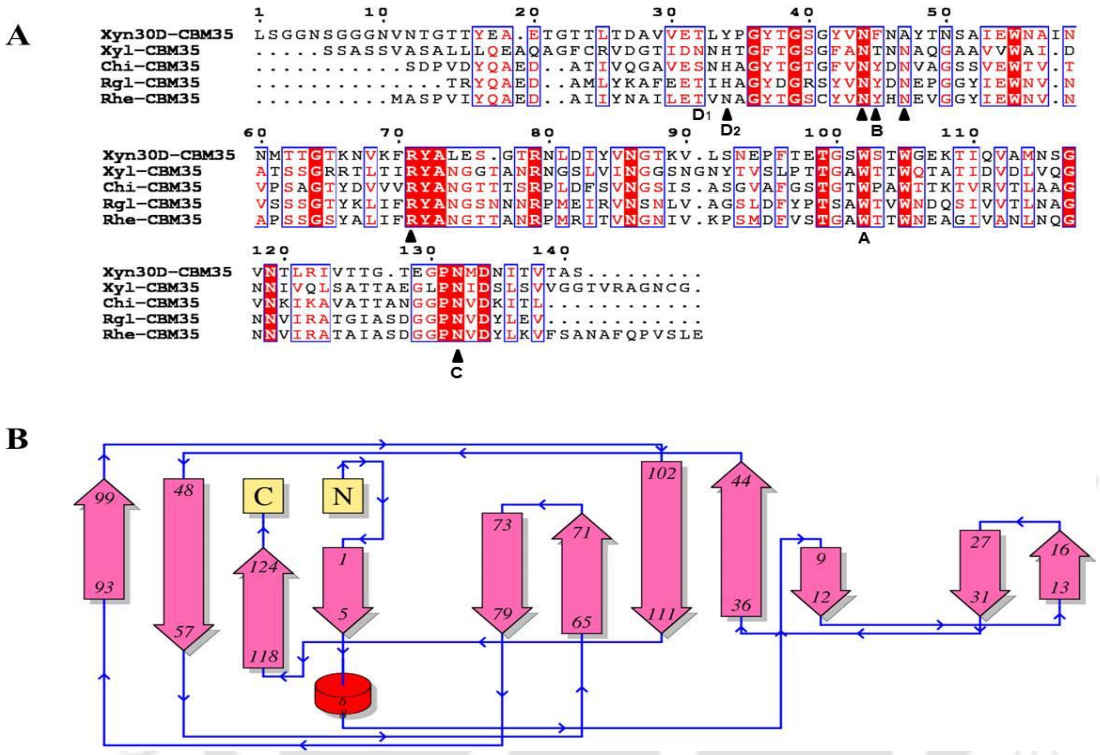


Fig. 5.1 (A) Primary structure alignment of *Rgl*-CBM35 with its homologues *viz.* Xyn30D-CBM35 from *Paenibacillus barcinonensis* (26), Xyl-CBM35 from *Cellvibrio japonicus* (9), Chi-CBM35 from *Amycolatopsis orientalis* (9) and Rhe-CBM35 from *Clostridium thermocellum* (9). Ligand binding site residues involved in binding β-D-Glc*p*A and Δ 4,5-Gal*p*A are marked as ▲ and regions A, B, C, D of VLS have been indicated. (B) Topology diagram of *Rgl*-CBM35 in which β-sheets are shown in pink arrows and α-helix as red cylinder.

However, often only one of the sites is functional (Abbott *et al*., 2014). Although, structural and mutational analyses have generated evidences pointing to presence of two ligand binding sites in some CBMs (Abbott *et al*., 2014). One of the two sites is present on a β-sheet forming a concave surface (concave face site; CFS)

while other is carried by loops interconnecting two β-sheets (variable loop site; VLS) (Correia *et al*., 2010).

Multiple sequence alignment revealed that β-D-Glc*p*A and Δ4,5-Gal*p* binding by *Rgl*-CBM35 may be attributed to conserved His19, Asn29, Tyr30, Asn31, Arg64, Trp87 and Asn118 residues (Fig. 5.1A). In Chi-CBM35 corresponding residues are His22, Asn32, Tyr33, Asn35, Arg67, Trp90 and Asn121. These residues are located in loops connecting β-sheets and constitute the VLS. In Chi-CBM35 Arg67 makes bidentate hydrogen bonds with *C6* carboxylate of β-D-Glc*p*A and Δ4, 5-Gal*p* while Tyr33, Asn32 and Asn35 make Ca^{2+} ion mediated hydrogen bonds with the *C6* carboxylate (Montanier *et al*., 2009). His22 and Asn121 interact with *O2* and *O3* of β-D-Glc*p*A and Δ4, 5-Gal*p*. Trp90 provides a hydrophobic platform that interacts with the sugar ring. Similar roles may be assigned to corresponding residues of *Rgl*-CBM35 which are: His19, Asn29, Tyr30, Asn31, Arg64, Trp87 and Asn118.

The family CBM6 and family CBM35 are recognised to be related based on similarities in amino acid sequence and conserved fold (Correia *et al*., 2010). VLS of family 6 CBMs has been considered to have five different regions A, B, C, D and E (Abbott *et al*., 2014). A close inspection of primary and tertiary structures unravelled the conservation of the regions A and C among the members of family 6 and 35 CBM (Fig. 5.1A). Region A comprises an aromatic residue (Trp87 in *Rgl*-CBM35) located in loops connecting β8 (Asn73-Leu79) and β9 (Gln93-Leu99) strands (Fig. 5.1A&B), which makes interactions with the sugar ring using CH-π bonds as also reported earlier (Abbott *et al*., 2014). The region B, located in β4 (Tyr27 - Tyr30) strand within family 35 CBM is highly variable in contrast to a conserved aromatic residue in Region B of family 6 CBM. The residue constituting region B, Tyr/Thr (Tyr30 in *Rgl*-

CBM35) interacts with a Ca^{2+} ion and is therefore involved in the determining ligand specificity. The region C asparagine (Asn118 in *Rgl*-CBM35) located in β11 (Asn108-Glu123) strand is structurally and functionally conserved. It is present at the base of ligand binding site and forms hydrogen bonds with *O3* and *O4* of sugar ring (Correia *et al*., 2010). Regions D and E are bifurcated in two sub-regions each D1, D2 and E1, E2, respectively (Fig. 5.1A). At the D1 position, the presence of a large polar residue or a small apolar residue governs the binding of CBM to the terminal or internal regions, respectively, of a polysaccharide. At D1 and D2 positions, the uronic acid binding CBM35s have the polar residues, Tyr16 and His19 in *Rgl*-CBM35. The uronic acid binding CBM35s do not display the characteristic aromatic residue of the region E (Valenzuela *et al*., 2012).

5.3.2 Protein melting analysis of *Rgl*-CBM35

The protein melting curve showed that *Rgl*-CBM35 starts unfolding at 70°C giving maximum unfolding at 82°C (Fig. 5.2). In the presence of 10 mM Ca^{2+} ions the unfolding of *Rgl*-CBM35 started at higher temperature, 80°C and the maximum unfolding occurred at 92°C. This showed that the Ca^{2+} ions impart thermostability to *Rgl*-CBM35. In the presence of 10 mM EDTA, *Rgl*-CBM35 started melting at 61°C, 10°C lower than that in its absence. In the presence of 10 mM Ca^{2+} ions the treatment with 10 mM EDTA did not decrease the melting temperature of *Rgl*-CBM35 and protein melting started only after 80°C, displaying the role of inherent Ca^{2+} ions in providing the thermostability.

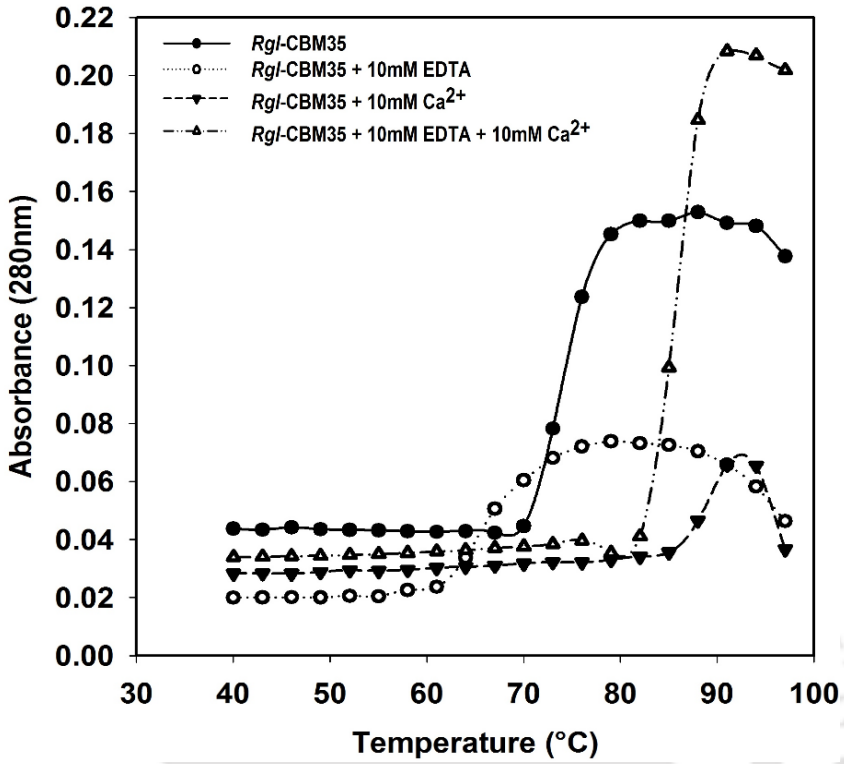


Fig. 5.2 Protein melting curve of *Rgl*-CBM35 with no additive (●),10 mM EDTA (○), 10 mM Ca^{2+} ions (▼) and in presence of both 10 mM EDTA and 10 mM Ca^{2+} ions (Δ).

5.3.3 Binding analysis *of Rgl*-CBM35 against soluble ligands

To comprehend the biological significance of *Rgl*-CBM35, its ability to bind various carbohydrate ligands found in plant cell was assessed by affinity gel electrophoresis (AGE), ITC. *Rgl*-CBM35 bound a wide range of ligands including polysaccharides, oligosaccharides and monosaccharides (Table 5.6). However, binding to barley-β-glucan, carob galactomannan, carboxymethyl cellulose, laminarin and polygalacturonic acid was not detected (Table 5.6). The AGE results displayed that *Rgl*-CBM35 bound soybean rhamnogalacturonan I, sugarbeet arabinan, potato galactan, birchwood xylan, beechwood xylan and rye arabinoxylan (Fig. 5.3). However, binding to soluble wheat arabinoxylan could not be detected by AGE.

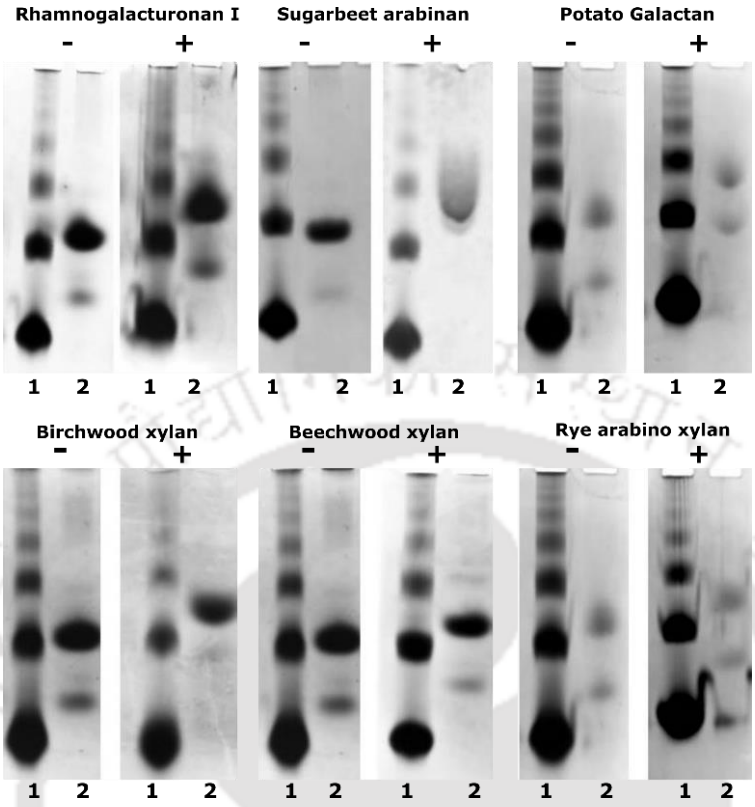


Fig. 5.3 Qualitative Affinity Gel Electrophoresis of *Rgl*-CBM35 using native polyacrylamide gels (7.5%, w/v) in presence (+) and absence (-) of various soluble plant polysaccharides (0.4%, w/v): Soybean rhamnogalacturonan I, Sugarbeet arabinan, Potato Galactan, Birchwood xylan, Bechwood xylan, Rye arabinoxylan. Lane 1- Bovine Serum Albumin, lane 2- *Rgl*-CBM35.

Table 5.6 Binding of *Rgl*-CBM35 with various carbohydrates.

Ligand	Binding	Technique
Soluble		
Soybean rhamnogalacturonan I	binding detected	AGE, ITC
Sugarbeet arabinan	binding detected	AGE, ITC
Potato galactan	binding detected	AGE, ITC
Birchwood xylan	binding detected	AGE
Beechwood xylan	binding detected	AGE
Rye arabinoxylan	binding detected	AGE
Wheat arabinoxylan	binding detected	ITC
Rhamnogalacturonan I oligosaccharides	binding detected	ITC
Pectin oligosaccharides	binding detected	ITC
β-D-Glucuronic acid	binding detected	ITC
Polygalacturonic acid	no binding	AGE
Carob galactomannan	no binding	AGE
Barley β-glucan	no binding	AGE
Carboxymethyl cellulose	no binding	AGE
Laminarin	no binding	AGE
Insoluble		
Wheat arabinoxylan	binding detected	SDS-PAGE
Oat spelt xylan	no binding	SDS-PAGE
Avicel	no binding	SDS-PAGE
Ivory nut mannan	no binding	SDS-PAGE

The quantitative assessment by ITC showed that *Rgl*-CBM35 displayed highest affinity towards RG I oligosaccharides (K_a = 1.2×10^5 M^{-1}) followed by pectic oligosaccharides (K_a = 2.1 ×10^4 M^{-1}) and β-D-Glc*p*A (K_a = 1.0×10^4 M^{-1}). The binding affinity of *Rgl*-CBM35 towards oligosaccharides and monosaccharide was higher than the polysaccharides used (Fig. 5.4, Table 5.7). The binding affinity of *Rgl*-CBM35 for soybean RG I (K_a = 5.4 ±0.6 × 10^3 M^{-1}) was 5 times higher than its affinity for sugarbeet arabinan (K_a = 1.0 ±0.5×10^3 M^{-1}) and 3 times higher than that for wheat arabinoxylan (1.8±0.9×10^3 M^{-1}) (Fig. 5.4, Table 5.7). Similar results with K_a values of the same order were also reported for other CBM35 (Correia *et al*., 2010).

Table 5.7 ITC analysis of *Rgl*-CBM35 binding to the plant carbohydrate ligands.

Protein	Ligand	K_a (M^{-1})	ΔG (kcal/mol)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)	N
Rgl-CBM35	RG I	5.4 ±0.6 ×10^3	-3.71	-3.6 ±0.4	0.11	-
Rgl-CBM35	Sugarbeet arabinan	1.1 ±0.5×10^3	-18.6	-20±30	-1.3	-
Rgl-CBM35	Potato galactan	weak binding	-	-	-	-
Rgl-CBM35	Wheat arabinoxylan	1.8 ±0.9×10^3	-8.3	-8.7±8.5	-0.36	-
Rgl-CBM35	RG I oligosaccharides	1.2 ±0.07 ×10^5	-6.36	-6.4 ±0.05	-0.04	1.09
Rgl-CBM35	Pectic oligosaccharides	2.1 ±0.27 ×10^4	-9.4	-9.66 ±0.4	-0.31	1.03
Rgl-CBM35	β-D-Glc*p*A	1.0 ±0.07 ×10^4	-8.1	-8.3 ±0.3	-0.24	1.05
Rgl-CBM35 + EDTA*****	RG I	1.5 ±0.2 ×10^5	-156	-175±1.2	-14.1	-
Rgl-CBM35 + EDTA*****	Sugarbeet arabinan	1.2 ±0.3 ×10^7	-91.7	-110 ±1.2	-8.4	-
Rgl-CBM35 + EDTA*****	Potato galactan	1.16±0.1×10^6	-88.63	-144±2.8	-11.37	-
Rgl-CBM35 + EDTA*****	Wheat arabinoxylan	8.67±2.4×10^5	-204	-222±8.2	-17.95	-
Rgl-CBM35 + EDTA*****	RGI oligosaccharides	no binding	-	-	-	-
Rgl-CBM35 + EDTA*****	Pectic oligosaccharides	no binding	-	-	-	-
Rgl-CBM35 + EDTA*****	β-D-Glc*p*A	no binding	-	-	-	-
Rgl-CBM35 N118A	RG I	no binding	-	-	-	-
Rgl-CBM35 N118A	Sugarbeet arabinan	no binding	-	-	-	-
Rgl-CBM35 N118A	Potato galactan	no binding	-	-	-	-
Rgl-CBM35 N118A	Wheat arabinoxylan	no binding	-	-	-	
Rgl-CBM35 N118A	RGI oligosaccharides	no binding	-	-	-	-
Rgl-CBM35 N118A	Pectic oligosaccharides	no binding	-	-	-	-
Rgl-CBM35 N118A	β-D-Glc*p*A	no binding	-	-	-	-
Rgl-CBM35 N118A + EDTA*	RG I	2.0±0.3 ×10^5	-64.5	-73±0.8	-5.5	-
Rgl-CBM35 N118A + EDTA*	Sugarbeet arabinan	1.6±0.1 ×10^6	-40.1	-43 ±0.2	-2.9	-
Rgl-CBM35 N118A + EDTA*	Potato galactan	7.5±1.6 ×10^5	-247.3	-279±9	-22.7	-
Rgl-CBM35 N118A + EDTA*	Wheat arabinoxylan	3.68±1.2 ×10^6	-143.95	-156.3±4.4	-12.35	-
Rgl-CBM35 N118A + EDTA*	RG I oligosaccharides	weak binding	-	-	-	-
Rgl-CBM35 Y37A/N118A + EDTA*	RG	2.13±0.18×10^6	-162	-176±1.5	-14	-
Rgl-CBM35 Y37A/N118A + EDTA*	Sugarbeet arabinan	weak binding	-	-	-	-
Rgl-CBM35 Y37A/N118A + EDTA*	Potato galactan	2.45±0.3×10^6	-117.9	-272±3.2	-22.10	-
Rgl-CBM35 Y37A/N118A + EDTA*	Wheat arabinoxylan	no binding	-	-	-	-
Rgl-CBM35 R69A/N118A + EDTA*	RG I	2.29±0.27×10^6	-237.1	-258±2.9	-20.9	-
Rgl-CBM35 R69A/N118A + EDTA*	Sugarbeet arabinan	weak binding	-	-	-	-
Rgl-CBM35 R69A/N118A + EDTA*	Potato galactan	2.44±0.4×10^6	-181.4	-231±3.9	-18.6	-
Rgl-CBM35 R69A/N118A + EDTA*	Wheat arabinoxylan	no binding	-	-	-	-

RG I: rhamnogalacturonan I from soybean.

RGI oligosaccharides and pectin oligosaccharides: unsaturated oligosaccharides containing Δ4,5-GalpA residues at the non-reducing end.

***** *Protein treated with 10 mM EDTA followed by buffer exchange to 50 mM Tris HCl (pH 8.5)*

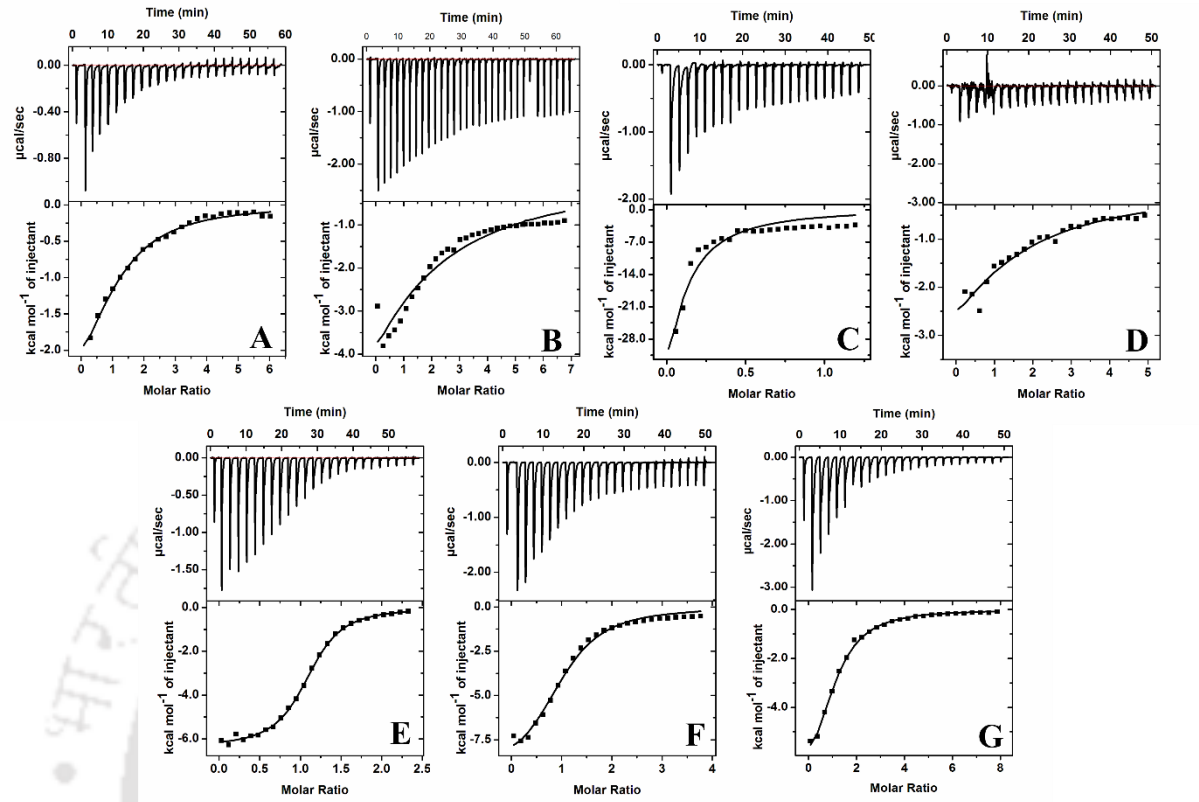


Fig. 5.4 Isothermal Titration Calorimetry of *Rgl*-CBM35 binding with various plant carbohydrate ligands. The upper half of each panel shows raw binding heats while the lower half of each panel shows integrated binding heats. From A-G, titration of *Rgl*-CBM35 with various ligands: (A) soybean rhamnogalacturonan I, (B) sugarbeet arabinan, (C) potato galactan, (D) wheat arabinoxylan, (E) Rhamnogalacturonan I oligosaccharides, (F) pectic oligosaccharides and (G) β-D-Glucuronic acid.

The binding affinity of *Rgl*-CBM35 towards birchwood xylan and beechwood xylan may be attributed to its interaction with β-D-Glc*p*A. Xyn30D-CBM35 was also reported to bind substituted xylans by interacting with β-D-Glc*p*A decorations on the xylan main chain (Valenzuela *et al*., 2012). The rationale behind binding of sugarbeet arabinan by *Rgl*-CBM35 may be to shuttle the RG lyase module towards α-1→5 linked arabinan side chains of RG I, and enabling it to eventually target the RG I main chain (O'Neill *et al*., 1996). Interestingly, *Rgl*-CBM35 bound potato galactan that contains β-1→4 linked β-D-Gal*p* residues but, did not bind carob galactomannan

which has decorations of α-anomer of galactose (α-D-Gal*p*) displaying functional selectivity towards the β-anomer of galactose present in side chains of RG I (Hachem *et al*., 2000). In contrast, *Ct*Gal-CBM35 form *Clostridium thermocellum* selectively bound α-D-Gal*p* decorations of carob galactomannan (Correia *et al*., 2010). *Rgl*-CBM35 also showed affinity for RG I oligosaccharides and pectin oligosaccharides containing Δ4,5-Gal*p*A residue at the non-reducing end. Such oligosaccharides are considered as indicators of PCW degradation (Montanier *et al*., 2009). *Rgl*-CBM35 has higher (20 times) affinity for RG I oligosaccharides than RG I polymer. This may be due to fact that Δ4,5-Gal*p*A residue of RG I oligosaccharides is probably functioning as a ligand for *Rgl*-CBM35 in addition to arabinan and galactan side chains of RG I. Others members of CBM35 family have exhibited similarly strong affinity for Δ4,5-Gal*p*A (Montanier *et al*., 2009). In natural habitat, recognition of such moieties by *Rgl*-CBM35 of *C. thermocellum* probably enables the CBM to target its cognate catalytic domain to regions of plant tissues already undergoing active degradation. Such a mechanism of targeting their cognate enzyme by CBM35s was experimentally tested and was reported earlier (Montanier *et al*., 2009). The results presented above show that *Rgl*-CBM35 has wider ligand specificity as compared to its closest homologues. Previously, the family 35 CBM members, *viz.* Rhe-CBM35, Chi-CBM35 and Xyl-CBM35 were reported to bind β-D-Glc*p*A and Δ4,5-Gal*p*A (Montanier *et al*., 2009). Another CBM35, Xyn30D-CBM35 from *Paenibacillus barcinonensis* was reported to bind glucuronoxylans and arabinoxylans mediated by β-D-Glc*p*A and α-L-Ara*f* (Valenzuela *et al*., 2012). Family 35 CBMs have also been reported to bind α-D-galactopyranose (α-D-Gal*p*) and β-D-mannopyranose (Correia *et*

al., 2010). It is intriguing to note that *Rgl*-CBM35 binds a variety of ligands *viz.* β-D-Glc*p*A, Δ 4, 5-Gal*p*A, α-L-Ara*f* and β-D-Gal*p*.

5.3.4 Binding analysis *of Rgl*-CBM35 against insoluble ligands

Qualitative binding analysis of *Rgl*-CBM35 with insoluble polysaccharides showed binding with insoluble wheat arabinoxylan (Fig. 5.5, i). However, *Rgl*-CBM35 did not bind other insoluble polysaccharides such as oat spelt xylan, avicel and ivory nut mannan (Fig. 5.5, ii-iv). The binding of *Rgl*-CBM35 with insoluble wheat arabinoxylan may be attributed to its affinity for α-L-Ara*f* .

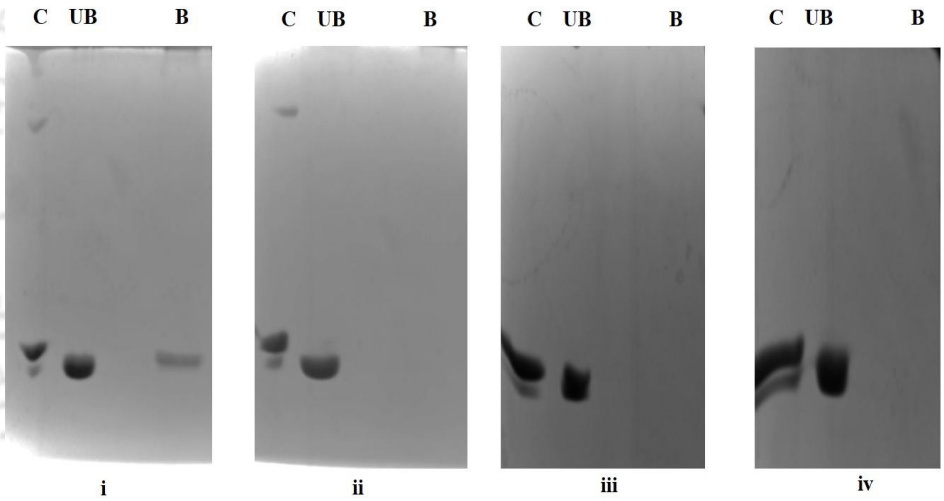


Fig. 5.5 Qualitative binding analysis of *Rgl*-CBM35 against (i) insoluble wheat arabinoxylan, (ii) oat spelt xylan, (iii) avicel and (iv) ivory nut mannan by SDS-PAGE gel (14%, w/v). Lane C -*Rgl*-CBM35, Lane UB -unbound *Rgl*-CBM35, Lane B - bound *Rgl*-CBM35.

5.3.5 Secondary structure analysis of *Rgl*-CBM35

The secondary structure prediction tool PsiPred, predicted that *Rgl*-CBM35 contains 51% β-sheets and rest the loops. Circular Dichroism (CD) analysis of *Rgl*-CBM35 confirmed that the structure is β-sheet rich comprising 44% β-sheets, 1.74% α-helices and rest the loops (Fig. 5.6).

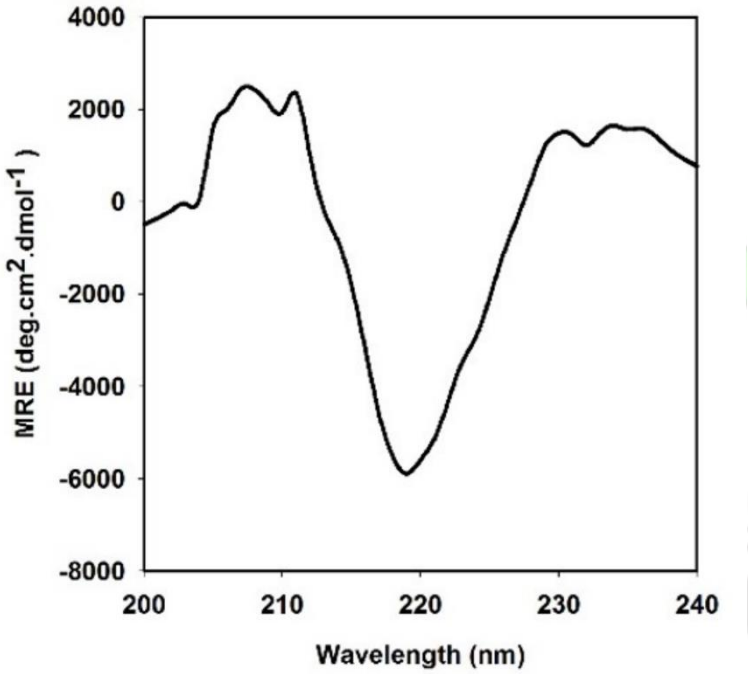


Fig. 5.6 Far UV-CD spectrum of *Rgl*-CBM35 for determining the secondary structure elements.

5.3.6 3D structure modelling and analysis of *Rgl*-CBM35

The 3D structure of *Rgl*-CBM35 was modelled in order to understand the structural features that allow recognition of a wide range of ligands (Fig. 5.7A&B). The modelled structure adopted a β-sandwich fold which is one of the most prevalent folds among CBMs (Abbott *et al*., 2014). Upon inspection of the molecular envelop of *Rgl*-CBM35 (Fig. 5.7C), it was observed that the ligand binding site has a tryptophan residue (Trp87), which can possibly make hydrophobic stacking interactions with sugar ring of ligands. This Trp87 is present in a shallow cove, capable of accommodating only monomeric sugars or a terminal residue of polymeric ligands (Fig. 5.7C). The modelled structure was validated using Verify_3D, Ramachandran plot program on the SAVES server and ProSa web tool. The modelled structure passed the Verify_3D assessment as the result indicated that more than 80% of

modelled residues scored ≥ 0.2 in the 3D/1D profile. Ramachandran plot developed for the modelled *Rgl*-CBM35 showed 87.9% residues to be in most favoured regions (A, B, L), 10.3% residues to be additionally allowed regions (a, b, l, p). ProSA-web results showed the Z-score of -4.65 for the modelled structure was within the range of scores usually found for native proteins of identical size.

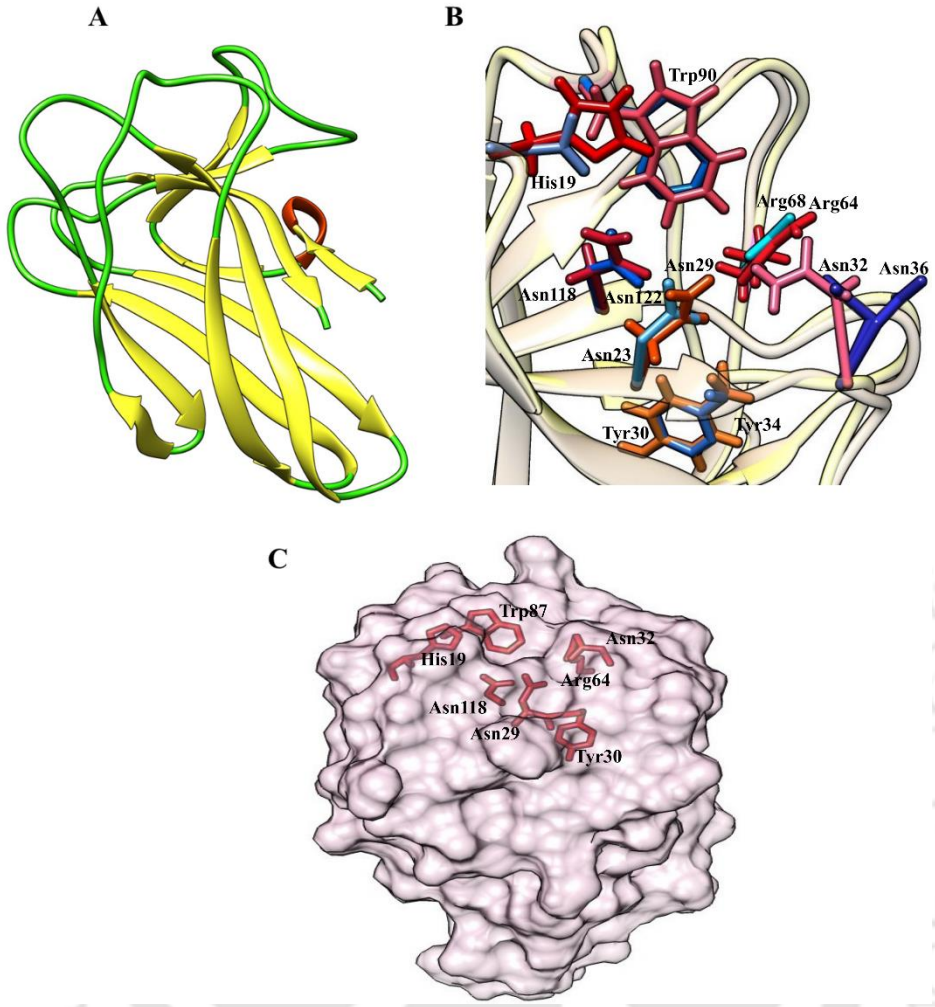


Fig. 5.7 (A) Cartoon representation of modelled *Rgl*-CBM35 structure showing anti-parallel β-sheets. α-helix is shown in red colour, β-strands are in yellow colour and loops are coloured green, (B) Structural superposition of modelled *Rgl*-CBM35 structure onto Rhe-CBM35 crystal structure (2W1W) shows the orientation of conserved ligand binding site (VLS) residues. *Rgl*-CBM35 residues are represented in shades of red colour while, Rhe-CBM35 residues are depicted in shades of blue colour and (C) Representation of the modelled *Rgl*-CBM35 surface showing the binding site residues. The ligand binding site residues are shown as sticks in red.

The modelled *Rgl*-CBM35 structure was composed of three anti-parallel β-sheets (A, B and C) and a short α-helix (Ala5-Asp7) that links the β-sheets A and B (Figs. 5.1 & 5.7A). The β-sheet A comprised β-strands β1 (Arg2-Gln4), β6 (Gly48-Ala57), β9 (Gln93-Leu99) and β11 (Asn108-Glu123) (Fig. 5.1B). The second β-sheet,

B comprised β-strand β2 (Met9-Tyr11), β5 (Tyr37 - Va44), β7 (Met66-Val70), β8 (Asn73-Leu79) and β10 (Glu102 - Gly110). The third β-sheet, C was composed of β3 (Ala13-Glu16) and β4 (Tyr27 - Tyr30). The β-sheets, A and C were arranged facing opposite to β-sheet B. Sheet A and B were elongated and twisted. The structure alignment results obtained using DALI server showed that modelled *Rgl*-CBM35 structure displays rmsd of 0.9Å with Rhe-CBM35 over 124 C_α atoms and rmsd of 0.8Å with Xyl-CBM35 over 124 C_α atoms.

5.3.7 Solution structure determination of *Rgl*-CBM35 by Small Angle X-ray Scattering

The solution conformations of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 were determined by SAXS. The scattering pattern of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 (Fig. 5.8A) showed a good linear fit at a small Q^2 region in the Guinier plot (Fig. 5.8B) and the calculated radius of gyration (R_g) was 2.12 nm, 1.65 nm and 2.29 nm respectively, indicating the predominant globular shape of *Rgl*-CBM35 in the solution. Alternatively, Guinier approximation of rod shaped *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 showed the R_c value of 0.68 nm, 0.60 nm and 0.70 nm, respectively. The linearity of Guinier plot in the low Q-region showed that the scattered intensities follow the Guinier law and suggested that the length of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 to be around 6.95 nm, 5.32 nm and 7.55 nm, respectively.

The molecular mass calculated from the DatMow program for *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 were 27.84 kDa, 14.29 kDa and 29.63 kDa, which clearly indicated that the EDTA-treated *Rgl*-CBM35

was monomeric in nature whereas, *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 were in dimeric form. Pair-wise distance distribution function profile of interatomic vectors [*P*(r)] showed an asymmetric bell shaped curve with maximum diameter of 8.5 nm, 4.0 nm and 9.5 nm for *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35, respectively (Fig. 5.8C). The analysis of Kratky plot of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 indicated that they are fully folded in solution (Fig. 5.8D).

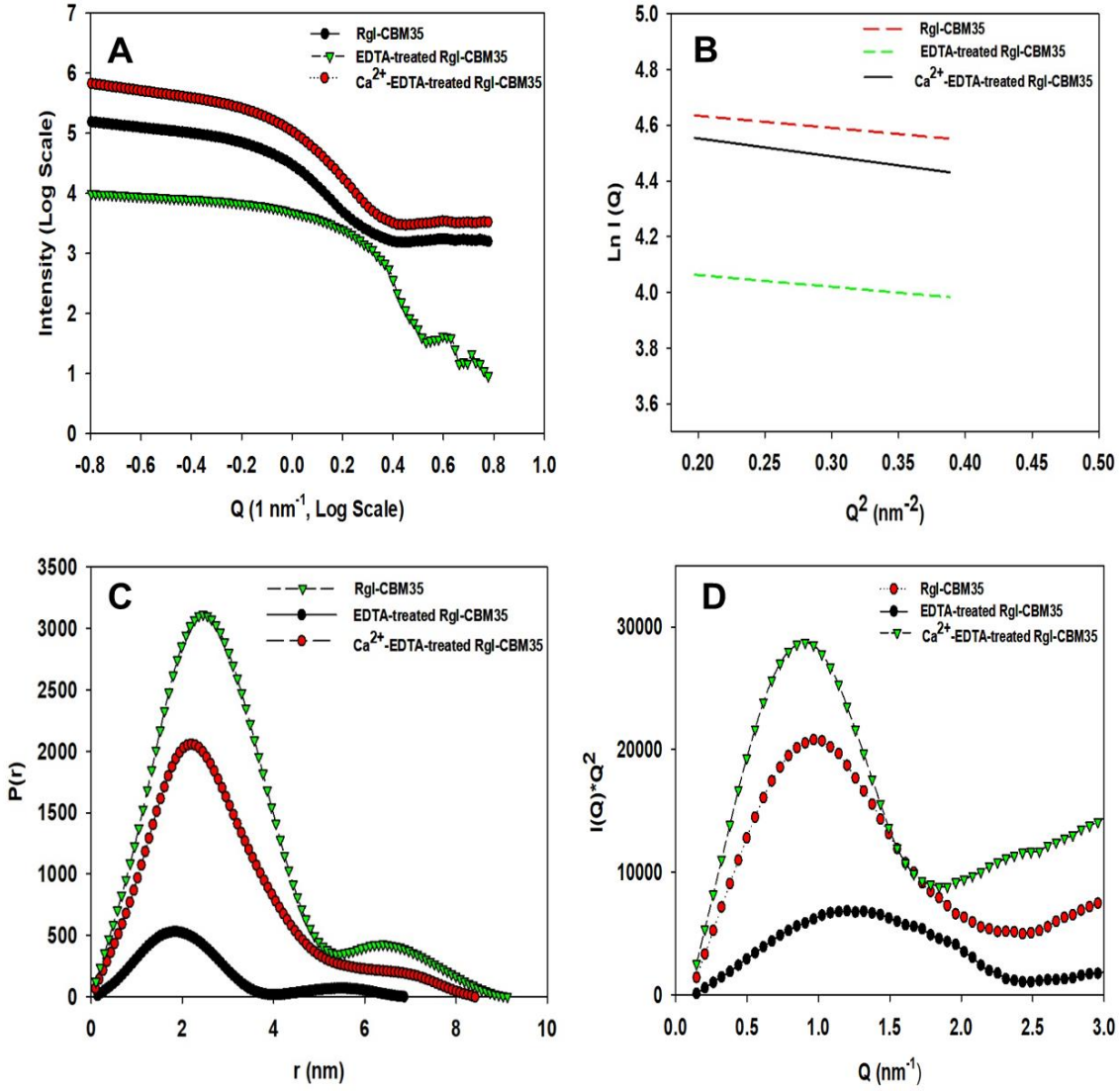


Fig. 5.8 (A) SAXS scattering intensity profile of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35, (B) Guinier plot of the SAXS intensities, (C) P(r) curve of the predominant scattering dataset of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 plotted as a function of r and (D) Kratky plot of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35.

The DRM of *Rgl*-CBM35 showed a bean like shape (Fig. 5.9A), while the DRM of EDTA-treated *Rgl*-CBM35 was compact and clearly revealed a single domain (Fig. 5.9C). However, Ca^{2+}-EDTA-treated *Rgl*-CBM35 showed a kidney bean like shape (Fig. 5.3.8 E). The *ab initio* derived DRM of *Rgl*-CBM35, EDTA-treated *Rgl*-CBM35 and Ca^{2+}-EDTA-treated *Rgl*-CBM35 superposed well with the modelled structure (Fig. 5.9B, D & F).

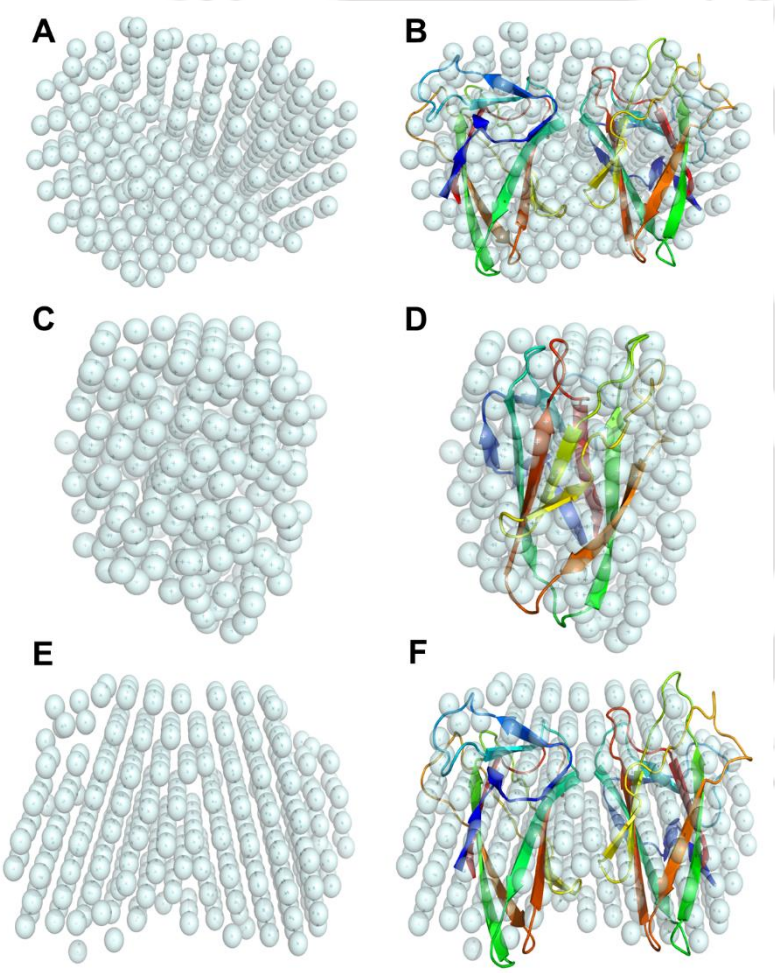


Fig. 5.9 Averaged molecular envelope of *Rgl*-CBM35 (A), EDTA-treated *Rgl*-CBM35 (C) and Ca^{2+}-EDTA-treated *Rgl*-CBM35 (E). Modelled structure of *Rgl*-CBM35 (B), EDTA-treated *Rgl*-CBM35 (D) and Ca^{2+}-EDTA-treated *Rgl*-CBM35 (F) superposed on SAXS data generated *ab initio* structure.

5.3.8 Construction of *Rgl*-CBM35 mutants by Site-directed mutagenesis

One single mutant, N118A (Fig. 5.10A, B & C) and two double mutants, Y37A/N118A and R69A/N118A (Fig. 5.10D, E & F) of *Rgl*-CBM35 were generated by site-directed mutagenesis. The gene encoding these mutants was cloned into pET28a(+) expression vector and the positive clones were screened by restriction digestion (Fig. 5.10). Sequencing of the positive clones done by Scigenom Labs Pvt. Ltd, India, confirmed the presence of mutation I the cloned genes (Fig. 5.11, 5.12 & 5.13). The mutant proteins were purified by immobilised metal ion chromatography (IMAC). The purified *Rgl*-CBM35 mutants were analysed by SDS-PAGE (14%, w/v) gel, which showed that the mutants had a molecular size of approximately, 16 kDa (Fig. 5.14).

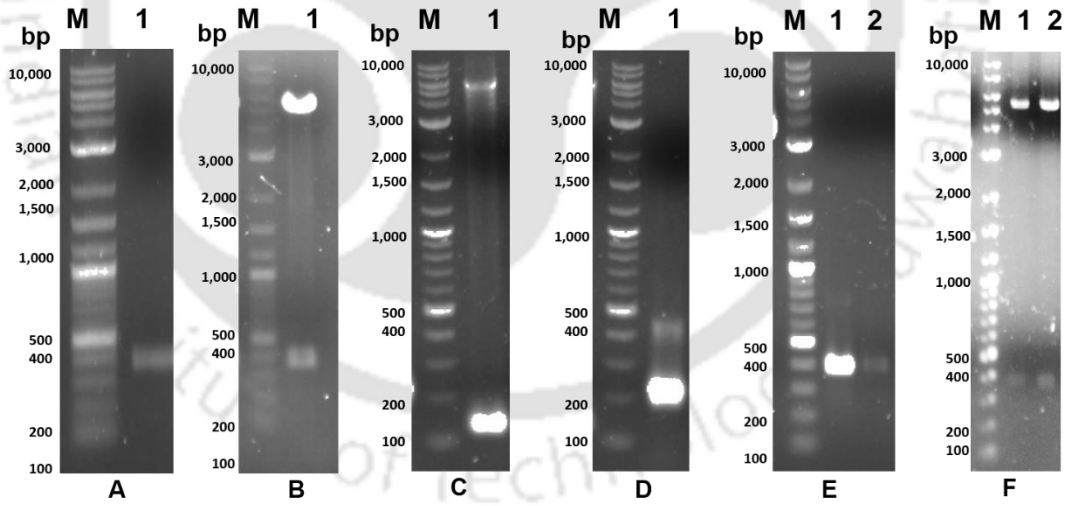
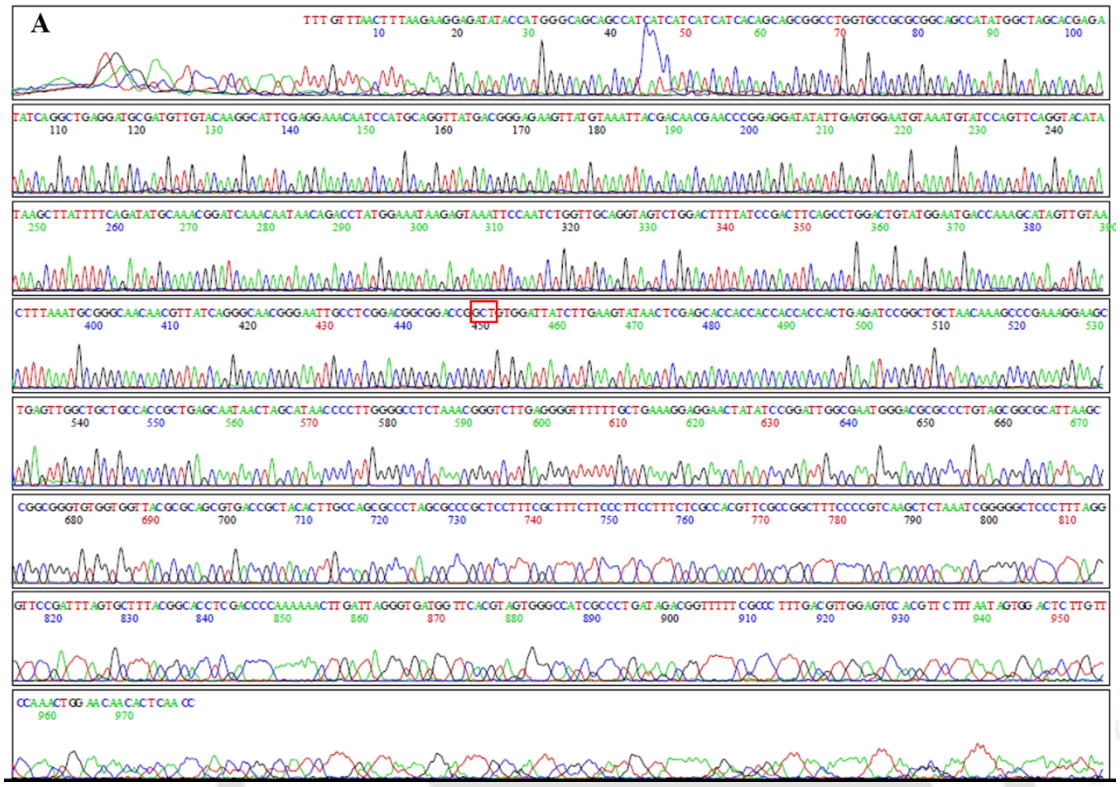


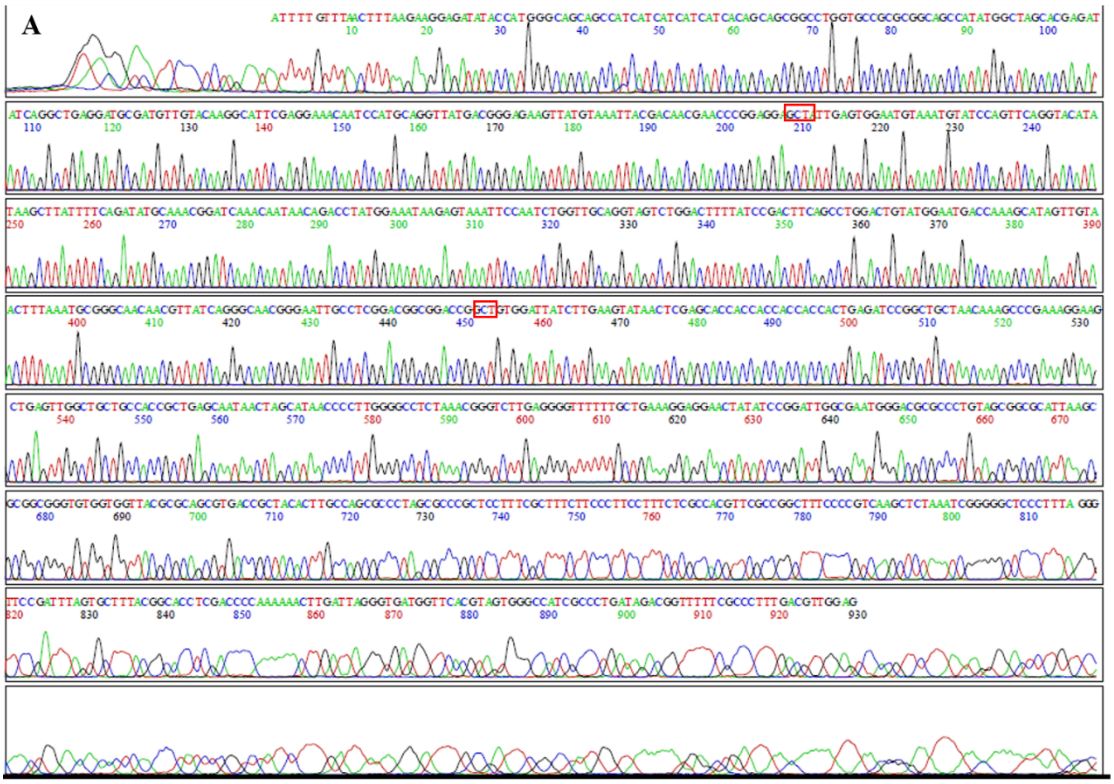
Fig. 5.10 Agarose gels (0.8%, w/v) showing: (**A**) PCR amplified gene (372 bp) encoding *Rgl*-CBM35 mutant, N118A (lane 1), (**B**) *Nhe*I-*Xho*I digested recombinant pET-28a plasmid containing genes encoding *Rgl*-CBM35 mutants, N118A (lane 1), **(C)** I PCR step amplicon (111bp) for Y37A/N118A mutagenesis **(**lane 1), **(D)** I PCR step amplicon (207 bp) for R69A/N118A mutagenesis **(**lane 1), **(E)** PCR amplified genes (372 bp) encoding *Rgl*-CBM35 mutants, Y37A/N118A (lane 1), R69A/N118A (lane 2) and **(F)** *Nhe*I-*Xho*I digested recombinant pET-28a plasmid containing genes encoding *Rgl*-CBM35 mutants, Y37A/N118A (lane 1), R69A/N118A (lane 2). Lane M- DNA marker (2 Log ladder, NEB).



B

TTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATCACA
GCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCACGAGATATCAGGCTGAGGATG
CGATGTTGTACAAGGCATTCGAGGAAACAATCCATGCAGGTTATGACGGGAGAAGTTATG
TAAATTACGACAACGAACCCGGAGGATATATTGAGTGGAATGTAAATGTATCCAGTTCAG
GTACATATAAGCTTATTTTCAGATATGCAAACGGATCAAACAATAACAGACCTATGGAAA
TAAGAGTAAATTCCAATCTGGTTGCAGGTAGTCTGGACTTTTATCCGACTTCAGCCTGGA
CTGTATGGAATGACCAAAGCATAGTTGTAACTTTAAATGCGGGCAACAACGTTATCAGGG
CAACGGGAATTGCCTCGGACGGCGGACCGGCTGTGGATTATCTTGAAGTATAACTCGAGC
ACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGG
CTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGA
GGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAG
CGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
CGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTT
TCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCA
CCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA
GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCA
AACTGGAACAACACTCAACC

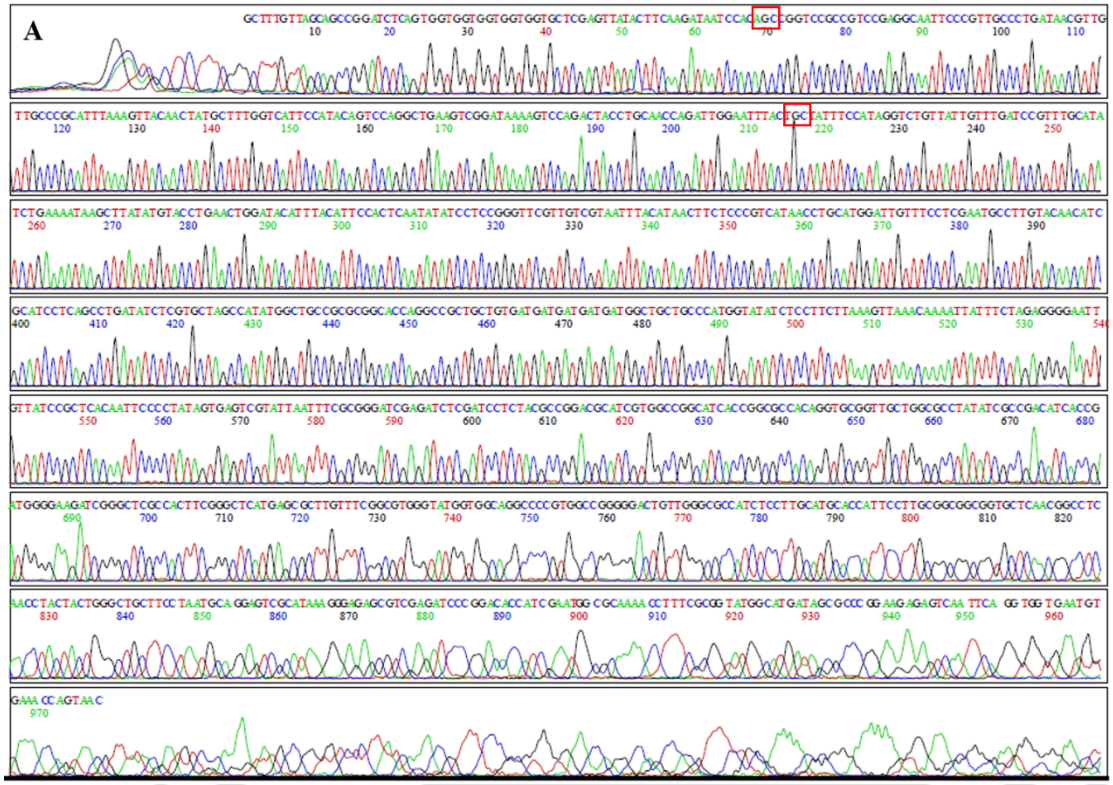
Fig. 5.11 DNA Sequencing of *Rgl*-CBM35 N118A mutant (A**)** Electropherogram showing the DNA sequencing result for cloned gene encoding *Rgl*-CBM35 N118A. Sequencing was done using T7 forward primer. (B) the deduced sequence. The mutated residues have been shown in red.



B

ATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATCA
CAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCACGAGATATCAGGCTGAGGA
TGCGATGTTGTACAAGGCATTCGAGGAAACAATCCATGCAGGTTATGACGGGAGAAGTTA
TGTAAATTACGACAACGAACCCGGAGGAGCTATTGAGTGGAATGTAAATGTATCCAGTTC
AGGTACATATAAGCTTATTTTCAGATATGCAAACGGATCAAACAATAACAGACCTATGGA
AATAAGAGTAAATTCCAATCTGGTTGCAGGTAGTCTGGACTTTTATCCGACTTCAGCCTG
GACTGTATGGAATGACCAAAGCATAGTTGTAACTTTAAATGCGGGCAACAACGTTATCAG
GGCAACGGGAATTGCCTCGGACGGCGGACCGGCTGTGGATTATCTTGAAGTATAACTCGA
GCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTT
GGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTT
GAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGT
AGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCC
AGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC
TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGG
CACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGA
TAGACGGTTTTTCGCCCTTTGACGTTGGAG

Fig. 5.12 DNA Sequencing of *Rgl*-CBM35 Y37A/N118A mutant (A**)** Electropherogram showing the DNA sequencing result for cloned gene encoding *Rgl*-CBM35 Y37A/N118A. Sequencing was done using T7 forward primer. (B) the deduced sequence. The mutated residues have been shown in red.



B

GCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTATACTTCAAGA
TAATCCACAGCCGGTCCGCCGTCCGAGGCAATTCCCGTTGCCCTGATAACGTTGTTGCCC
GCATTTAAAGTTACAACTATGCTTTGGTCATTCCATACAGTCCAGGCTGAAGTCGGATAA
AAGTCCAGACTACCTGCAACCAGATTGGAATTTACTGCTATTTCCATAGGTCTGTTATTG
TTTGATCCGTTTGCATATCTGAAAATAAGCTTATATGTACCTGAACTGGATACATTTACA
TTCCACTCAATATATCCTCCGGGTTCGTTGTCGTAATTTACATAACTTCTCCCGTCATAA
CCTGCATGGATTGTTTCCTCGAATGCCTTGTACAACATCGCATCCTCAGCCTGATATCTC
GTGCTAGCCATATGGCTGCCGCGCGGCACCAGGCCGCTGCTGTGATGATGATGATGATGG
CTGCTGCCCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATT
GTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCG
ATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGC
GCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGC
GCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATC
TCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGC
TGCTTCCTAATGCAGGAGTCGCATAAAGGGAGAGCGTCGAGATCCCGGACACCATCGAAT
GGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGTGGT
GAATGTGAAACCAGTAAC

Fig. 5.13 DNA Sequencing of *Rgl*-CBM35 R69A/N118A mutant (A**)** Electropherogram showing the DNA sequencing result for cloned gene encoding *Rgl*-CBM35 R69A/N118A. Sequencing was done using T7 reverse primer. (B) the deduced sequence. The mutated residues have been shown in red.

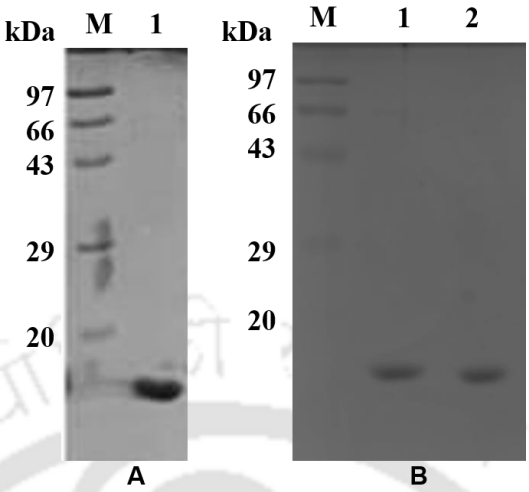


Fig. 5.14 SDS-PAGE (14% w/v) gel showing (A) purified *Rgl*-CBM35 N118A mutant, (B) purified *Rgl*-CBM35 Y37A/N118A mutant (lane 1), purified *Rgl*-CBM35 R69A/N118A mutant (lane 2) Lane M: Protein marker (Bangalore GeNei, India).

5.3.9 Identification of α-L-Ara*f* and β-D-Gal*p* binding sites

The multiple sequence analysis established that *Rgl*-CBM35 contains a conserved site for binding β-D-Glc*p*A and Δ4,5-GalA*p*. *Rgl*-CBM35 also bound sugarbeet arabinan and potato galactan which are the polymers of *α*-L-Ara*f* and β-D-Gal*p* residues, respectively (Fig. 5.3 & 5.4 and Table 5.7). However, a specific *α*-L-Ara*f* biding site has not been elucidated in any of the homologue of *Rgl*-CBM35. Moreover, no other *Rgl*-CBM35 homologues possess the β anomer of Gal*p* binding site except a distantly related CBM, *Ct*CBM35-Gal, which binds α anomer of Gal*p* (Correia *et al*., 2010).

The tertiary structures of previously reported CBMs were explored in order to identify the *α*-L-Ara*f* and β-D-Gal*p* binding site(s) on *Rgl*-CBM35. The recognition of axial *O4* of both L-arabinose and D-galactose by *Ct*CBM35-Gal (Correia *et al*., 2010) and *Ct*CBM62 (Montanier *et al*., 2011) was critical. The modelled *Rgl*-CBM35 structure also presented a shallow binding pocket similar to *Ct*CBM35-Gal and

*Ct*CBM62. As discussed earlier in the section 5.3.1 (Fig. 5.1A) that the asparagine residue, present in the region C, in CBM6 and CBM35, is involved in forging hydrogen bond with *O4* of sugar ring. Therefore, the corresponding Asn118 of *Rgl*-CBM35 was selected as a plausible candidate for site-directed mutagenesis and was replaced by an alanine residue. This residue is present in the conserved ligand binding site of *Rgl*-CBM35 that binds β-D-Glc*p*A and Δ4,5-Gal*p*A (Figs. 5.1A & 5.7B). Affinity gel electrophoresis and ITC analyses revealed that N118A mutant does not bind sugarbeet arabinan and potato galactan (Fig. 5.15A&B, Table 5.7). The ITC analysis showed that N118A also does not bind soybean rhamnogalacturonan I, wheat arabinoxylan, RG I oligosaccharides, pectin oligosaccharides and β-D-Glc*p*A (Fig. 5.15B, Table 5.7) confirming that Asn118 is involved in the ligand binding.

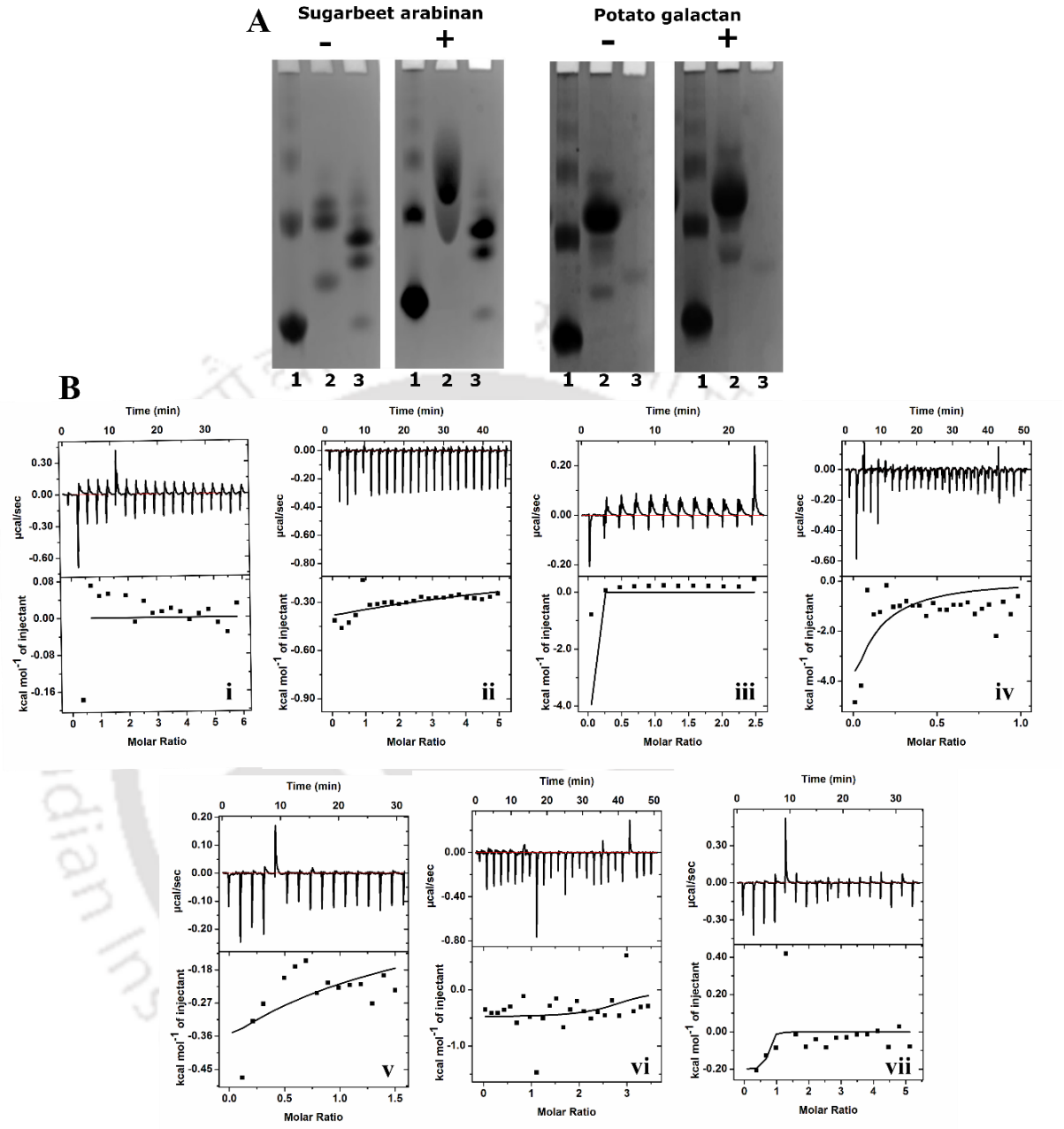


Fig. 5.15 (A) Qualitative binding analysis by Affinity Gel Electrophoresis of wild-type *Rgl*-CBM35 and N118A mutant using native polyacrylamide gels (7.5%, w/v) in presence (+) and absence (-) of sugarbeet arabinan (0.4%, w/v) and potato galactan (0.4%, w/v). Lane 1- Bovine Serum Albumin, lane 2- wild type *Rgl*-CBM35, Lane 3- N118A. (B) Isothermal Titration Calorimetry analysis of N118A mutant binding interaction with: (i) soybean rhamnogalacturonan I, (ii) sugarbeet arabinan, (iii) potato galactan, (iv) wheat arabinoxylan, (v) RG I oligosaccharides, (vi) pectin oligosaccharides and (vii) β-D-Glc*p*A. The upper half of both panels shows raw binding heats while the lower half of each panel shows integrated binding heats.

5.3.10 Role of Calcium ions in ligand binding by *Rgl*-CBM35

The Ca^{2+} ion is strictly required by some CBMs for the ligand binding, while, some CBMs do not require Ca^{2+} ion for ligand binding (Montanier *et al.*, 2011). The multiple sequence alignment showed conserved Ca^{2+} ions binding residues therefore, *Rgl*-CBM35 was expected to bind Ca^{2+} ions. The binding of *Rgl*-CBM35 after treatment with 10 mM EDTA to unsaturated RG I oligosaccharides (containing Δ4,5-Gal*p*A), pectic oligosaccharides (containing Δ4,5-Gal*p*A) and β-D-Glc*p*A was obliterated because, EDTA removed the inherent Ca^{2+} ions (Fig. 5.16A, Table 5.7). This indicated that Ca^{2+} ions are indispensable for the recognition of these ligands. The Ca^{2+} ion has been reported to make critical polar contacts with *C6* carboxylate of β-D-Glc*p*A and Δ4,5-Gal*p*A in CBM35 (Montanier *et al.*, 2011). On contrary, the EDTA-treated *Rgl*-CBM35 exhibited significant binding against soybean rhamnogalacturonan I, sugarbeet arabinan, potato galactan and wheat arabinoxylan as demonstrated by ITC (Fig. 5.16A, Table 5.7). Moreover, the migration of EDTA-treated *Rgl*-CBM35 was also retarded on the native-PAGE gel containing sugarbeet arabinan or potato galactan, showing, that its binding affinity was not disrupted (Fig. 5.16B). On Native-PAGE gel the EDTA-treated *Rgl*-CBM35 displayed an additional band with longer migration than the EDTA-untreated *Rgl*-CBM35. This may be due the separation of some dimeric molecules into monomers. EDTA-treated *Rgl*-CBM35 displayed stronger binding affinity (20-10000 times) with polysaccharides as compared to *Rgl*-CBM35 (Table 5.7). This may be due to the reason that EDTA treatment of *Rgl*-CBM35 exposes a second site at the concave surface that allows the binding of more residues of polymeric chain leading to higher enthalpy change and high K_a as discussed in the next paragraph.

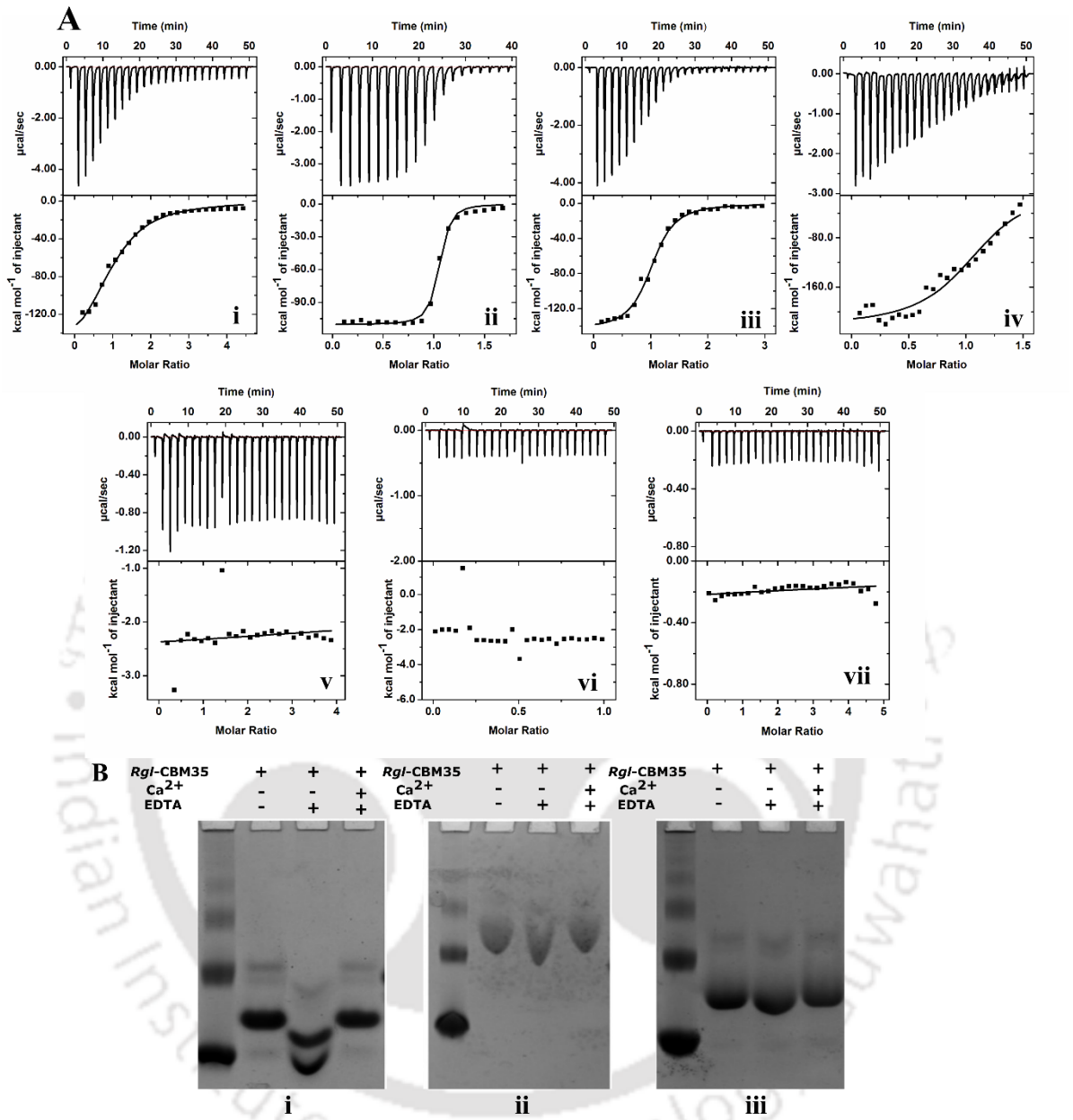


Fig. 5.16 Binding analysis of EDTA treated *Rgl*-CBM35 N118A against soluble polysaccharides. (A) Isothermal Titration Calorimetry analysis of (i) *Rgl*-CBM35-EDTA vs CBM35 vs soybean rhamnogalacturonan I, (ii) *Rgl*-CBM35-EDTA vs sugarbeet arabinan, (iii) *Rgl*-CBM35-EDTA vs potato galactan, (iv) *Rgl*-CBM35-EDTA vs wheat arabinoxylan, (v) *Rgl*-CBM35-EDTA vs rhamnogalacturonan I oligosaccharides, (vi) *Rgl*-CBM35-EDTA vs pectin oligosaccharides and (vii) *Rgl*-CBM35-EDTA vs β-D-Glc*p*A. (B) Affinity Gel Electrophoresis of *Rgl*-CBM35 using native polyacrylamide gels (7.5%, w/v) in absence (i) of sugarbeet arabinan and potato galactan and presence of (0.4%, w/v) sugarbeet arabinan (ii) and potato galactan (iii). BSA- Bovine Serum Albumin, Lane 1- *Rgl*-CBM35, Lane 2- *Rgl*-CBM35 incubated with 10 mM EDTA for 90 min at 25°C, Lane 3- *Rgl*-CBM35 incubated with 10 mM EDTA for 90 min at 25°C followed by addition of 10 mM $CaCl_2$ and further incubation for 30 min.

Surprisingly, the EDTA-treated mutant *Rgl*-CBM35-N118A also displayed appreciable binding affinity against soybean rhamnogalacturonan I, sugarbeet arabinan, potato galactan, wheat arabinoxylan and weak affinity for RG I oligosaccharides as observed by ITC analysis (Fig. 5.17A, Table 5.7). AGE also demonstrated that the EDTA-treated-N118A binds sugarbeet arabinan and potato galactan (Fig. 5.17B (i-iii)). The EDTA-treated *Rgl*-CBM35-N118A also displayed an additional band with longer migration than the EDTA-untreated *Rgl*-CBM35-N118A similar to the EDTA-treated *Rgl*-CBM35. The migration of this additional band was retarded in presence of ligands (Fig. 5.17B).

SAXS analysis suggested that *Rgl*-CBM35 exists as a dimer but after EDTA treatment, it exists in monomeric form as discussed in section 5.3.7 (Fig. 5.9B&D). The effect of EDTA treatment on *Rgl*-CBM35 was reversed upon addition of Ca^{2+} ions (Fig. 5.9F). The dimerization of *Rgl*-CBM35 may not be attributed to disulphide linkage as the protein lacks cysteine residues. Metal ion (Ca^{2+}) mediated dimerisation has been reported in case of a CBM, *Ct*CBM62 from *C. thermocellum* (Montanier *et al.*, 2011). Several functional groups are present on the protein surface that may also dictate the oligomerization by metal ion coordination (Salgado *et al.*, 2009). Therefore, from the SAXS results it can be inferred that the Ca^{2+} ions are involved in the mediation of *Rgl*-CBM35 dimerization.

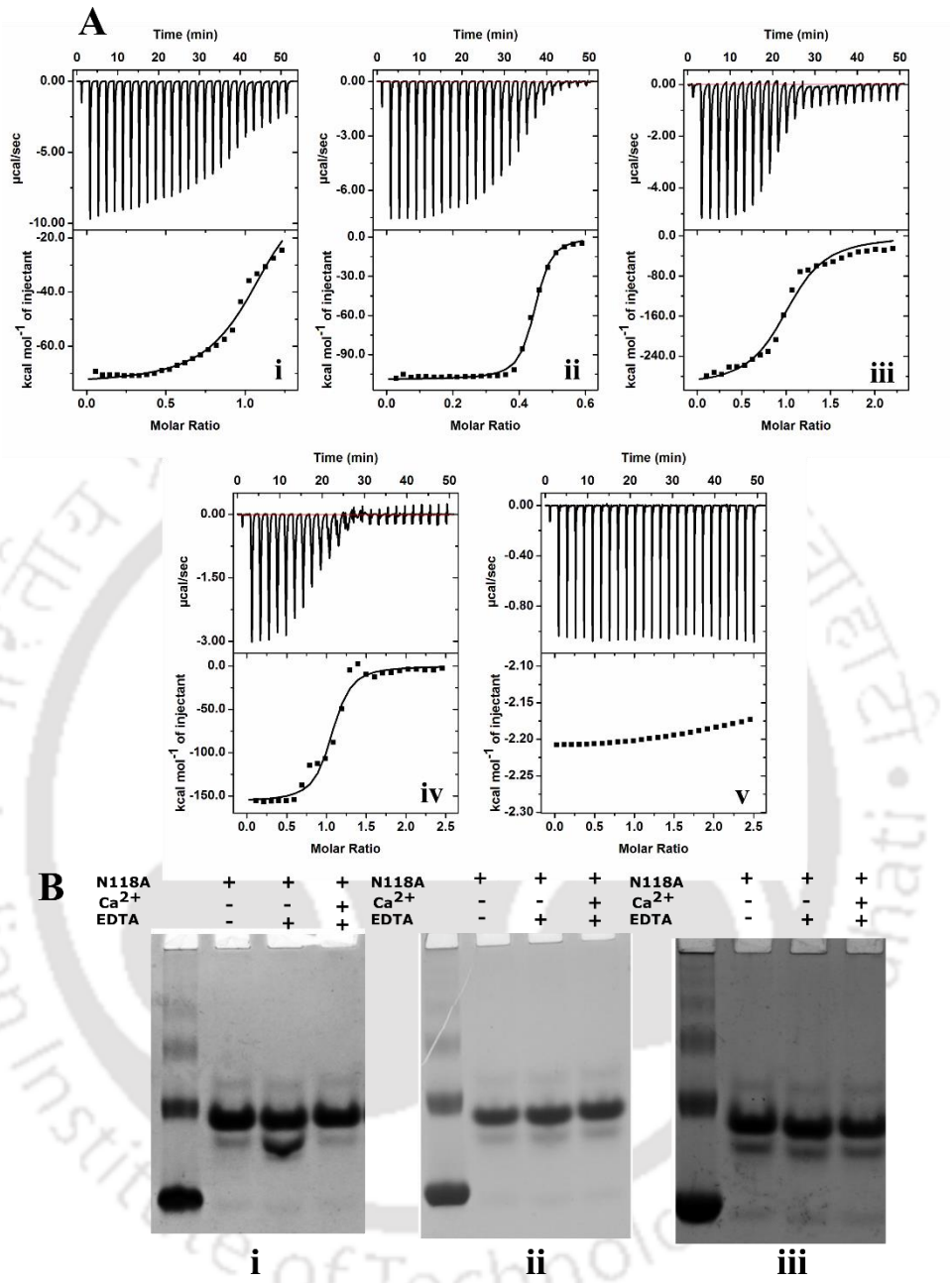


Fig. 5.17 Binding analysis of EDTA treated *Rgl*-CBM35 N118A against soluble polysaccharides. (A) Isothermal Titration Calorimetry analysis of (i) *Rgl*-CBM35 N118A-EDTA vs soybean rhamnogalacturonan I, (ii) *Rgl*-CBM35 N118A-EDTA vs sugarbeet arabinan, (iii) *Rgl*-CBM35 N118A-EDTA vs potato galactan, (iv) *Rgl*-CBM35 N118A-EDTA vs wheat arabinoxylan and (v) *Rgl*-CBM35 N118A-EDTA vs rhamnogalacturonan I oligosaccharides. (B) Affinity Gel Electrophoresis of N118A mutant of *Rgl*-CBM35 using native-PAGE gels (7.5%, v/v) in (i) absence of ligand and in the presence of (0.4%, w/v) (ii) sugarbeet arabinan and (iii) potato galactan. Lanes: BSA- Bovine Serum Albumin, Lane 1- N118A, Lane 2- N118A incubated with 10 mM EDTA for 90 min at 25°C, Lane 3- N118A incubated with 10 mM EDTA for 90 min at 25°C followed by addition of 10 mM $CaCl_2$ and further incubation for 30 min.

The binding of EDTA-treated *Rgl*-CBM35-N118A with arabinan and galactan indicated that, apart from the conserved ligand binding site, *Rgl*-CBM35 is also endowed with another binding site for arabinan and galactan. The lowering of degree of oligomerisation (dimeric to monomeric form) of the N118A mutant after EDTA treatment (on removal of Ca^{2+} ions) probably exposed the second site for arabinan and galactan binding. The surface exposed site probably binds more than one ligand residue of polymeric chain. This leads to high enthalpy change and high binding constants (Table 5.7).

5.3.11 Identification of second ligand binding site

In pursuit of another binding site that could accommodate arabinan and galactan, the surface of *Rgl*-CBM35, especially its concave face was probed. The aromatic amino acid residues are considered the hallmark of the ligand binding sites in CBMs. The modelled structure of *Rgl*-CBM35 was compared with two other CBM the structures *viz.* *Cm*CBM6-2 (*Cellvibrio mixtus*) and *Zg*LamC$_{CBM6}$ (*Zobellia galactanivorans*) (Henshaw *et al.*, 2004; Jam *et al.*, 2016). These two CBMs also possess two ligand binding sites, one at the variable loop site (VLS) and another at the concave face site (CFS). Their CFS has an aromatic residue and a polar amino acid residue placed opposite to each other. Therefore, Tyr37 and Arg69 present in *Rgl*-CBM35 having a similar arrangement on the concave face were selected as possible ligand binding residues (Fig. 5.18). Hence, Tyr37 and Arg69 residues of N118A mutant of *Rgl*-CBM35 were replaced with an alanine residue.

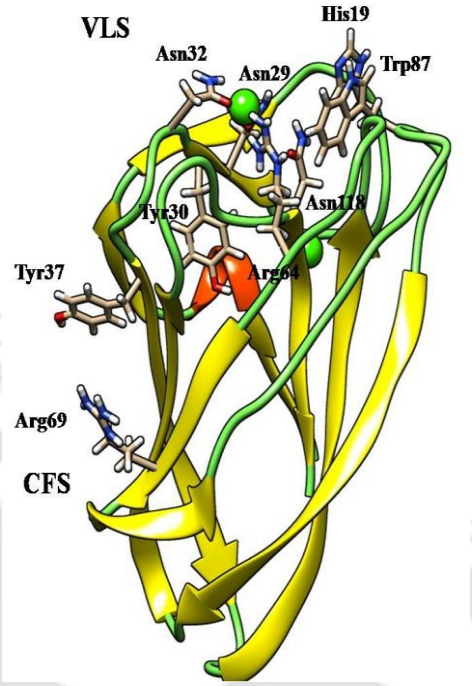


Fig. 5.18 Cartoon representation of modelled *Rgl*-CBM35 structure showing the variable loop site (VLS) and the concave face site (CFS). The ligand binding residues of each site have been labelled. α-helix is shown in red colour, β-strands are in yellow colour and loops are in green colour.

The EDTA-treated double mutants, Y37A/N118A and R69A/N118A of *Rgl*-CBM35 showed dramatically reduced binding affinity towards sugarbeet arabinan (Fig. 5.19). Double mutants, Y37A/N118A and R69A/N118A showed no affinity for wheat arabinoxylan (Fig. 5.19). This established the presence of a second binding site that binds arabinan. There are no previous reports available for family CBM35 having two ligand-binding sites.

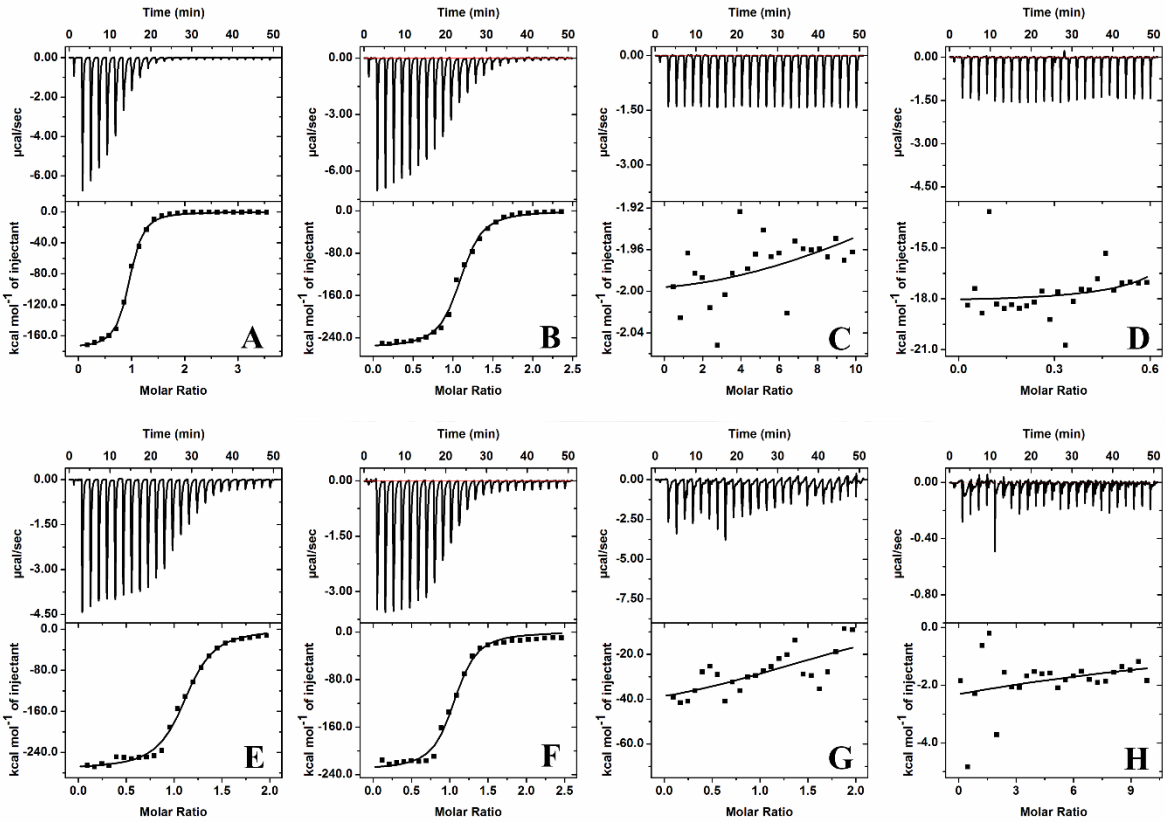


Fig. 5.19 Isothermal Titration Calorimetry analysis of (A) Y37A/N118A-EDTA vs soybean rhamnogalacturonan I, (B) R69A/N118A-EDTA vs soybean rhamnogalacturonan I, (C) Y37A/N118A-EDTA vs sugarbeet arabinan, (D) R69A/N118A-EDTA vs sugarbeet arabinan, (E) Y37A/N118A-EDTA vs potato galactan and (F) R69A/N118A-EDTA vs potato galactan, (G) Y37A/N118A-EDTA vs wheat arabinoxylan and (F) R69A/N118A-EDTA vs wheat arabinoxylan.

The above results connote that Ca^{2+} ions act as a switch that turns off the functioning of the second, surface exposed ligand-binding site of *Rgl*-CBM35. Unpredictably, these two EDTA- treated double mutants continued to bind potato galactan and soybean rhamnogalacturonan I (Fig. 5.19). The two EDTA- treated double mutants showed affinity towards soybean rhamnogalacturonan I, probably due the presence of galactan side chains. Therefore, the potato galactan binding residues remained elusive. The residues interacting with galactan are most likely to be surface exposed, spanning a considerable area over the *Rgl*-CBM35 surface rather than being

clustered together as in a conventional ligand binding site. Such a ligand binding site spanning over an area has been recently discovered in CBMs from *Ruminococcus flavefaciens* (Venditto *et al.,* 2016). The amino acid residues responsible for calcium mediated dimerization could not be determined. An earlier study had also unsuccessfully attempted to elucidate the residues orchestrating calcium mediated dimerization of *Ct*CBM62 (Montanier *et al.,* 2011).

The reports on pectin degrading enzymes associated with a CBM are scarce as compared with cellulases and hemicellulases. *Rgl*-CBM35 is the only second CBM reported to be associated with a rhamnogalacturonan lyase. The only other CBM known to be associated with a rhamnogalacturonan lyase belongs to family CBM3a and that binds cellulose (McKie *et al.,* 2001). It could be thought that pectin is the most accessible of all PCW polymers and its degradation is not much of a challenge. However, some amounts of pectic polymers associated with cellulose and/or hemicellulose are also are present in relatively inaccessible regions of cell wall. It has been demonstrated that the pectin degradation could be facilitated in such regions by the CBMs which target the cellulose or hemicellulose associated with the pectin (Hervé *et al.,* 2010). β-D-Glc*p*A, α-L-Ara*f* and β-D-Gal*p* are often present as decorations on hemi-cellulosic polymers. Therefore, the ability of *Rgl*-CBM35 to target these residues may enable extensive pectin degradation by *C. thermocellum* cells. RG I, albeit an important component of the PCW is present in low amount as compared with other polysaccharides. In the nature, probably the oligomerisaton of *Rgl*-CBM35 brings its associated catalytic module (*Ct*RGL) (Dhillon *et al.,* 2016) into close proximity ensuring efficient degradation and recycling of RG I. To the best of our knowledge only two pectin targeting CBMs are known YeCBM32 (*Yersinia*

enterolitica) and CBM77$_{RfPL1/9}$ (*Ruminococcus flavefaciens*), but, both of these primarily bind homogalacturonan (Venditto *et al.,* 2016). *Rgl*-CBM35 is a singular example of a CBM that has evolved to target RG I component of pectin by the virtue of its multi-ligand specificity.

5.4 Conclusions

Carbohydrate Binding Modules (CBMs) targeting cellulose, xylan and mannan have been reported, however, a CBM targeting rhamnogalacturonan I (RG I) has never been identified. We had studied earlier a rhamnogalacturonan lyase (*Ct*RGL) from *Clostridium thermocellum* that was associated with a family 35 CBM, *Rgl*-CBM35. In this study we show that *Rgl*-CBM35 displays binding with β-D-glucuronic acid (β-D-Glc*p*A), Δ4,5-anhydro-D-galactopyranosyluronic acid (Δ4,5-Gal*p*A), rhamnogalacturonan I, arabinan, galactan, glucuronoxylans and arabinoxylans. *Rgl*-CBM35 contains a conserved ligand binding site in the loops known for binding β-D-Glc*p*A and Δ4,5-Gal*p*A moiety of unsaturated RG I and pectic-oligosaccharides. Mutagenesis revealed that Asn118 plays an important role in binding β-D-Glc*p*A, Δ4,5-Gal*p*A, sugarbeet arabinan and potato galactan at its conserved ligand binding site present in surface exposed loops. EDTA-treated *Rgl*-CBM35 showed no affinity towards β-D-Glc*p*A and Δ4,5-Gal*p*A underscoring Ca^{2+} mediated ligand recognition. Contrastingly, the EDTA-treated *Rgl*-CBM35 and its mutant N118A displayed affinity for sugarbeet arabinan and potato galactan. The curtailed affinity of Y37A/N118A and R69A/N118A double mutants towards sugarbeet arabinan emphasized the presence of a second ligand binding site. *Rgl*-CBM35 is the first CBM reported to primarily target RG I and also is the first member of family 35 CBM possessing at least two ligand binding sites.

References

Abbott, D. W., van Bueren, A. L. (2014) Using structure to inform carbohydrate binding module function. Current Opinion in Structural Biology, 28: 32-40.

Bolam, D. N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B., Gilbert, H. J. (2004) X4 modules represent a new family of carbohydrate-binding modules that display novel properties. Journal of Biological Chemistry, 279(22): 22953-22963.

Chakraborty, S., Fernandes, V. O., Dias, F. M., Prates, J. A., Ferreira, L. M., Fontes, C. M., and Centeno, M. S. (2015) Role of pectinolytic enzymes identified in *Clostridium thermocellum* cellulosome. PloS one. 10(2): e0116787.

Correia, M. A., Abbott, D. W., Gloster, T. M., Fernandes, V. O., Prates, J. A., Montanier, C., Dumon, C., Williamson, M. P., Tunnicliffe, B. B., Ziyuan, L., Flint, J. E., Gideon, D., Henrissat, B., Cutinho, P., Carlos, M. G. A. Fontes, Gilbert, H. (2010) Signature active site architectures illuminate the molecular basis for ligand specificity in family 35 carbohydrate binding module. Biochemistry. 49(29): 6193-6205.

Dhillon, A., Fernandes, V. O., Dias, F. M., Prates, J. A., Ferreira, L. M., Fontes, C. M., Centeno, M. S., Goyal, A. (2016) A New Member of Family 11 Polysaccharide Lyase, Rhamnogalacturonan Lyase (*Ct*RGL*f*) from *Clostridium thermocellum*. Molecular Biotechnology, 58(4): 232-240.

Duan, C. J., Feng, Y. L., Cao, Q. L., Huang, M. Y., Feng, J. X. (2016) Identification of a novel family of carbohydrate-binding modules with broad ligand specificity. Scientific Reports, 6: 19392.

Dvortsov, I. A., Lunina, N. A., Chekanovskaya, L. A., Schwarz, W. H., Zverlov, V. V., Velikodvorskaya, G. A. (2009) Carbohydrate-binding properties of a

separately folding protein module from β-1, 3-glucanase Lic16A of *Clostridium thermocellum*. Microbiology, 155(7): 2442-2449.

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

Gilbert, H. J., Knox, J. P., Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Current Opinion in Structural Biology, 23(5): 669-677.

Gouet, P., Courcelle, E., Stuart, D. I. (1999) ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics, 15(4): 305-308.

Hachem, M. A., Karlsson, E. N., Bartonek-Roxå, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., Williamson, M. P., and Holst, O. (2000) Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. Biochemical Journal, 345(1): 53-60.

Henshaw, J. L., Bolam, D. N., Pires, V. M., Czjzek, M., Henrissat, B., Ferreira, L. M., Fontes, C., Gilbert, H. J. (2004) The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. Journal of Biological Chemistry, 279(20): 21552-21559.

Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., Knox, J. P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. Proceedings of National Academy of Sciences, 107(34): 15293-15298.

Jam, M., Ficko-Blean, E., Labourel, A., Larocque, R., Czjzek, M., Michel, G. (2016) Unraveling the multivalent binding of a marine family 6 carbohydrate-binding module with its native laminarin ligand. FEBS Journal, 283: 1863-1879.

Ke, S. H., Madison, E. L. (1997) Rapid and efficient site-directed mutagenesis by single-tube 'megaprimer' PCR method. Nucleic Acids Research, 25(16): 3371-3372.

Kelly, S. M., Jess, T. J., Price, N. C. (2005) How to study proteins by circular dichroism. Biochimica et Biophysica Acta, 1751(2): 119-139.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography, *36: 1277*–1282.

Liithy, R., Bowie, J. U., Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. Nature, 356(6364): 83-85.

Lombard, V., Golaconda, R. H., Drula, E. (2014) The Carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research, 42: D490-D495.

McKie, V. A., Vincken, J. P., Voragen, A. G. (2001) A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. Biochemical Journal, 355: 67–177.

Miyanaga, A., Koseki, T., Miwa, Y., Mese, Y., Nakamura, S., Kuno, A., Fushinobu, S. (2006) The family 42 carbohydrate-binding module of family 54 α-L-arabinofuranosidase specifically binds the arabinofuranose side chain of hemicellulose. Biochemical Journal, 399(3): 503-511.

Montanier, C., Van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis., R. J., Grondin, G. G., Ghinet, M. G., Gloster, T. M.,

Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., Kerovuo, J., Brzezinski, R., Fontes, C. M., Davies, G. J., Boraston, A. B., Gilbert, H. J. (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proceedings of National Academy of Sciences, 106(9): 3065-3070.

Montanier, C. Y., Correia, M. A., Flint, J. E., Zhu, Y., Baslé, A., McKee, L. S., Prates, J., Polizzi, S., Coutinho, P., Lewis, R., Henrisatt, B., Fontes, C., Gilbert, H. (2011) A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. Journal of Biological Chemistry, 286(25): 22499-22509.

Moreira, S., Castanheira, P., Casal, M., Faro, C., Gama, M. (2010) Expression of the functional carbohydrate-binding module (CBM) of human laforin. Protein Expression and Purification, 74(2): 169-174.

O'Neill, M. A., Warrenfeltz, D., Kates, K. (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. Journal of Biological Chemistry, 271: 22923-22930.

Perez-Iratxeta, C., Andrade-Navarro, M. A. (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. BMC Structural Biology, 8(1): 1.

Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., Svergun, D. I. (2007) ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. Journal of Applied Crystallography, 40: S223–S228.

Petoukhov, M. V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V. Svergun, D.I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography, 45: 342-350.

Ribeiro, T., Santos-Silva, T., Alves, V. D., Dias, F. M., Luís, A. S., Prates, J. A., Ferreira, L., Romão, M., Fontes, C. M. (2010) Family 42 carbohydrate-binding modules display multiple arabinoxylan-binding interfaces presenting different ligand affinities. Biochimica et Biophysica Acta, 1804(10): 2054-2062.

Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L., Tezcan, F. A. (2009) Control of protein oligomerization symmetry by metal coordination: C2 and C3 symmetrical assemblies through Cu^{II} and Ni^{II} coordination. Inorganic Chemistry, 48(7): 2726-2728.

Svergun, D. I., Petoukhov, M. V., Koch, M. H. J. (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophysical Journal. 80: 2946-2953.

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

Valenzuela, S. V., Diaz, P., Pastor, F. J. (2012) Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. Applied and Environmental Microbiology, 78(11): 3923-393.

Venditto, I., Luis, A. S., Rydahl, M., Schückel, J., Fernandes, V. O., Vidal-Melgosa, S., Bule, P., Goyal, A., Pires, V. M. A., Dourdao, C. G., Ferreira, L. M., Coutinho., P., Henrisatt., B., Knox., J. P., Baslé, A., Najmudin, S., Gilbert, H.,

Willats, W. G. T., Fontes, C. M. G. A (2016) Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. Proceedings of National Academy of Sciences, 113(26): 7136-7141.

Wiederstein, M., Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35: W407-W410.

Chapter 6

Application of recombinant rhamnogalacturonan layse, *Ct*RGL*f* from *Clostridium thermocellum* in bio scouring of cotton fabric and degumming of jute fibers

6.1 Introduction

Polysaccharides *viz*., cellulose, hemicellulose and pectin are the major components of plant cell wall (PCW) (Mohnen, 2008). Pectin is predominantly present in primary cell wall and middle lamella and is composed of three interconnected polysaccharides, homogalacturonan (HG), rhamnogalacuronan I (RG I) and rhamnogalacuronan I (RG II) (O'Neill *et al*., 1996). PCW carbohydrates are of great industrial importance, for instance they are the major components of plant material derived fibers such as cotton, jute, flax and ramie (BeMiller, 2007). Plant fibers are used in ropes, paper, textiles and other woven goods. Cotton is mostly used for clothing, tents/tarpaulins etc. After cotton, jute is the next most commercially important plant fiber. Jute is used to produce food grade bags, sacks, carpet backing cloth, handicrafts and blankets. The primary cell wall of cotton fibers contains approximately, 52% cellulose, 12% pectin, 12% proteins and 17% of other

organic/inorganic compounds (Etters, 1999). The wax component of primary cell wall contains fatty alcohols, fatty acids, cholesterin and other hydrocarbons. These waxy compounds associated with pectin impart hydrophobic nature to the cotton fibers (Etters, 1999). Similarly, jute fibers also contain a coating of pectin (0.2% -1.5%) and waxes (0.4% -1%) which provides hydrophobicity to the fibers (Kozłowski, 2012).

To be able to use the cotton fibers in textiles and other goods, cotton fabric is exposed to a series of processing steps, which are performed in aqueous medium (de Melo da Silva *et al*., 2017). Therefore, to increase its hydrophilicity and dyeing efficiency, it is essential to remove pectin and associated wax from cotton fibers (Klug-Santner *et al*., 2006). The process used to remove non-cellulosic components of cotton fibers is termed as Scouring. During conventional scouring, the cotton fabric is treated with sodium hydroxide at 100°C (Sawada *et al*., 1998). This process is effective in improving hydrophilicity of cotton. But, is energy intensive and generates toxic effluents. If the process is not stringently controlled, it damages the cellulose fibrils and due to excess removal of wax, the fabric may develop a rough hand (Etters, 1999). Further, the alkali has to be neutralised by acetic acid leading to formation of sodium acetate, which can discolour the fabric. Therefore, extensive washing is necessary after alkali scouring (Etters, 1999). In order to reduce the consumption of water, cut the energy cost and eliminate toxic effluents, the enzyme mediated scouring called bio-scouring is being explored as an alternative to chemicals (Etters, 1999; Chakraborty *et al*., 2017; Das *et al*., 2012). Several studies have reported the use recombinant polygalacturonanases as well as enzymes from natural isolates in bio-scouring of cotton (Klug-Santner *et al*., 2006; Mohnen, 2008; Morozova *et al*., 2006). In a recent study, a combination of pectin lyase, cellulose and lipase was found to be

equally efficient in scouring cotton compared to conventional scouring (de Melo da Silva *et al*., 2017). However, the use of a rhamnogalacturonan I degrading enzyme has never been reported, earlier.

Jute fibers are processed to remove pectin and associated wax during a pre-processing step called degumming before they are used for spin yarns. Degumming of jute is traditionally achieved by method called retting in which decorticated jute stem is immersed in ponds for several days (Das *et al*., 2012; Duan *et al*., 2016). During retting polygalacturonases, pectate lyases and pectin esterases depolymerise pectin that causes the removal of wax (Das *et al*., 2012). Traditional retting process is slow, labour intensive and causes water pollution (Haque *et al*., 2002). Alternative chemical degumming process of bast fibers are undesirable owing to their adverse effects on the environment (Saikia *et al*., 2009). Enzymatic retting of bast fibers has been suggested to be faster and reproducible (Gillespie *et al*., 1990). Recently, degumming of ramie fibers has been demonstrated by using recombinant enzymes (Zhou *et al*., 2016; Zhou *et al*., 2017). Degumming of ramie fibers has been extensively studied. However, degumming of jute by recombinant enzymes is not known. Recently, in two separate studies, the use of mixed cultures of bacteria belonging to *Bacillus* genus and *Pectobacterium sp.* DCE-01 for degumming of jute was reported (Das *et al*., 2012; Duan *et al*., 2016). In bio-scouring and enzymatic degumming process, pectin degrading enzymes are most desirable. In this study, the use of a rhamogalacturonan lyase, *Ct*RGL*f* from *Clostridium thermocellum* was explored for bioscouring of cotton and enzymatic degumming of jute at bench scale. *Ct*RGL*f* is primarily a rhamogalacturonan lyase with considerable polygalacturonase activity (Dhillon *et al*.,

2016). This report shows for the first time applicability of a rhamogalacturonan lyase in bio-scouring and degumming process.

6.2 Materials and Methods

6.2.1 Production of *Ct*RGL*f*

The recombinant rhamnogalacturonan lyase, (*Ct*RGL*f*) cloned and expressed earlier was used in the present study (Dhillon *et al*., 2016). *E. coli* BL21 (DE3) cells harboring the gene encoding *Ct*RGL*f* were grown in LB medium supplemented with kanamycin (50 µg/ml) and used for production of crude *Ct*RGL*f* as described earlier (Dhillon *et al*., 2016). Briefly, a 500 ml culture of *E. coli* BL21 (DE3) cells was grown at 37°C and 180 rpm till the mid exponential phase (A_{600} = 0.6). The culture was then induced with isopropyl-β-D-thiogalactopyranoside (1 mM) further incubated at 37°C and 180 rpm for 20 h. After incubation, the cells were centrifuged at 9000g, 4°C for 10 min. The cell pellet was washed and subsequently re-suspended in 10 ml of 50 mM Tris-HCl buffer (pH 8.5). The cells were then sonicated for 12 min (5 s on and 15 s off pulse; 33% amplitude) using ultra-sonicator (Vibra Cell, Sonics, U.S.A). The sonicated cells were centrifuged at 20,000g and 4°C for 50 min to obtain crude *Ct*RGL*f* in the supernatant. The protein concentration was assayed by method of Bradford using BSA as standard (Bradford, 1976).

6.2.2 Rhamnogalacturonan lyase activity assay

Enzyme activity of crude *Ct*RGL*f* was assayed at 70°C for 5 min using soybean rhamnogalacturonan (Megazyme, Ireland) as substrate. The reaction mixture (125 µl) for enzyme activity assay comprised 1% (w/v) soybean rhamnogalacturonan I dissolved in 50 mM Tris-HCl buffer (pH 8.5) containing 3 mM CaCl2 and 20 µl of crude *Ct*RGL*f*. The enzyme activity was measured as reported earlier (Dhillon *et al*., 2016).

6.2.3 Bio-scouring of cotton fabric

6.2.3.1 Desizing of cotton fabric

Plain-woven coarse cotton fabric (2.5×6.5 cm; 0.1 g) was used for bio-scouring experiments. Cotton fabric was soaked in water and then rinsed after 10 min to remove any dust or dirt. 1 mg of α-amylase (Sigma-Aldrich, India) dissolved in 100 ml of sodium phosphate buffer (pH 6.8) was used for desizing the cotton fabric pieces (5 x 0.1 g) at 40°C and 100 rpm for 1 h.

6.2.3.2 Bioscouring

The desized and washed cotton fabric was used for bioscouring experiments. To optimise the concentration of crude *Ct*RGL*f* required for bio-scouring the washed fabric pieces were treated with different concentrations (3 mg/ml; 0.8U/ml, 6 mg/ml; 1.6U/ml, 9 mg/ml; 2.4U/ml, 12 mg/ml; 3.2U/ml) of *Ct*RGL*f* in individual test tubes (15 mm x 125 mm). The 12 mg /ml stock of enzyme was appropriately diluted using 50 mM Tris-HCl buffer (pH 8.5) to make the final volume to 2.5 ml and the final $CaCl_2$ concentration was 3 mM. The tubes were incubated at 60°C and 100 rpm for 1 h. The cotton fabric immersed in 2.5 ml of 50 mM Tris-HCl buffer (pH 8.5) containing 3 mM $CaCl_2$ without the enzyme was used as a negative control. After incubation, the fabric pieces were rinsed with water and then dried at 60°C for 12 h. To optimise the bio-scouring time period, the pieces of cotton fabric were treated with 2.5 ml of *Ct*RGL*f* (12 mg/ml; 3.2 U/ml) in 50 mM Tris-HCl buffer (pH 8.5) containing 3 mM $CaCl_2$ and incubated at 60°C and 100 rpm for different time periods (15, 30, 45, 60 and 90 min). The increase in water absorption capacity was determined by measuring the time taken by the cotton fabric to absorb a drop (50 µl) of water

(Chakraborty *et al*., 2017; Morozova *et al*., 2006). The bio scouring experiments were performed in triplicates.

6.2.4 Enzymatic degumming of jute fibers

The jute fibers of approximately, 1.5 cm length were used for degumming experiments. 10 mg of jute fibers contained in 2 ml micro-centrifuge tubes were treated with different *Ct*RGL*f* concentrations (3 mg/ml; 0.8U/ml, 6 mg/ml; 1.6U/ml, 9 mg/ml; 2.4U/ml, 12 mg/ml; 3.2U/ml) at 60°C and 100 rpm for 1 h. The 12 mg/ml stock of enzyme was appropriately diluted using 50 mM Tris-HCl buffer (pH 8.5) to make the final volume to 1 ml, with a final $CaCl_2$ concentration of 3 mM. The crude *Ct*RGL*f* was removed by pipette, the fibers were washed 3-5 times with water and dried at 60°C for 12 h. In a parallel experiment jute fibers incubated in micro-centrifuge tubes (2 nos.) containing 1 ml of 50 mM Tris-HCl buffer (pH 8.5) containing 3 mM $CaCl_2$ were used as negative control. Jute fibers treated with 5 g/l sodium hydroxide were uses as a positive control. To optimise the degumming time, the jute fibers were treated with 1 ml of *Ct*RGL*f* (12 mg/ml; 3.2 U/ml) and incubated at 60°C and 100 rpm for different time periods (30 min, 1h, 2h and 3h). The extent of degumming was analysed by imaging the treated and un-treated fibers under Field Emission Scanning Electron Microscope (FESEM) (Sigma, Zeiss, Germany). The degumming experiments were performed in duplicates.

6.3. Results and Discussion

6.3.1 Increase in wet-ability of *Ct*RGL*f* treated cotton fabric

The time taken by the cotton fabric to absorb a drop of water without the treatment of *Ct*RGL*f* enzyme was 40 min (Fig. 6.1). The cotton fibers treated with crude *Ct*RGL*f* showed enhanced wet-ability. The time taken by the cotton fabric to absorb a drop of water decreased from 8 min to ~ 30 s with the increase in concentration of crude *Ct*RGL*f* from 3 mg/ml to 12 mg/ml, respectively (Fig. 6.1). Cotton fabric treated with 9 mg/ml of crude *Ct*RGL*f* took ~ 34 s to absorb a drop of water, while those treated with 6 mg/ml and 3 mg/ml of crude *Ct*RGL*f* took ~ 1.5 min and 8 min, respectively. 12 mg/ml (3.2 U/ml) of crude *Ct*RGL*f* was found to be best concentration for bio-scouring of cotton fabric (Fig. 6.1).

The *Ct*RGL*f* untreated cotton fabric took 40 min to absorb a drop of water (Fig. 6.2). The time taken by the cotton fabric to absorb a drop of water decreased from 25 min to 30 s with the increase in duration of bio-scouring from 15 min to 45 min, respectively (Fig. 6.2). After 45 min there was no further decrease in the time to absorb the water droplet. The cotton fabric treated for 15 and 30 min took 25 min and 1.2 min, respectively. The crude *Ct*RGL*f* treatment resulted in 98% reduction in time taken (from 40 min to 30 s) to absorb a water drop after 45 min of bio-scouring (Fig. 6.2). The recombinant pectate lyase from *Clostridium thermocellum* treatment resulted a decrease from 15 min to 3 min (85%) in the time taken to absorb a water drop by the cotton fabric (Chakraborty *et al*., 2017). In an another study that utilized a combination of commercial pectinase, cellulase and lipase, the time taken to absorb a water drop by cotton fabric was decreased (by 87%) from 1.8 min to 14 s (de Melo da Silva *et al*., 2017).

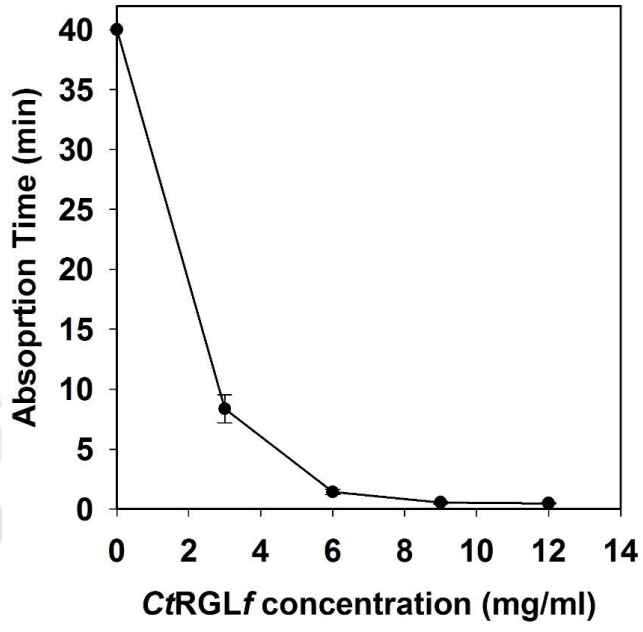


Fig. 6.1 Effect of crude *Ct*RGL*f* enzyme concentration on hydrophilicity of cotton fabric. Bio-scouring was carried out at 60°C and 100 rpm for 60 min. *Ct*RGL*f* untreated cotton fabric took 40 min to absorb a drop of water.

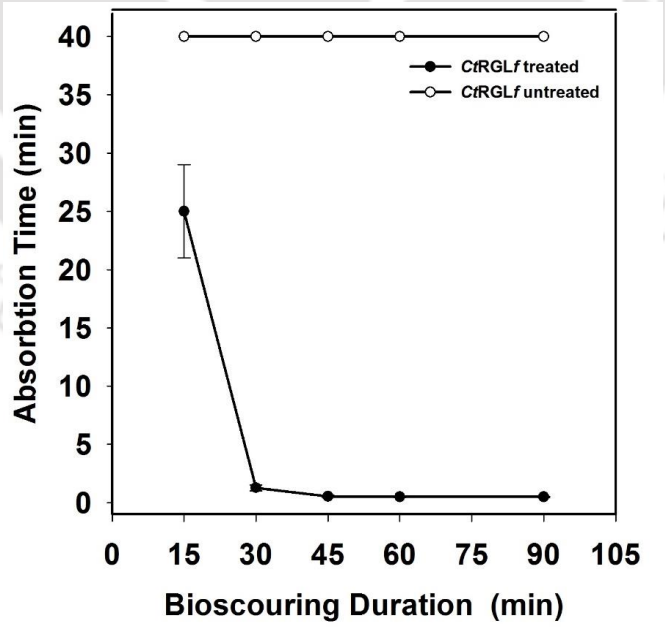


Fig. 6.2 Effect of bio-scouring time duration on hydrophilicity of cotton fabric. Bio-scouring was carried out at 60°C and 100 rpm. *Ct*RGL*f* untreated cotton fabric took 40 min to absorb a drop of water.

6.3.2 Enzymatic degumming of jute fibers

The digital images of enzymatically degummed, chemically degummed and enzyme untreated jute fibers show variation in colour (Fig. 6.3). The enzymatically and chemically degummed fibers appeared light in colour as compared with the enzyme untreated fibers (Fig. 6.3). Different concentrations of crude *Ct*RGL*f* were used for degumming the jute fiber and analysed by FESEM. The FESEM images of untreated jute fibers revealed that the surface was rough due to the presence of pectin associated gummy substances (Fig. 6.4A). Removal of pectin associated gummy substances by sodium hydroxide made the fibers surface smooth (Fig. 6.4B). The gummy substances from jute fibers were most efficiently removed by 12 mg/ml (3.2 U/ml) of *Ct*RGL*f* (Fig. 6.4C). FESEM images revealed that *Ct*RGL*f* treatment of jute fibers made their surface smooth. 1 h treatment of jute fibers by enzyme was found to be sufficient as the surface of fibers even after 3h of enzyme treatment were similar to the fibers treated for 1 h (Fig. 6.5).

Fig. 6.3 Digital images of (A) *Ct*RGL*f* untreated jute fibers (negative control), (B) chemical (NaOH) treated jute fibers (positive control) and (C) *Ct*RGL*f* treated jute fibers.

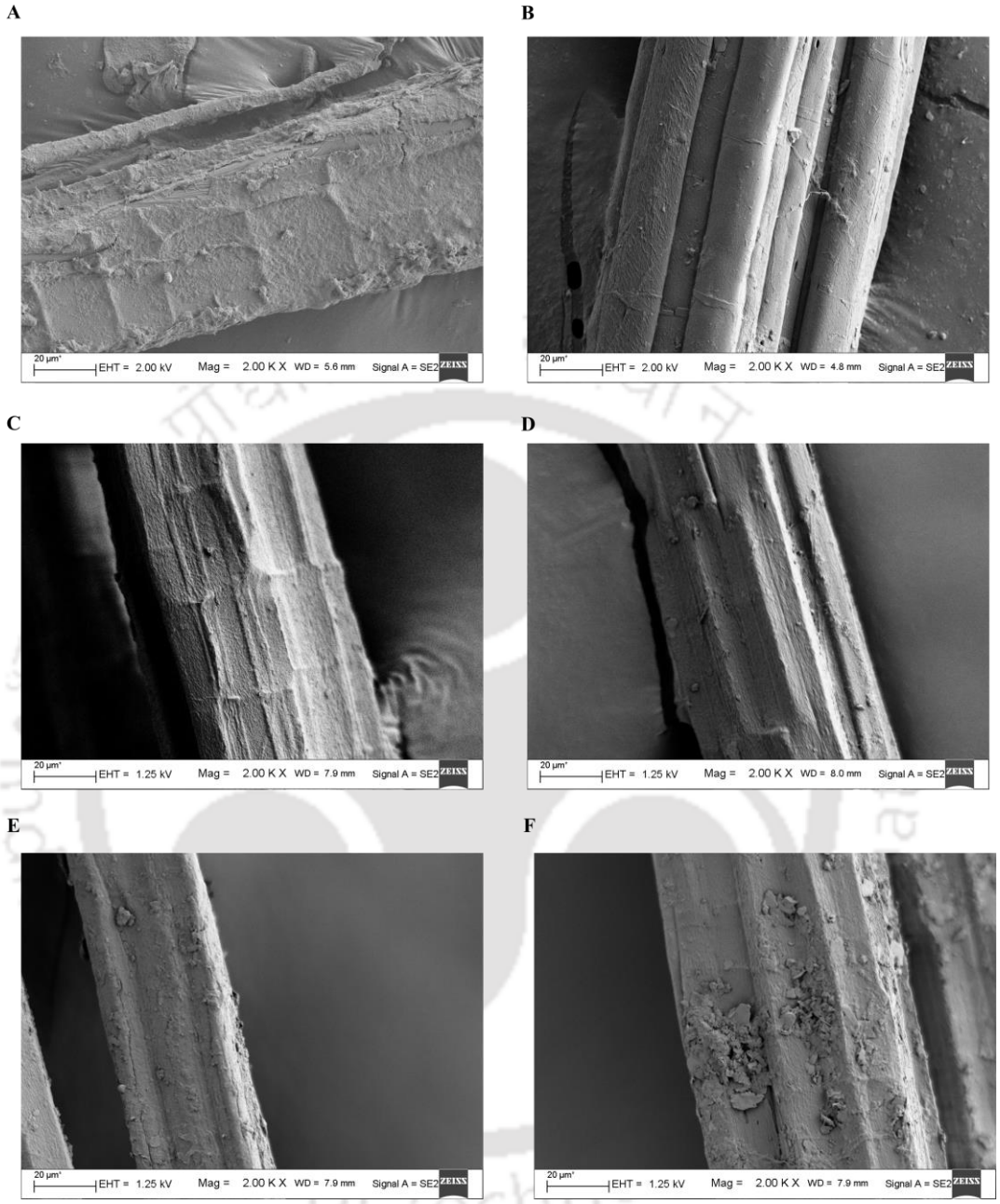


Fig. 6.4 FESEM images of jute fibers, degummed enzymatically using (A) *Ct*RGL*f* untreated jute fibers (control), (B) chemically (NaOH) degummed jute fibers (positive control), (C) *Ct*RGL*f* (12 mg/ml; 3.2 U/ml), (D) *Ct*RGL*f* (9 mg/ml; 2.4 U/ml), (E) *Ct*RGL*f* (6 mg/ml; 1.6 U/ml) and (F) *Ct*RGL*f* (3 mg/ml; 0.8 U/ml) treated jute fibers.

The removal of pectin-associated gum from jute fibers by *Ct*RGL*f* was comparable with that of chemical (NaOH; 5g/l) treated fibers. Surprisingly, the use of

any crude or pure enzymes for degumming of jute has never been explored. Only a few recent studies have focused on microbial degumming of jute fibers using bacterial isolates (Das *et al*., 2012; Duan *et al*., 2016). Microbial degumming process involves 15 h to 15 days. However, the time period for degumming, 1 h, reported in the present study is much less. The degumming of ramie was carried out using recombinant pectate lyase, where, 4 h of degumming time was used (Zhou *et al*., 2017).

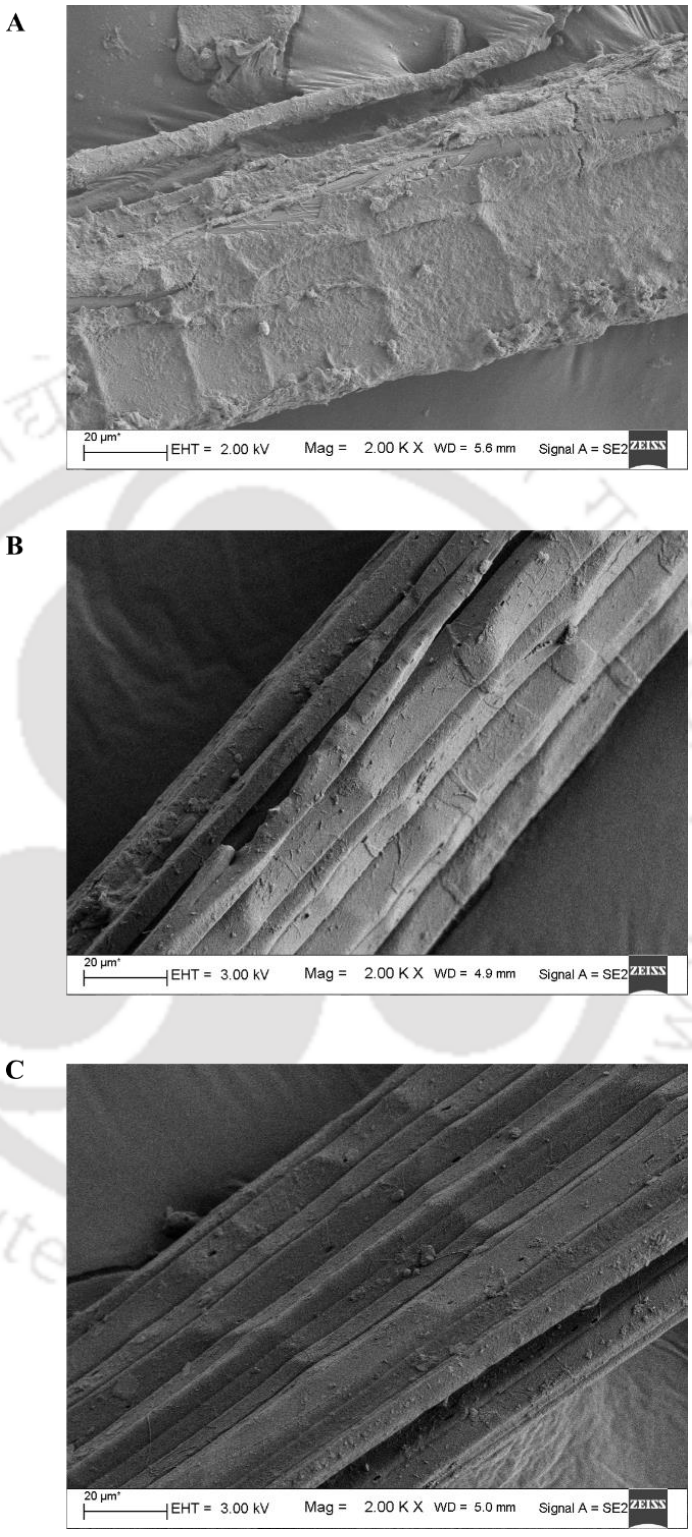


Fig. 6.5 FESEM images of (A) *Ct*RGL*f* untreated jute fibers, (B) jute fibers degummed using *Ct*RGL*f* (12 mg/ml; 3.2 U/ml) for 1 h and (C) jute fibers degummed using *Ct*RGL*f* (12 mg/ml; 3.2 U/ml) for 3 h.

6.4 Conclusions

The scouring of cotton and degumming of jute fiber involves intensive consumption of water and chemicals, therefore enzymatic processes are emerging in order to reduce the use of both. In this study, the recombinant rhamnogalacturonan lyase, *Ct*RGL*f* was used for bio scouring of cotton fabric and degumming of jute fibers. The best concentration for both bio-scouring of cotton and degumming of jute was found to be 12 mg/ml (3.2 U/ml) of crude *Ct*RGL*f*. The cotton fabric treated with 12 mg/ml (3.2 U/ml) of crude *Ct*RGL*f* took 30 s to absorb a drop of water, while the *Ct*RGL*f* untreated fabric took 40 min. The crude *Ct*RGL*f* reduced the water absorption time of cotton fabric by 98% after 45 min of treatment, where as, in the industries, the usual scouring time is 1 h. The FESEM images showed that the crude *Ct*RGL*f* (12 mg/ml; 3.2 U/ml), effectively removed the waxy compounds from the surface of jute fibers after 1h of treatment. These results show that *Ct*RGL*f* has a potential application for textile industry. To best of our knowledge, this is the first study on the use of rhamnogalacturonan lyase for bio scouring of cotton fabric and degumming of jute fiber.

References

BeMiller, J. N. (2007) Cell Walls: Economic Significance. Handbook of Plant Science, 1:283.

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

Chakraborty, S., Rao, J. M., Goyal, A. (2017) Immobilization of recombinant pectate lyase from *Clostridium thermocellum* ATCC-27405 on magnetic nanoparticles for bioscouring of cotton fabric. Biotechnology Progress, 33(1): 236-244.

Das, B., Chakrabarti, K., Ghosh, S., Majumdar, B., Tripathi, S., Chakraborty, A. (2012) Effect of efficient pectinolytic bacterial isolates on retting and fiber quality of jute. Industrial Crops and Products, 36(1): 415-419.

de Melo da Silva, L. G., de Oliveira, D., Ulson de Souza, A. A., Guelli Ulson de Souza, S. M. (2017) Study and application of an enzymatic pool in bioscouring of cotton knit fabric. Candian Journal of Chemical Engineering, 9999:1-8.

Dhillon, A., Fernandes, V. O., Dias, F. M., Prates, J. A., Ferreira, L. M., Fontes, C. M., Goyal, A. (2016) A new member of family 11 polysaccharide lyase, rhamnogalacturonan lyase (*Ct*RGL*f*) from *Clostridium thermocellum*. Molecular Biotechnology, 58(4): 232-240.

Duan, S., Feng, X., Cheng, L., Peng, Y., Zheng, K., Liu, Z. (2016) Bio-degumming technology of jute bast by *Pectobacterium* sp. DCE-01. AMB Express, 6:86-92.

Etters, J. N. (1999) Cotton preparation with alkaline pectinase: an environmental advance. Textile Chemist and Colorist and American Dyestuff Reporter, 1(3): 33-36.

Gillespie, A. M., Keane, D., Griffin, T. O., Tuohy, M. G., Donaghy, J., Haylock, R. W., Coughlan, M. P. (1990) The application of fungal enzymes in flax retting and the properties of an extracellular polygalacturonase from *Penicillium capsulatum*. Biotechnology in Pulp and Paper Manufacture. Butterworth-Heinemann, Stoneham: 211-219.

Haque, M. S., Ahmed, Z., Asaduzzaman, M., Quashem, M. A., Akhter, F. (2002) Distribution and activity of microbial population for jute retting and their impact on water of jute growing areas of Bangladesh. Pakistan Journal of Biological Sciences, 5:704-6.

Kozłowski, R. M. (Ed.). (2012) Handbook of natural fibers: Types, properties and factors affecting breeding and cultivation. Elsevier, 1: 32-34.

Klug-Santner, B. G., Schnitzhofer, W., Vršanská, M., Weber, J., Agrawal, P. B., Nierstrasz, V. A., Guebitz, G. M. (2006) Purification and characterization of a new bioscouring pectate lyase from *Bacillus pumilus* BK2. Journal of Biotechnology, 121(3): 390-401.

Mohnen, D. (2008) Pectin structure and biosynthesis. Current Opinion in Plant Biology, 11(3): 266-277.

Morozova, V. V., Semenova, M. V., Salanovich, T. N., Okunev, O. N., Koshelev, A. V., Bubnova, T. V., Krichevskii, G. E., Timatkov, A. G., Barysheva, N.V., Sinitsyn, A. P. (2006) Application of neutral-alkaline pectate lyases to cotton fabric boil off. Applied Biochemistry and Microbiology, 42: 603-608.

O'Neill, M. A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A. G., 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. Journal of Biological Chemistry, 271: 22923-22930.

Saikia, R., Boruah, P., Samanta, R. (2009) Microbial degumming of decorticated ramie and its fibre characteristics. Indian Journal of Fiber and Textile Research, 34: 187-190.

Sawada, K., Tokino, S., Ueda, M., Wang, X. Y. (1998) Bioscouring of cotton with pectinase enzyme. Coloration Technology, 114(11): 333-336.

Zhou, C., Ye, J., Xue, Y., Ma, Y. (2015) Directed evolution and structural analysis of alkaline pectate lyase from the alkaliphilic bacterium *Bacillus* sp. strain N16-5 to improve its thermostability for efficient ramie degumming. Applied and Environmental Microbiology, 81(17): 5714-5723.

Zhou, C., Xue, Y., Ma, Y. (2017) Cloning, evaluation and high-level expression of a thermo-alkaline pectate lyase from alkaliphilic *Bacillus clausii* with potential in ramie degumming. Applied Microbiology and Biotechnology, 101(9): 3663-3676.

Journal Publications

Accepted:

1. **Arun Dhillon** and Arun Goyal. (**2017**) Structure modeling and characterization of a rhamnogalacturonan lyase (CtRGL) from *Clostridium thermocellum*. **Journal of Proteins & Proteomics**, 8(4), 183-194.
2. **Arun Dhillon,** Vania O. Fernandes, Fernando M.V. Dias, José A.M. Prates, Luis M.A. Ferreira, Carlos M.G.A. Fontes, M.S.J. Centeno and Arun Goyal. (**2016**) A new member of family 11 polysaccharide lyase, rhamnogalacturonan lyase (*Ct*RGL*f*) from *Clostridium thermocellum*. **Molecular Biotechnology**, 58(4), 232-240.

Submitted:

3. **Arun Dhillon** and Arun Goyal. (**2017**) Bio-scouring of cotton fabric and enzymatic degumming of jute fibres by a thermo-alkaline recombinant rhamnogalacturonan lyase, *Ct*RGL*f* from *Clostridium thermocellum.* (under review)
4. **Arun Dhillon**, Kedar Sharma, Vikky Rajulapati and Arun Goyal. (**2017**) The multi-ligand binding first family 35 Carbohydrate Binding Module (CBM35) of *Clostridium thermocellum* targets rhamnogalacturonan I. (**s**ubmitted)

Conferences/Symposia/Meetings

1. **Arun Dhillon** and Arun Goyal. (2017) Recombinant rhamnogalacturonanan lyase (*Ct*RGL*f*) from *Clostridium thermocellum* and its use in textile processing. 12th Carbohydrate Bioengineering Meeting, April 23-26, 2017, Vienna, Austria.
2. **Arun Dhillon** and Arun Goyal. (2016) A novel family 35 Carbohydrate Binding Module (*Rgl*-CBM35) from *Clostridium thermocellum* binds rhamnogalacturonan I. 57th International Annual Conference of The Association of Microbiologists of India, Nov 24-27, 2016, Gauhati University and IASST, Guwahati, Assam India.
3. **Arun Dhillon,** Vania O. Fernandes, Fernando M.V. Dias, José A.M. Prates, Luis M.A. Ferreira, Carlos M.G.A. Fontes, M.S.J. Centeno and Arun Goyal. (2015) Biochemical characterization and determination of substrate cleavage pattern of a recombinant rhamnogalacturonan lyase, *Ct*RGL from *Clostridium thermocellum*. 56th Annual Conference of Association of Microbiologists India (AMI), December 5-8 2015, Jawaharlal Nehru University, New Delhi.
4. Priyanka Nath, Anil Kumar Verma, **Arun Dhillon**, Kedar Sharma, and Arun Goyal. (2015) Identification of promising functional residues capable of introducing endo-xylanase activity into an exo-acting arabinofuranosidase (*Ct*43Ara*f*) with enhanced activity: An *in silico* approach. 56th Annual Conference of Association of Microbiologists India (AMI), December 5-8 2015, Jawaharlal Nehru University, New Delhi.
5. **Arun Dhillon** and Arun Goyal. (2014) Biochemical characterization of recombinant rhamnogalacturonanlyase (*Ct*RGL), a family 11 Polysaccharide Lyase (PL11) from *Clostridium thermocellum.* 11th BRSI convention & International Conference on Emerging Trends in Biotechnology at JNU New Delhi, India.
6. **Arun Dhillon** and Arun Goyal. (2014) Cloning, expression and characterization of rhamnogalacturonan lyase, a family 11 Polysaccharide Lyase (PL11) from *Clostridium thermocellum*. 27thInternational Carbohydrate Symposium. January 12-17, 2014, Indian Institute of Science, Bangalore.
7. Rishikesh Shukla, **Arun Dhillon** and Arun Goyal. (2011) Purification and characterization of *Leuconostoc mesenteroides* NRRL B-1149 sucrose hydrolysing enzymes and enzyme synthesized oligosaccharides. 52nd Annual Conference, Association of Microbiologists India (AMI), November 3-6, 2011, Panjab University, Chandigarh.

VITAE

The author was born on August 28, 1988 in the city of Nagpur, (Maharashtra). He passed the Secondary Examination (10th Class) conducted by Central Board of Secondary Education, Delhi in 2004 and Higher Secondary Examination (12th Class) conducted by Central Board of Secondary Education, Delhi in 2006. He completed B.Tech. (Biotechnology) from Career Institute of Technology and Management, Faridabad, affiliated to Maharishi Dayanand University, Rohtak (Haryana) in July, 2011.

Mr. Arun Dhillon joined the Ph.D. program in July, 2011 at Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781 039, Assam, India. He successfully completed the course work with 6.85/10 CPI. He received Institute Fellowship (IIT Guwahati) from Aug 2011 to Jul 2016, under the scheme run by the Ministry of Human Resource and Development (MHRD), New Delhi. From December 2016 to April, 2017 he received fellowship from a DBT sponsored project in Department of Biosciences & Bioengineering under Prof. Arun Goyal. He delivered the open (PhD Synopsis) Seminar on October 16, 2017 and presented his thesis work before the Doctoral Committee and his performance was satisfactory. He submitted the PhD thesis in October 2017.