Cloning, expression, purification, biochemical, functional and structure characterization of an endoglucanase of family GH5_4 (*Rf*GH5_4) from *Ruminococcus flavefaciens* FD-1 v3 and its application in lignocellulosic biomass conversion

PhD Thesis

by

Parmeshwar Gavande

January 2024

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI – 781039, ASSAM, INDIA

Cloning, expression, purification, biochemical, functional and structure characterization of an endoglucanase of family GH5_4 (*Rf*GH5_4) from *Ruminococcus flavefaciens* FD-1 v3 and its application in lignocellulosic biomass conversion

A Thesis

Submitted in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy

by

Parmeshwar Gavande

Under supervision of

Professor Arun Goyal

January 2024

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, biochemical, functional and structure characterization of an endoglucanase of family GH5_4 (*Rf*GH5_4) from *Ruminococcus flavefaciens* FD-1 v3 and its application in lignocellulosic biomass conversion"** 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.

9th January, 2024 *Parmeshwar Gavande*

(186106008)

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

CERTIFICATE

It is certified that the work described in this thesis entitled **"Cloning, expression, purification, biochemical, functional and structure characterization of an endoglucanase of family GH5_4 (*Rf*GH5_4) from *Ruminococcus flavefaciens* FD-1 v3 and its application in lignocellulosic biomass conversion"** by **Parmeshwar Gavande (Roll No. 186106008)** 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.

9-01-2024

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.

Dedicated

to my

Dada-Aai

(Father-Mother)

&

Mhatarimay

(Grandmother)

Acknowledgements

As is the beautiful feeling for Sun, owing his thanks to the mighty night, which selflessly stood by him throughout the challenging darkness till the rise; so is the inexpressible delightedness of mine towards my ***Thesis Supervisor Prof. Arun Goyal*** *for grooming me in whatever I am today, and for all my achievements. His teachings were critical in defining the correct course of my doctoral journey which would not be sustained successfully, had his instructions, directions and suggestions not been honing me towards a better researcher, every rising day. I wish to be indebted forever in my life for his path-showing teachings. These would be surely spanning over my research-academic career as well as being some of the important pillars of my principles towards life, in the times to come.*

This a pleasant moment for me to thank my ***Doctoral Committee*** *(DC) members-* ***Dr. Rajkumar P. Thummer*** *(Chairman, DC),* ***Prof. Vijayanand S. Moholkar*** *and* ***Dr. Priyadarshi Satpati*** *whose valuable suggestion during the annual progress reviews seminars (APS) improved the present thesis to have a better scientific temperament. It will be among my best cherishable memories to recall the dutiful efforts put by honourable committee members to successfully conduct my State-of-The-Art Seminar (SOAS) when the COVID19 pandemic was on its highest pick in May 2020. I am sincerely grateful towards my DC members for their constructive and critical evaluation of my doctoral journey.*

It is indeed a pleasure to thank ***Department of Biosciences and Bioengineering (BSBE and Central Instrumentation Facility (CIF), Indian Institute of Technology Guwahati*** *for providing the world-class research facilities which paved the way for successful completion of my Thesis work and allowed me to stand out in the international competition.*

I am sincerely grateful towards ***Prof. Rakhi Chaturvedi, Head of the Dept. of BSBE*** *for her constant encouragement and striving for providing the best possible research facilities to the Department's Doctoral Research Scholars. I am immensely thankful to all the former HoDs of the Department for keeping the competitive facilities and overall research ambience in the Department. I am especially thankful to* ***Prof. Latha Rangan, former HoD of Dept. of BSBE*** *for giving me the opportunity in*

christening the name for the Departmental newsletter, "Awalokan" and starting the inclusive programs like Annual Departmental Retreat-Biotech Express which inculcated the feeling of togetherness among us, the research scholars, when it was highly required.

I owe my sincere thanks towards my senior lab colleagues Dr. Vikky Rajulapathi, Dr. Sumitha J. Banu, Dr. Kedar Sharma, Dr. Krishan Kumar, Dr. Shweta Singh, Dr. Priyanka Nath, Dr. Abhijeet Thakur, Dr. Kaustubh C. Khaire for helping me every now and then while performing the experiments in the laboratory.

I cannot forget to thank all my labmates- Jebin, Robin, Premeshworii, Madhulika, Ardhendu, Vishwanath, Akshita, Shreya, Aishwarya, Bipasha, Nazneen and Gurleen- with whom I have some of the best cherishable memories of the IITG campus life. I wish all the best to my new juniors Sushruta, Akshay and Aswini to achieve the best in their Doctoral study. It was a great journey for me to be part of the Carbohydrate Enzyme Biotechnology Laboratory (CEBL) at IIT Guwahati.

I owe my thanks to all my buddies at Kameng hostel, especially, Krishnakant and Nikhil who have been the real stressbusters during my ups and downs of the PhD journey.

I am really thankful to Ministry of Education (erstwhile MHRD) for providing me the GATE fellowship for my Doctoral research which made my financial base concrete and thus allowed me to focus and streamline my research activities.

I am really thankful to Dr. Ravishankar Ramachandran, Principal Scientist and his team members, CSIR-Central Drug Research Institute, Lucknow, India for providing the SAXS facility and Data collection. I am also thankful to IIT Guwahati for Param-Ishan Supercomputing facility and Centre for Environment, IIT Guwahati, for providing the DLS facility.

And this not a feeling of thankfulness; rather, I owe them so much I cannot count in volume- the ones who raised me, inculcated the morals and ethics within me, stood by me throughout my life and are the happiest person in this world on my every small

achievement- these are my parents: ***Dada-Aai****, my sisters-****Ratna Akka, Swati Akka, Sapna Akka and Pranjali****. They make me complete.... I recall my memories with my grandmother,* ***Mhatarimay*** *who has been residing at God's house since 2017. I always cherished her as my biggest stalwart and my cheerleader.*

I am thankful to all those who motivated, supported me directly and indirectly during my PhD tenure.

Thank you, ***God****, for giving me this wonderful life, without your blessings my journey would have been incomplete… Thank you,* ***Destiny,*** *for the beautiful teachings and life experiences… I wish, I will succeed to the best, with your teachings, in serving this great land of* ***Bharatwarsha*** *while practicing the principle of* ***Vasudhaiva Kutumbakam****.*

Parmeshwar Vitthal Gavande,

January, 2024

SYNOPSIS

The present work entitled as "Cloning, expression, purification, biochemical, functional and structure characterization of an endoglucanase of family GH5_4 (*Rf*GH5_4) from *Ruminococcus flavefaciens* FD-1 v3 and its application in lignocellulosic biomass conversion" has been divided into 5 chapters.

Significance of the work

Efficient hydrolysis of plant cell wall polysaccharides has been fine-tuned over millions of years in ecological niches that are subjected to intensive selective pressures exemplified by the rumen of mammalian herbivores. A cohort of rumen anaerobic bacteria assembles their plant cell wall degrading apparatus into multi-protein complexes termed as cellulosomes. *R. flavefaciens* is one of the few rumen bacteria which degrade plant cell wall, particularly cellulose, efficiently. It is the also a member of important group of bacteria which contains organized and advanced form of cellulosome (Venditto *et al.*, 2016) As can be concluded from review of the published literature, having complete genome sequence of *R. flavefaciens* and its analysis, we now have whole set of its putative and annotated endoglucanases (Berg Miller *et al.*, 2009). There are at least 14 glycoside hydrolases of family GH5 in this bacterium's genome but they have merely been completely characterized biochemically and structurally. In the present state, fungal cellulases dominate the industrial cellulase market. However, it is found that k_{cat} and k_{cat}/K_m of these enzymes are very low, whereas K_m for cellulosic substrates are quite high. So, it is not exaggeratory to say that we are still in search of an efficient endoglucanase for the purpose of bioethanol production. *Rf*GH5_4, a putative endoglucanase, is naturally present in the genome of *R. flavefacien*s FD-1 v3

as a part of a multienzyme complex, *Rf*GH$_{1/2}$ (GenBank Accession No. WP_009984467.1). Being from a rumenophilic bacterium, it was hypothesized to be an efficient endoglucanase for deconstruction of plant biomass for the purpose of bioethanol production. Moreover, being a member of subfamily 4, it was proposed to be a polyspecific/multifunctional cellulase/endoglucanase hydrolysing various glucan like cellulose, xyloglucan, lichenan and β-glucan. Molecular docking studies of *Rf*GH_4 with various cellulosic and hemicellulosic ligands would reveal its active-site interactions with the substrates. Molecular dynamics simulation study of the *Rf*GH_4 protein with cellulosic ligand would help in demonstrating its stability and compactness with the cellulosic substrates. Characterization by other physical techniques likes Thin layer chromatography, MALDI-MS, HPLC and Circular Dichroism (CD) would reveal the information about its structure-function relationship, mechanism of catalysis as well as multifunctionality, active-site interactions and behaviour in the solution form. Solution structure studies by SAXS would shed light on its intermolecular interactions, monodispersity of the protein and consequently its stability. Therefore, along with being a potential candidate as a biocatalyst for bioethanol production from lignocellulose, it could be potentially employed for various other applications like textile, food and pulp industries, pharmaceutical sector, modification of cellulose surface, plant growth modulation, production of xyloglucan oligosaccharides and xyloglucan-specific probes (Rashmi and Siddalingamurthy, 2018).

Chapter 1 gives the introduction and overview of the literature. Cellulose is the major component of lignocellulosic plant biomass and most abundant polysaccharide present on earth. Endo-β-1,4-glucanase (EC 3.2.1.4) is produced by microorganisms like fungi, bacteria and archaea. It randomly catalyzes the β-1,4-glycosidic bond

between glucose monomers of cellulose chain and produce short chain cello-oligosaccharides. Endo-β-1,4-glucanase is classified under Glycoside Hydrolase (GH) group in Carbohydrate Enzyme database. Endo-β-1,4-glucanases from different sources are distributed over various GH families such as GH5-GH9, GH12, GH16, GH17, GH26, GH44, GH45, GH48, GH51, GH74, GH124 and GH148 based on the amino acid sequence similarity and their mode of action. The structure of endoglucanases possesses either $(\alpha/\beta)_8$ TIM-barrel (triose phosphate isomerase) or β-jelly-roll topology or $(\alpha/\alpha)_6$-barrel. The active-site of endoglucanase shows characteristic open groove, where the cellulosic chain is randomly cleaved to give cello-oligosaccharides through either retaining or inverting mechanism. Most of the endoglucanases contain one or more Carbohydrate Binding Module/s (CBM) which help the enzyme in attaching the recalcitrant cellulose thereby assisting in hydrolysis. Endoglucanases having multifunctionality, wide pH and optimum temperature ranges and stability with notable catalytic efficiency are required for various industrial, bioenergy, food and feed applications. Thus, this chapter includes the literature review of various GH families having endoglucanases, details of the bacterium *Ruminococcus flavefaciens* and description of few of GHs from it.

Chapter 2 describes the cloning of the gene encoding *Rf*GH5_4 using recombinant DNA technology techniques, its expression in protein form and purification to homogeneity for the purpose of further characterization. The gene encoding *Rf*GH5_4 (GenBank Accession Number - WP_009984467.1) endoglucanase was cloned from bacterium *R. flavefaciens* FD-1 v3. *Rf*GH5_4 is part of the wild-type multimodular complex, *Rf*GH51/2 placed at its N-terminal. The PCR amplified fragment of gene encoding *Rf*GH5_4 showed a band of approx. 1.0 kb (1047 bp), The *Nhe*I-*Xho*I digested

PCR DNA fragment and linearized pET-28a(+) were ligated using T4 DNA ligase. The ligated mixture was transformed into *E. coli* TOP10 cells and colonies were evaluated for positive clone confirmation through restriction digestion of the isolated recombinant plasmid. The recombinant plasmid was then transformed into *E. coli* BL21 (DE3) expression cells. The positive colonies were screened for expression of protein of interest, *Rf*GH5_4 IPTG induction. The recombinant protein, *Rf*GH5_4 was purified to homogeneity using the Affinity Chromatography (IMAC) which yielded total 6.97 mg of the protein at 1.74 mg/mL concentration. The total purified protein obtained from the culture pellet of 100 mL expressed *E. coli* BL21 (DE3) cells was 14.0 mg/g of dry cell weight. The SDS-PAGE analysis of the purified *Rf*GH5_4 protein on 12% (w/v) gel showed the molecular weight of approximately 41 kDa which corroborated to that of its theoretically estimated value. The purified *Rf*GH5_4 protein was used for further biochemical, functional and structural characterization.

Chapter 3 describes the biochemical and functional characterization of recombinant endoglucanase, *Rf*GH5_4 from *R. flavefaciens* FD-1 v3 which showed maximum activity at pH 5.5 and 55°C. It was stable between pH 5.0-8.0, retaining 85% activity and between 5°C - 45°C, retaining 75% activity, after 60 min. *Rf*GH5_4 displayed maximum activity (U/mg) against barley β-D-glucan (665) followed by carboxymethyl cellulose (450), xyloglucan (343), konjac glucomannan (285), phosphoric acid swollen cellulose (86), beechwood xylan (21.7) and carob galactomannan (16), thereby displaying the multi-functionality. Catalytic efficiency ($mL.mg^{-1}s^{-1}$) of *Rf*GH5_4 against carboxymethyl cellulose (146) and konjac glucomannan (529) was significantly high. TLC and MALDI-TOF-MS analyses of

*Rf*GH5_4 treated hydrolysates of cellulosic and hemicellulosic polysaccharides displayed oligosaccharides of degree of polymerization (DP) between DP2-DP11. TLC, HPLC and Processivity-Index analyses revealed *Rf*GH5_4 to be a processive endoglucanase as initially, for 30 min it hydrolysed amorphous cellulose to cellotetraose followed by persistent release of cellotriose and cellobiose. This study established *Rf*GH5_4 as a multifunctional endoglucanase prospecting potential application in renewable energy.

Chapter 4 elaborates on the structure characterization of endoglucanase, *Rf*GH5_4 with various cellulosic and hemicellulosic ligands by molecular docking and molecular dynamics simulation followed by conformational dynamics in solution form. Multifunctional endoglucanase, *Rf*GH5_4 from *Ruminococcus flavefaciens* showed $(\beta/\alpha)_8$-TIM barrel structure by homology modeling. Glu168 and Glu292 residues acted as general acid and base during catalysis. Circular Dichroism results showed, 40.83% α-helices, 13.84% β-strands and 45% random turns-coils for *Rf*GH5_4 corroborated with predictions by PSIPRED and SOPMA. Molecular Dynamics simulations of *Rf*GH5_4 for 100 ns showed RMSD, 0.71 nm while for *Rf*GH5_4-Cellopentaose complex was 0.55 nm, confirming that the binding of cellulosic ligand stabilizes its structural fold. *Rf*GH5_4 showed strong affinity towards cellulosic ligands having higher degree of polymerization such as cellohexaose (-11.70 kcal/mol) and cellodecaose (-12.64 kcal/mol). Interestingly, complex hemicellulosic ligands such as XLLG of xyloglucan also showed higher affinity (-13.2 kcal/mol) and accommodated at *Rf*GH5_4 active-site. Its catalytic cleft was broad enough to accommodate and hydrolyze various cellulosic and hemicellulosic ligands like XLLG of xyloglucan setting the basis of multifunctionality of *Rf*GH5_4. Loops L2, L3 and L4 having Trp58

formed barrier at active-site of *Rf*GH5_4 were responsible for processivity. *Rf*GH5_4 showed monodispersed state in solution form at 2.5 mg/mL and a rattle-toy shape as analyzed by Small Angle X-ray Scattering (SAXS) as well as Dynamic Light Scattering (DLS). Zeta potential of *Rf*GH5_4 was -16 mV indicating its higher stability in solution form. This study showed the structural insights of multifunctional *Rf*GH5_4 endoglucanase which could be beneficial for generation lignocellulosic bioethanol and in health, prebiotic and food sector.

Chapter 5, describes the application of *Rf*GH5_4 for the deconstruction and saccharification of various alkali pre-treated lignocellulosic biomasses and its synergy with other cellulase enzymes during cellulose hydrolysis. A new cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 for saccharification of various pre-treated agricultural biomasses involving recombinant endoglucanase-*Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 v3), cellobiohydrolase-*Ct*CBH5A and β-glucosidase-*Ct*GH1 from *Clostridium thermocellum* was formulated. For this, NaOH+autoclaving based alkali pre-treatment of lignocellulosic biomasses (2%, w/v NaOH, Cotton Main Stalk, ptCMS and Cotton Small Branches, ptCSB), (1%, w/v NaOH, Sugarcane Bagasse, ptSBG, *Sorghum durra* stalk, ptSDR, Finger Millet Stalk, ptFMS and Maize Leaves, ptMZL) was carried out. Alkali NaOH based pre-treatment of lignocellulosic biomasses was found to be effective for removal lignin and reduction of hemicellulose. This was inferred from the notable increment in the crystallinity index (*CrI*) of pre-treated biomasses as compared with the untreated ones with maximum shift in *CrI* recorded for ptCSB (41.7 from 25.4 of raw CSB) and ptSDR (44.2 from 28.4 of raw SDR). Comparative FE-SEM analyses of alkali pre-treated biomasses showed rough surface and texture against smooth surface of untreated ones further indicating the increased

surface area and effectiveness of alkali pre-treatment. *Rf*GH5_4 was applied in the deconstruction of delignified lignocellulose, wherein it yielded maximum TRS (mg/g) from ptSDR (72), ptFMS (62) and ptSBG (38). The TLC and MALDI-TOF MS analyses of these *Rf*GH5_4 hydrolysates saccharified from pre-treated biomasses showed release of mixture of higher DP cellooligosaccharides (DP2-DP11) from pre-treated biomasses demonstrating its capability to hydrolyse lignocellulose. The cocktail of *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 at ratio 50:50:100 U/g showed synergy against pure cellulosic substrate, WFP1 and delignified lignocellulose, ptSDR where this cocktail yielded 26.6 mg/g_{WFP1}, 62.0 mg/g_{ptSDR} of TRS and 7.5 mg/g_{WFP1}, 28.8 mg/g_{ptSDR} of D-glucose, respectively. The release of D-glucose from WFP1 and ptSDR by *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 was visualised through TLC analysis, confirming the synergy among *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1. These results showed that the new cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 is capable of hydrolysing delignified biomass like ptSDR. *Rf*GH5_4 could tolerate upto 20% (v/v) of ethanol till 96 h at 30°C suggesting its suitability for SSF. This prospects *Rf*GH5_4 and the newly formulated cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 as a potential cellulase toolbox for bioethanol production. Further statistical optimization of ratio of individual enzyme in cellulase cocktail could enhance the TRS and D-glucose yield. Moreover, the notable synergy observed between *Rf*GH5_4 and *Ct*CBH5A for cellulose conversion may be explored and a bifunctional chimeric cellulase, *Rf*GH5_4-*Ct*CBH5A can be constructed to make the saccharification process time saving and cost-effective.

References

Berg Miller, M. E., Antonopoulos, D. A., Rincon, M. T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B., & Coutinho, P. M. (2009). Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PloS One*, *4*(8), e6650.

Rashmi, R., & Siddalingamurthy, K. R. (2018). Microbial xyloglucanases: A comprehensive review. *Biocatalysis and Biotransformation*, *36*(4), 280–295.

Venditto, I., Luis, A. S., Rydahl, M., Schückel, J., Fernandes, V. O., Vidal-Melgosa, S., Bule, P., Goyal, A., Pires, V. M. R., & Dourado, C. G. (2016). Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proceedings of the National Academy of Sciences*, *113*(26), 7136–7141.

CONTENTS

Chapter 1

General Introduction and Review of Literature

1.1 Introduction

Life thrived and flourished on Earth because nature has established interesting ways of recycling the huge carbon and energy stored in the organic biomass. Carbohydrates, as monomers, oligomers, polymers or even as glycol-conjugates, have had a very vital role in all sorts of life forms (Venditto *et al.*, 2016). Plant biomass is the prominent stock of various carbohydrate polymers like cellulose and hemicellulose. Primary plant cell wall of most vascular plants usually consists of cellulose and hemicellulose as its structural skeleton along with lignin (Klemm *et al.*, 2005; Gilbert 2010) as depicted in **Fig. 1.1**. Cellulose is the most abundant carbohydrate polymer on the planet Earth which contains D-glucose as subunits linked through β-1,4 glycosidic linkages (Singh *et al.*, 2020). Hemicellulose is a group of carbohydrate polymers that contains a wide range of polymers like mannan, xylan, arabinan, pectin, xyloglucan, glucomannan, glucuronoxylan, arabinoxylan, galactan, galactomannan and

arabinogalactan, etc (Kuhad *et al.*, 2006; Liu *et al.*, 2008). Cellulose is observed to have the highest share among different polysaccharides of the plant cell wall. It ranges from 15-30% in the primary cell wall of dicots and grass, whereas it accounts for roughly 50% in their secondary cell wall composition (O'Neill and York, 2003).

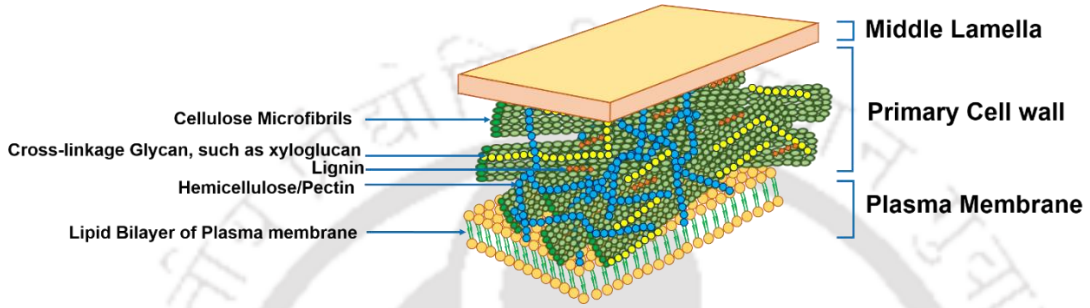

Fig. 1.1 Composition of primary plant cell wall (PCW): A trans-sectional view. Cellulose is stacked as microfibrils in the PCW.

Remnants of harvested crops or dead plants containing cellulose are required to be converted to their monomeric forms like D-glucose, to serve the purpose of recycling in the nature. D-glucose is one of the most important building blocks and the sources of energy for different life forms as its role can be observed in animal physiology. The biocatalysts deconstructing the cellulose and hemicellulose to monomeric sugars thus play crucial role in balancing the carbon between the living and non-living worlds. The enzymes hydrolyzing cellulose popularly known as cellulases, which break the β-1,4 glycosidic bonds present between the two D-glucose monomers, are grouped under a wider term, the glycoside hydrolases (GH). However, cellulase is not a single type of enzyme; rather it stands for a group of enzymes which synergistically hydrolyze cellulose polymer to D-glucose.

Cellulase enzyme system accommodates essentially three different hydrolases namely, endo-β-1,4-glucanase or endoglucanase (EC 3.2.1.4), exoglucanase or

cellobiohydrolase (EC 3.2.1.91) and β-glucosidase (EC 3.2.1.21). Endoglucanase plays an irreplaceable role through random hydrolysis in saccharifying the plant biomass to cellooligosaccharides which are further processed by exoglucanase and β-glucosidase to produce cellobiose and monomeric D-glucose, respectively. Thus, endoglucanase stands out to be indispensable amongst the cellulases in the hydrolysis of cellulose and lignocellulosic plant biomass. The increasing demand for energy and depletion of non-renewable sources of energy like petroleum has triggered energy generation from renewable sources. Easily accessible and plentiful lignocellulosic plant and agriculture waste serves the renewable energy strategy at the forefront to generate renewable fuels like bioethanol. And, the bioethanol production is only possible with the role of endoglucanases. Moreover, endoglucanases serve wide range of applications in food, feed, fruit juice and paper industries.

Given the prominent role of endoglucanases in deconstruction of lignocellulose and aforementioned industries, it becomes pertinent to study their structural and functional basis of catabolism and efficiency. Exploration of catalytically and biochemically efficient endoglucanases from various biological resources is underway for the production of bioethanol (Demain *et al.,* 2005; Wu *et al.*, 2018).

1.2 Cellulolytic enzymes

1.2.1 Cellulase

Cellulases are defined by their ability to cleave the β-1,4-glycosidic bonds of glycan polymers. They are synthesized by microorganisms during their growth on cellulosic materials (Saranraj *et al.,* 2012).

1.2.2 Cellulase evolution and conservation in nature

The cellulose produced by plants is not seen accumulated on Earth as it is efficiently hydrolyzed back to its monomer, D-glucose by naturally present cellulases. On the contrary, the overall rate of cellulose synthesis by plants through the photosynthesis was recorded to be around 10^5 tons/h (Henrissat, 1994). These facts underline the importance of cellulases for biomass recycling. The need of recycling the recalcitrant lignocellulosic biomass back in the nature necessitated the evolution and emergence of cellulase enzymes in the nature. They play the pivotal role in balancing the carbon cycle in nature. Endoglucanases have retained the function of cellulose hydrolysis in various domains of life and thus sustained the test of time and evolution. Nonetheless, they have gone through the divergence where a variety of endoglucanase families can be found having different structural basis but performing the same function of endo-hydrolysis of cellulose.

1.2.3 Endo-β-1,4-glucanase

Endo-β-1,4-glucanase or endo-1, 4-β-D-glucan glucanohydrolase (EC 3.2.1.4) is responsible for hydrolyzing the internal glycosidic linkages of cellulose chain thereby decreases its length (**Fig. 1.2**). Endoglucanases are produced by cellulolytic fungi, bacteria including those in the gut ecosystem of termite and ruminants. Plants also produce endoglucanases (such as endoglucanases from glycoside hydrolase family, GH9). Endoglucanases cleave at random at the internal amorphous cellulose regions of the cellulose chain and generate oligosaccharides of various lengths (Sangrila and Maiti, 2013). It is generally active at amorphous regions of natural cellulose, phosphoric acid-swollen amorphous cellulose (PASC), soluble derivatives of cellulose such as carboxymethyl cellulose (CMC-Na) and cellooligosaccharides. Endo-β-1,3/1,4-

glucanase (EC 3.2.1.73) is one more important type of endoglucanase (which cleaves mixed-linked β–glucan) along with endo-β-1,4-glucanase (EC 3.2.1.4) which also plays significant role in cellulose and hemicellulose deconstruction. Endoglucanases show excellent diversity i.e. divergence in their structure-function relationship, as can be observed through different CAZy families showing endoglucanase activity (EC 3.2.1.4). The structure may be different but they still perform the same function of random/processive cellulose hydrolysis.

1.2.4 Exoglucanase

Exoglucanase (cellobiohydrolase) or 1,4-β-D-glucan cellobiohydrolase (EC 3.2.1.91) attacks the crystalline ends of cellulose to yield cellobiose (Saranraj *et al.*, 2012). Exoglucanases catalyze the reducing or non-reducing ends of cellulose polysaccharide chains and releases either glucose (glucanohydrolase) or cellobiose (cellobiohydrolase) (**Fig. 1.2**). Reducing end is the endpoint of cellulose polymer with free anomeric carbon whereas, the non-reducing end has its anomeric carbon engaged in the β-glycoside linkage. Reducing end acts as reducing agent as it can donate an electron pair to the electron acceptor during a redox reaction. Reducing end has accessible anomeric -OH group which is capable of forming new β-glycosidic linkage. Exoglucanases are active on crystalline cellulose substrates such as Avicel, amorphous cellulose like PASC and cellooligosaccharides but are inactive against cellobiose. However, some exoglucanases catalyze the removal of cellobiose from cellooligosaccharides or from *p*-nitro phenyl-β–D-cellobioside (EC 3.2.1.74) but are inactive against amorphous cellulose or CMC-Na.

1.2.5 β-glucosidase

The released cellobiose is hydrolyzed into two D-glucose molecules by β-glucosidase. β-glucosidases hydrolyze cellobiose and soluble cellooligosaccharides from non-reducing end (EC 3.2.1.21) to D-glucose (**Fig. 1.2**). These are inactive against crystalline or amorphous cellulose (Saranraj *et al.*, 2012).

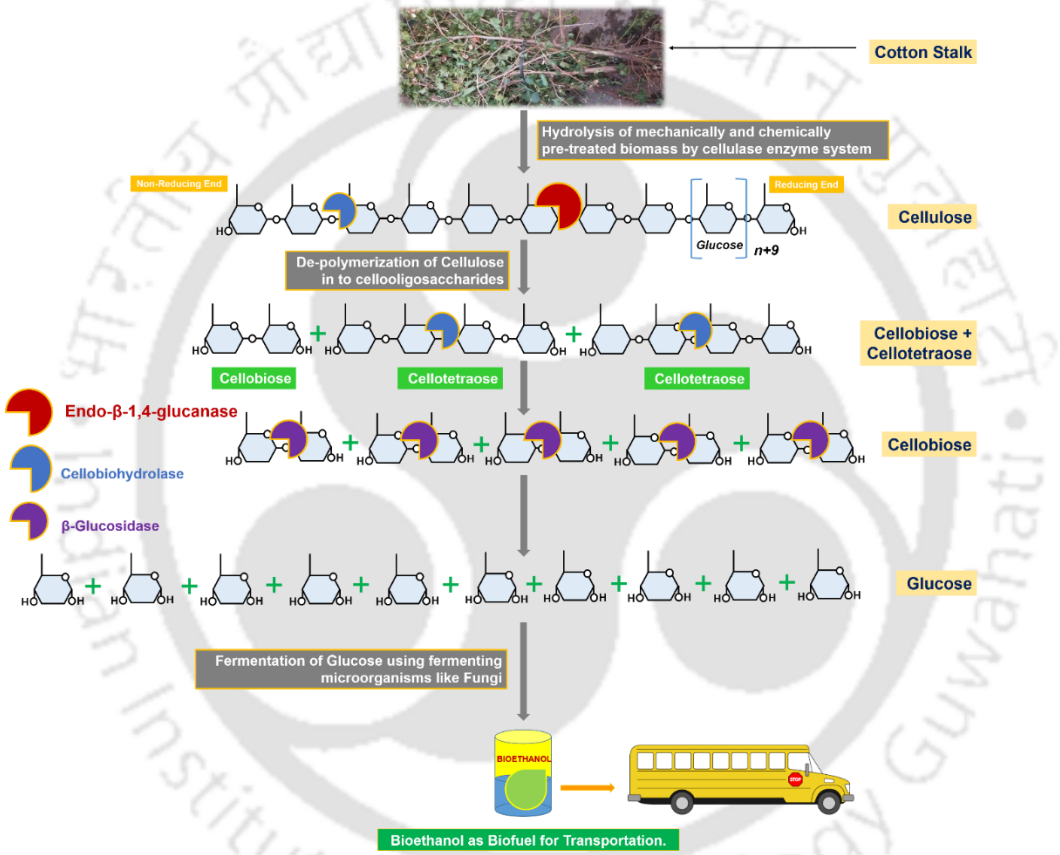

Fig. 1.2 Role of endo-β-1,4-glucanase in cellulose deconstruction. Endoglucanase hydrolyzes the cellulose to cellooligosaccharides which are further hydrolyzed by cellobiohydrolase to cellobiose. Cellobiose is hydrolyzed to two glucose monomers by β-glucosidase.

1.2.6 Cellulosome

Cellulosome is the enzyme machinery comprising multi-enzyme complexes produced by only a select group of anaerobic bacteria like *Ruminococcus flavefaciens*, *Clostridium thermocellum* (Artzi *et al.*, 2017) and anaerobic fungi like

Neocallimastigomycota yeast (Gilmore *et al.*, 2020). The cellulosome is a well-organized structure, mostly extracellular cell surface cellulase complex which can proficiently hydrolyze the compact and recalcitrant cellulose microfibrils. The insoluble cellulosic or hemicellulosic substrate is hydrolyzed into soluble oligosaccharides and monosaccharides by the extracellular cellulosome. These soluble forms of simple sugar are then absorbed inside by the microbial cell for the purpose of regular metabolism. A cellulosome is made up of two components, the Type-I Dockerin-linked enzymes and Cohesin-linked scaffoldin proteins (**Fig. 1.3**). Cohesins form strong interactions with the dockerin. Scaffoldins are bound to cell-wall of the bacterium and often contain a carbohydrate binding module, CBM (Artzi *et al.*, 2017) which binds to the cellulose polymer. Type-II Dockerin connects two different Scaffoldins and also attaches the primary scaffoldins to the anchoring protein in the cell wall as found in cellulosome of *Clostridium thermocellum.*(Artzi *et al.*, 2017). Interestingly, a recently characterized Type-III Dockerin found in the cellulosome of *R. flavefaciens* can attach the Scaffoldins to the microbial cell wall as well as adhere the CAZymes to the Scafoldins.

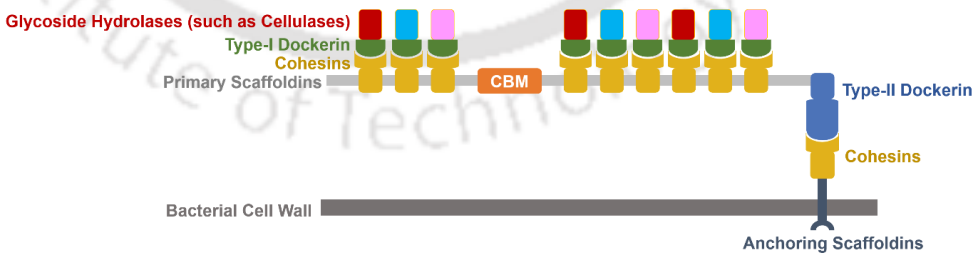

Fig. 1.3 Schematic diagram of cell-bound cellulosome complex of bacterium *Clostridium thermocellum.*

1.3 Endoglucanases belong to various GH families

Endoglucanases are grouped under different GH families based on the similarity of amino acid sequence and mode of action in the Carbohydrate-Active Enzyme

(CAZy) (Cantarel *et al.*, 2009) database (http://www.cazy.org/Glycoside-Hydrolases.html). Glycoside hydrolases are part of a large group of carbohydrate modifying enzymes, or CAZymes. The GH families which harbor the endoglucanases are: GH5, GH6, GH7, GH8, GH9, GH10, GH12, GH16, GH17, GH26, GH44, GH45, GH48, GH51, GH74, GH124 and GH148 (**Fig. 1.4**). GH10, GH26, GH51 families have only few members having endoglucanase activity, whereas GH families such as GH5, GH9 and GH48 have a large number of endoglucanases.

Fig. 1.4 CAZy database showing the GH families harboring endoglucanases, based on Cantarel *et al.*, 2009.

1.3.1 GH5 family

The GH5 family, previously known as cellulase family A, is quantitatively, one of the largest GH families and belongs to erstwhile Clan, GH-A (Henrissat *et al.*, 1989). It is known to show variety of specificities, notably endoglucanase (cellulase), endomannanase as well as exoglucanase, exomannanase and β-mannosidase (Aspeborg *et al.*, 2012). Interestingly, no human enzyme has been reported in this family. It harbors more than three thousand GH5 enzyme sequences. Around 80% of the public domain sequences in CAZy database were classified into 51 subfamilies (GH5_1 to GH5_53, excluding GH5_3 and GH5_6 which have been merged into GH5_4 and GH5_5 respectively). GH5 consists of enzymes with around 18 different identified enzyme

activities. Subfamilies 1, 2, 4, 5, 8, 22 and 26 of GH5, predominantly contain the endoglucanase enzymes.

1.3.2 GH6 family

Only endoglucanase (EC 3.2.1.4) and cellobiohydrolase (EC 3.2.1.91) activities have been reported for this family. It was previously known as "Cellulase Family B" *Cel6A* (Henrissat *et al.*, 1989) from *Hypocrea jecorina* (erstwhile, *Trichoderma reesei*) is a representative endoglucanase from GH6 family. The 3D structure of GH6 family endoglucanases is made up of α/β barrel which is not standardized like $(\alpha/\beta)_8$-TIM barrel. For example, cellobiohydrolase, CBH II of family GH6 has only seven stranded central β barrel and the active-site is made up of larger open groove (Meinke *et al.*, 1995). The members of GH6 family follow inversion type of catalysis through general acid-base hydrolysis mode.

1.3.3 GH7 family

This is an erstwhile GH-B family of cellulases. Although, the member of GH7 family are known for their cellobiohydrolase activity at reducing end of cellulose (EC 3.2.1.76) unlike EC 3.2.1.9, notable endoglucanase activity (EC 3.2.1.4) is also registered for GH7. Being one of the first ever characterized GH7 endoglucanases, EG I (Cel7B) of fungus *Fusarium oxysporum* is an exemplary model for GH7 endoglucanases (Sulzenbacher *et al.*, 1996; Aranda Soto, 2019)

1.3.4 GH8 family

GH8 endoglucanases have $(\alpha/\alpha)_6$ barrel structures. Previously it was called as GH Clan-M or cellulase family D. Majority of GH8 members are endo-acting GHs. All of the GHs from GH8 are from bacteria and there is no member reported from eukaryote or archaea (Henrissat *et al.*, 1989). Like in GH5, GH8 members also show diverse

enzyme activities such as endoglucanase (EC 3.2.1.4), chitosanase (EC 3.2.1.132), and xylanase (EC 3.2.1.8). GH8 is further subdivided into three subfamilies based on the position of general base. General acid for subfamily GH8a is glutamate, whereas general base is aspartate; both of which act as catalytic residues (Domínguez *et al.*, 1996).

1.3.5 GH9 family

The GH9 family, previously known as "Cellulase Family E" is known for endoglucanases (EC 3.2.1.4) having one or more carbohydrate binding module (CBM) alongwith the catalytic module (Kumar *et al.*, 2019). Some of the GH9 endoglucanases hydrolyze cellulose processively. However, CBM often is essential for GH9 to be functional (Petkun *et al.*, 2015). The catalytic module of the GH5 endoglucanases is rigidly attached to CBM3c at C-terminal end as established from the study of cellulase E4 from *Thermomonospora fusca* (Sakon *et al.*, 1997). At present, all the known plant endoglucanases are part of GH9 family. The GH9 family is further divided into two subgroups, E1 and E2. The subgroup, E1 contains mostly the eubacterial cellulases, whereas E2 consists of endoglucanases from all the life domains like bacteria, plant, fungi, earthworm, etc (Tomme *et al.*, 1995). All the GH9 processive endoglucanases are part of E1 subgroup. Currently, 150 endoglucanases are characterized from GH9 family and 20 enzyme structures have been solved.

1.3.6 GH12 family

GH12 CAZymes are part of Clan C GHs which also has family GH11 glycoside hydrolases. GH12 endoglucanases were previously called family H cellulases. Family GH12 is another GH family alongwith subfamily GH5_4 of family GH5 having xyloglucan specific endoglucanases. They are usually endoglucanases alongwith

possessing xyloglucan-endo-hydrolase/endoglucanase (EC 3.2.1.151) and β-1,3;1,4-mixed linkage hydrolytic activity (Gottschalk *et al.,* 1998). GH12 endoglucanases have β-jelly roll as the secondary structure. Both the catalytic residues, general acid and general base are glutamate for family GH12.

1.3.7 GH44 family

This is the erstwhile cellulase family J of glycoside hydrolases. Member of GH44 family are broad specificity endoglucanases acting on lichenan, xylan, CMC-Na and to some extent, on avicel. However, these endoglucanases as well have prime activity on xyloglucan (Najmudin *et al.,* 2006). What makes special to GH44 endoglucanases is their higher rate of hydrolysis on tetra-saccharides, cellooligosaccharides like cellohexaose, longer cello-oligomers. Comparatively, they have low reaction rate on shorter cellooligosaccharides. When acting on cellohexaose, GH14 endoglucanases tend to generate more of cellotetraose and cellobiose rather that two cellotriose molecules. Moreover, cellotetraose proportion is higher than cellobiose during the product accumulation. This is thus the processive behavior of GH44 endoglucanases, a special mechanism of hydrolysis of endoglucanases as discussed later in the chapter. GH44 endoglucanases show TIM-barrel like domain and a β-sandwich domain as their secondary structure. Ca^{2+} and Zn^{2+} have been found as bound metal ligand at the active-site of Cel44a endoglucanase from *Clostridium thermocellum* (Tomme *et al.,* 1995; Kitago *et al.,* 2007).

1.3.8 GH45 family

This was historically known as cellulase family K and is part of Clan A of CAZymes. GH45 endoglucanases (EC 3.2.1.4) predominantly hydrolyze soluble β-1,4-glucans (Schülein *et al.,* 1998). They share structural similarity with plant and bacterial

expansins. Most of industrial textile-detergent industries utilize GH45 endoglucanases for fiber modification. The secondary structure of GH45 endoglucanases is composed of 6-stranded β-barrel which is allied by one more β-strand.

1.3.9 GH48 family

This is the most important endoglucanase family as far as the cellulosomes are concerned. Every bacterial or fungal cellulosomal endoglucanase is grouped under GH48 family. This family is part of cellulase clan M cellulases. Major substrate for GH48 endoglucanase is amorphous or crystalline cellulose which are preferred by them over soluble glucan like CMC-Na. GH48 show notable broad substrate specificity which is regarded crucial for synergy during the hydrolysis of natural cellulose. GH48 family contains processive as well as random β-1,4-endoglucanase, cellobiohydrolase and chitinase activities. The activity of GH48 endoglucanases is heavily hampered by the presence of cellobiose (Parsiegla *et al.,* 2008). The activity of processive endoglucanase, *CelS/S8* of *Clostridium thermocellum* was found stabilized by Ca^{2+} ions. Glutamate is general acid whereas aspartate (and less often, tyrosine) is the nucleophile (general base), deprotonating the water molecule which attacks the anomeric carbon. GH48 follow $(\alpha/\alpha)_6$-barrel secondary structure topology. *Cel48F* from *Clostridium cellulolyticum* is another processive endoglucanase from family GH48. Aspartate as a nucleophile for GH48 family endoglucanases was first experimentally established through crystal structure analysis of *Cel48A* of bacterium *Thermobifida fusca*, respectively (Kostylev and Wilson, 2011).

1.4 Synergism of endo-β-1,4-glucanase with exoglucanase and β-glucosidase

The most important characteristic of cellulase enzymes is their synergy (**Fig. 1.2**), where product of one enzyme is the substrate for subsequent action by the other enzyme

(Bhat and Bhat, 1997). It is a phenomenon that makes the otherwise difficult and energy consuming process of deconstruction of recalcitrant cellulose, rapid. The cellulose chain is polymorphic having certain regions highly ordered and crystalline whereas other regions are amorphous and comparatively disordered. Endoglucanase prefers the amorphous region for hydrolysis converting the cellulose into cellooligosaccharides. Cellobiohydrolase I (CBH I) or exoglucanase cleaves these amorphous region cellooligosaccharides into cellobiose from the reducing end (Ganner *et al.,* 2012). Endoglucanase and CBH I are thus specific for amorphous cellulose region. Contrary to this, cellobiohydrolase II (CBH II) shows preference for free reducing ends of cellooligosaccharides from crystalline region. However, the action of endoglucanase and CBH I on nearby amorphous region is a prerequisite for the functioning of CBH II. The crystalline cellulose region is made accessible to CBHII by the action of endoglucanase and CBH I through their actions on the amorphous cellulose. CBH II can thus show pronounced activity on crystalline cellulose region with the help of endoglucanase. The produced cellooligosaccharides by endoglucanase and cellobiose produced by CBH I and CBH II display half part of synergism. The synergy of cellulases is accomplished by the biocatalyst, the β glucosidase (BGL) (Rani *et al.,* 2014). The cellobiose produced by CBH acts as the substrate for BGL. The cellobiose is hydrolyzed into two glucose monomers by BGL. Glucose is the final product of lignocellulosic deconstruction process which can then be converted to bioethanol through fermentation process.

1.5 Endo-β-1,4-glucanase producing microorganisms

A variety of microorganisms take part in cellulose hydrolysis with an aid of a cellulase system (**Fig. 1.5**). The best-characterized fungal cellulase systems are,

Phanerochaete chrysosporium, Fusarium solani, Penicillum funiculosum, Talaromyces emersonii, Trichoderma koningii and *Trichoderma reesei* (Sajith *et al.*, 2014). Some of the aerobic cellulolytic bacteria having cellulase systems are *Cellulomonas* spp., *Cellvibrio* spp., *Microbispora bispora* and *Thermomonospora* spp. (**Fig. 1.5**). There are efficient anaerobic cellulolytic bacteria such as: *Acetivibrio cellulolyticus, Bacteroides cellulosolvens, Bacteriodes succinogenes, Clostridium thermocellum, Ruminococcus albus* and *Ruminococcus flavefaciens* (Kuhad *et al.*, 2006). Fungi are the prominent cellulase producers which hydrolyze around 80% of the total cellulose present on earth (Sajith *et al.*, 2014). Fungi have conventionally been considered the main cellulase producing microorganisms, though this belief has been heavily broken by studies on bacteria and actinomycetes which have also shown remarkably efficient endoglucanases (**Table 1.1**). Major fungal genera harboring cellulases are *Trichoderma, Aspergillus, Fusarium, Penicillium, Selerotium* and *Cladosporium* of which, the genera such as *Trichoderma* and *Aspergillus* are well explored for industrial use for cellulases. The most popular fungus for industrial cellulase production is *Trichoderma reesei*. Its cellulase systems such as *EGIII (Cel5A)* are even capable of hydrolyzing native and crystalline cellulose. Fungal genera, *Trichoderma* and *Aspergillus* are known to be efficient cellulase producers and the crude enzymes produced by these microorganisms are available commercially (**Table 1.2**).

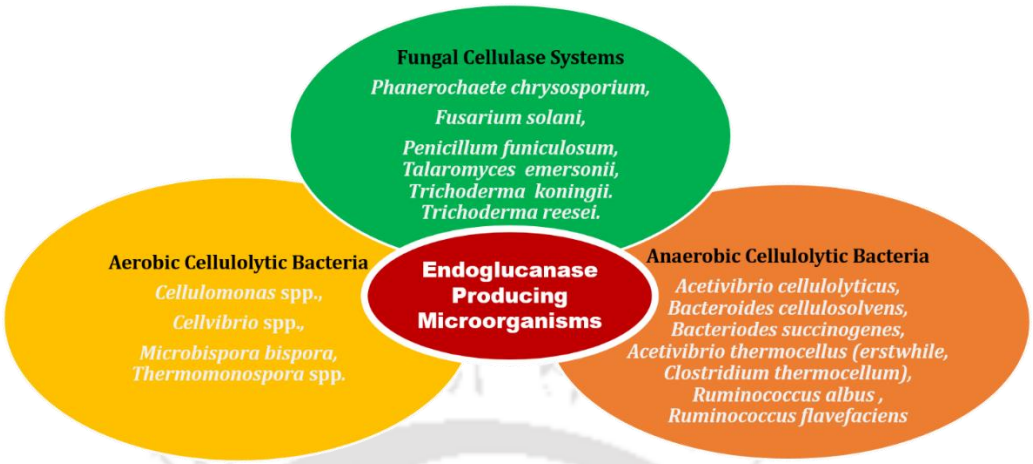

Fig. 1.5 Some of the microorganisms producing endoglucanase.

1.6 Biochemical properties, kinetics and catalytic efficiency of endoglucanases

Optimal temperature and pH is crucial for the endoglucanases to hydrolyze the cellulosic substrate. Important endoglucanases from different microbes with their biochemical and kinetic properties are listed in **Table 1.1**. It can be observed from the **Table 1.1** that, endoglucanases prefer acidic environment for cellulose hydrolysis having the optimum pH between 4.0 and 5.5. The reason behind acidic pH favored by endoglucanases is that the protonic environment help stabilize the 3D structure which allows the catalytic site (often the acidic catalytic amino acid residues) to hydrolyze the cellulose polymer efficiently.

Optimum temperature of endoglucanases varies and depend upon the source. Mesophilic sources give endoglucanases with optimal temperature near or below 50°C, whereas endoglucanases from thermophilic bacteria and fungi normally efficiently hydrolyze glucan at higher temperature ranging between 60-90°C (**Table 1.1**). However, thermophilic endoglucanases active at higher temperature are nowadays preferred in biomass conversion industry for their cost-effectiveness. Higher optimum temperature of endoglucanases increases the efficiency of biomass conversion process

Table 1.1. Kinetic and biochemical properties of various endoglucanases.

Enzyme	Substrate	Organism	Optima	Kinetics	Reference
Endoglucanase, *Egst*	CMC-Na, Medium Visc.	*Scytalidium thermophilum*	pH 5.5, 60°C, 66 kDa	*SA* – 246 K_m – 10.5 K_{Cat}/K_m – 45.2	Meleiro *et al.*, 2018
β-1,4-endoglucanase, *rMt-egl*	CMC-Na	*Myceliophthora thermophila BJA*	pH 10, 50°C, 47 kDa	*SA* – 119.4 K_m – 5 K_{Cat}/K_m – 204	Phadtare *et al.*, 2017
β-1,4-1,3-endoglucanase, *At*GH12, (GH12)	Barley β-glucan	*Aspergillus terreus*	pH 5.0, 55°C, 24 kDa	*SA* – 256.14 K_m – 2.21	Segato *et al.*, 2017
β-1,4-endoglucanase, *TaEGII*	CMC-Na	*Trichoderma atroviride*	pH 5.0, 60°C, 44.23 kDa	*SA* – 119.4 K_m – 0.014 K_{Cat}/K_m – 6.50	Huang *et al.*, 2017
β-1,4-endoglucanase, *PvCel5A*	β-glucan	*Penicillium verruculosum*	pH 4.5, 40°C	*SA* – 60 K_m - 20 K_{Cat}/K_m – 8.25	Dos Santos *et al.*, 2015
Bifunctional Cellulase, *CtCel5E*, (GH5)	CMC-Na	*Clostridium thermocellum*	pH 5.0, 50°C	*SA* – 736 K_m – 2.1 K_{Cat}/K_m – 12.40	Yuan *et al.*, 2015
β-1,4-endoglucanase, *EG2*	β-glucan	*Penicillium verruculosum*	pH 5.0, 50°C, 36.2 kDa	*SA* – 75 K_m – 22.8 K_{Cat}/K_m – 6.6	Merzlov *et al.*, 2015
β-1,4-endoglucanase, *TeEgl5A*, (GH5)	CMC-Na	*Talaromyces emersonii* CBS394.64	pH 4.5, 90°C, 37 kDa	*SA* - 624 K_m – 31.7 K_{Cat}/K_m – 51.6	Wang *et al.*, 2014
β-1,4-endoglucanase, *TmCel12A* (GH12)	β-glucan	*Thermotoga maritima*	Wild type, pH 5.0, 85°C	*SA* – NA K_m - 2.08 K_{Cat}/K_m – 402	Cheng *et al.*, 2012
Carboxymethyl cellulase (CMCase)	CMC-Na	*Aspergillus oryzae cmc-1*	pH 4.4, 55°C, 25 kDa	*SA* – 92 K_m – 20 K_{Cat}/K_m – 17.8	Javed *et al.*, 2009
β-1,4-endoglucanase, *Egl499*	CMC-Na	*Bacillus subtilis* JA18	pH 5.8, 60°C, Wild type	*SA* – 284 K_m – 11.4 K_{Cat}/K_m-21.6	Wang *et al.*, 2009
Endoglucanase, *CbEG*	CMC-Na	*Cellulomonas biazotea*	pH 5.5, 40°C	*SA* – 217 K_m – 3.5 K_{Cat}/K_m – 62	Rajoka *et al.*, 2004
Endoglucanase, *EG III*	CMC-Na	*Trichoderma reesei*	pH 5.0, 30°C, 48 kDa	*SA* – 351 K_m – 5 K_{Cat}/K_m – 7.33	Macarron *et al.*, 1993

SA = Specific Activity (U/mg), K_m = mg.mL^{-1}, K_{Cat}/K_m = mL.mg^{-1}.s^{-1}, CMC-Na = Carboxymethyl Cellulose NA = Not Available.

in the renewable energy sector as they remain thermostable during the longer duration of saccharification (Nieves *et al.*, 1997). Some of the important commercially available cellulases are listed in **Table 1.2** which highlights the industrial and global demand for endo-β-1,4-glucanases. Thermostability of endoglucanase at the reaction temperature is an important factor given its role in time-consuming cellulose conversion processes. A good endoglucanase should be stable in the temperature range of 30-35℃ for long durations if it is prospected for simultaneous saccharification and fermentation process (SSF). Because, fungi ferment the produced glucose in the temperature range, 30- 35℃. The saccharification and fermentation reaction runs from several days. Therefore, endoglucanase should be thermally stable in this temperature range as well as at its optimum temperature.

Endoglucanases from thermophilic microorganisms are thermostable and help reduce the cost of pretreatment which are also preferred at industrial level processes. It was reported that arginine, histidine, glutamate and methionine are the important residues, which impart thermostability to thermophilic endoglucanases, whereas glutamine and serine are important for stability of mesophilic endoglucanases (Yennamalli *et al.*, 2011). Moreover, ionic interaction of the enzyme in solution contribute towards the thermostability of endoglucanases. Some metal ions such as Ca^{2+} also stabilize the endoglucanases at higher temperature as observed for endoglucanase, *CelS/S8* of GH48 from *Clostridium thermocellum* (Parsiegla *et al.*, 2008).

The efficiency of endoglucanase is evaluated by calculating the catalytic efficiency (k_{cat}/K_m) and minimum substrate concentration to achieve half of the maximum reaction rate (K_m) and maximum achievable rate of the reaction (V_{max}). The fungal endoglucanases so far characterized are higher in number. This is due to their

Table 1.2. Commercially available endoglucanases.

Enzyme	Substrate	Organism	Optima	*SA*	Manufacturer
PRO-E0021, (β-1,3:1,4-endoglucanase, GH5)	β-glucan	*Clostridium thermocellum NCIB 10682*	pH 7.0, 60°C	900	Prozomix Ltd., UK
E-CELTM, (β-1,4-endoglucanase,GH5)	CMC-Na	*Thermotoga maritima*	pH 6.0, 80°C	245	Megazyme Ltd., Ireland
E-LICHN (β-1,3:1,4-endoglucanase, GH16)	Barley β-glucan	*Bacillus subtilis*	pH 6.5, 60°C	230	Megazyme Ltd., Ireland
E-CELAN (Endo-cellulase, GH12)	CMC-Na	*Aspergillus niger*	pH 4.5, 40°C,	80	Megazyme Ltd., Ireland
E-CELTR (β-1,4-endoglucanase, GH7)	CMC-Na	*Trichoderma longibrachiatum*	pH 4.5, 70°C	70	Megazyme Ltd., Ireland
E-CELTE (β-1,4-endoglucanase, GH5)	CMC-Na	*Talaromyces emersonii*	pH 4.5, 80°C	65	Megazyme Ltd., Ireland
E-CELBA (β-1,4-endoglucanase, GH5)	CMC-Na	*Bacillus amyloliquefaciens*	pH 6.0, 40°C	60	Megazyme Ltd., Ireland
Spezyme#3 (β-1,4-endoglucanase)	CMC-Na	*Trichoderma reesei*	pH 4.0, 50°C	25	Genencor Intl., USie
Cellulase AP30K (β-1,4-endoglucanase)	CMC-Na	*Trichoderma reesei*	pH 4.5, 60°C	21	Amano Enzyme, USA
E-CELTH (β-1,4-endoglucanase, GH6)	CMC-Na	*Thermobifida halotolerens*	pH 8.5, 80°C	16	Megazyme Ltd., Ireland
ACCELLERASE 1500 (β-1,4 endoglucanase)	CMC-Na	*Trichoderma reesei*	pH 4.8, 50°C	2.8	DuPont De Nemours, Inc. USA
C2730-Celluclast (β-1,4-endoglucanase)	Bacterial Cellulose	*Trichoderma reesei*	pH 6, 52°C	0.7	Sigma-Aldrich Corp. / Novozyme Corp., USA
C1184-Cellulase (β-1,4-endoglucanase)	Cellulose	*Aspergillus niger*	pH 5, 37°C	0.3	Sigma-Aldrich Corp. / Novozyme Corp., USA
CZ0381-Cellulase 5B (β-1,4-endoglucanase, GH5)	β-glucan, Xyloglucan	*Caldicellulosiruptor saccharolyticus*	pH 6.5, 80°C	NA	NZYTech Ltd., Portugal
CZ0218-Cellulase 12A, (β-1,4-endoglucanase, GH12)	Barley β-glucan, CMC-Na	*Thermotoga maritima*	pH 6.0, 85°C	NA	NZYTech Ltd., Portugal
Cellulase 6A-CZ0294, (β-1,4-endoglucanase, GH6)	Barley β-glucan, CMC-Na	*Podospora anserina*	pH 5, 40°C	NA	NZYTech Ltd., Portugal

SA = Specific activity (U/mg), NA- Not Available.

believed robustness and higher performance. However, many bacterial cellulolytic enzymes have also been studied. More than 700 endoglucanases (EC 3.2.1.4) from

various microorganisms have been reported in the BRENDA database (https://www.brenda-enzymes.org), the internationally recognized comprehensive enzyme information system. It can be observed from **Table 1.1**, that the catalytic efficiency (k_{cat}/K_m) of endoglucanase, *Egst* from *Scytalidium thermophilum*, *EG III* from *Trichoderma reesei*, *CtCel5E* from *Clostridium thermocellum* and *Egl499* from *Bacillus subtilis* JA18 are significantly high, which categorize them as highly efficient endoglucanases.

1.7 Structure of endo-β-1,4-glucanases

Advancements in the X-ray crystallographic (X-Ray Diffraction, XRD) and NMR techniques have led to unravelling of 3D structures of cellulases from various organisms (Lombard *et al.*, 2014). The catalytic module of GH5 endoglucanases fold in $(\beta/\alpha)_8$ TIM-barrel motif (Ducros *et al.*, 1995; Dos Santos *et al.*, 2015). The active-site of endoglucanase contains an open groove unlike the tunnel-like structure of cellobiohydrolases (Segato *et al.*, 2014). The open groove can accommodate the amorphous cellulose chain in a random manner, whereas the tunnel of cellobiohydrolases allow only reducing or non-reducing ends of the cellulose polymer. Endoglucanases from family GH5, GH6, GH7, GH8, GH9, GH12, GH44, GH45, GH48, GH51, GH61 and GH124 show the open groove at their active-site. However, endoglucanases are present in various GH families and thus differ in their structure. GH5 endoglucanases contain compact 8-fold $(\beta/\alpha)_8$ barrel as the core of the 3D-structure. Two acidic residues, often the glutamates, act as the catalytic residues. One acts as proton donor whereas the other one behaves as a nucleophile (Segato *et al.*, 2014).

Apart from two catalytic residues, other substrate binding sites present at the active-site groove also play important role in the catalysis. The subsites -4 to +3 rich in Tryptophan residues were observed in the endoglucanase, *Cel5A* of *Thermoascus aurantiacus* (Leggio and Larsen, 2002). GH7 and GH12 families on the other hand represent endoglucanases having the topology of β-jelly roll. However, all of them form an extended open groove and cellulose chain is accessed randomly by the catalytic residues thus cleaving the β-1,4 glycosidic bond and generating the cellooligosaccharides. The open cleft active-site of GH9 endoglucanase can have at least 6 carbohydrate binding sites from -4 to +2 at their active-site pocket (Sakon *et al.*, 1997). The CBM3c of processive GH9 cellulases is aligned with the active-site cleft. **Fig. 1.6A-a** shows the 3D structure for *EGII*/*Cel5A* of *Thermoascus aurantiacus* from GH5 family $(\beta/\alpha)_8$ barrel) whereas **Fig. 1.6B-b** shows *EGIII/Cel12A* (β-jelly-roll) from *Trichoderma reesei*. These enzymes are example of divergent evolution where they have different structures but perform same catalytic function.

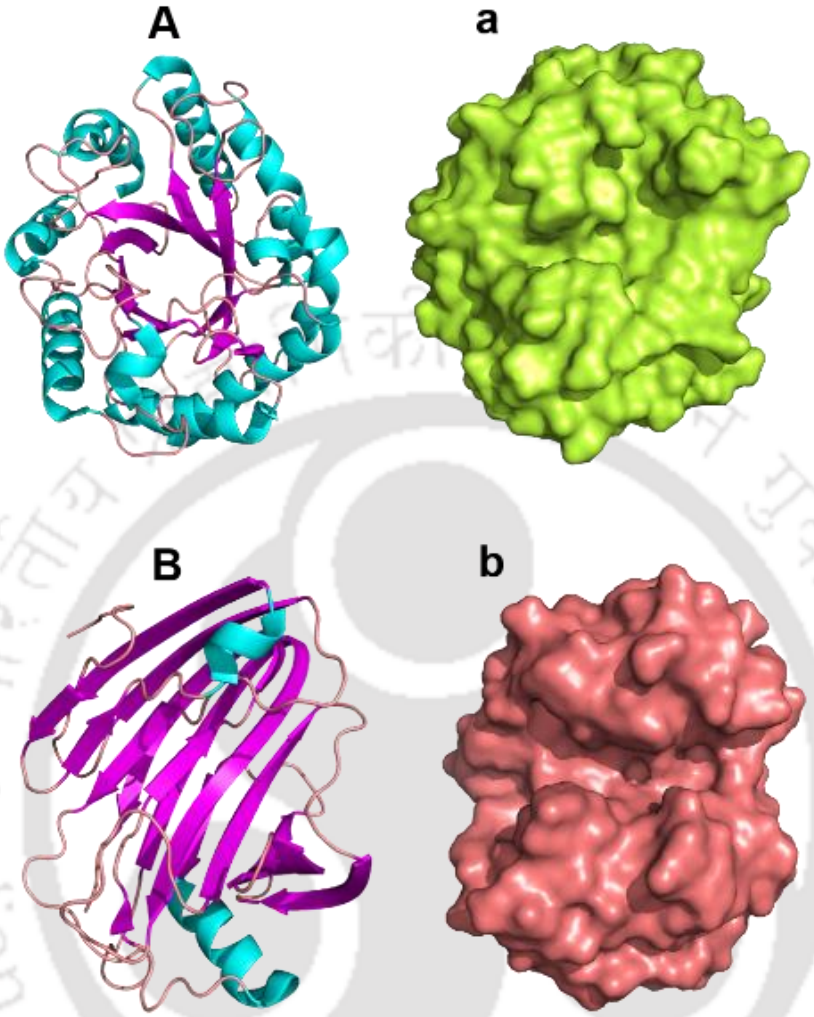

Fig. 1.6 3D structure of endoglucanases from GH families showing secondary structures (A, B) and surface view (a, b). Aa) Endoglucanase, *EGII*/*Cel5A* (PDB ID: 1GZJ) of *Thermoascus aurantiacus* from GH5 family $(\beta/\alpha)_8$ barrel), Bb) Endoglucanase, *EGIII/ Cel12A* (PDB ID: 1OA2) (β-jelly-roll) from *Trichoderma reesei.*

1.7.1 Mechanism of cellulose hydrolysis in endoglucanases

The essential function of endoglucanase is the breaking of the β-1,4-D-glycosidic bond. There are two mechanisms for cleavage of the glycosidic bond, either retaining or inverting (Koshland Jr, 1953; Sinnott, 1990). The catalytic amino acids present at the active-site of the endoglucanase perform one or two nucleophilic substitutions at the anomeric carbon of the D-glucose (Henrissat, 1994). In the retaining version of

catalysis, the configuration of anomeric β-carbon is conserved whereas, it is altered to α-carbon in the inverting type. Thus, in inverting type, the released monomer (D-glucose) now has the α-anomeric carbon unlike its polymeric chain cellulose which, by default contains β-anomeric glucose.

Moreover, the retaining mechanism takes to trans-glycosylation whereas, the inverting mechanism always leads to complete hydrolysis of the β-glycosidic linkages (Sinnott, 1990). The distance between the two catalytic residues is more than 5.5 Å in inverting type of hydrolysis whereas, it is below 5.5 Å in retaining type. Members of family GH5 cleave β-glycosidic bonds by retaining mechanism through acid-base hydrolysis in which, two strictly conserved carboxylates (usually, two glutamate residues) located at the C-terminal of β strands 4 and 7, catalyze the reaction (Davies and Henrissat, 1995). In other words, both the catalytic residues are in the close vicinity of each other (distance ≤ 5.5 Å). The GH9 endoglucanase follow the inverting type mechanism of hydrolysis as elucidated for *Cel9A* endoglucanase from *Thermobifida fusca* (Li *et al.*, 2007). *Cel9A* of GH9 family has catalytic residues glutamate and two aspartates. The reaction takes place as general acid-base catalysis, where glutamate acts as the general acid and one of the aspartate residue acts as a general base. The second aspartate of the catalytic triad is involved in binding the water molecule.

1.8 Multifunctionality of endoglucanases

Endoglucanases are predominantly involved in leaving the cellulose polymer into oligomers, called cellooligosaccharides. Irrespective of the source, endoglucanases are considered as the biocatalysts for hydrolyzing β-1,4 linkages of cellulose. However, recent studies have established that some endoglucanases can even hydrolyze certain

hemicellulosic linkages. Such endoglucanases are recognized as multifunctional endoglucanases. The other linkages which multifunctional endoglucanases can break are, mixed β-1,4;1,3 linkages of β-D-Glucan and β-1,4 cellulose chain substituted with xylose monomers i.e., xyloglucan, glucomannan having main linkage of β-1,4 D-mannose monomers with occasional presence of D-glucose in the main chain, xylan having β-1,4 linked D-xylose monomers and the β-1,4 D-mannose linkage of galactomannan (Aspeborg *et al.*, 2012).

Most of the Multifunctional endoglucanases have been classified under subfamily 4 of GH5 family in CAZy database. The Subfamily 4 (erstwhile, A3+A4 clan) is the set of polyspecific endo-β-1,4-glucanases. So far, four enzyme activities have been reported for subfamily GH5_4. They hydrolyze soluble glycan like CMC-Na, 1,3:1,4-mixed glucan like β-glucan and xyloglucan (Aspeborg et al. 2012). This is the single subfamily of GH5 which hydrolyzes xyloglucan (EC 3.2.1.151) (Yaoi *et al.*, 2005). So, these endoglucanases can be attributed as xyloglucan-specific / hydrolyzing endo-β-1,4-glucanase or endo-xyloglucanase (EC 3.2.1.4 and EC 3.2.1.151), licheninase (EC 3.2.1.73), and less often xylanase (EC 3.2.1.8). The Subfamily GH5_4 members are typically extracellular bacterial enzymes. Although, some members of subfamily GH5_4 come from ciliates and fungi multifunctional endoglucanases are predominantly from the ruminophilic microorganisms. The licheninase enzyme activity (EC 3.2.1.73) has also been described in this subfamily 4 of GH5 family (Foong and Doi, 1992).

1.8.1 Broad substrate specificity of various endoglucanases

Various multifunctional endoglucanases have been reported for their broad substrate specificity. A broad specificity endoglucanase of family GH12, *BLXG12* from

Bacillus licheniformis showed hydrolysis of tamarind xyloglucan (53 $U.mol^{-1}$ where, U = 1 mole of glucose. min^{-1}), konjac glucomannan (10 $U.mol^{-1}$), CMC-Na (21 $U.mol^{-1}$). *BLXG12* also showed trace amount of activity on Avicel and curdlan polymer (β-1,3-D-glucose linkage) from *Alcaligenes faecalis* (Gloster *et al.*, 2007). Recently, multifunctional endoglucanases, *EG5C* and *EG5C-1* from *Bacillus subtilis* BS-5 was reported to hydrolyze CMC-Na, β-D-Glucan, PASC, *p*NPC, Avicel and Filter Paper (Wu *et al.*, 2018). The k_{cat}/K_m of *EG5C* and *EG5C-1* against CMC-Na and PASC was recorded to be 38 and 1.6 $mL^{-1}.mg^{-1}.s^{-1}$, respectively. Endoglucanases, *EngB* and *EngD* from *Clostridium cellulovorans* also showed multifunctionality which hydrolyzed pure micro granular cellulose (51 and 37 U/mg, respectively) and microcrystalline avicel (20 and 16 U/mg, respectively) (Foong and Doi, 1992). Activity of *EngB* and *EngD* was also reported on CMC-Na (82 and 86 U/mg, respectively) and lichenan (853 and 555 U/mg, respectively). The multifunctionality of an enzyme is not a very common property of endoglucanases. Aromatic residues like Tryptophan near the active-site protruding outwardly are believed to be responsible for substrate binding, substrate selectivity and multifunctionality of endoglucanases (Glasgow *et al.*, 2020). Certain endoglucanases are of bi-functional nature, only showing major activity alongwith an additional accessary activity. An alkaline bifunctional endoglucanase, *Pgl5A* of subfamily GH5_4 from *Paenibacillus* sp. S09 showed 94.5 U/mg of activity against xyloglucan and a negligible activity (0.5 U/mg) against CMC-Na (Cheng *et al.*, 2019). Broad substrate specificity is uncommon in GH9 family. Nonetheless, a glycoside hydrolase from GH9 family from rice was reported to show activity on cellulose as well as hemicellulose (mixed-β-1,3;1,4-D-glucan, arabinoxylan, and glucomannan) of plant cell wall (Yoshida and Komae, 2006).

1.8.2 Importance of multifunctional endoglucanases

Multifunctional endoglucanases are of great importance for various industrial applications. Their capacity to deconstruct different hemicellulosic polymers reduces the need of individual hemicellulases in the renewable energy sector. A single endoglucanase can act as β-1,4-glucano hydrolase, β-1,3;1,4-D-glucano hydrolase, lichenase (or licheninase), xyloglucanase, xylanase, glucomannanase and galactomannanase as observed for endoglucanase, *CtCel9D-Cel44A* (erstwhile, endoglucanase *CelJ*) from *Clostridium thermocellum* (Najmudin *et al.*, 2006). Moreover, the generation of various hemicellulosic oligosaccharides by these multifunctional endoglucanases could be prospected for feed, food and pharmaceutical industry.

1.9 Processivity of endoglucanases

Endoglucanase performs random cleavage of β-1,4 D-glycosidic linkage of cellulose. So, oligosaccharides of various degrees of polymerization (DP) are generated during the cellulose hydrolysis. This essentially means that endoglucanase binds the cellulose chain randomly to cleave it off. The initial products released from cellulose are often mixture of various oligosaccharides of higher degree of polymerization (DP) like cellohexaose (DP6), cellopentaose (DP5) or cellotetraose (DP4). However, certain endoglucanases show a characteristic binding to cellulose chain where they processively (or sequentially) generate the similar type of cellooligosaccharides consistently (Zheng and Ding, 2013). These endoglucanases are called as processive endoglucanase. Processive endoglucanases consistently release either only cellotetraose or only cellobiose or cellobiose-cellotriose together from the cellulose polymer chain. This is called processivity of endoglucanases. Processivity is usually

the property of cellobiohydrolases (CBH). So, endoglucanases showing processivity thus become noteworthy. Processive endoglucanase, *CcCel9A* from *Clostridium cellulosi* of family GH9 was demonstrated to consistently release cellotetraose (**Fig. 1.7**) from cellulose (Zhang *et al.*, 2018). *EG1* from fungus *Volvariella volvacea* was recorded to be a processive endoglucanase as it showed the release of cellotriose and cellobiose from PASC (Zheng and Ding, 2013). *EG5C-1* of *Bacillus subtilis* BS-5 is also a processive endoglucanase which releases cellotriose and cellobiose, consistently (Wu *et al.*, 2018). Open groove catalytic cleft with a barrier loop and aromatic amino acids located at the entrance of the catalytic cleft have been demonstrated for the processive behavior of the endoglucanases such as *EG5C-1* (F238) of *B. subtilis* BS-5. Cellobiose producing processive endoglucanases are important for the biomass conversion industry as they reduce the need of cellobiohydrolase enzyme during the saccharification thereby proving as a multifunctional yet cost-effective endoglucanase.

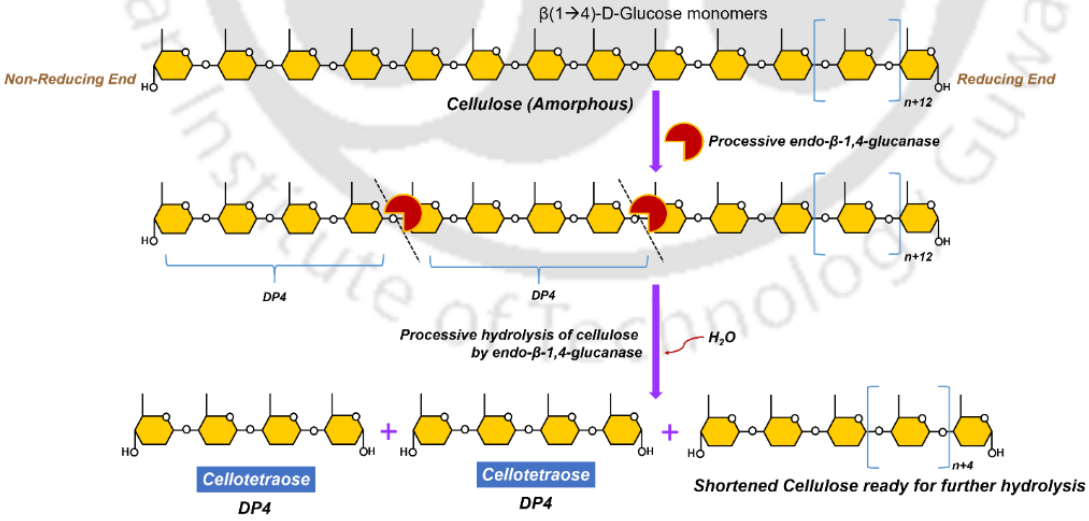

Fig. 1.7 Deconstruction of cellulose by processive endo-β-1,4-glucanase through consistent release of cellotetraose.

1.10 Applications of endoglucanases

The most important use of endoglucanases is in the renewable energy sector to produce lignocellulosic bioethanol. However, they have some major applications in other industries as well (**Fig. 1.8**). Many of the endoglucanases are multifunctional as discussed in **Section 1.8**. This imparts a special attribute to these enzymes thereby increasing their industrial importance. While performing various functions like xyloglucanase, licheninase, glucanase and xylanase; canvas of applications for these endoglucanases broadens significantly (Venditto *et al.*, 2016). An endoglucanase with xyloglucanase function can be sought for preparation of cellulose fiber modification. Moreover, carbohydrate-based surfactants synthesis could be carried out with the xyloglucan specific endoglucanases.

Most recently, endoglucanases are being employed for conversion of plant biomass into chemicals and biofuels such as bioethanol (Dotsenko *et al.*, 2016). In paper and pulp industry, addition of xyloglucanase treated xyloglucan decreases surface friction,

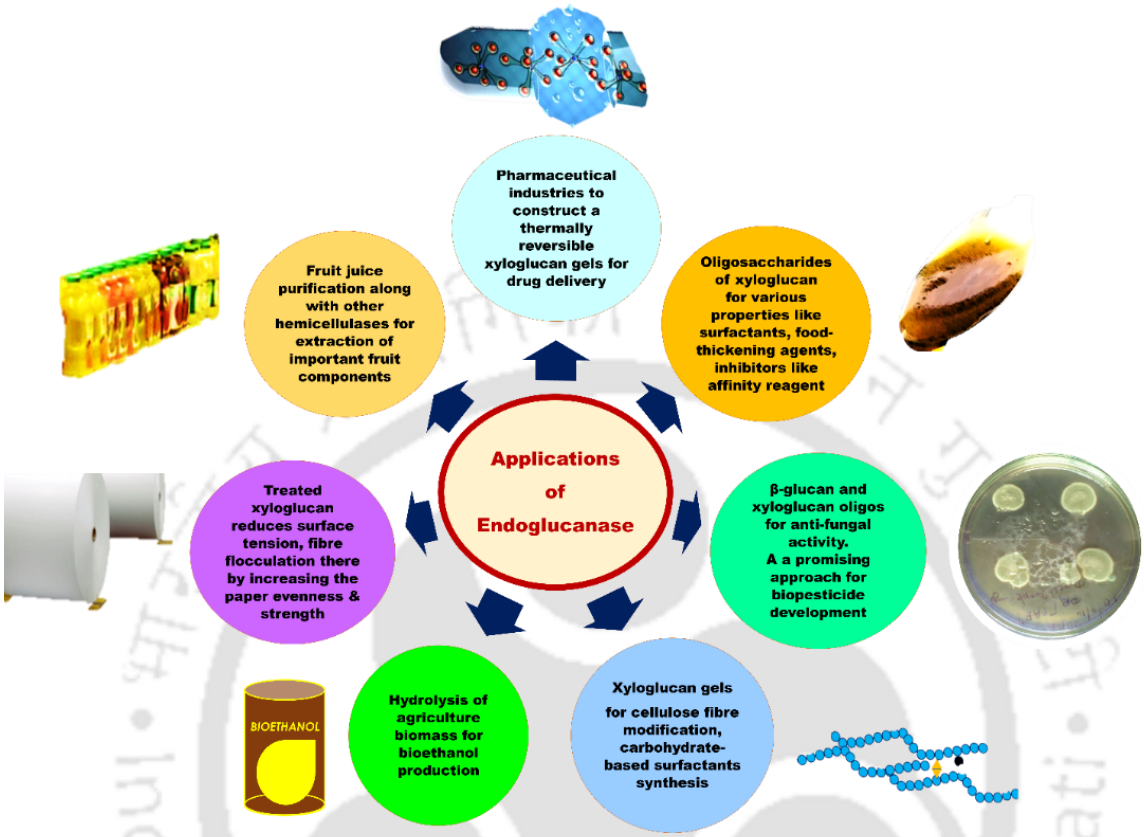

Fig. 1.8 Applications of endoglucanases in biofuel production, paper mill, fruit juice, pharmaceutical and food industry.

fiber flocculation thereby increasing the paper evenness and strength (Oksanen *et al.*, 2011). Endoglucanases also have applications in fruit juice purification alongwith other hemicellulases used for extraction of important fruit components and to reduce the processing tenure (Vincken *et al.*, 1996; Bhat, 2000). Besides, endoxyloglucanases are employed for brewing and alteration of xyloglucan in food and feed industry. Endoxyloglucanases are also being explored for their usage in pharmaceutical industries to construct a thermally reversible xyloglucan gels for the purpose of drug delivery (Miyazaki *et al.*, 1998). Combination of β-glucan and xyloglucan, generated oligosaccharides by using xyloglucanases has been reported to show anti-fungal activity, which is a promising approach for biopesticide development (Sidrah *et al.*,

2014). Oligosaccharides produced from xyloglucan using endoxyloglucanases have been observed for various properties like surfactants, food-thickening agents, inhibitors like affinity reagents (Rashmi and Siddalingamurthy, 2018).

Another emerging application of endoglucanases is in the field of biorefineries (Annamalai *et al.*, 2016). The cost of lignocellulosic bioethanol (LBE) production is quite high given the extensive steps involved in the pretreatment of lignocellulosic biomass. In order to make the LBE production cost-effective, bio-refinery approach is well versed nowadays. Along with LBE, different commercially important products such as Microfibrillated cellulose (MFC), Glutamic acid, 1;2-Butanediol, Itaconic acid, and xylitol are generated from lignocellulosic biomass in a bio-refinery. This way the overall cost of LBE production can be drastically reduced. Endoglucanases play the crucial role in a bio-refinery. They deconstruct the pretreated agricultural biomass into cellooligosaccharide which is thus available for the downstream processing to produce LBE. Moreover, hydrolysis of cellulose in the biomass loosens the overall hemicellulose present in the cell wall which is thus readily available for other hemicellulase enzymes for hydrolysis. Endoglucanase are thus important for a bio-refinery.

Thus, endoglucanase plays a pivotal role in the hydrolysis of the cellulose polymer. As the world is heading towards energy crisis, the application of endoglucanases in renewable energy sector is pertinent. The great diversity of endoglucanases led to their classification in various GH families as clubbed under CAZy database based on their sequence similarity. Family GH5, GH8, GH9 and GH12 harbors most of the endoglucanases. Endoglucanases follow divergent evolution where they perform similar function although representing different structures.

Endoglucanases are found in all the three domains of life like bacteria, archaea, fungi and plants. Acidic environment is generally preferred by endoglucanases for the catalysis of β-1,4 linkage of cellulose. Higher catalytic efficiency of endoglucanases is found in *Clostridium thermocellum* and *Trichoderma reesei*. Endoglucanase structure can be of $(\beta/\alpha)_8$-TIM barrel, β-jelly roll or $(\alpha/\alpha)_6$ barrel. Usually, endoglucanases follow general acid-base catalysis with the two glutamate residues acting as the catalytic residues at the active-site. Endoglucanases are often multifunctional as they show broad substrate specificity towards hemicellulosic substrates along with cellulose. Some of the GH5, whereas all the GH9 endoglucanases hydrolyze the cellulose processively thereby generating cellobiose along with other cellooligosaccharides. This property of endoglucanases reduces the need of cellobiohydrolase during saccharification, thus making the biomass conversion process cost-effective. Endoglucanases are applied in bioethanol production, paper mill and food processing industries.

1.11 Microorganism of interest: *Ruminococcus flavefaciens* FD-1 v3

The rumen of mammalian herbivores is among the most abundant sources of cellulolytic microorganisms. The plant cell wall is a major source of energy for ruminant herbivores and its degradation in the rumen is largely dependent upon the activity of three species of cellulolytic bacteria- *Ruminococcus flavefaciens*, *Ruminococcus albus* and *Fibrobacter succinogenes*. It is generally agreed that ruminal cellulolysis is carried out primarily by these three species of bacteria (Dehority, 1993; Hungate, 2013). Ruminants can utilize cellulose, which constitutes their major source of carbon, once it has been hydrolysed and fermented by rumen bacteria. Cellulosic biomass is hydrolysed in the rumen by a complex of cellulolytic enzymes which is made up of three components: endo-β-1,4-glucanase, exo-β-1,4-glucanase, and β-

glucosidase. The endoglucanase is able to degrade soluble cellulose, whereas the exoglucanase can degrade insoluble cellulose. *R. flavefaciens*, an anaerobic, mesophilic & Gram-positive bacterium (**Fig. 1.9**), is one of the most important species of rumen bacteria and is capable of degrading plant cell walls (Pettipher and Latham, 1979a). Taxonomical classification of *R. flavefaciens* was described by (Liu *et al.*, 2008) has been summarized in **Table 1.3**. *R. flavefaciens* degrades crystalline cellulose more efficiently than *R. albus*. The rumen cellulolytic bacterium *R. flavefaciens* adhere immediately and firmly to fibrous plant particles and degrade grass and straw cell-wall polysaccharides faster than the other ruminal cellulolytic species (Latham *et al.*, 1978).

In an invited review, researchers have concluded that at least two mechanisms, cellulosome-like complexes and carbohydrate epitopes of the glycol-calyx layer, are involved in the adhesion of *R. flavefaciens* to cellulose (Miron *et al.*, 2001). Genome sequencing of *R. flavefaciens* FD-1 has revealed many remarkable facts about the bacterium such as being a large repertoire of cellulolytic enzymes. It has been concluded that it harbours at least 14 genes (catalytic modules) of GHs and 68 various CBMs. In this list, family GH5 members dominate the count (Berg Miller *et al.*, 2009). This shows the large diversity of CAZymes harboured by *R. flavefaciens* FD-1. These studies make *R. flavefaciens* a suitable candidate for harvesting potential efficient cellulases for the purpose of various industrial processes like bioethanol production.

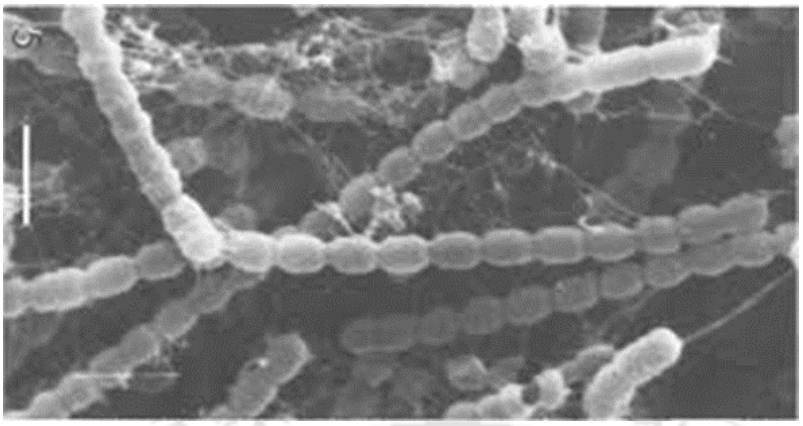

Fig. 1.9 Microscopic image of *Ruminococcus flavefaciens* (Latham *et al.*, 1978).

Table 1.3. Taxonomy of *R. flavefaciens* (Liu *et al.*, 2008).

Kingdom	Bacteria
Phylum	Firmicutes
Class	Clostridia
Order	Clostridiales
Family	Ruminococcaceae
Genus	*Ruminococcus*
Species	*flavefaciens*
Strain	FD-1 v3

1.12 Various Glycoside Hydrolases from *Ruminococcus flavefaciens*

Few cellulase proteins from *R. flavefafciens* has been partially characterized either with the help of random cloning or purification from the cultured microorganism itself. A study investigated the cellulase complex from *R. flavefaciens* as a crude preparation and concluded that the most active enzymes present in the complex were of the exo-β-1,4-glucanase type (Pettipher and Latham, 1979b). However, it was not purified any of the enzymes present or investigate the number and types of cellulolytic enzymes in the different molecular weight complexes. The major enzyme type was exo-β-1,4-glucanase (cellobiohydrolase, CBH), based on the products of cellulolytic

activity (cellobiose, cellotriose, and a small amount of glucose) and the changes in viscosity versus reducing sugar. Pettipher and Latham in 1979 later purified and characterized cellulases and xylanases from *R. flavefaciens* where they reported K_m of 0.432 mg/ml of a CMCase enzyme. The reported CMCase was optimally active at pH 6.4 and at 45ºC with specific activity of 14 µmol/min/mg (U/mg). The first report on cloned cellulase from *R. flavefaciens* was published in 1987 where a gene of *R. flavefaciens* FD-1 was cloned in the vector pEcoR251 to form the recombinant plasmid pMEB200. The cloned endoglucanase gene showed carboxymethyl cellulase enzyme activity but no degradation of Avicel or filter paper, thus showing that it is purely an endoglucanase (Barros and Thomson, 1987). However, it was not characterised biochemically and structurally. A group of researchers cloned an endoglucanase (*CelA*) from *R. flavefaciens* FD-1 which had shown 0.0025 U/mg specific activity against (Wang *et al.,* 1993). Another group reported for the first time, the dockerin-like sequences in *endA* (GH5) gene thus revealing the existence of cellulosome in *R. flavefaciens* (Kirby *et al.,* 1997). In a different study , it was described how *endB,* a GH44 endoglucanase from *R. flavefaciens* 17 binds to the cellulose with the help of a novel CBM (Rincón *et al.,* 2001).

Howard and White reported a bi-functional endoglucanase from *R. flavefaciens* which showed specific activity of 0.006 U/mg against CMC (Howard and White, 1990). A licheninase (β-1-3,1-4-glucanase) from *R. flavefaciens* was expressed by Flint et al. in *E. coli* (Flint *et al.,* 1989). Two xylanases, viz. *xynA* and *xynB* from *R. flavefaciens* were biochemically characterized which were found to be efficiently hydrolysing larchwood xylan. Specific activities of these enzyme were noted to be 870 and 210 U/mg, respectively (Flint *et al.,* 1991). Exo-β-1,4-glucanase (*ExoA*) characterized from

R. flavefaciens using *p*-nitrophenyl-o-D-cellobioside (NPC) as a substrate was found to be Ca^{2+} dependent. It has K_m of 3.08 mM and V_{max} of 0.298 U/mg and no activity against CMC was observed (Gardner *et al.,* 1987). Two endoglucanases, namely A and B purified from *R. flavefaciens* were found to be showing activity against both CMC and xylan. Endoglucanase A (*EndA*) showed specific activity of 15.5 U/mg and 12.2 U/mg against CMC and xylan, respectively. 10 U/mg and 3.5 U/mg of specific activities were noted for Endoglucanase B (*EndB*) (Doerner and White, 1990). In another study, one of the clones with 9 kb inserts showed the activities of all three cellulases. In addition, two of the 4 kb-insert clones and one 9 kb-insert clone of cellulase were found to degrade crystalline Avicel (Huang *et al.,* 1989). Earlier, a group of researcher characterized a β-1,3-glucanase from *R. flavefaciens* using the 15-mer laminarin as a substrate showing the K_m of 0.37 mg/mL. It had optimum pH and temperature of 6.8 and 55^oC, respectively (Erfle and Teather, 1991). Two xylanases expressed in *E. coli* from *R. flavefaciens,* namely, *Xyl-A* and *Xyl-C* showed specific activitiy of 320 and 260 U/mg on xylan (Garcia-Campayo *et al.,* 1993). Moreover, in an other study, researchers characterised an endomannanase (GH5_7) from *R. flavefaciens* FD-1 v3 which efficiently hydrolysed Locust bean galactomannan (298.5 U/mg), Konjac glucomannan (256 U/mg) and Carob galactomannan (177 U/mg) (Goyal *et al.,* 2019). It acted optimally at pH 6.0 and 60^oC and showed V_{max} of 398 U/mg and K_m of 0.92 mg/mL (Goyal *et al.,* 2019). Moreover, researchers elucidated the crystal structures of two novel GH5 xyloglucanases (XEGs) retrieved from a rumen microflora metagenomic library (*Uncultured bacterium*), in the native state and in complex with xyloglucan-derived oligosaccharides (Dos Santos *et al.,* 2015). Protein structure of native XEG5A and XEG5B (PDB ID – 4W8A) was determined by single anomalous dispersion (SAD)

method and refined at 1.72 Aº resolution. The oligosaccharide found in the XEG5A complex revealed, for the first time, the positive sub sites of a GH5 XEG. This complex also showed that the pocket-like topology of the +1 subsite is essential for the binding of a xylosyl side chain, which probably has an important role in substrate recognition by GH5 XEGs. The oligosaccharide found in the XEG5B complex along with extensive structural analyses revealed the structural basis for XXXG hydrolysis of xyloglucan, which involves a modification in the −1 sub site, enabling the accommodation of a xylosyl moiety at this position. During the course of the present doctoral work, a group of scientist structurally characterized 10 members of subfamily GH5_4 from different organisms such as an endoglucanase (PDB ID: 6XSU a.k.a. *Rf*GH5_4 from *R. flavefaciens* (Glasgow *et al.,* 2020) where they elucidated the $(\beta/\alpha)_8$-TIM barrel nature of these family GH5 glycoside hydrolases. This study shed light on how the aromatic amino acids surrounding the active-site of subfamily GH5_4 members are responsible for their broad substrate specificity or multifunctionality. Some other solved protein structure from subfamily GH5_4 include JEP (Resolution - 1.4 Aº) from *Paenibacillus pabuli* (Gloster *et al.,* 2007) and 4V2X (Resolution – 1.64 Aº) from *Bacillus halodurans* C-125 (Venditto *et al.,* 2016). These enzymes highlight the diversity of microbial xyloglucan-active cellulases (endoglucanases) available in nature.

1.13 Enzyme of interest: Endoglucanase, *Rf*GH5_4 from *R. flavefaciens*

The proposed enzyme of interest for this study is glycoside hydrolase from rumen bacterium *R. flavefaciens* with putative endoglucanase activity as predicted by its BLAST (Basic Local Alignment Search Tool) analysis. It showed 43.38 % sequence identity and 90% query cover with 1EDG_A protein of *Ruminiclostridium cellulolyticum* H10. It is a member of subfamily 4 of CAZyme family GH5 and thus

would be called (*Rf*GH5_4) henceforth. *Rf*GH5_4 (348 amino acids) is a part of modular enzyme *Rf*GH5$_{1/2}$ of *R. flavefaciens* (**Fig. 1.10**). It is located at the N-terminal end of *Rf*GH5$_{1/2}$, while C-terminal domain is a putative endo-mannanase (*Rf*GH5_7) (Venditto *et al.*, 2016). The two catalytic modules are separated by family 80 carbohydrate binding module (*Rf*CBM80). *Rf*GH5_4 is listed in the CAZy database as a part of *Rf*GH5$_{1/2}$ module, having CAZy name *Rf*GH5-1/2 (GenBank Accession No. WP_009984467.1). Being a member of subfamily 4 of GH5, it is proposed to be a multifunctional endoxyloglucanase i.e. endoglucanase degrading the xyloglucan, lichenan and mixed linked β-glucan.

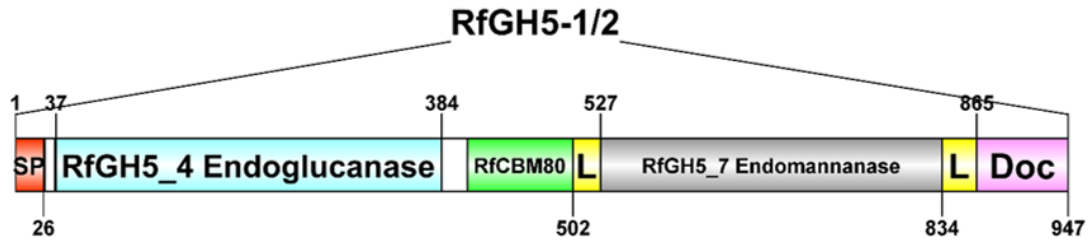

Fig. 1.10 Molecular architechutre of multimodular enzyme, *Rf*GH5$_{1/2}$ of *R. flavefaciens* FD-1 v3 showing *Rf*GH5_4 module (Light Blue) at its N-terminal after the signal peptide (Red).

1.14 Knowledge gap on *Rf*GH5_4 Endoglucanase

In spite of considerable reports on studies of cellulases/endoglucanases from *R. flavefaciens*, their thorough biochemical, functional and structural characterization lags behind as compare to its other bacterial and fungal counterparts. Being the backbone of hydrolysis of lignocellulosic biomass in herbivores' rumen, it is armed with the efficient endoglucanases and other cellulases. Genome sequencing and its analysis also assured about the huge reservoir at least 14 cellulolytic genes in the genome of *R. flavefaciens.* However, no cellulolytic enzyme from *R. flavefaciens* has been represented as

characterized in the various internationally recognized enzyme databases like Enzyme Commission (EC), BRENDA, ExPASy and Prosite. Carbohydrate Active Enzyme (CAZy) database reports 3 protein sequence from *R. flavefaciens* in the bacterial domain. GenBank accession no. for these predicted proteins are WP_009984467.1, and CAA36207.1 from *R. flavefaciens* FD-1 v3 and CAB05881.1 from *R. flavefaciens* 17). These protein sequences have been categorized under subfamily 4 of GH5 except CAA36207.1 which is a CellodextrinaseA (*CelA*) belonging to subfamily 37. So far, total 614 glycoside hydrolase members of GH5 family have been characterized from different microorganisms. Earlier, homology modelling, molecular docking and Small Angle X-ray Scattering (SAXS) based analysis of a GH5 enzyme, mannanase, *Rf*GH5_7 from *R. flavefaciens* has been reported (Goyal *et al.*, 2019). It is a (α/β)8 TIM-barrel motif and falls under subfamily 7, showing endomannanase activity. Glasgow *et al.*, 2020 revealed the structural properties and substrate specificities of *Rf*GH5_4 enzyme. In this study, *Rf*GH5_4 was primarily shown to be a Glucomannanase. Moreover, the enzyme assays were performed at 30 °C and not at the optimized reaction conditions which leaves the gap for establishing its true enzyme kinetics as well as functional properties as a cellulase (Endoglucanase). It is also clear that ruminal cellulases such as those from of *R. flavefaciens* would be of potential industrial importance and are needed to be studied well in depth. Also, the structural behaviour of *Rf*GH_4 protein in solution form as a a function of protein concentration is not yet explored. Molecular docking, molecular dynamics simulations of *Rf*GH_4 with cellulosic ligands is also yet to be studied.

Comprehensive comparison of endoglucanases from various sources presented in **Table 1.1**, shows that *Thermotoga maritima, Talaromyces emersonii, Scytalidium*

thermophilum, Trichoderma reesei, Cellulomonas biazotea, Aspergillus oryzae CMC-1 and *Clostridium thermocellum* harbour endoglucanases with high specific activity as compared to other enlisted cellulases. However, when the catalytic efficiency (k_{cat}/K_m) of these enzymes is considered for CMC as a substrate, only few of them emerge as an efficient endoglucanase. These are namely, endoglucanase rMt-egl of *Myceliopthora thermophila* BJA (204 $mL.mg^{-1}.s^{-1}$), endoglucanase of *Cellulomonas biazotea* (62 $mL.mg^{-1}.s^{-1}$), TeEgl5A of *Talaromyces emersonii* CBS394.64 (51.6 $mL.mg^{-1}.s^{-1}$), Egst of *Scytalidium thermophilum* (45.25 $mL.mg^{-1}.s^{-1}$), Egl499 of *Bacillus subtilis* JA18 (21.7 $mL.mg^{-1}.s^{-1}$) and CMCase of *Aspergillus oryzae* cmc-1 (17.8 $mL.mg^{-1}.s^{-1}$). The most studied and industrially explored organism for cellulases i.e. *Trichoderma reesei* was found to be having the low catalytic efficiency for its endoglucanase, EG III (Cel12A, ~7.33 $mL.mg^{-1}.s^{-1}$). Although efficiency of *Trichoderma reesei* cellulases is considered as a combine effect of endoglucanase, cellobiohydrolase and β-glucosidase collectively; its endoglucanase alone is revealed to be less efficient than its bacterial counterparts. Also, the cellulase derived from *T. reesei* and most other fungi are not stable at high pH or high temperatures. Most of the industrial cellulase candidates showed specific activity less than a 100 U/mg (**Table 1.2**).

This survey of cellulase enzymes also suggests that there is still a need of finding more efficient endoglucanases for industrial practices which would work efficiently, have better catalytic efficiency, low K_m and high k_{cat} and k_{cat}/K_m. Xyloglucan-degrading endoglucanases like *Rf*GH5_4 could facilitate the access of cellulose in plant cell wall and also contribute to release C5/C6 sugars (D-Glucose) that can be further used in fermentation. These enzymes have great importance in the conversion of lignocellulose into fermentable sugars, especially when dicotyledons are the source of the biomass,

which has high xyloglucan content (Menon *et al.*, 2010; Dos Santos *et al.*, 2015) available sources of dicotyledonous as well as monocotyledonous residual biomass are agriculture wastes such as Cotton stalk, Pigeon pea stalk, sorghum durra stalk, finger millet stalk, maize leaves as well sugarcane bagasse from sugar industry. Thus, *Rf*GH5_4 was hypothesized to be one of the efficient endoglucanase from *R. flavefaciens* with prospected efficient lignocellulose saccharification for bioethanol production.

1.15 Significance of the work

Efficient hydrolysis of plant cell wall polysaccharides has been fine-tuned over millions of years in ecological niches that are subjected to intensive selective pressures exemplified by the rumen of mammalian herbivores. A cohort of rumen anaerobic bacteria assembles their plant cell wall degrading apparatus into multi-protein complexes termed as cellulosomes. *R. flavefaciens* is one of the few rumen bacteria which degrade plant cell wall, particularly cellulose, efficiently. It is the also a member of importanr group of bacteria which contains organized and advanced form of cellulosome (Venditto *et al.*, 2016) As can be concluded from review of the published literature, having complete genome sequence of *R. flavefaciens* and its analysis, we now have whole set of its putative and annotated endoglucanases (Berg Miller *et al.*, 2009). There are at least 14 glycoside hydrolases of family GH5 in this bacterium's genome but they have merely been completely characterized biochemically and structurally. In the present state, fungal cellulases dominate the industrial cellulase market. However, it is found that k_{cat} and k_{cat}/K_m of these enzymes are very low, whereas K_m for cellulosic substrates are quite high. So, it is not exaggeratory to say that we are still in search of an efficient endoglucanase for the purpose of bioethanol production. *Rf*GH5_4, a

putative endoglucanase, is naturally present in the genome of *R. flavefaciens* FD-1 v3 as a part of a multienzyme complex, *Rf*GH$_{1/2}$ (GenBank Accession No. WP_009984467.1). Being from a rumenophilic bacterium, it was hypothesized to be an efficient endoglucanase for deconstruction of plant biomass for the purpose of bioethanol production. Moreover, being a member of subfamily 4, it was proposed to be a polyspecific/multifunctional cellulase/endoglucanase hydrolysing various glucan like cellulose, xyloglucan, lichenan and β-glucan. Molecular docking studies of *Rf*GH_4 with various cellulosic and hemicellulosic ligands would reveal its active-site interactions with the substrates. Molecular dynamics simulation study of the *Rf*GH_4 protein with cellulosic ligand would help in demonstrating its stability and compactness with the cellulosic substrates. Characterization by other physical techniques likes Thin layer chromatography, MALDI-MS, HPLC and Circular Dichroism (CD) would reveal the information about its structure-function relationship, mechanism of catalysis as well as multifunctionality, active-site interactions and behaviour in the solution form. Solution structure studies by SAXS whould shed light on its intermolecular interactions, monodispersity of the protein and consequently its stability. Therefore, along with being a potential candidate as a biocatalyst for bioethanol production from lignocellulose, it could be potentially employed for various other applications like textile, food and pulp industries, pharmaceutical sector, modification of cellulose surface, plant growth modulation, production of xyloglucan oligosaccharides and xyloglucan-specific probes (Rashmi and Siddalingamurthy, 2018).

1.16 Specific Objectives

1. Cloning, expression and purification of endoglucanase (*Rf*GH5_4) from *Ruminococcus flavefaciens* FD-1 v3.
2. Biochemical, biophysical and functional characterization of *Rf*GH5_4.
3. Molecular docking and Molecular Dynamics simulation based *in silico* structure characterization of *Rf*GH5_4 with various carbohydrate ligands.
4. Solution structure study of *Rf*GH5_4 by Small Angle X-Ray Scattering (SAXS).
5. Application of *Rf*GH5_4 in combination with other cellulolytic enzymes for saccharification of lignocellulosic biomass.

References

Annamalai N, Rajeswari MV, Balasubramanian T (2016) Endo-1, 4-β-glucanases: Role, applications and recent developments. *Microbial Enzymes in Bioconversions of Biomass* 37–45

Aranda Soto NB (2019) Aproximación a la validación de método bioinformático de termoestabilización, basado en la enzima exo-celobiohidrolasa I Cel7C del hongo *Phanerochaete chrysosporium*, mediante reemplazo de aminoácidos por cisteínas en su dominio catalítico y su expresión e. *Repositorio Academico*, de la *Univesidad de Chile* 1-88

Artzi L, Bayer EA, Moraïs S (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology* 15:83–95

Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology* 12:1–16

Barros ME, Thomson JA (1987) Cloning and expression in Escherichia coli of a cellulase gene from *Ruminococcus flavefaciens*. *Journal of Bacteriology* 169:1760–1762

Berg Miller ME, Antonopoulos DA, Rincon MT, Band M, Bari A, Akraiko T, Hernandez A, Thimmapuram J, Henrissat B, Coutinho PM (2009) Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PloS One* 4:e6650

Bhat MK (2000) Cellulases and related enzymes in biotechnology. *Biotechnology Advances* 18:355–383

Bhat MK, Bhat S (1997) Cellulose degrading enzymes and their potential industrial applications. *Biotechnology Advances* 15:583–620

Cantarel BL, Coutinho PM, 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

Cheng R, Cheng L, Wang L, Fu R, Sun X, Li J, Wang S, Zhang J (2019) Characterization of an alkali-stable xyloglucanase/mixed-linkage β-glucanase Pgl5A from *Paenibacillus* sp. S09. *International Journal of Biological Macromolecules* 140:1158–1166

Cheng Y-S, Ko T-P, Huang J-W, Wu T-H, Lin C-Y, Luo W, Li Q, Ma Y, Huang C-H, Wang AH-J (2012) Enhanced activity of *Thermotoga maritima* cellulase 12A by mutating a unique surface loop. *Applied Microbiology and Biotechnology* 95:661–669

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

Dehority BA (1993) Laboratory manual for classification and morphology of rumen ciliate protozoa. *CRC Press*

Demain AL, Newcomb M, Wu JHD (2005) Cellulase, clostridia, and ethanol.

Microbiology and Molecular Biology Reviews 69:124–154

Doerner KC, White BA (1990) Assessment of the endo-1, 4-beta-glucanase components of *Ruminococcus flavefaciens* FD-1. *Applied and Environmental Microbiology* 56:1844–1850

Domínguez R, Souchon H, Lascombe M-B, Alzari PM (1996) The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. *Journal of Molecular Biology* 257:1042–1051

Dos Santos CR, Cordeiro RL, Wong DWS, Murakami MT (2015) Structural Basis for Xyloglucan Specificity and α-d-Xyl p (1→ 6)-d-Glc p Recognition at the− 1 Subsite within the GH5 Family. *Biochemistry* 54:1930–1942

Dotsenko AS, Gusakov A V, Rozhkova AM, Sinitsyna OA, Nemashkalov VA, Sinitsyn AP (2016) Effect of N-linked glycosylation on the activity and other properties of recombinant endoglucanase IIa (Cel5A) from *Penicillium verruculosum*. *Protein Engineering, Design and Selection* 29:495–502

Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe H-P, Belaich J-P, Davies GJ, Haser R (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* 3:939–949

Erfle JD, Teather R (1991) Isolation and properties of a (1, 3)-beta-D-glucanase from *Ruminococcus flavefaciens*. *Applied and Environmental Microbiology* 57:122–129

Flint HJ, McPHERSON CA, Bisset J (1989) Molecular cloning of genes from *Ruminococcus flavefaciens* encoding xylanase and beta (1-3, 1-4) glucanase activities. *Applied and Environmental Microbiology* 55:1230–1233

Flint HJ, McPherson CA, Martin J (1991) Expression of two xylanase genes from the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 cloned in pUC13. *Microbiology* 137:123–129

Foong FC, Doi RH (1992) Characterization and comparison of *Clostridium cellulovorans* endoglucanases-xylanases EngB and EngD hyperexpressed in *Escherichia coli*. *Journal of Bacteriology* 174:1403–1409

Ganner T, Bubner P, Eibinger M, Mayrhofer C, Plank H, Nidetzky B (2012) Dissecting and reconstructing synergism: in situ visualization of cooperativity among cellulases. *Journal of Biological Chemistry* 287:43215–43222

Garcia-Campayo V, McCrae SI, Zhang J-X, Flint HJ, Wood TM (1993) Mode of action, kinetic properties and physicochemical characterization of two different domains of a bifunctional (1→ 4)-β-d-xylanase from *Ruminococcus flavefaciens* expressed separately in *Escherichia coli*. *Biochemical Journal* 296:235–243

Gardner RM, Doerner KC, White BA (1987) Purification and characterization of an exo-beta-1, 4-glucanase from *Ruminococcus flavefaciens* FD-1. *Journal of Bacteriology* 169:4581–4588

Gilbert HJ (2010) The biochemistry and structural biology of plant cell wall

deconstruction. *Plant Physiology* 153:444–455

Gilmore SP, Lillington SP, Haitjema CH, de Groot R, O'Malley MA (2020) Designing chimeric enzymes inspired by fungal cellulosomes. *Synthetic and Systems Biotechnology* 5:23–32

Glasgow EM, Kemna EI, Bingman CA, Ing N, Deng K, Bianchetti CM, Takasuka TE, Northen TR, Fox BG (2020) A structural and kinetic survey of GH5_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis. *Journal of Biological Chemistry* 295:17752–17769

Gloster TM, Ibatullin FM, Macauley K, Eklof JM, Roberts S, Turkenburg JP, Bjørnvad ME, Jørgensen PL, Danielsen S, Johansen KS (2007) Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *Journal of Biological Chemistry* 282:19177–19189

Gottschalk TE, Mikkelsen JD, Nielsen JE, Nielsen KK, Brunstedt J (1998) Immunolocalization and characterization of a β-1, 3-glucanase from sugar beet, deduction of its primary structure and nucleotide sequence by cDNA and genomic cloning. *Plant Science* 132:153–167

Goyal D, Kumar K, Centeno MSJ, Thakur A, Pires VMR, Bule P, Fontes CMGA, Goyal A (2019) Molecular cloning, expression and biochemical characterization of a family 5 glycoside hydrolase first endo-mannanase (RfGH5_7) from *Ruminococcus flavefaciens* FD-1 v3. *Molecular Biotechnology* 61:826–835

Henrissat B (1994) Cellulases and their interaction with cellulose. *Cellulose* 1:169–196

Henrissat B, Claeyssens M, Tomme P, Lemesle L, Mornon J-P (1989) Cellulase families revealed by hydrophobic cluster analysi. *Gene* 81:83–95

Howard GT, White BA (1990) Cloning in escherichia coli of a bi-functional cellulase/xylanase enzyme from *Ruminococcus flavefaciens* FD-1. *Animal Biotechnology* 1:95–106

Huang C-M, Kelly WJ, Asmundson R V, Yu P-L (1989) Molecular cloning and expression of multiple cellulase genes of *Ruminococcus flavefaciens* strain 186 in *Escherichia coli*. *Applied Microbiology and Biotechnology* 31:265–271

Huang X, Li Q, Chen X, Fan J, Xu X, Sun X, Li D, Zhao H (2017) Expression and characteristics of an endoglucanase from *Trichoderma atroviride* (TaEGII) in Saccharomyces cerevisiae. *Applied Biochemistry and Biotechnology* 182:1158–1170

Hungate RE (2013) The rumen and its microbes. *Elsevier*

Javed MR, Rashid MH, Nadeem H, Riaz M, Perveen R (2009) Catalytic and thermodynamic characterization of endoglucanase (CMCase) from *Aspergillus oryzae* cmc-1. *Applied Biochemistry and Biotechnology* 157:483–497

Kirby J, Martin JC, Daniel AS, Flint HJ (1997) Dockerin-like sequences in cellulases and xylanases from the rumen cellulolytic bacterium *Ruminococcus flavefaciens*. *FEMS Microbiology Letters* 149:213–219

Kitago Y, Karita S, Watanabe N, Kamiya M, Aizawa T, Sakka K, Tanaka I (2007) Crystal structure of Cel44A, a glycoside hydrolase family 44 endoglucanase from *Clostridium thermocellum*. *Journal of Biological Chemistry* 282:35703–35711

Klemm D, Heublein B, Fink H, Bohn A (2005) Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie International Edition* 44:3358–3393

Koshland Jr DE (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews* 28:416–436

Kostylev M, Wilson DB (2011) Determination of the catalytic base in family 48 glycosyl hydrolases. *Applied and Environmental Microbiology* 77:6274–6276

Kuhad RC, Singh A, Eriksson K-EL (2006) Microorganisms and enzymes involved in the degradation of plant fiber cell walls. *Biotechnology in the Pulp and Paper Industry* 45–125

Kumar K, Singal S, Goyal A (2019) Role of carbohydrate binding module (CBM3c) of GH9 β-1, 4 endoglucanase (Cel9W) from *Hungateiclostridium thermocellum* ATCC 27405 in catalysis. *Carbohydrate Research* 484:107782

Latham MJ, Brooker BE, Pettipher GL, Harris PJ (1978) *Ruminococcus flavefaciens* cell coat and adhesion to cotton cellulose and to cell walls in leaves of perennial ryegrass (*Lolium perenne*). *Applied and Environmental Microbiology* 35:156–165

Leggio L Lo, Larsen S (2002) The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Letters* 523:103–108

Li Y, Irwin DC, Wilson DB (2007) Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Applied and Environmental Microbiology* 73:3165–3172

Liu C, Finegold SM, Song Y, Lawson PA (2008) Reclassification of *Clostridium coccoides, Ruminococcus hansenii, Ruminococcus hydrogenotrophicus, Ruminococcus luti, Ruminococcus productus* and *Ruminococcus schinkii* as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydroge*. *International Journal of Systematic and Evolutionary Microbiology* 58:1896–1902

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42:D490--D495

Macarron R, Acebal C, Castillon MP, Dominguez JM, De la Mata I, Pettersson G, Tomme P, Claeyssens M (1993) Mode of action of endoglucanase III from *Trichoderma reesei*. *Biochemical Journal* 289:867–873

Meinke A, Damude HG, Tomme P, Kwan E, Kilburn DG, Miller RC, Warren RAJ, Gilkes NR (1995) Enhancement of the Endo-β-1, 4-glucanase Activity of an Exocellobiohydrolase by Deletion of a Surface Loop (∗). *Journal of Biological Chemistry* 270:4383–4386

Meleiro LP, Carli S, Fonseca-Maldonado R, da Silva Torricillas M, Zimbardi ALRL, Ward RJ, Jorge JA, Furriel RPM (2018) Overexpression of a cellobiose-glucose-halotolerant endoglucanase from *Scytalidium thermophilum*. *Applied Biochemistry and Biotechnology* 185:316–333

Menon V, Prakash G, Rao M (2010) Enzymatic hydrolysis and ethanol production using xyloglucanase and *Debaromyces hansenii* from tamarind kernel powder: galactoxyloglucan predominant hemicellulose. *Journal of Biotechnology* 148:233–239

Merzlov DA, Zorov IN, Dotsenko GS, Denisenko YA, Rozhkova AM, Satrutdinov AD, Rubtsova EA, Kondratieva EG, Sinitsyn AP (2015) Properties of enzyme preparations and homogeneous enzymes—endoglucanases EG2 *Penicillium verruculosum* and LAM *Myceliophthora thermophila*. *Biochemistry (Moscow)* 80:473–482

Miron J, Ben-Ghedalia D, Morrison M (2001) Invited review: adhesion mechanisms of rumen cellulolytic bacteria. *Journal of Dairy Science* 84:1294–1309

Miyazaki S, Suisha F, Kawasaki N, Shirakawa M, Yamatoya K, Attwood D (1998) Thermally reversible xyloglucan gels as vehicles for rectal drug delivery. *Journal of Controlled Release* 56:75–83

Najmudin S, Guerreiro CIPD, Carvalho AL, Prates JAM, Correia MAS, Alves VD, Ferreira LMA, Romão MJ, Gilbert HJ, Bolam DN (2006) Xyloglucan is recognized by carbohydrate-binding modules that interact with β-glucan chains. *Journal of Biological Chemistry* 281:8815–8828

Nieves RA, Ehrman CI, Adney WS, Elander RT, Himmel ME (1997) Survey and analysis of commercial cellulase preparations suitable for biomass conversion to ethanol. *World Journal of Microbiology and Biotechnology* 14:301–304

O'Neill MA, York WS (2003) The composition and structure of plant primary cell. *The Plant Cell Wall* 8:1–54

Oksanen A, Timo R, Retulainen E, Salminen K, Brumer H (2011) Improving wet web runnability and paper quality by an uncharged polysaccharide. *Journal of Biobased Materials and Bioenergy* 5:187–191

Parsiegla G, Reverbel C, Tardif C, Driguez H, Haser R (2008) Structures of mutants of cellulase Cel48F of *Clostridium cellulolyticum* in complex with long hemithiocellooligosaccharides give rise to a new view of the substrate pathway during processive action. *Journal of Molecular Biology* 375:499–510

Petkun S, Grinberg IR, Lamed R, Jindou S, Burstein T, Yaniv O, Shoham Y, Shimon LJW, Bayer EA, Frolow F (2015) Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: structural and functional significance of the intermodular linker. *PeerJ* 3:e1126

Pettipher GL, Latham MJ (1979a) Characteristics of enzymes produced by *Ruminococcus flavefaciens* which degrade plant cell walls. *Microbiology* 110:21–27

Pettipher GL, Latham MJ (1979b) Production of enzymes degrading plant cell walls

and fermentation of cellobiose by *Ruminococcus flavefaciens* in batch and continuous culture. *Microbiology* 110:29–38

Phadtare P, Joshi S, Satyanarayana T (2017) Recombinant thermo-alkali-stable endoglucanase of *Myceliopthora thermophila* BJA (rMt-egl): biochemical characteristics and applicability in enzymatic saccharification of agro-residues. *International Journal of Biological Macromolecules* 104:107–116

Rajoka MI, Ashraf Y, Khalid AM (2004) Kinetic of improved production and carboxymethyl cellulose hydrolysis by an endo-glucanase from a derepressed mutant of *Cellulomonas biazotea*. *Biotechnology Letters* 26:1329–1333

Rani V, Mohanram S, Tiwari R, Nain L, Arora A (2014) Beta-glucosidase: key enzyme in determining efficiency of cellulase and biomass hydrolysis. *Journal of Bioprocessing and Biotechniques* 5:197

Rashmi R, Siddalingamurthy KR (2018) Microbial xyloglucanases: A comprehensive review. *Biocatalysis and Biotransformation* 36:280–295

Rincón MT, McCrae SI, Kirby J, Scott KP, Flint HJ (2001) EndB, a multidomain family 44 cellulase from *Ruminococcus flavefaciens* 17, binds to cellulose via a novel cellulose-binding module and to another *R. flavefaciens* protein via a dockerin domain. *Applied and Environmental Microbiology* 67:4426–4431

Sajith S, Sreedevi S, Priji P, Unni KN, Benjamin S (2014) Production and partial purification of cellulase from a novel fungus, *Aspergillus flavus* BS1. *Annals of Microbiology* 64:763–771

Sakon J, Irwin D, Wilson DB, Karplus PA (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology* 4:810–818

Sangrila S, Maiti TK (2013) Cellulase production by bacteria: A review. *British Microbiology Research Journal* 3:235–258

Saranraj P, Stella D, Reetha D (2012) Microbial cellulases and its applictions. *International Journal of Biochemistry and Biotechnology Science* 1:1–12

Schülein M, Kauppinen MS, Lange L, Lassen SF, Andersen LN, Klysner S, Nielsen JB (1998) Characterization of fungal cellulases for fiber modification. *ACS Publications*

Segato F, Damásio ARL, de Lucas RC, Squina FM, Prade RA (2014) Genomics Review of Holocellulose Deconstruction by Aspergilli. *Microbiology and Molecular Biology Reviews* 78:588.

Segato F, Dias B, Berto GL, de Oliveira DM, De Souza FHM, Citadini AP, Murakami MT, Damásio ARL, Squina FM, Polikarpov I (2017) Cloning, heterologous expression and biochemical characterization of a non-specific endoglucanase family 12 from *Aspergillus terreus* NIH2624. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1865:395–403

Sidrah H, Maimoona A, Asma S, Salman N, Nadeem H, Fauzia Y (2014) Exploring β-glucan activities with xyloglucanase as hydrolytic enzyme, anti tumor and anti

infectant through molecular docking studies. *Pakistan Journal of Public Health* 4:16–21

Singh S, Dhillon A, Goyal A (2020) Enhanced catalytic efficiency of *Bacillus amyloliquefaciens* SS35 endoglucanase by ultraviolet directed evolution and mutation analysis. *Renewable Energy* 151:1124–1133

Sinnott ML (1990) Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews* 90:1171–1202

Sulzenbacher G, Driguez H, Henrissat B, Schülein M, Davies GJ (1996) Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. *Biochemistry* 35:15280–15287

Tomme P, Warren RAJ, Gilkes NR (1995) Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology* 37:1–81

Venditto I, Luis AS, Rydahl M, Schückel J, Fernandes VO, Vidal-Melgosa S, Bule P, Goyal A, Pires VMR, Dourado CG (2016) Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proceedings of the National Academy of Sciences* 113:7136–7141

Vincken J-P, Beldman G, Messen WMA, Voragen AGJ (1996) Degradation of apple fruit xyloglucan by endoglucanase. *Carbohydrate Polymers* 29:75–85

Wang K, Luo H, Bai Y, Shi P, Huang H, Xue X, Yao B (2014) A thermophilic endo-1, 4-β-glucanase from *Talaromyces emersonii* CBS394. 64 with broad substrate specificity and great application potentials. *Applied Microbiology and Biotechnology* 98:7051–7060

Wang W, Reid SJ, Thomson JA (1993) Transcriptional regulation of an endoglucanase and a cellodextrinase gene in *Ruminococcus flavefaciens* FD-1. *Microbiology* 139:1219–1226

Wang Y, Yuan H, Wang J, Yu Z (2009) Truncation of the cellulose binding domain improved thermal stability of endo-β-1, 4-glucanase from *Bacillus subtilis* JA18. *Bioresource Technology* 100:345–349

Wu B, Zheng S, Pedroso MM, Guddat LW, Chang S, He B, Schenk G (2018) Processivity and enzymatic mechanism of a multifunctional family 5 endoglucanase from *Bacillus subtilis* BS-5 with potential applications in the saccharification of cellulosic substrates. *Biotechnology for Biofuels* 11:1–15

Yaoi K, Nakai T, Kameda Y, Hiyoshi A, Mitsuishi Y (2005) Cloning and characterization of two xyloglucanases from *Paenibacillus* sp. strain KM21. *Applied and Environmental Microbiology* 71:7670–7678

Yennamalli RM, Rader AJ, Wolt JD, Sen TZ (2011) Thermostability in endoglucanases is fold-specific. *BMC Structural Biology* 11:1–15

Yoshida K, Komae K (2006) A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. *Plant cell Physiol* 47:1541–1554

Yuan S-F, Wu T-H, Lee H-L, Hsieh H-Y, Lin W-L, Yang B, Chang C-K, Li Q, Gao J, Huang C-H (2015) Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*. *Journal of Biological Chemistry* 290:5739–5748

Zhang K-D, Li W, Wang Y-F, Zheng Y-L, Tan F-C, Ma X-Q, Yao L-S, Bayer EA, Wang L-S, Li F-L (2018) Processive degradation of crystalline cellulose by a multimodular endoglucanase via a wirewalking mode. *Biomacromolecules* 19:1686–1696

Zheng F, Ding S (2013) Processivity and enzymatic mode of a glycoside hydrolase family 5 endoglucanase from *Volvariella volvacea*. *Applied and Environmental Microbiology* 79:989–996

Chapter 2

Cloning, expression and purification of endoglucanase, *Rf*GH5_4 of family GH5 from *Ruminococcus flavefaciens* FD-1 v3

2.1 Introduction

Cellulose is the most abundant polysaccharide on earth, primarily present in the plant biomass (Singh *et al.*, 2020). It is an unbranched polysaccharide made up of glucose units linked by β-1,4-linkage. Cellulose can be hydrolysed to monosaccharide (glucose) and subsequently converted to bioethanol (Kumar *et al.*, 2019). Therefore, cellulosic plant biomass is seen as a significant renewable energy source. Cellulose is hydrolysed synergistically to cellooligosaccharides, cellobiose and glucose by various bacterial or fungal cellulase enzymes. These are namely, endoglucanase (EC 3.2.1.4), exoglucanase (EC 3.2.1.91) and β-glucosidase (EC 3.2.1.21). Endoglucanase randomly hydrolyse cellulose chain to cellooligosaccharides, whereas exoglucanase produces cellobiose entities (Liu *et al.*, 2021). The produced cellooligosaccharides and cellobiose

are then hydrolysed to D-glucose by β-glucosidase (Lynd *et al.,* 2002). The CAZy classification of glycoside hydrolases (GH) is based on the amino acid sequence of enzyme and function (Lombard *et al.,* 2013; Drula *et al.,* 2022). The glycoside hydrolase family 5 (GH5) is one of the largest GH families. The subfamily 4 (GH5_4) of family GH5 is known for cellulases showing broad substrate specificity. The enzymes from the subfamily GH5_4 can display multifunctional activities by acting as endoglucanase (EC 3.2.1.4), Lichenase/licheninase (EC 3.2.1.73), xylanase (EC 3.2.1.8) and xyloglucanase (EC 3.2.1.151) (Aspeborg *et al.*, 2012).

Inhabitant bacteria of herbivorous rumen have been facing cellulosic pressure since time immemorial. These rumen loving microorganisms have evolved efficient machinery, i.e., the cellulosomes, to deconstruct the cellulose. *Ruminococcus flavefaciens* FD-1 v3 is an anaerobic, mesophilic and a Gram-positive bacterium. FD-1 v3 strain of *R. flavefaciens* was isolated from bolus earlier (Bryant *et al.*, 1958). As discussed in Chapter 1, *R. flavefaciens* resides in the intestine of monogastric mammalian herbivores as one of its natural microbiome (Venditto *et al.*, 2016). It is the most populated cellulolytic inhabitant of the animal rumen. Genome sequencing analysis of *R. flavefaciens* FD-1 v3 uncovered a vast array of cellulosomal genes. These are putatively considered versatile for the efficient degradation of lignocellulose (Rincon *et al.*, 2010). As introduced in Chapter 1, the current enzyme of interest, *Rf*GH5_4, is a part of the putative multienzyme complex (cellulosome), *Rf*GH5$_{1/2}$ from *R. flavefaciens* FD-1 v3 (Venditto *et al.,* 2016). The another module of *Rf*GH5$_{1/2}$, *Rf*GH5_7 was earlier characterized as an efficient endomannanase (Goyal *et al.*, 2019). In the present investigation, the role of the catalytic module, *Rf*GH5_4 of *Rf*GH5$_{1/2}$ has been investigated. The gene encoding a putative endoglucanase, *Rf*GH5_4 from *R.*

flavefaciens FD-1 v3 (GenBank Accession Number - WP_009984467.1) was cloned in pET-28a(+) expression vector and transformed in *E. coli* BL-21 expression cells and purified to homogeneity as analysed by SDS-PAGE gel (12%, w/v) which was utilised for further downstream experiments such as biochemical and functional analyses.

2.2 Materials and Methods

2.2.1 Bacterial strains and vectors

The expression vector, pET-28a(+) was used for cloning and expression of the gene encoding *Rf*GH5_4. *E. coli* TOP10 (Novagen®) competent cells were used for amplification of recombinant plasmid containing pET-28a(+) vector and the gene insert encoding *Rf*GH5_4. *E. coli* BL21 (DE3) cells from Novagen®, Merck KGaA, Darmstadt, Germany were used for expression of the *Rf*GH5_4 protein.

2.2.2 The chemicals and kits

Taq DNA polymerase for PCR amplification and restriction enzymes *Nhe*I and *Xho*I were procured from New England Biolabs (NEB), USA. The oligonucleotide primers of *Rf*GH5_4 were procured from GCC Biotech Pvt. Ltd., India. Trizma base, Ethidium bromide (EtBr), Bradford reagent G-250, DNAse-RNAse free water (pH 8.0), mini-prep plasmid isolation kit and gel extraction kit, antibiotics (Kanamycin) and components of polyacrylamide gel electrophoresis were obtained from Sigma-Aldrich Co. LLC, USA. T4 DNA ligase was purchased from Takara Bio Inc., Japan. Coomassie Brilliant Blue R250 for protein staining, Luria Bertani medium (LB), disodium ethylenediaminetetraacetate (EDTA), D-glucose, sodium hydroxide (NaOH) and sodium dodecyl sulphate (SDS) were purchased from HiMedia Laboratories Pvt. Ltd., India. Methanol was procured from Merck, India. The 0.1-10 kb DNA ladder was purchased from NEB, USA, whereas 12 kb High Molecular Weight DNA ladder was

obtained from BioBharati LifeScience Pvt. Ltd., India. Protein ladder with Coomassie BrilliantBlue staining in 12% (w/v) SDS-PAGE gel was procured from BioBharati LifeScience Pvt. Ltd., India.

2.2.3 PCR amplification and cloning of gene encoding RfGH5_4

The 1047 bp gene encoding *Rf*GH5_4 (GenBank Accession No. WP_009984467.1) from *R. flavefaciens* FD-1 v3 without the signal peptide as shown in **Fig. 1.10** in Chapter 1 was amplified by PCR, using its primary amplicon, cloned in pHTP1 vector, as the template, which was generously provided by NZYTech Genes and Enzymes Ltd., Lisbon, Portugal. The primers containing *Nhe*I and *Xho*I restriction sites in forward and reverse primers, respectively were used for PCR amplification. The forward and reverse primers used for PCR are given in the **Table 2.1**.

Table 2.1. Primer sequences used for cloning gene of *Rf*GH5_4 from *R. flavefaciens*.

Construct	Primer sequence*
*Rf*GH5_4	Forward: F-5'-GGTGCTAGCGCTTCCAACATGACCGCAAGCCAGATCAC-3' (T_m : 71°C) Reverse: R-5'-ACCCTCGAGTTATACTCCGAGTACTTCCATCATCTTGTTGAC-3' (T_m : 67°C)

**Underlined nucleotides indicate the restriction sites for NheI and XhoI in the F and R primers, respectively.*

The reaction composition and the PCR cycle conditions for amplification of gene mentioned in the **Tables 2.2 and 2.3**, respectively. PCR amplification was performed by the initial denaturation at 95°C for 3 min followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s, cyclic extension at 68°C for 1 min followed by the final extension at 68°C for 10 min (**Table 2.3**). PCR amplification was performed by using a thermal cycler (Applied Biosystems, GeneAmp® PCR System 9700).

Table 2.2. PCR mix for amplification of gene encoding *Rf*GH5_4 from *R. flavefaciens.*

PCR components	Volume (µL)	Final concentration
Reaction buffer (10x)	6.0	1x
dNTP mix (100 mM)	0.5	0.2 mM
Forward primer (15 µM)	1.8	0.45 µM
Reverse primer (15 µM)	1.8	0.45 µM
Sigma water, pH 8.0	48.4	--
*Rf*GH5_4 Template (35 ng/µL)	1.0	0.58 ng/ µL
Taq DNA polymerase (NEB, 1 U/µL)	0.5	0.02 U/µL
Total	60.0	--

Table 2.3. PCR cycles for amplification of gene encoding *Rf*GH5_4 from *R. flavefaciens.*

Steps		Time
I. Denaturation at 95°C		3 min
II. 30 cycles of	i) Denaturation at 95°C	30 s
	ii) Annealing at 60°C	30 s
	iii) Extension at 68°C	1.5 min
III. Final extension at 68°C		10 mi
IV. Store	4°C	

2.2.4 Agarose gel electrophoresis of PCR amplicons of RfGH5_4 gene

The amplicons of gene encoding *Rf*GH5_4 generated using PCR were run on a 0.8% (w/v) agarose gel along with a DNA Marker (10 kbp DNA Ladder, NEB). 0.4 g of agarose was dissolved in 50 mL 1x TAE (Tris-Acetate EDTA) buffer (as described in **Section 2.2.5**) by using microwave oven till the solution became transparent. The dissolved agarose solution was allowed to cool until it was warm enough to be as a liquid (~50°C). 5 µL of EtBr from 5.0 mg/mL stock was added into the above 0.8% agarose solution and mixed well. The solution was poured into the gel casting tray along with inserted well-comb and kept undisturbed for 30 min for the solidification. The PCR amplified DNA was mixed with the 5x DNA loading dye in 4:1 ratio (4 µL DNA:

1 µL Loading Dye) and loaded on to the agarose gel. The DNA amplicon was electrophoresed in the gel tank (Sigma-Aldrich, USA) by using a constant voltage (50 Volts) for 45 min in 200 mL of 1X TAE buffer. The DNA bands intercalated by EtBr were illuminated under UV light in a gel documentation system (Gel Doc™ XR System, Bio-Rad Laboratories Inc., USA).

2.2.5 Preparation of 10x TAE buffer and DNA loading dye

A stock solution of 1 litre 10X TAE (Tris-Acetate EDTA) buffer was prepared by dissolving 400 mM Tris-base (Trizma), 200 mM glacial acetatic acid and 10 mM EDTA in water and the pH was adjusted at 8.0 (Russell and Sambrook 2001). A 5x solution of DNA loading dye was prepared by mixing the components mentioned below in **Table 2.4.** The final pH of the DNA loading dye was adjusted to pH 8.0. The DNA and loading dye were mixed in a 4:1 ratio.

Table 2.4. Composition of 5x DNA loading dye.

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)
Final pH	8.0

2.2.6 DNA extraction from agarose gel

The PCR amplified DNA was purified from agarose gel using a gel extraction kit (Sigma GenElute™ Gel Extraction Kit). The protocol for DNA purification from agarose gel was provided by the manufacturer as described below. The extracted DNA

was eluted in 20 µL elution buffer provided in the Gel Extraction Kit (Sigma-Aldrich, USA).

2.2.6.1 DNA gel extraction protocol using silica-based columns:

1. The PCR or plasmid DNA amplicon band was excised from gel using sharp sterile scalpel and transferred to an empty 1.5 mL micro centrifuge tube (MCT) whose weight was already noted. The MCT tube with gel was again weighed and the difference between empty and filled tube was considered as the gel weight.
2. Three volume of Gel Solubilisation Solution was added to every 1 volume of gel (100 mg ~ 100 µL) in the MCT tube and incubated at 50°C for 10min (or until the gel slice has completely dissolved).
3. 1 gel volume of 100% isopropanol was added to the MCT tube and mixed until it became homogenous.
4. GenElute Binding Column G (DNA binding column) was placed in 2 mL collection tube provided with the kit. 500 µL of the Column Preparation Solution was added to the binding column and centrifuged at 13,000***g*** for 1 min.
5. The above solution containing PCR-amplified or plasmid DNA (700 µL) was added to DNA binding columns and centrifuged at 13,000***g*** for 1 min at room temperature and the flow through was discarded.
6. 700 µL of Wash Solution was added to the binding column and Centrifuged at 13,000***g*** for 1 minute. Centrifugation was repeated again for 1 minute without any additional wash solution to remove excess ethanol.
7. Now the column containing bound DNA was placed in a fresh 1.5 mL sterile MCT tube. 30 µL of DNAse free water (Sigma-Aldrich Co. LLC, USA) or elution buffer (10 mM Tris-Cl, pH 8.5) was added at the centre of the column. The column was

incubated for 2 min at room temperature and centrifuged at 13,000***g*** for 1 min.

8. The PCR amplified or plasmid DNA was eluted from QIAquick spin columns was collected in 1.5 mL sterile MCT tube. The DNA was stored at - 20°C for further use.

2.2.7 Restriction Digestion of the PCR amplicons of RfGH5_4

The PCR DNA amplicons of *Rf*GH5_4 purified by gel extraction method which contained the restriction sites for *Nhe*I and *Xho*I was digested as per the reaction set up given in **Table 2.5.** The reaction mixture was incubated for 1.5 h, at 37°C using a water bath (Grant Instruments Ltd., Cambridge, England). The reaction was stopped by placing it on ice. The digested PCR fragments were electrophoresed using a 0.8% (w/v) agarose gel. The portion of gel showing illumination of *Rf*GH5_4 DNA fragments of desired size UV illumination were sliced carefully; the DNA fragments were purified using gel extraction method as mentioned in **Section 2.2.6** and eluted in 30 µL elution buffer.

Table 2.5. Reaction set up of restriction enzyme (RE) digestion of *Rf*GH5_4 amplicons.

Restriction Enzyme digestion set up	**Volume, 1× (**µL)
Reaction buffer (10×)	1.0
PCR DNA amplicons (400 ng/µL)	1.0
DNAse free water	6.0
*Nhe*I (10 U/µL)	1.0
*Xho*I (20 U/µL)	1.0
Total	10.0

2.2.8 Restriction Digestion of pET-28a(+) expression vector for the cloning of PCR amplicons of the gene encoding RfGH5_4

The 5369 bp (5.3 kbp) long pET-28a(+) vector (24 ng/µL) was digested using the restriction enzyme system of *Nhe*I and *Xho*I to generate the overhangs suitable for its ligation with the PCR amplicons of *Rf*GH5_4. The details of the reaction setup are shown in **Table 2.6**. The digestion mixture was incubated at 37°C for 1.5 h, the digested

vector was electrophoresed on 0.8% agarose gel and purified by gel extraction method as given in the **Section 2.2.6**. The digested vector was eluted in 20 µL of elution buffer.

Table 2.6. Reaction set up of the RE digestion of pET-28a(+) expression vector.

Restriction Enzyme digestion set up of pET-28a(+)	**Volume, 1X (µL)**
Reaction buffer (10X)	2.5
pET-28a(+) vector (50 ng/µL)	15.0
DNAse free water	6.0
*Nhe*I (10 U/µL)	0.5
*Xho*I (20 U/µL)	1.0
Total	25.0

2.2.9 Ligation of NheI-XhoI digested PCR amplicons with pET-28a(+) vector

The *Nhe*I-*Xho*I RE digested *Rf*GH5_4 PCR amplicons and pET-28a(+) vector were ligated using T4 DNA ligase enzyme. The insert: vector molar ratio was kept at 3:1 in the reaction mixture. The amounts of *Nhe*I-*Xho*I digested DNA insert of the gene encoding *Rf*GH5_4 required for cloning was calculated using the following formula (Engler and Richardson 1982). The ligation reaction set up used for RE digested PCR amplicon and pET-28a(+)vector is given in **Table 2.7**. The ligation reaction was incubated at 16°C for 18h.

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

$$\frac{72ng \times 1.0\ kbp}{5.369\ kbp} \times 3:1 = \ 40.2\ ng\ of\ Insert\ Rf\text{GH5\_4}\ DNA\ amplicon.$$

Table 2.7. Ligation reaction set up of *Rf*GH5_4 PCR amplicons and pET-28a(+).

Ligation reaction set up	**Volume, 1X (µL)**
Reaction buffer (10X)	1.0
pET-28a(+) vector (24 ng/µL)	3.0 (72 ng)
RE Digested *Rf*GH5_4 DNA Insert (26.1 ng/µL)	1.5 (40.2 ng)
DNAse free water	3.5
T4 DNA Ligase (350 U/µL)	1.0
Total	10

2.2.10 Preparation of culture medium

The appropriate amount (g) of commercially available LB medium (HiMedia Laboratories Pvt. Ltd., India) in deionized water was used for growing the *E. coli* cells containing recombinant plasmid. The composition of the LB medium is given in the **Table 2.8**. The 25.13 g of LB medium was dissolved in 800 mL deionized water. The pH was adjusted to 7.2 (if required) and final volume was made to 1 L using deionized water. 100 mL of LB medium was distributed to individual 250 mL aseptically capped Erlenmeyer flasks 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 the microbial inoculation. 10 mL culture tubes of LB medium were also prepared in a similar way.

Table 2.8. Composition of Luria-Bertani medium (Russell and Sambrook 2001).

Components of LB Broth	Final concentration (%, w/v)
Tryptone	1.0
Yeast extract powder	0.5
Sodium chloride	1.0

2.2.11 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 2.2.10**, cooled to around 50-55°C and required amount of antibiotics (Kanamycin, 50 µg/mL) was added aseptically under laminar air flow (LAF). 25 mL of medium supplemented with antibiotics were poured into the sterile petri plates and allowed to stay for 15- 20 min to solidify. The solidified LB-Agar plates were kept overnight at 37°C for sterility test and to check the presence of any contamination. These sterile media containing petri plates were used for transformation and other experiments.

2.2.12 Preparation of E. coli TOP10 competent cells by $CaCl_2$ method

Day 1

1. 100 µL of culture of *E. coli* TOP10 from glycerol stocks were inoculated into 10.0 mL LB medium and grown at 37ºC at 180 rpm for overnight.
2. 0.1 M $CaCl_2$, 0.1 M $MgCl_2$ and 0.1 M $CaCl_2$ solution with 15% glycerol was prepared and filter-sterilized via 0.22 µm membrane filter in laminar-air flow (LAF) and stored at 4°C.

Day 2

3. 1.0 mL of grown culture from the above was taken and inoculated into 100 mL LB medium in 250 mL conical flask. The flask was incubated at 37ºC, 180 rpm till cell Optical Density (OD) reached 0.4 at 600 nm.
4. When the absorbance (280 nm) reached at 0.4, the culture was chilled on ice for 15 min.
5. The culture was centrifuged at 2710***g*** at 4ºC for 10 min.
6. The supernatant was carefully removed and the pellet was gently resuspended in 5.0 mL of sterile, ice chilled 0.1 M $CaCl_2$ solution.
7. The resuspended cells were centrifuged again at 2710***g*** at 4ºC for 10 min.
8. The supernatant was carefully removed and the pellet was gently resuspended 5.0 mL of sterile, ice chilled 0.1 M $MgCl_2$ solution.
9. The resuspended cells were centrifuged again at 2710***g*** at 4ºC for 10 min.
10. The supernatant was carefully removed and the pellet was gently resuspended in 1.0 mL of sterile, ice chilled 0.1 M $CaCl_2$ solution containing 15% (v/v) glycerol. The 100 µL of competent cells were aliquoted into sterile 1.5 mL microcentrifuge tube and kept at -80ºC for further use.

2.2.13 Transformation of ligated DNA using E. coli TOP10 competent cells

The ligated DNA (recombinant plasmid DNA) of *Rf*GH5_4 was transformed using *E. coli* TOP10 competent cells. The steps followed in transformation protocol are as follows:

1. The MCT tube containing competent cell (100 µL) was taken out from -80°C and kept on ice for 5 min, then 10 µL of ligation mixture was added to it.
2. The tube was gently tapped 4-5 times and kept on ice for 30 min.
3. The cells were heat-shocked at 42°C for 40s.
4. The cells were immediately transferred to ice for 5 min.
5. 1.0 mL of LB medium previously incubated at 37°C was added to the transformed cells.
6. The transformed cells were incubated at 37°C in an incubator at 200 rpm for 1h.
7. The cells were harvested by centrifugation at 5000***g*** at 25ºC for 3 min.
8. 1.0 mL supernatant was discarded and the cell pellet was then resuspended in remaining 100 µL supernatant.
9. The cells were spread on LB agar plates supplemented with Kanamycin at a final concentration 50 µg/mL.
10. The LB agar plates were incubated at 37°C for overnight.
11. The transformation efficiency of the competent cells was calculated by using the following equation,

Transformation Efficiency:

$$= \frac{Number\ of\ Colony\ Forming\ Units\ on\ LB\ agar\ plate}{\mu g\ of\ Plasmid\ DNA}$$

$$= No.\ of\ CFUs/\mu g$$

Transformation efficiency is defined as the number of Colony Forming Units (CFUs) which would be produced by transforming 1 µg of plasmid into a given volume of competent cells. The colonies were picked at random from LB agar plates and individually inoculated in 10 mL LB medium supplemented with 50 µg/mL Kanamycin and incubated at 37ºC, 180 rpm for 12h for the isolation of plasmid DNA containing gene encoding *Rf*GH5_4 to check for the positive clones.

2.2.14 Isolation of recombinant plasmid DNA containing gene encoding RfGH5_4

The recombinant plasmid DNA containing the gene encoding *Rf*GH5_4 was isolated from *E. coli* TOP10 cells by the Plasmid miniprep isolation kit (Sigma-Aldrich Co. LLC, USA) using the manufacturer's protocol for the storage and transformation in expression host cells. The isolated plasmid was eluted in 30 µL of sterile DNAse free water.

Plasmid isolation protocol by plasmid miniprep kit:

1. 10 mL of 12 h, overnight grown *E. coli* TOP10 cells containing recombinant plasmid with *Rf*GH5_4 gene were pelleted by centrifugation at ≥12,000***g*** for 1 min.
2. Bacterial pellet was resuspended with 200 µL of the Resuspension Solution.
3. Resuspended cells were lysed by adding 200 µL of the Lysis Solution and by inverting it 6-7 times.
4. The cell debris was precipitated by adding 350 µL of the Neutralization/Binding Solution. The cell debris was precipitated by centrifuging at ≥12,000***g*** or maximum speed for 10 min.
5. The column was prepared by passing 500 µL of the Column Preparation Solution through GenElute Miniprep Binding Column at 12,000***g*** for 1 min.
6. The cleared lysate was transformed from step 3 to the column prepared in step 4 and

centrifuge at 12,000***g*** for 30 seconds to 1 min.

7. 750 µL of the diluted Wash Solution was added the column and centrifuged at 12,000***g*** for 1 min.
8. 50 µL of Elution Solution was added to the column and centrifuged at 12,000***g*** for 1 minute. The eluted DNA was stored at -20°C for further use.

2.2.15 Evaluation of recombinant plasmid DNA for identification of positive clones

The purified plasmid DNA as mentioned above was digested by RE system of *Nhe*I and *Xho*I enzymes as per the protocol given in **Sections 2.2.7 and 2.2.8**. The reaction mixture was set up as given in **Table 2.9**. The digested fragments *viz.* pET-28a(+) vector and the insert DNA of *Rf*GH5_4 were visualized under UV trans-illuminator at 260 nm. The digested fragments (insert and vector) were evaluated whether the size of insert DNA and vector pET-28a(+), were similar to the expected size. This way, the positive clones for the respective recombinant derivatives were confirmed. The glycerol stocks of *E. coli* TOP10 cells harbouring the recombinant plasmid were prepared by keeping the final concentration of glycerol to 25% (v/v). The glycerol stocks in the Cryo vials of 2 ml were stored at -80°C.

Table 2.9. RE digestion set up of recombinant plasmid DNA of pET-28a(+)-*Rf*GH5_4.

RE Digestion set up	**1X (µL)**
Ligation buffer (10X)	2.0
DNAse Free Water	6.5
Recombinant Plasmid DNA (35.5 ng)	10
*Nhe*I (10 U/µL)	0.5
*Xho*I (20 U/µL)	1.0
Total	20

2.2.16 Transformation of recombinant plasmid in E. coli BL21 (DE3) cells for expression of RfGH5_4 protein

Total 1.0 µL of isolated recombinant plasmid DNA as mentioned in earlier was transformed into the 100 µL *E. coli* BL21 (DE3) competent cells as per the protocol given in **Section 2.2.13**. The cells containing recombinant plasmid DNA, pET-28a(+)-*Rf*GH5_4 in their cytosol were spread on LB agar plates with 50 g/mL Kanamycin. The plates were incubated overnight at 37ºC for 12 h to observe the growth of positive colonies. The discrete colonies were independently picked and inoculated in the 10 mL LB medium having 50 µg/mL of Kanamycin.

2.2.17 Expression of recombinant protein, RfGH5_4

The *E. coli* BL21 (DE3) cells transformed with recombinant plasmid, with pET-28a(+) vector and the gene encoding *Rf*GH5_4 were freshly inoculated in 10 mL LB medium supplemented with 50 µg/mL Kanamycin at 37ºC and 180 rpm for 12 h. The expression of enzyme, *Rf*GH5_4 by *E. coli* BL21 (DE3) cells was carried out in 400 mL LB medium containing 50 µg/mL kanamycin. The cells were grown until mid-exponential phase till the absorbance at 600 nm (A_{600}) reached ~0.6 after which a final concentration of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to it. The cells were allowed to grow further at 24°C and 180 rpm for 16 h expression of the recombinant protein. The aliquots, 200 µL of un-induced and induced cells samples were taken and centrifuged at 13,000***g*** and 4ºC for 2 min and the supernatant LB medium was discarded. The cell pellet was resuspended in 40 µL distilled water and the protein expression was checked on SDS-PAGE by using a 12% (w/v) gel.

2.2.18 Analysis of recombinant protein, RfGH5_4 by Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

2.2.18.1 Preparation of acrylamide solution

Acrylamide solution (30%, w/v) was prepared by mixing two components. First 0.8 g of bis-acrylamide was weighed and dissolved in 50 mL of deionized water collected at 18 MΩ (Millipore, Milli-Q water purification system) using a magnetic stirrer (IKA, C-MAG HS7). After completely dissolving, bis-acrylamide, 29.2 g of acrylamide was added to it and stirred on a magnetic stirrer till the solution became clear. The final volume was adjusted to 100 mL with ultra-pure water by keeping the measuring cylinder (100 mL) wrapped in aluminium foil as acrylamide is light sensitive. The final acrylamide solution (30%, w/v) was filtered (Whatman Paper No. 1) under stored in an amber colour bottle at 4°C.

2.2.18.2 Polymerization of SDS-PAGE gel

The resolving gel and stacking gels were prepared by following the protocols from (Russell and Sambrook, 2001) using the composition as shown in **Tables 2.10** and **2.11**. The resolving gel was prepared by adding all the components sequentially as mentioned in **Table 2.10**, in a 25 mL glass container, by keeping acrylamide concentration at 12% (w/v). Similarly, the stacking gel (4%, w/v) was prepared by dissolving all the components as mentioned in **Table 2.11**.

Table 2.10. Components of Resolving gel of SDS-PAGE.

Components	12% Gel Volume (mL)
Acrylamide Solution (30%, w/v)	4.0
Deionized water	0.7
SDS (10%, w/v)	1.0
Glycerol (50%, v/v)	1.0
1.5 M Tris-HCl (pH 8.8)	3.3
Ammonium Persulphate (10%, w/v)	0.1
TEMED	10 µL

Table 2.11. Components of Stacking gel of SDS-PAGE.

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
Ammonium Persulphate (10%, w/v)	0.05
TEMED	5 µL

2.2.18.3 Preparation of SDS-PAGE running buffer

The electrophoresis running (tank) buffer is composed of Tris-base (25 mM), glycine (192 mM) and 0.1 % (w/v) of SDS, with a final of pH 8.3. The SDS-PAGE gels were run using a 1X running or tank buffer prepared from the 10X stock solution as described in **Table 2.12**. The final pH of the buffer was adjusted to 8.3. The 10X buffer was filtered (Whatman Paper No. 1) and stored at 4°C.

Table 2.12. Composition 10X Tris-Glycine running buffer.

Components	Final Concentration (10X buffer)
Tris-base	0.250 M
Glycine	1.92 M
SDS	1.0 % (w/v)

2.2.18.4 Preparation of SDS-PAGE Sample Loading buffer

The protein samples to be analysed using the SDS-PAGE were mixed with the 1X sample loading buffer (pH 6.8). 5X sample loading buffer was prepared by dissolving all the components as described in **Table 2.13** and the pH of the buffer was adjusted to 6.8. The components were dissolved sequentially as mentioned in **Table 2.13** to make 5X sample buffer. However, the final concentration while loading to an SDS-PAGE gel was always kept 1X by mixing 4 volumes of sample (protein) with 1 volume of 5X sample buffer.

Table 2.13. Composition of 5X sample loading buffer (Laemmli, 1970).

Components	Final concentration (5X buffer)
Tris-HCl (pH 6.8)	62.5 mM
Glycerol	20.0 (%, v/v)
SDS	2.0 (%, w/v)
Bromophenol Blue	0.025 (%, w/v)
β-mercaptoethanol	5.0 (%, w/v)

2.2.18.5 Preparation of SDS-PAGE gel staining and de-staining solutions

The SDS-PAGE gel staining solution (100 mL) was prepared by dissolving 250 mg or 0.25% (w/v) of Coomassie Brilliant Blue (CBB R-250) dye in 50 mL of deionized water in an amber colour bottle by keeping on a magnetic stirrer for overnight. The solution was filtered (Whatman Paper No. 1), then 40 mL of methanol and 10 mL of glacial acetic acid were added to finally make the ratio 5:4:1 (deionized water: methanol: glacial acetic acid). The de-staining solution was prepared by dissolving deionized water: methanol: glacial acetic in 5:4:1 ratio. The gel was gently removed from the SDS-PAGE assembly after the run was complete, washed with water and immersed in the staining solution by immersion for 5 min at 30℃ (Room temperature). The stained gel was further developed by heating with staining solution in the Microwave oven (LG Grill Intellowave, LG Electronics, India) for 2 min. In the gel de-staining process, initially, the excessive CCB R-250 was removed by microwave heating (at 60% power) for 2 min of the gel immersed in the deionized water with 5 changes. The gel was destained by periodic change of fresh deionized water after it accumulated the stain. Final de-staining of the gels was performed by immersing it in de-staining solution under gentle shaking condition with change of de-staining solution every 30 min, until the protein bands were distinctly visible. The SDS-PAGE gel image

was captured by using white (visible) light illumination in a gel documentation system (Gel Doc™ XR System, Bio-Rad Laboratories Inc., USA).

2.2.19 Purification of recombinant protein, RfGH5_4

The recombinant protein, *Rf*GH5_4 was expressed intracellularly along with His_6-tag in *E. coli* BL21 (DE3) as mentioned in **Section 2.2.17** in a 100 mL LB medium. The expressed cells were pelleted down by centrifugation at 8000***g*** at 4°C for 15 min. The buffers mentioned in **Table 2.14** were used for the purification process. The cell pellet was resuspended in the 5 mL 50 mM sodium phosphate buffer (pH 7.4), containing 50 mM imidazole and 300 mM NaCl and subjected to ultra-sonication (Sonics, Vibra cell) for 30 min at 4 s ON and 7 s OFF pulse with 33% amplitude. The lysed cells were centrifuged at 15,000***g*** at 4°C for 50 min. The supernatant called Cell Extract (CE) was filtered through 0.45 µm membrane using a syringe filter. The purification of *Rf*GH5_4 was performed by Immobilized Metal Ion Affinity Chromatography (IMAC) as discussed below.

Table 2.14.Composition of buffers used for purification of recombinant protein, *Rf*GH5_4 by affinity purification (IMAC).

Buffers*	**Composition**
Equilibration buffer	50 mM Sodium-phosphate, pH 7.4 300 mM NaCl, 50 mM Imidazole
Elution buffer	50 mM Sodium-phosphate, pH 7.4 300 mM NaCl, 500 mM Imidazole
Cleaning buffer	50 mM Tris-HCl, pH 8.0 500 mM NaCl, 50 mM EDTA

**Prior to use, all the buffers and deionized water were degassed using vacuum filter (Millipore, India), Membrane Filter assembly (Tarson Products Ltd, India) and Magnetic stirrer (REMI MS-500, REMI Elektrotechnik Ltd, India)*

Purification protocol for recombinant Rf*GH5_4 *using HiTrap IMAC column:

1. The clean HiTrap chelating Sepharose column, 5.0 mL (HiTrap Chelating HP, GE Healthcare, USA) stored in 20% (v/v) ethanol at 4°C was connected to P-1 Peristaltic pump (Amersham Biosciences (Cytiva LifeSciences), UK, Made in Sweden). It was charged using 10.0 mL of 0.1 M $NiSO_4$ solution at the flow-rate of 2 mL/min. The unbound Ni^{2+} ions were washed away with 2-3 volumes (~15 mL) of deionized water.
2. The column was equilibrated with 5 volumes of equilibration buffer (**Table 2.14**).
3. The filtered cell extract of recombinant protein was loaded on to the column at a flow rate of 0.5 mL/min.
4. The column was then washed with 20 column volumes (100 mL) of equilibration buffer to remove the unbound proteins. Bradford reagent based qualitative protein detection assay (20 µL protein flow-through and 80 µL Bradford reagent) by analysing the development of blue colour in the Last Column Wash (LCW) was performed to assure that no unbound protein is left behind in the column except the expected bound protein of interest.
5. The retained protein of interest was then eluted with 10 mL elution buffer (50 mM sodium phosphate, pH 7.4, 300 mM NaCl, 500 mM imidazole) and 1 mL fractions were collected. The qualitative Bradford based as mentioned in Step 4 was performed for all the fractions collected and those showing higher intensity of blue colour were mixed and dialysed.
6. The eluted protein *Rf*GH5_4, after carefully packing in the semipermeable dialysis membrane bag (12.4 kDa pore size, HiMedia Laboratories Pvt. Ltd., India) was

dialyzed against 1000 mL of 50 mM sodium phosphate buffer (pH 7.4) at 4℃ to remove 300 mM NaCl salt and 500 mM imidazole present with four times change of dialysis buffer at every 2 h.

2.2.19.1 Cleaning of the used HiTrap IMAC column

The HiTrap chelating Sepharose column, 5.0 mL (HiTrap Chelating HP, GE Healthcare, USA) was washed with 5 volumes of Cleaning buffer (**Table 2.14**) followed by 5 volumes of 1 M NaOH. The column was incubated with 1 M NaOH followed by the washing using the degassed deionized water at 4 mL /min flow-rate, until the complete removal of 1 M NaOH. The removal of NaOH in the washed column was manually checked by confirming the absence of sliminess. The cleaned column was stored in 20% (v/v) ethanol at 4℃ until future use.

2.2.19.2 Analysis of purified protein samples by SDS-PAGE

The dialysed protein fractions (20 µL) and other protein samples (40 µL) from the cells showing the steps from expression to purification (Un-induced cells, Induced cells, Sonicated pellet, Cell Extract and Last Column Wash) were individually mixed with the 5X sample loading buffer (5 µL for purified and 10 µL for other protein samples), as discussed in **Section 2.2.18.4** to make it to the final concentration of 1X. The samples were gently mixed with the help of Vibrator (IKA®, MS 3-Digital, China) and boiled for exactly 5 min using a boiling water-bath (Grant Instruments, England). 15 µL of each boiled sample was loaded on a 4% (w/v) SDS-PAGE Stacking gel. The protein mass marker (3 µL) was also loaded on the gel in a separate Lane for evaluation of molecular weight of the purified protein of interest. The gel was run in a SDS-PAGE Vertical Protein Electrophoresis Gel Equipment, Mini PROTIAN® Tetra Cell System (Bio-Rad Laboratories, Inc., Singapore) using 1X Tris-Glycine running buffer. The

voltage was initially kept at 60 V until the proteins entered from stacking gel into the SDS-PAGE 12% (w/v) resolving gel where the constant Voltage of 100 V was applied. The successfully run SDS-PAGE gel was stained with Coomassie Brilliant Blue R-250, as discussed in the **Section 2.2.18.5**, to observe the homogenous protein bands of interest in the purified fractions.

2.2.19.3 Determination of molecular mass of purified RfGH5_4 by Hedrick plot

The molecular mass of *Rf*GH5_4 was further validated by Hedrick plot (Hedrick and Smith, 1968; Hames, 1998) by measuring the Relative Mobility (a.k.a. Retention factor, R_f) of the standard protein molecular mass marker on the SDS-PAGE gel (12%, w/v). The R_f was plotted against the Log_{10} of their respective Molecular mass (Mw) in Da. The relative mobility was calculated as follows:

$$\text{Relative Mobility} = \frac{\text{Migration of the respective protein band (cm)}}{\text{Final migration of Bromophenol Blue dye front (cm)}}$$

The relative mobility of *Rf*GH5_4 (represented as '*y*' in **Fig. 2.5**) was derived using the Hedrick plot to find it respective Log_{10} [Mw]. The final molecular mass was deduced by taking the Antilog of the calculated Log_{10} [Mw] value for *Rf*GH5_4.

2.2.20 Protein estimation by Bradford method

Bradford method was used for estimating the concentration of purified protein by using Bovine Serum Albumin (BSA) as the standard (Bradford, 1976). Bovine serum albumin (BSA) purchased from Sigma-Aldrich Co. LLC, USA. was used as standard protein. Twenty (20) µL of BSA within concentration range, 15-180 µg/mL were incubated with 200 µL of Bradford reagent for 15 min in dark environment in the 96-well titre plate. A standard plot of OD at 595 nm versus different concentration of BSA, as mentioned above, was prepared to calculate the 1 OD equivalent µg/mL of the BSA protein. The protein concentration of the purified protein samples was calculated in the

similar way. All the reactions were performed in triplicates and the mean of it was plotted.

Preparation of Bradford reagent:

Bradford assay estimates the amount of protein in a solution by using the spectral properties of Coomassie Brilliant Blue G-250 (Bradford, 1976). 100 mg (Final 0.01%, w/v) Coomassie Brilliant Blue G-250 was weighed and dissolved in 50 ml 95% ethanol(in an amber colour bottle). 100 mL 85% (w/v) phosphoric acid was added to it. A magnetic bead was placed inside the bottle and the contents were mixed properly by keeping on magnetic stirrer until the dye completely dissolved. The dye was finally diluted to 1000 mL with deionized water, filtered (Whatman, No. 1 paper) under dark conditions and stored at 4°C. Commercial Bradford reagent (Sigma-Aldrich Co. LLC, USA) was also utilized for the protein content determination. The amount of recombinant protein was estimated using the following equation,

$$[Protein] = \frac{\Delta A595 \times V \times C}{v}$$

Where:
A_{595} *= Change in the OD of the Sample against Blank*
V = Volume of the Protein-buffer mixture (mL)
C = 1 OD Equivalent of the BSA from standard plot (mg/mL)
v *= Volume of the enzyme solution used for assay (mL).*

2.2.21 Protein estimation by UV-visible spectrophotometer

The concentration of purified protein, *Rf*GH5_4 was also determined from its corresponding absorbance at 280 nm (A_{280}) using the equation below (Layne 1957; Stoscheck 1990). The absorbance was measured after proper dilution of the protein using a spectrophotometer (Gene Quant, GE Healthcare UK Ltd., UK) having a path

length of 1 cm. The molar extinction co-efficient 78630 $M^{-1}cm^{-1}$ for *Rf*GH5_4 was used for calculating the concentration.

$$[\,Rf\text{GH5\_4}] = \frac{A280 \times Molecular\ weight\ (Dalton) \times Dilution\ Factor}{Extinction\ Coefficient\ (M^{-1}cm^{-1}) \times Path\ Length\ (1\ cm)}$$

2.2.22 Purified Protein per gram of dry cell weight of expressed pellet.

In order to calculate the amount of protein (mg) per gram (g) of dry cell weight, the total cell pellet obtained after centrifugation of 100 mL *E. coli* BL21 (DE3) cells containing expressed protein of interest (*Rf*GH5_4) was weighed. Total volume (mL) and concentration (mg/mL) of eluted purified protein was recorded as mentioned in **Section 2.2.20 and 2.2.21**. The total mg of protein recorded was divided by total g of dry cell pellet to obtain mg of protein/g of dry cell weight (mg/g).

2.3 Results and Discussion

2.3.1 Sequence analysis of RfGH5_4 in the molecular architecture of RfGH5$_{1/2}$

The complete genome sequence analysis of *R. flavefaciens* FD-1 v3 reported 14 glycoside hydrolases of family GH5 (Berg Miller *et al.*, 2009). As shown in **Fig. 1.10** of Chapter 1, *Rf*GH5_4 is located at the N-terminal of *Rf*GH5$_{1/2}$ (37-384 amino acids) after a signal peptide (1-26 amino acids) and unknown sequence (27-36 amino acids), while the C-terminal contains an endo-mannanase (*Rf*GH5_7, 527-834 amino acids) (Venditto *et al.*, 2016; Goyal *et al.*, 2019). The two catalytic modules are separated by the family 80 carbohydrate binding module (*Rf*CBM80, 407-502 amino acids). The C-terminal of *Rf*GH5$_{1/2}$ ends with a type I dockerin domain (865-947 amino acids) attached with *Rf*GH5_7 through a linker sequence (835-864 amino acids). *Rf*GH5_4 is listed in the CAZy database as a part of *Rf*GH5$_{1/2}$ (GenBank Accession No. WP_009984467.1).

2.3.2 *PCR amplification of the gene encoding Rf*GH5_4

The 1047 bp DNA sequence encoding the protein, *Rf*GH5_4 without signal peptide and linker was amplified from the genomic DNA of *R. flavefaciens* FD1 v3 by PCR using the parameters mentioned in **Section 2.2.3**. Agarose gel showed the PCR amplicon of approximate size 1.0 kb (**Fig. 2.1**). The PCR amplicon was purified from the gel using gel extraction protocol as described in **Section 2.2.6** and was stored at -20℃ for downstream cloning experiments.

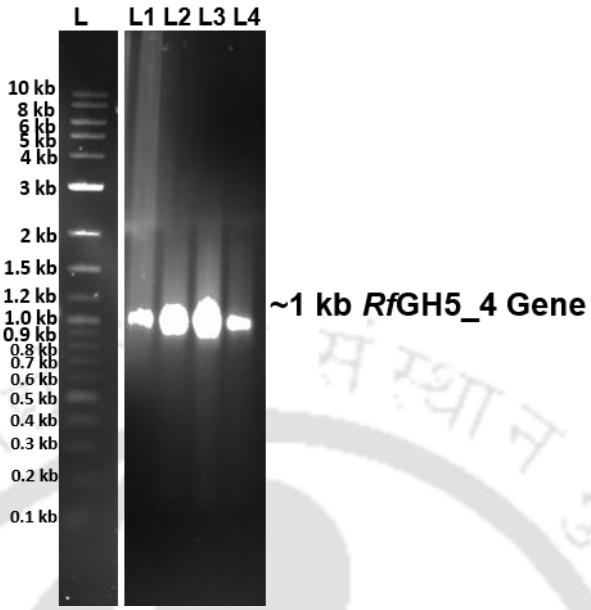

Fig. 2.1 Agarose gel showing PCR amplified fragment of amplicons of gene encoding *Rf*GH5_4. Lane L: DNA marker (0.1-10 kb DNA Ladder, *NEB*), Lane 1 to 4: PCR amplicons of *Rf*GH5_4.

2.3.3 Restriction enzyme digestion of PCR amplicon of gene encoding RfGH5_4

The 5' and 3' ends (*w.r.t* to N- to C-terminal) of the *Rf*GH5_4 PCR amplicon contained the restriction sites for *Nhe*I and *Xho*I, respectively. Thus, the amplified fragment of *Rf*GH5_4 was digested with *Nhe*I and *Xho*I restriction enzymes to generate the sticky ends. Simultaneously, pET-28a(+) vector was also digested using the same restriction enzymes. *Nhe*I and *Xho*I digested pET-28a(+) vector and *Rf*GH5_4 were visualized on 1.0% (w/v) agarose gel. The digested pET-28a(+) vector and *Rf*GH5_4 were purified from gel using extraction protocol as described in **Section 2.2.6** and used for cloning of *Rf*GH5_4 amplicon into pET-28a(+) vector.

2.3.4 Cloning, expression and purification of recombinant RfGH5_4 protein

2.3.4.1 Ligation of NheI-XhoI digested RfGH5_4 amplicons to pET-28a(+)

The digested fragments of *Rf*GH5_4 were ligated to pET-28a(+) vector using T4 DNA ligase enzyme and the resulted recombinant plasmid was used for

transformation in *E. coli* TOP10 cells. The ligation was confirmed by the digestion of isolated recombinant plasmid by *Nhe*I and *Xho*I restriction enzymes where the digested products were seen on 0.8% (w/v) agarose gel.

2.3.4.2 Transformation of E. coli TOP10 competent cells with recombinant plasmid DNA

E. coli TOP10 competent cells with transformation efficiency of $4.8x10^5$ CFUs/µg of plasmid DNA were transformed with the recombinant plasmid DNA of pET-28a(+) as mentioned in **Section 2.2.13**. The transformed cells were grown on LB agar plates containing kanamycin (50 µg/mL) and incubated at 37°C for 12 h. Total 6 colonies were observed on LB agar plate showing the transformation efficiency of $3.3x10^4$ CFUs/µg. 2 single colonies were randomly picked and grown overnight at 37°C in 10 mL LB medium in a glass test tube containing 50 µg/mL kanamycin, respectively to purify the recombinant plasmid of pET-28a(+)-*Rf*GH5_4.

2.3.4.3 Isolation of plasmid DNA and screening of positive clones

The recombinant plasmid DNA from 2 transformed colonies was extracted using the plasmid isolation protocol described in **Section 2.2.14**. The isolated plasmid DNA was shown on 1% (w/v) agarose gel. The undigested pET-28a (+) vector was also run on the agarose gel to identify the recombinant plasmids. The recombinant plasmid was digested with *Nhe*I and *Xho*I restriction enzymes to confirm the cloning of *Rf*GH5_4 in to pET-28a(+) vector. The digested recombinant plasmid was run on the agarose gel, wherein presence of two DNA bands matching the size of pET-28a(+) vector (5369 bp or 5.3 kb) and *Rf*GH5_4 (1047 bp or approx. 1.0 kb) were shown. The size of digested recombinant pET-28a(+) vector matched with the independently digested linearized version of pET-28a(+) vector at around 5.4 kb. The undigested pET-

28a(+) vector on the other hand showed two visible bands which is the characteristic pattern of circular plasmids due to coiling and supercoiling. Thus, the restriction digestion confirmed the cloning of gene encoding *Rf*GH5_4 into pET-28a(+) vector (**Fig. 2.2**). The plasmid DNA from positive clones were sequenced (Scigenom Labs Pvt. Ltd, India) and aligned with the wild type sequence, wherein no undesired mutation in the gene encoding *Rf*GH5_4 protein were detected (**Fig. 2.3**).

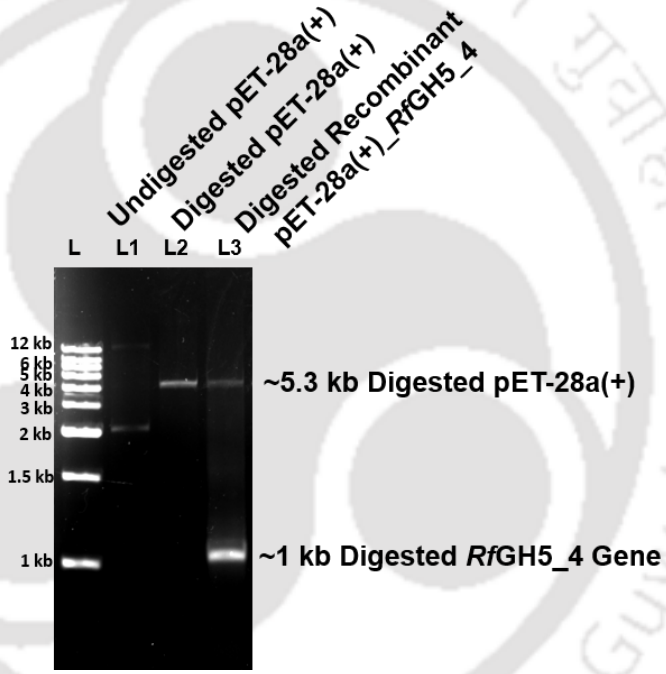

Fig. 2.2 Agarose gel showing undigested pET-28a(+) vector and *Nhe*I-*Xho*I digested pET-28a(+) as well as digested recombinant plasmid containing gene encoding *Rf*GH5_4. Lane L: DNA marker (12 kb DNA Ladder, BioBharati LifeSciences), Lane 1: Undigested pET-28a(+), L2: Digested pET-28a(+) and L3: Digested recombinant pET-28a(+)_*Rf*GH5_4 plasmid showing two distinct bands of pET-28a(+) and *Rf*GH5_4.

a > DNA Sequence of gene encoding *Rf*GH5_4 protein by T7-Forward Primer

GTAGAGCGGTCATTCCCCTCTAGAAATAATTTTGTTTACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCA
TCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCGCTTCCAACATGACCGCAAGCC
AGATCACTCAGGAAATGAAGATCGGCTGGAACCTCGGCAACACACTTGACGCATACGGCGGAACAGCAAAGG
GTGTCGCTACCGAGACATCATGGGGCAATCCAAAGACAACTAAGGCTATGATCGATGCTGTAAAGGCTAAGGGC
TTCAACACAGTTCGTATCCCTACAACATGGTTCCCCCATCTCGACGGAAGCAATAATATCGATTCCGCATGGATGG
CACGTGTAAAGGAAGTAGTTGACTACTGTATCGACAACGATATGTACGTTATCCTCAACCTCCACCATGAAGAAT
GGGTAGACCGTCCTGATCTTGGTTCAGCTTACAACGAGATGCAGCCAAAGTTCACAAAGATATGGTCACAGATC
GCTGATGCATTCAAGGATTACGATCAGCACCTTATCTTCGAGAGCATGAACGAGCCAAGAGCAAAGGGCACCG
ACCATGAATGGTGGGGACCTCAGCAGAGCGAAGTTGATACCATCAACAAGCTCAATCAGGACTTCGTTAAGCTC
ATCAGAAGCTCAGGCTCCGCAAACAACAAGAACCGTCTTCTCATGATCCCGGGTTACTGTGCATCCGCTGACCA
GACAATGTTCTCAAAGATCCAGGTACCTTCAGACAATATGGTCGCTATTTCCATCCACGCATATGTACCTTACGAC
TTCACTATGAATACTGCTGTTTCTGACCACAGCACATTCTCAACCAAGTATTCAAGCAGCCTTAACCAGACTCTCG
AAAGCATCAGAAATACTTTCACTTCAAAGGGAAATCCCTGTTGTTATCGGCGAATTCGGTACTTCTAACTTCGGT
AACACCGAGGCAAGAGTAAAGTGGCTGAACAGTATATGAGAACACTAAGAAATGGGTATCCCCTGTGTACTCTG
GGACAACACGTTATCAATGCACTTCATCAGCAGTGAATGCCACGGCTATCTCACAGAGCAATCTTACATGGTACA
CGAAAGCGACTGTGTCACAGATGATGAGTACTCGAGTATACTCAGCTCGGCTGAGTTCGTCGACAGCCTGCCGC
CGCAATCGGAACTGATCATCAGCATGCATTGAAAATCCGACTGCTAACAAGTCCGGAAAGAAAAGCGTTGATTG
AGC

b >DNA Sequence of gene encoding *Rf*GH5_4 protein by T7-Reverse Primer

ATCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAG
CTTGTCGACGGACCTCAGCGGAAGCTGAGGTTATACTCCGAGTACTTCCATCATCTTGTTGACAACAGGTTCGC
TTTCGGTGTACCATGTAAGATTGCTTCTGTTGAGATAGCCGTGGCATTCACCTGCTGATGAAGGTGCATTGATAA
CGTTGTTGTCCCAGAGTACACAGGGGATACCCATTTCCTTAGTTGTTCTCATATACTGTTCAGCCCACTTTACTCT
TGCCTCGGTGTTACCGAAGTTAGAAGTACCGAATTCGCCGATAACAACAGGGATTCCCTTTGAAGTGAAAGTAT
TTCTGATGCTTTCGAGAGTCTGGTTAAGGCTGCTTGAATACTTGGTTGAGAATGTGCTGTGGTCAGAAACAGCA
GTATTCATAGTGAAGTCGTAAGGTACATATGCGTGGATGGAAATAGCGACCATATTGTCTGAAGGTACCTGGATC
TTTGAGAACATTGTCTGGTCAGCGGATGCACAGTAACCCGGGATCATGAGAAGACGGTTCTTGTTGTTTGCGG
AGCCTGAGCTTCTGATGAGCTTAACGAAGTCCTGATTGAGCTTGTTGATGGTATCAACTTCGCTCTGCTGAGGTC
CCCACCATTCATGGTCGGTGCCCTTTGCTCTTGGCTCGTTCATGCTCTCGAAGATAAGGTGCTGATCGTAATCCTT
GAATGCATCAGCGATCTGTGACCATATCTTTGTGAACTTTGGCTGCATCTCGTTGTAAGCTGAACCAAGATCAGG
ACGGTCTACCCATTCTTCATGGTGGAGGTTGAGGATAACGTACATATCGTTGTCGATACAGTAGTCAACTACTTCC
TTTACACGTGCCATCCATGCGGAATCGATATTATTGCTTCCGTCGAGATGGGGGAACCATGTTGTAGGGATACGA
ACTGTGTTGAAGCCCTTAGCCTTTACAGCATCGATCATAGCCTTAGTTGTCTTTGGATTGCCCATGATGTCTCGGT
AGCGACACCCTTTGCTGTCGCGTATGCGTCAAGTGTGTGCCGAGGTTCAGCGATCTCATTTCTGATGATCTGCTG
CCGTCATGTGAGCCCTCAGCCCTGCTGAAGCGCTGCCTGGATGATGATGATGATGCTGCTGCATGGCAATCTCTC
TAAGGTAACAAATATTTCCTAAAGAGGGGCAAAG

c >Final Deduced DNA Sequence of gene encoding *RfGH5_4* protein (5'-3')

GTAGAGCGGTCATTCCCCTCTAGAAATAATTTTGTTTACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCA
TCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCGCTTCCAACATGACCGCAAGCC
AGATCACTCAGGAAATGAAGATCGGCTGGAACCTCGGCAACACACTTGACGCATACGGCGGAACAGCAAAGG
GTGTCGCTACCGAGACATCATGGGGCAATCCAAAGACAACTAAGGCTATGATCGATGCTGTAAAGGCTAAGGG
CTTCAACACAGTTCGTATCCCTACAACATGGTTCCCCCATCTCGACGGAAGCAATAATATCGATTCCGCATGGATG
GCACGTGTAAAGGAAGTAGTTGACTACTGTATCGACAACGATATGTACGTTATCCTCAACCTCCACCATGAAGAA
TGGGTAGACCGTCCTGATCTTGGTTCAGCTTACAACGAGATGCAGCCAAAGTTCACAAAGATATGGTCACAGAT
CGCTGATGCATTCAAGGATTACGATCAGCACCTTATCTTCGAGAGCATGAACGAGCCAAGAGCAAAGGGCACC
GACCATGAATGGTGGGGACCTCAGCAGAGCGAAGTTGATACCATCAACAAGCTCAATCAGGACTTCGTTAAGC
TCATCAGAAGCTCAGGCTCCGCAAACAACAAGAACCGTCTTCTCATGATCCCGGGTTACTGTGCATCCGCTGAC
CAGACAATGTTCTCAAAGATCCAGGTACCTTCAGACAATATGGTCGCTATTTCCATCCACGCATATGTACCTTACG
ACTTCACTATGAATACTGCTGTTTCTGACCACAGCACATTCTCAACCAAGTATTCAAGCAGCCTTAACCAGACTCT
CGAAAGCATCAGAAATACTTTCACTTCAAAGGGAATCCCTGTTGTTATCGGCGAATTCGGTACTTCTAACTTCGG
TAACACCGAGGCAAGAGTAAAGTGGGCTGAACAGTATATGAGAACAACTAAGGAAATGGGTATCCCCTGTGTA
CTCTGGGACAACAACGTTATCAATGCACCTTCATCAGCAGGTGAATGCCACGGCTATCTCAACAGAAGCAATCT
TACATGGTACACCGAAAGCGAACCTGTTGTCAACAAGATGATGGAAGTACTCGGAGTATAACCTCAGCTTCCGC
TGAGGTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAA
AGCCCGAAAGAAGCGATGTCTA

Fig. 2.3 DNA Sequencing results of gene encoding *Rf*GH5_4 recorded using (a) T7-forward and (b) T7-reverse primers. The figure (c) shows final 5'-3' deduced sequence of *Rf*GH5_4.

2.3.5 Expression and purification of recombinant protein

The *E. coli* BL21 (DE3) competent cells were transformed with the recombinant pET-28a(+) plasmids having the gene encoding *Rf*GH5_4 protein. With 150 transformants observed on the LB agar plate, the transformation efficiencies of *E. coli* BL21 (DE3) cells was $8.2x10^5$ CFUs/µg. The colonies were selected randomly and grown in 10 mL LB medium containing 50 µg/mL Kanamycin at 37°C, as described in **Section 2.2.17**. The cells (1%, v/v) from freshly grown 37°C inoculum was used for in a 100 mL LB medium. The cells were induced for expression of recombinant protein, *Rf*GH5_4 with 1 mM IPTG when the cell density showed the Absorbance, A_{600} of 0.4-0.6 which indicated the cells are in mid exponential phase. The cells were further grown at 24°C and 180 rpm for 16-18 h and then pelleted down at 6000 x g and 4°C for 15 min. The pellets were store at -20°C until further purification process. The recombinant proteins were purified by immobilised metal ion affinity chromatography (IMAC) as described in **Section 2.2.19** and then dialysed for removal of imidazole and sodium chloride. The purified protein, *Rf*GH5_4 was further analysed by SDS-PAGE using a 12% (w/v) gel. The theoretical molecular mass of *Rf*GH5_4 was calculated to be 41.18 kDa which is in corroboration with that observed on the SDS-PAGE gels (**Fig. 2.4**).

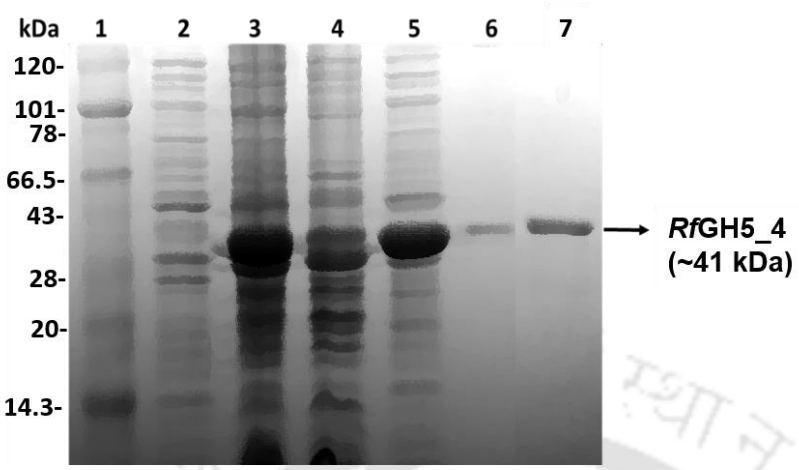

Fig. 2.4 SDS-PAGE analysis of *Rf*GH5_4 purification steps on 12% (w/v) gel. Lanes, 1- Protein mass marker, 2 - Un-Induced *E. coli* BL21 (DE3) cells, 3: Induced *E. coli* BL21 (DE3) cells at 1 mM IPTG, 4 - Induced Cell Pellet (after sonication), 5 - Cell-Free Extract, 6 - Last Column Wash, 7 - Showing band of purified protein *Rf*GH5_4 (*Approx*. 41 kDa).

2.3.6 Determination of molecular weight by Hedrick plot

The relative mobility of *Rf*GH5_4 was also calculated from Hedrick plot which was deduced to be approx. 40.63 kDa (**Fig. 2.5**) and found in agreement with its theoretical molecular mass (41.18 kDa).

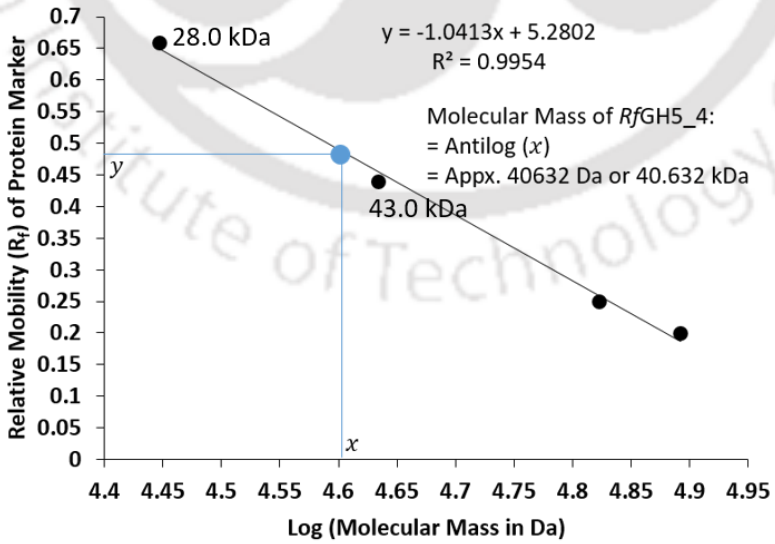

Fig. 2.5 Calculation of the approximate molecular mass of *Rf*GH5_4 protein using its relative mobility on the 12% (w/v) SDS-PAGE on the Hedrick plot.

2.3.7 Protein estimation of expressed and purified RfGH5_4 enzyme

The amount of purified recombinant protein obtained from 100 mL of grown *E. coli* BL21(DE3) cultures was estimated using the standard curve of BSA generated by Bradford method (**Fig. 2.6**) as well as UV-Visible spectroscopy by recording the OD at 595 (OD_{595}) and 280 nm (A_{280}), respectively, as discussed in **Section 2.2.20, 2.2.21 and** shown in **Table 2.15**. The concentration of purified protein by UV-Visible method was found to be 1.74 mg/mL which corroborated with that of Bradford method, whereas the total amount of recombinant purified protein present in eluted 4 mL fractions was 6.97 mg.

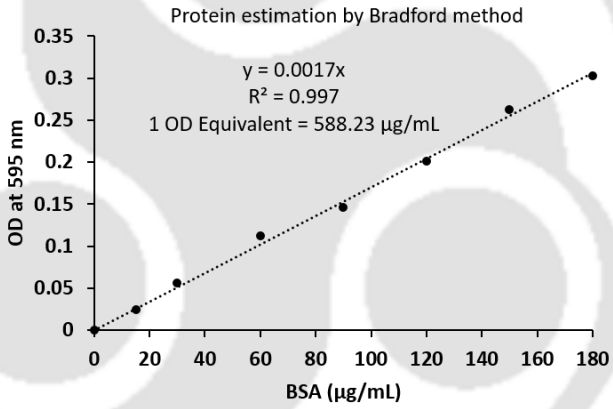

Fig. 2.6 Standard curve of BSA by Bradford method (Bradford, 1976) was generated for the protein estimation of the purified samples of *Rf*GH5_4.

Table 2.15. Purified recombinant proteins obtained from 100 mL culture.

Recombinant Protein	Protein Concentration (mg/mL)	Volume of the purified protein (mL)	Total yield of the purified protein (mg/100 mL of expressed cells)	mg of Protein per g of dry cell weight (mg/g)
*Rf*GH5_4	1.74	4.0	6.97	14.0

The dry cell weight of the 100 mL expressed cell pellet was 0.5 g. So, the total purified protein obtained was 14.0 mg/g of cell dry weight.

2.4 Conclusion

The gene encoding *Rf*GH5_4 (GenBank Accession Number - WP_009984467.1) endoglucanase was cloned from bacterium *R. flavefaciens* FD-1 v3. *Rf*GH5_4 is part of the wild-type multimodular complex, *Rf*GH5$_{1/2}$ placed at its N-terminal. The PCR amplified fragment of gene encoding *Rf*GH5_4 showed a band of approx. 1.0 kb (1047 bp), The *Nhe*I-*Xho*I digested PCR DNA fragment and linearized pET-28a(+) were ligated using T4 DNA ligase. The ligated mixture was transformed into *E. coli* TOP10 cells and colonies were evaluated for positive clone confirmation through restriction digestion of the isolated recombinant plasmid. The recombinant plasmid was then transformed into *E. coli* BL21 (DE3) expression cells. The positive colonies were screened for expression of protein of interest, *Rf*GH5_4 IPTG induction. The recombinant protein, *Rf*GH5_4 was purified to homogeneity using the Affinity Chromatography (IMAC) which yielded total 6.97 mg of the protein at 1.74 mg/mL concentration. The total purified protein obtained from the culture pellet of 100 mL expressed *E. coli* BL21 (DE3) cells was 14.0 mg/g of dry cell weight. The SDS-PAGE analysis of the purified *Rf*GH5_4 protein on 12% (w/v) gel showed the molecular weight of approximately 41 kDa which corroborated to that of its theoretical estimate. The purified *Rf*GH5_4 protein was used for further biochemical, functional and structural characterization.

References

Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology* 12:1–16

Berg Miller ME, Antonopoulos DA, Rincon MT, Band M, Bari A, Akraiko T, Hernandez A, Thimmapuram J, Henrissat B, Coutinho PM (2009) Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PLoS One* 4:e6650

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Annals of Biochemistry* 72:248–254

Bryant MP, Small N, Bouma C, Robinson IM (1958) Characteristics of ruminal anaerobic cellulolytic cocci and *Cillobacterium cellulosolvens* n. sp. *Journal of Bacteriology* 76:529–537

Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research* 50:D571–D577

Engler MJ, Richardson CC (1982) 1 DNA Ligases. In: *The enzymes. Elsevier*, pp 3–29

Hames BD, (1998) Gel electrophoresis of proteins a practical approach, Editor, *OUP Oxford*; 197: 3rd Edition

Hedrick JL and Smith AJ (1968) Size and charge isomer separation and estimation of molecular weights of proteins by disc gel electrophoresis. *Archives of Biochemistry and Biophysics* 126:155-64

Goyal D, Kumar K, Centeno MSJ, Thakur A, Pires VMR, Bule P, Fontes CMGA, Goyal A (2019) Molecular cloning, expression and biochemical characterization of a family 5 glycoside hydrolase first endo-mannanase (*Rf*GH5_7) from *Ruminococcus flavefaciens* FD-1 v3. *Molecular Biotechnology* 61:826–835

Kumar K, Singal S, Goyal A (2019) Role of carbohydrate binding module (CBM3c) of GH9 β-1, 4 endoglucanase (Cel9W) from *Hungateiclostridium thermocellum* ATCC 27405 in catalysis. *Carbohydrate Research* 484:107782

Laemmli UK (1970) SDS-page Laemmli method. *Nature* 227:680–685

Layne E (1957) Spectrophotometric and turbidimetric methods for measuring proteins, *Elsevier* 71:447-454.

Liu P, Li A, Wang Y, Cai Q, Yu H, Li Y, Peng H, Li Q, Wang Y, Wei X (2021) Distinct Miscanthus lignocellulose improves fungus secreting cellulases and xylanases for consistently enhanced biomass saccharification of diverse bioenergy crops. *Renewable Energy* 174:799–809

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*. 42(D1):D490-5

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews* 66:506–577

Rincon MT, Dassa B, Flint HJ, Travis AJ, Jindou S, Borovok I, Lamed R, Bayer EA, Henrissat B, Coutinho PM (2010) Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS One* 5:e12476

Russell DW, Sambrook J (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Cold Spring Harbor, NY

Singh S, Dhillon A, Goyal A (2020) Enhanced catalytic efficiency of *Bacillus amyloliquefaciens* SS35 endoglucanase by ultraviolet directed evolution and mutation analysis. *Renewable Energy* 151:1124–1133.

Stoscheck CM (1990) Quantitation of protein. In: *Methods in enzymology*. Elsevier, 6: pp 50–68

Venditto I, Luis AS, Rydahl M, Schückel J, Fernandes VO, Vidal-Melgosa S, Bule P, Goyal A, Pires VMR, Dourado CG (2016) Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proceedings of the National Academy of Sciences* 113:7136–7141

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 3

Biochemical and functional characterization of highly efficient, processive and multifunctional endoglucanase, *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 v3

3.1 Introduction

The conversion of lignocellulosic biomass to bioethanol currently faces a bottleneck which is the unavailability of catalytically efficient and stable endoglucanases. Moreover, the lignocellulosic biomass contains hemicellulose along with the cellulose. It is highly advantageous if the endoglucanase hydrolyzes more than one cellulosic and hemicellulosic substrate. Earlier, an alkali stable bifunctional β-glucanase, *Pgl5A* of subfamily 4 of family GH5 from *Paenibacillus* sp. S09 was characterized that showed 94.5 U/mg of specific activity on tamarind xyloglucan, however, its activity against CMC-Na was negligible, 0.5 U/mg (Cheng *et al.*, 2019). Another multifunctional alkali tolerant family GH5 endoglucanase, *Thrcel5A*, from *Thermoactinospora rubra YIM 77501T* was reported to show 85.7 U/mg specific activity against CMC-Na and 22.9 U/mg against beechwood xylan, respectively at the

optimum pH of 8.5 and 60℃ (Yin *et al.*, 2020). An endoglucanase *SoCel5* from bacterium *Stegonsporium opalus* from family GH5 showed a specific activity of 350 U/mg and catalytic efficiency of 77 $mL.mg^{-1}s^{-1}$ on carboxymethyl cellulose (CMC-Na) at pH 6 and 60°C (Zheng *et al.*, 2019). Another endoglucanase, *CelR5* of family GH5 from rhizosphere metagenomic library showed the specific activity of 15 U/mg and k_{cat} of 9.7 s^{-1} against CMC-Na (Wierzbicka-Woś *et al.*, 2019). Moreover, a thermophilic microbial consortium developed on rice straw from vermicompost reported only 20 U/mg of the endoglucanase activity (Gavande *et al.*, 2021). However, the lignocellulose deconstruction demands more efficient endoglucanases with higher catalytic efficiency, having a broad range of substrate specificity, pH stability and thermostability. Multifunctional and efficient endoglucanases could reduce the cost of bioethanol production, increase the productivity, thereby by making the bioethanol industry economically sustainable. The aim of this study was to determine biochemical and functional properties of an endoglucanase, *Rf*GH_4 from a ruminant gut bacterium *Ruminococcus flavefaciens* FD-1 v3. The gene encoding *Rf*GH_4 was cloned, expressed and purified in *E. coli* BL21(DE3) cells, as described in Chapter 2.

This chapter describes the biochemical parameters such as pH, temperature, stability and kinetic parameters like K_m, V_{max}, k_{cat}, catalytic efficiency (k_{cat}/K_m) of the endoglucanase (*Rf*GH_4). These properties play a pivotal role in evaluating the importance of a cellulase in the renewable energy sector. The study also demonstrates the multifunctionality of *Rf*GH_4 on various lignocellulosic substrates. The mode of processive hydrolysis of amorphous cellulose by *Rf*GH5_4 was also deciphered.

3.2 Materials and Methods

3.2.1 The substrates, chemicals and kits

Barley β-D-Glucan, sodium salt of carboxymethyl-cellulose (CMC-Na), Avicel®, birchwood xylan, konjac glucomannan, carob galactomannan, D-glucose, D-xylose, D-mannose and dihydroxybenzoic acid (DHB) were procured from Sigma-Aldrich Co. LLC., USA. Isopropyl-β-D thiogalactopyranoside (IPTG), Kanamycin, hydroxyethyl cellulose (HEC), lichenan were also purchased from Sigma-Aldrich Co. LLC., USA. Xyloglucan was purchased from Megazyme International, Ireland. Beechwood xylan was purchased from SRL India Pvt. Ltd. Cellulose powder, ethylene diamine tetraacetic acid (EDTA), ethylene glycol tetraacetic acid (EGTA), 2-(*N*-morpholino) ethane sulfonic acid (MES) buffer, potassium sodium tartrate, sodium bicarbonate, sulphuric acid, acetic acid glacial, extrapure ethanol and protein dialysis tube (pore size, 12-14 kDa) were purchased from Hi-Media Laboratories Pvt. Ltd., India. Phosphoric Acid Swollen Cellulose (PASC), also known as (a.k.a.) Regenerated Amorphous Cellulose (RAC), was prepared from Avicel® (henceforth referred to as Avicel). Sodium carbonate anhydrous, ammonium molybdate, phenol, ortho-phosphoric acid (minimum, 85%) was procured from Merck Life Sciences Pvt. Ltd., India. n-Butanol was obtained from Sisco Research Laboratories Pvt. Ltd., India. Sodium sulphate and sodium arsenate was purchased from Thermo Fisher Scientific India Pvt. Ltd., India. The GOD-POD kit for D-glucose estimation was procured from Coral Clinical Systems (Tulip Diagnostics Pvt. Ltd.), India. The thin layer chromatography plate (TLC Silica gel 60 F_{254}, 20 × 20 cm) was purchased from Merck KGaA, Darmstadt, Germany.

3.2.2 Enzyme assay of RfGH5_4

The reaction mixture (100 µL) consisted of 1% (w/v) of the polysaccharide as the substrate (CMC-Na, HEC, barley β-D-glucan, lichenan, xyloglucan, carob galactomannan and konjac glucomannan) dissolved in 20 mM citrate-phosphate buffer (pH 5.5) and 10 µL purified *Rf*GH5_4 enzyme (5 µg/mL). The reaction mixture was incubated at different temperatures including 55°C for an optimized time period of 2 min. For Avicel and cellulose powder, the reaction was carried out at 30°C with the rotation of 200 rpm for 2 h as reported earlier (Zhang *et al.*, 2006). A temperature of 30°C was selected for the assay of Avicel and cellulose powder since the enzyme, *Rf*GH5_4 is not thermally stable at 55°C when incubated for more than 10 min. The released reducing sugar was quantified by the earlier reported methods (Nelson, 1944; Somogyi, 1945; Somogyi, 1952) for calculating the enzyme activity of *Rf*GH5_4. Freshly prepared Solution D (100 µL, as described in **Section 3.2.2.1**) was added to 100 µL of reaction mixture followed by boiling for 20 min. 100 µL of Solution C was added to the boiled mixture and the reaction volume was made up to 1000 µL by adding 700 µL of water. The total reaction mixture was vortexed and centrifuged at 10,000***g*** for 10 min at room temperature to separate the unhydrolysed polymers from the soluble components. The Optical Density (OD) at 500 nm (A_{500}) was measured by UV-Visible spectrophotometer (GeneQuant 1300, GE Healthcare, UK). D-glucose was employed as a standard to calculate the enzyme activity of *Rf*GH5_4 for which a standard curve of D-glucose under the above mentioned identical assay conditions was generated (**Section 3.2.2.1**). All the assays were performed in triplicate sets.

3.2.2.1 Reagents for reducing sugar estimation by Nelson-Somogyi method

The reagents for estimation of reducing sugar were prepared as per the protocols given earlier (Nelson, 1944; Somogyi, 1945).

Reagent A:

Sodium carbonate anhydrous (Na_2CO_3) : 6.25 g
Potassium sodium tartrate ($C_4H_4O_6KNa \cdot 4H_2O$): 6.25 g
Sodium bicarbonate ($NaHCO_3$): 5.0 g
Sodium sulphate anhydrous (Na_2SO_4): 50.0 g

The above mentioned chemicals were dissolved in 100 mL of the deionized water and the final volume was adjusted to 250 mL. The solution was filtered through Whatman Paper No. 1 and stored at 30°C.

Reagent B

Reagent B was prepared by dissolving 15 g of copper sulfate ($CuSO_4$) in 50 mL deionized water and one or two drops of concentrated sulphuric acid (H_2SO_4) were added to it. The final volume was made up to 100 mL with deionized water and the solution was filtered through Whatman Paper No. 1 and stored at 30°C.

Reagent C

Reagent C was prepared in two steps under dark. First of all, 2.5 g of ammonium molybdate [$(NH_4)_6Mo_7O_{24}$] was dissolved in 45 mL of deionized water in a 100 mL beaker and 2.1 mL of concentrated sulphuric acid (H_2SO_4) was added to it. In an another beaker, 0.3 g of sodium arsenate (Na_3AsO_4) was dissolved in 2.5 mL of deionized water. Now, this solution was added to ammonium molybdate solution and the contents were mixed (total volume was around 50 mL). The solution was filtered through Whatman Paper No. 1 under dark conditions and stored at 37°C. It was used after 24 h incubation.

Reagent D

This solution was prepared freshly, just before the onset of the experiment, by mixing Reagent A and with Reagent B in a ratio of 1:25 i.e. for 1 mL of Reagent A, 40 µL of Reagent B was added.

3.2.2.2 Generation of standard plot of D-glucose, D-mannose and D-xylose by Nelson-Somogyi method

The reaction products of cellulase enzymes in cellulose hydrolysis are D-Glucose and other reducing sugar equivalents like cellooligosaccharides. So, the standard plot using D-glucose was prepared by varying the concentration from 10 µg/mL to 160 µg/mL by Nelson-Somogyi method (Nelson, 1944; Somogyi, 1945). The reaction mixture (100 µL) containing D-glucose in 20 mM sodium phosphate buffer, pH 7.0 was incubated at 50°C for 15 min and then 100 µL of solution D (**Section 3.2.2.1**) was added to it. The reaction mixture was then heated in boiling water bath for 20 min and cooled. The solution C (100 µL) was added (as descirbed in **Section 3.2.2.1**) and the contents were mixed. Then 700 µL of deionized water was added to make the final volume to 1000 µL. The OD at 500 nm (OD_{500}) was measured using UV-Visible spectrophotometer (MultiSkan SkyHigh, Thermo Fisher Scientific Inc., USA) against a Blank reaction prepared without enzyme using 20 mM sodium phosphate buffer, pH 7.0. A standard plot of OD_{500} versus D-glucose concentration (µg/mL) was generated (**Fig. 3.1a**) and 1 OD equivalent of D-glucose (µg/mL) was calculated. Similarly, the standard curves of D-mannose (30 µg/mL - 100 µg/mL, **Fig. 3.1b**) and D-xylose (10 µg/mL - 210 µg/mL, **Fig. 3.1c**) were generated.

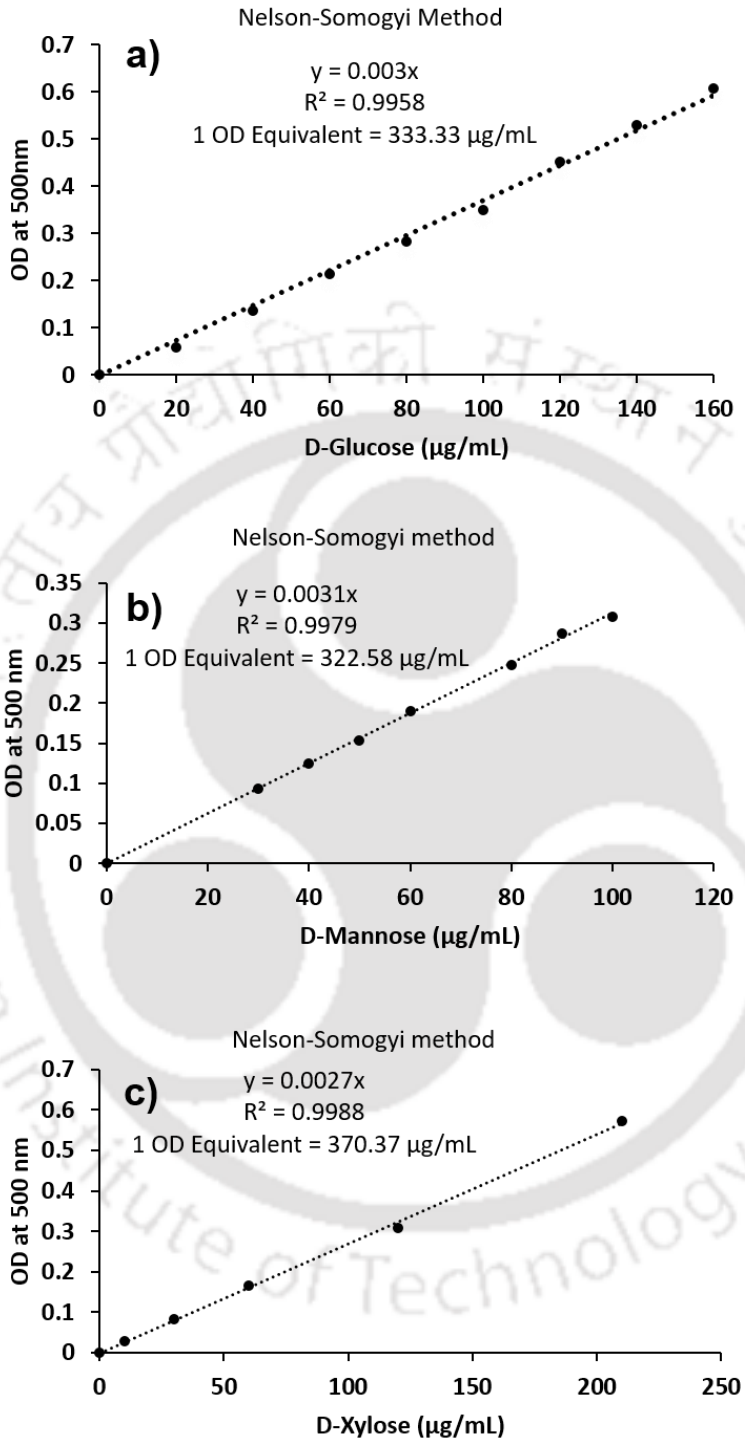

Fig. 3.1 Standard curves of a) D-glucose, b) D-mannose and c) D-xylose for estimation of reducing sugar equivalents by Nelson-Somogyi method (Nelson, 1944; Somogyi, 1945).

3.2.2.3 Standard curve of carbohydrate content by phenol-sulphuric acid method

To estimate the total carbohydrate content obtained in the in-house generated polysaccharides such as Phosphoric Acid Swollen Cellulose (PASC) a.k.a. Regenerated Amorphous Cellulose (RAC, as described in **Section 3.2.3.1**), the standard curve of carbohydrate equivalents (D-glucose) was established using Phenol-Sulphuric acid method as per the protocols reported earlier (Fox and Robyt, 1991). The detailed protocol is given below:

Protocol for carbohydrate content estimation by Phenol-Sulphuric acid method:

1. Phenol reagent was prepared in water with the final concentration of 5% (w/v).
2. D-glucose dilutions (25 µL) of various concentrations (0.1 mg/mL-5.0 mg/mL) were distributed in a clean microtitre plate. A reaction Blank containing water was set up as a reaction Control.
3. Equal volume of freshly prepared Phenol 5% (w/v, 25 µL) was added in these glucose samples.
4. The reaction mixture was gently mixed for 30 s.
5. The reaction was carefully placed on ice.
6. Concentrated sulphuric acid (98% pure, 125 µL) was added to this mixture.
7. The reaction mixture was incubated in a water bath (Grant Instruments Ltd., Cambridge, England) at 80°C for 30 min.
8. The reaction mixture was allowed to cool down and the OD_{490} nm was recorded.
9. All the reactions were performed in triplicate set and Mean ±SD was plotted to generate the standard curve of OD_{490nm} against concentration of D-glucose (mg/mL) and 1 OD equivalent was obtained (**Fig. 3.2**).

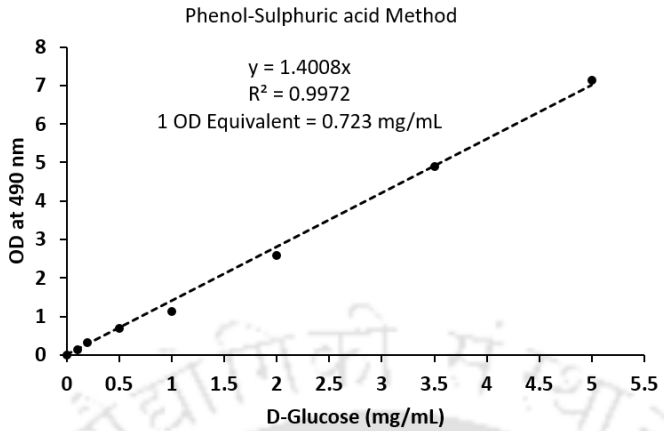

Fig. 3.2 Standard curve for estimation of Carbohydrate equivalents (D-glucose) by Phenol-Sulphuric acid method.

3.2.2.4 Standard curve of D-glucose by Glucose OxiDase-PerOxiDase (GOD-POD) method

For exclusive determination of presence of glucose in the enzyme reaction mixtures, the standard curve of D-glucose was generated by GOD-POD method using Glucose Oxidase-Peroxidase (GOD-POD) reagent L1 (Coral Clinical Systems a division of Tulip Diagnostics Pvt. Ltd., India) as per the protocol developed earlier (Raabo and Terkildsen, 1960). Sequentially, a range of varied concentration of D-glucose (100 to 800 µg/mL) solution was prepared in the deionized water. All the glucose dilutions (2 µL) were distributed in the fresh and clean plane bottom 96-well titre plate. A blank using equal volume of deionized water (2 µL) was set up. The GOD-POD reagent L1 (200 µL) was added to the reaction mixtures and Blank. The reaction mixtures were incubated at 37°C for 15 min and OD at 505 nm (OD_{505}) was recorded against the Blank. The standard curve of D-glucose was plotted for OD_{505} against the respective concentration of D-glucose to get the 1 OD equivalent of glucose (**Fig. 3.3**).

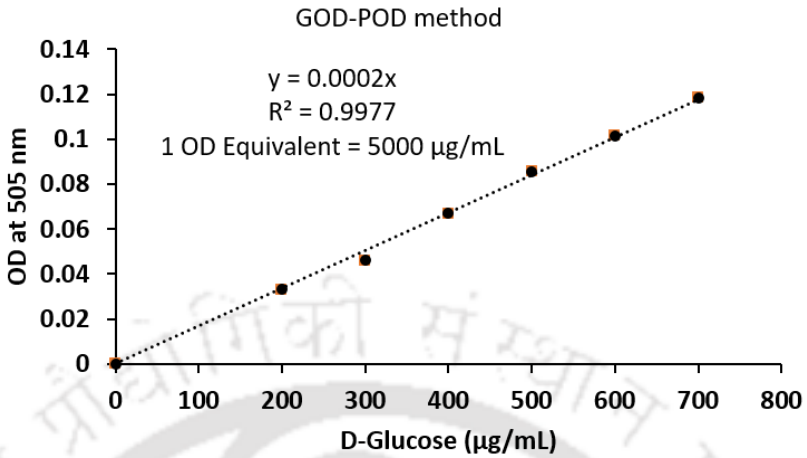

Fig. 3.3 Standard curve of D-glucose by GOD-POD method.

3.2.2.5 Calculation of enzyme activity of RfGH5_4

The enzyme activity of *Rf*GH5_4 was determined first in U/mL followed by calculating its specific activity in U/mg. Specific activity was defined as µmol of D-glucose produced from the substrate per minute (U) per mg of the enzyme (µmol/min/mg or U/mg) under the optimized reaction conditions. The enzyme activity was calculated as formulated below,

$$\text{Enzyme Activity (U/mL)} = \frac{\Delta A_{500} \times C \times V}{180 \times t \times v}$$

Where,

ΔA_{500} = Change in absorbance of the sample at 500 nm,

C = 1 OD equivalent D-glucose concentration from standard plot,

V = Volume of the reaction mixture (mL),

180 = Molecular mass of the D-glucose,

v = Volume of enzyme taken in assay (mL) for reducing sugar estimation.

3.2.3 Substrate specificity of RfGH5_4

The enzyme activity of *Rf*GH5_4 was determined against different cellulosic carbohydrate polymers like β-1,4-glucans namely, sodium salt of

Carboxymethylcellulose (CMC-Na), Hydroxymethylcellulose (HEC), Phosphoric Acid Swollen Cellulose (PASC), Avicel and cellulose powder. *Rf*GH5_4 was also tested for its activity on hemicellulosic polymers like mixed β-1,3;1,4 linked glucans (barley β-D-glucan and lichenan), tamarind xyloglucan, konjac glucomannan, carob galactomannan and xylans (Beechwood and Birchwood). Respective 1 OD reducing sugar equivalents other than D-glucose were used for calculating the enzyme activity against aforementioned noncellulosic substrates. These 1 OD reducing sugar equivalents were namely, D-xylose for xylan and D-mannose for galactomannan. The enzyme assay was performed and the specific activity was calculated as described in the **Section 3.2.2**.

3.2.3.1 Preparation of Phosphoric Acid Swollen Cellulose

To investigate the enzyme activity of *Rf*GH5_4 against amorphous cellulose, Phosphoric Acid Swollen Cellulose (PASC a.k.a. Regenerated Amorphous Cellulose, RAC) was prepared with the method given earlier (Zhang *et al.*, 2006) which is described below.

Reagents required:

Microcrystalline Avicel (Sigma-Aldrich Co. LLC, USA) - 200 mg,

Deionized water – 500 mL

Ice-cold Phosphoric acid (H_3PO_4, 86.2% v/v) – 10 mL

Sodium Carbonate (Na_2CO_3, 2 M) – 0.5 mL

Sodium Azide (0.5%, w/v) – 5 mL

Protocol for generation of PASC:

1. The microcrystalline cellulose (Avicel, 200 mg) was suspended in 0.6 mL deionized water for 5 min, in a 50 mL centrifuge tube to make it wet.
2. Ice-cold Phosphoric acid (H_3PO_4, 86.2% v/v- 10 mL) was slowly added to the above suspension followed by vigorous stirring to make the final concentration 83.2% (v/v). The cellulose mixture turned transparent after 5 min.
3. The transparent cellulose mixture was incubated on ice for 1 h with occasional stirring.
4. Ice-cold deionized water (10 mL) was added to the transparent cellulose followed by vigorous stirring. This step was repeated 4 times to add total 40 mL of the ice-cold deionized water. A whitish and cloudy precipitate of cellulose was observed after this step.
5. The cellulose mixture was centrifuged at 5000***g*** at 4°C for 20 min.
6. The supernatant was discarded and white pellet was retained.
7. The white pellet was resuspended in the ice-cold deionized water (50 mL) and stirred for 30 s.
8. The resuspended cellulose mixture was centrifuged at 5000***g*** at 4°C for 20 min and supernatant was discarded. This step was repeated 4 times to completely remove the phosphoric acid. The pellet was retained.
9. Immediately, 2 M sodium carbonate (0.5 mL) was added to neutralize the residual phosphoric acid.
10. Ice-cold deionized water (45 mL) was added to the above mixture and stirred for 30 s.

11. The resuspended cellulose mixture was centrifuged at 5000***g*** at 4℃ for 20 min and supernatant was discarded. This step was repeated 2 times or until the pH reached 5.0 to 7.0. The final pellet was retained.
12. Deionized water (5 mL) to the regenerated pellet to make the cellulose slurry.
13. The regenerated homogenous cellulose slurry (PASC or RAC) was stored at 4℃ in the refrigerator.
14. The total concentration of carbohydrate content of the regenerated PASC was estimated by Phenol-Sulphuric acid assay as described in **Section 3.2.2.3.** The total concentration of prepared PASC was estimated to be 10 mg/mL.
15. Sodium azide (0.5%. w/v) was added to the PASC slurry to prevent fungal contamination.
16. The prepared PASC was stable for weeks at 4℃ which was vortexed before dispensing for the experiments as required and used in the biochemical characterization of *Rf*GH5_4 enzyme.

3.2.4 Biochemical characterization of RfGH5_4

The effect of pH on the enzyme activity of *Rf*GH5_4 was studied by using 20 mM buffers of the following pH range: sodium citrate-phosphate (pH 3.0-7.0), sodium phosphate (pH 5.8-8.0). The reaction mixture (100 µL) of different pHs containing 1% (w/v) CMC-Na and 10 µL (5 µg/mL) of the enzyme was incubated at 55℃ for 2 min. The released reducing sugar was estimated as described in **Section 3.2.2** and the enzyme activity was calculated.

The effect of temperature on the activity of *Rf*GH5_4 was studied by incubating the 100 µL reaction mixture in 20 mM citrate-phosphate buffer (pH 5.5) for 2 min and by varying from temperature from 30℃ to 80℃.

The pH stability of *Rf*GH5_4 was performed by incubating the 100 µL of enzyme at various pH in the buffers: 20 mM sodium citrate-phosphate (pH 3.0-7.0), 20 mM sodium phosphate (pH 5.8-8.0), 20 mM MES (pH 5.5-6.7) and 20 mM Tris-HCl (pH 7.5-9.0), at 30°C for 60 min. The enzyme (10 µL) was taken in 100 µL reaction mixture and was assayed at an optimum pH (5.5) and temperature (55°C) for 2 min. The thermostability of *Rf*GH5_4 was analyzed by incubating 100 µL of the enzyme (5 µg/mL) at different temperatures (5°C-70°C) for 60 min. 10 µL of this incubated enzyme was used for assay in a 100 µL reaction mixture as mentioned earlier to calculate the specific activity at optimum pH (5.5) and temperature (55°C). The highest enzyme activity obtained in pH stability and thermostability experiments was considered as 100%, in order to calculate the relative pH stability and thermostability of *Rf*GH5_4, respectively. All the assays were performed in triplicate sets.

3.2.5 Kinetic parameter analysis of RfGH5_4 against polysaccharides

The kinetic parameters of *Rf*GH5_4 were evaluated by the Michaelis-Menten equation under steady-state conditions using different substrates. The enzyme (5 µg/mL) was incubated with different concentration of carbohydrate substrates, namely, barley β-D-glucan, CMC-Na, tamarind xyloglucan, lichenan and konjac glucomannan in a 100 µL reaction volume under 20 mM citrate-phosphate buffer (pH 5.5) at 55°C temperature for 2 min. The equivalent concentration of substrates was taken as blank and all the reactions were performed in triplicates. The enzyme activity was calculated as mentioned in **Section 3.2.2.** A double-reciprocal plot i.e. Lineweaver-Burk plot (Lineweaver and Burk 1934) of the respective Michaelis-Menten equation (Johnson and Goody 2011) was generated by using GraphPad Prism 6 to determine the kinetic

parameters, namely K_m, V_{max}, turnover number (k_{cat}) and catalytic efficiency (k_{cat}/K_m, $mL.mg^{-1}.s^{-1}$).

3.2.6 Effect of metal ions, chelating agents and additive on RfGH5_4 activity

The enzyme, *Rf*GH5_4 (5 µg/mL final concentration) in a 100 µL reaction mixture volume containing 1%, w/v CMC-Na in 20 mM citrate-phosphate buffer (pH 5.5) and 2, 5 or 10 mM concentration of metal ions or additives/chelating agents (EDTA, EGTA, SDS, Urea or guanidine hydrochloride) was incubated at 55℃ and the reducing sugar was estimated by following the protocol reported earlier (Sharma *et al.*, 2018). The metal ions used for the assay were: Mg^{2+} ($MgCl_2$), Ca^{2+} ($CaCl_2$), Na^+ (NaCl), K^+ (KCl), Mn^{2+} ($MnCl_2$), Li^+ (LiCl), Co^{2+} ($CoCl_2$), Ni^{2+} ($NiSO_4$), Fe^{2+} ($FeSO_4$), Zn^{2+} ($ZnCl_2$) and Cu^{2+} ($CuSO_4$). The effect of urea (100 mM), guanidine hydrochloride (10 mM) and SDS (0.1%, w/v), Triton X-100 (0.5%, v/v), Tween-80 (1%, v/v), dimethyl sulfoxide (DMSO, 1%, v/v) and glycerol (1, 2, 4, 8 and 10%, v/v) on *Rf*GH5_4 enzyme activity was also studied. A blank for each reaction containing 1% (w/v) CMC-Na without any additive was used. The enzyme activity of *Rf*GH5_4 under optimal reaction conditions, without any additive was considered as 100%, to calculate the relative activity of enzyme with additive.

3.2.7 Protein melting analysis of RfGH5_4

The temperature at which *Rf*GH5_4 melts was deduced by subjecting 50 µg/mL of enzyme in a freshly prepared 20 mM MES buffer, pH 5.5 to increasing temperature from 25℃ to 100℃ with an increment of 1℃ per min on a UV-visible spectrophotometer (Varian, Cary 100 Bio) attached with the Peltier temperature controller. The change in absorbance at 280 nm (A_{280}) was recorded at varying temperatures, thereby generating a curve of absorbance against the temperature. The

effect of 10 mM K^+ or Ca^{2+} ions on the stability and melting temperature of *Rf*GH5_4 was also analysed. An independent experiment was carried out by incubating the enzyme at different temperatures in the presence of 10 mM EDTA to examine its effect on the enzyme structural properties.

3.2.8 Catalytic mechanism of RfGH5_4

The hydrolyzed products of CMC-Na by *Rf*GH5_4 were analyzed using TLC to unveil the mode of action of enzyme by following the standard protocol as reported earlier (Kumar *et al.,* 2019). 100 µL reaction mixture containing 90 µL of 1 % (w/v) CMC-Na dissolved in 20 mM citrate-phosphate buffer (pH 5.5) and 10 µL of *Rf*GH5_4 (50 µg/mL) was incubated at 30°C for different time intervals (1, 2, 5, 10, 15, 30, 45 and 60 min as well as 2, 4, 6, 12 and 24 h). The reaction was stopped with 3X reaction volume of 100% ethanol (300 µL) followed by centrifugation at 10000***g*** for 10 min at room temperature to pellet down the unhydrolysed CMC-Na polymers. The alcoholic supernatant containing released reducing sugars (cellooligosaccharides length) was recovered in a fresh 1.5 mL microcentrifuge tube and kept for drying at 50°C for 12 h. The dried residues were dissolved in a final volume of 10 µL of deionized water. Finally, 1 µL of this concentrated hydrolysate of each reaction and 0.5 µL (10 mg/mL) mixture of standards (D-glucose, cellobiose, cellotriose and cellotetraose) were loaded on to the TLC plate. The TLC plates were dried and kept in glass developing chamber. The chromatogram was developed as reported previously (Kumar *et al.,* 2019), wherein *n*-butanol:acetic acid:water in the ratio of 2:1:1 was used as the mobile phase. The developed TLC plate was dried and the oligosaccharides generated by *Rf*GH5_4 were visualized by spraying the visualizing solution (sulphuric acid:methanol, 5:95%, v/v and α-naphthol, 0.5%, w/v) followed by drying in the hot air oven at 70°C. The 2 µL of

the hydrolysate from each reaction was subjected to estimation of reducing sugar using the method as described in **Section 3.2.2.2**.

3.2.9 MALDI-TOF MS and TLC analyses of RfGH5_4 hydrolyzed products

The hydrolysis of various cellulosic and hemicellulosic substrates by *Rf*GH5_4 was confirmed by Matrix-Assisted Laser Desorption Ionization – Time-of-Flight Mass Spectrometry (MALDI-TOF MS) and TLC analyses. As a whole, 100 µL reaction mixture containing 90 µL of 1% (w/v) substrate (barley β-D-glucan, CMC-Na, lichenan, tamarind xyloglucan, HEC, konjac glucomannan, avicel, cellulose powder, PASC, beechwood xylan, birchwood xylan and carob galactomannan) in 20 mM sodium-phosphate buffer, pH 5.5 and 10 µL of *Rf*GH5_4 (50 µg/mL) was incubated at 30°C for 2 h. The hydrolysate for each substrate was prepared as discussed in **Section 3.2.8** and the chromatogram was generated.

For mass spectrometric analysis, 1 µL of hydrolysate produced was mixed with 1 µL of Dihydroxybenzene (DHB, 10 mg/mL) matrix (acetonitrile: water in 50:50% ratio, v/v and 0.1%, v/v trifluoroacetic acid, TFA). The analysis was performed as reported earlier, wherein peaks generated from DHB matrix in the lower *m/z* range (100-350) were eliminated to assign the peaks obtained from cellooligosaccharides (Son and Cha, 2014). MALDI-TOF MS was performed on time-of-flight mass spectrometer Autoflexspeed (Bruker Daltonics, Bremen, Germany) in positive reflectron mode. It was provided with the accelerating voltage of 4.9x2400 V. The time delay for pulse ion extraction and laser shot frequency was set at 120 ns and 2000 Hz, respectively. Two thousand shots per sample were recorded. The *m/z* range was plotted from 350 so that interference of high-intensity DHB matrix signals could be minimized.

3.2.10 Processivity of RfGH5_4

The mode of action of *Rf*GH5_4 was further investigated to get the insights of whether the enzyme was hydrolyzing the cellulosic substrates in a processive (consistent release of two short oligosaccharides) or non-processive manner (random release of higher oligosaccharides). Total 100 µL reaction mixture containing 1% (w/v) PASC (10 mg/mL) in 20 mM citrate-phosphate buffer, pH 5.5 and 10 µL of 50 µg/mL *Rf*GH5_4 enzyme was incubated at 30°C for different time intervals, 1, 2, 5, 10, 15, 30 min and 1, 4, 12, and 24 h. By prospecting the role of *Rf*GH5_4 in Simultaneous Saccharification and Fermentation (SSF), 30°C temperature was selected for the reaction. The Processivity Index (PI) by was calculated by estimating the total reducing sugar in soluble fraction and the reducing ends in insoluble fractions as discussed in the **Section 3.2.2**, as reported earlier (Zheng and Ding, 2013). Following formula was used to deduce the PI,

$$\textbf{Processivity Index (PI)} = \frac{\textit{Reducing sugar in Soluble fraction}}{\textit{Reducing ends in Insoluble fraction}}$$

Moreover, the TLC analysis was also performed for these samples to assess the processivity of *Rf*GH5_4, as described in **Section 3.2.8**. Total 1 µL of hydrolysate was loaded on to the TLC plate followed by its development by using the methods described in **Section 3.2.8**. The oligosaccharide products generated during hydrolysis of PASC were further analysed by High Performance Liquid Chromatography (HPLC) system (Model- Prominence Modular UFLC, Shimadzu Corporation, Japan), where MetaCarb 67C analytical reverse phase column (300x6.5 mm) at 85°C with double distilled water as the mobile phase was used at 0.5 mL/min flow rate. The chromatogram generated for each reaction sample was analysed for the presence of oligosaccharides using Lab

Solution software (Shimadzu, Japan) and compared with the chromatogram of cellooligosaccharide (G1, G2, G3 and G4) standards using detected at specific retention time (min) intervals.

3.3 Results and Discussion

3.3.1 Substrate specificity of RfGH5_4

The maximum enzyme activity exhibited by *Rf*GH5_4 was against a β-1,3; 1,4 mixed linked polysaccharide, barley β-D-glucan (665 U/mg) followed by CMC-Na (450 ±8 U/mg), lichenan (393 ±8 U/mg), xyloglucan (343 ±1 U/mg), HEC (333 ±3 U/mg) and konjac glucomannan (285 ±12 U/mg). The various polysaccharides hydrolyzed by *Rf*GH5_4 are listed in **Table 3.1**. Interestingly, *Rf*GH5_4 also hydrolyzed Phosphoric Acic Swollen Cellulose, PASC (86.6 ±2.6 U/mg), cellulose powder (0.32 ±0.02 U/mg) and microcrystalline cellulose, Avicel (0.2 ± 0.01 U/mg) though to a less extent as compared with soluble CMC-Na. Similar results were obtained for endoglucanases from *Clostridium cellulovorans* where multifunctional endoglucanases, *EngB* and *EngD* hydrolyzed CMC (82 and 86 U/mg, respectively) along with lichenan (853 and 555 U/mg, respectively) at pH 6.0 and 37°C in a 12 h incubation assay (Foong and Doi, 1992). *EngB* and *EngD* also hydrolysed pure micro granular cellulose (51 and 37 U/mg, respectively) and microcrystalline Avicel (20 and 16 U/mg, respectively). However, a recombinant endoglucanase, *pMEB200* from *R. flavefaciens* FD-1 showed activity against CMC-Na (290 µg of glucose equivalent.mg^{-1}.min^{-1}), but no hydrolysis of avicel was observed (Barros and Thomson, 1987). On conversion of the hydrolytic yield of *Rf*GH5_4 to same unit, this (290 µg of glucose equivalent.mg^{-1}.min^{-1}) was found be quite lower than the value, 8167 ±232 µg of glucose equivalent.mg^{-1}.min^{-1} obtained by *Rf*GH5_4 against CMC-Na with also detectable activity against Avicel (0.2 ± 0.01 U/mg). Similarly, the endoglucanase, *EglA* from *Bacillus pumilus* hydrolyzed only CMC-Na (8.25 U/mg), but could not act upon avicel and xylan (Lima *et al.*, 2005). Surprisingly, *Rf*GH5_4 also hydrolyzed hemicellulosic substrates with significant

activities such as beechwood xylan (21.7 21.7 ±4.5 U/mg), birchwood xylan (9.7 ±3.3 U/mg) and carob galactomannan (16 ±1 U/mg), displaying its broad substrate specificity. However, the hydrolysis of above mentioned hemicellulosic substrates by *Rf*GH5_4 was significantly lower, when compared with the cellulosic substrates (**Table 3.1**).

Table 3.1. Substrate specificity of *Rf*GH5_4 from *R. flavefaciens* FD-1 v3.

Substrate (1.0%, w/v)	Specific Activity (U/mg)
Barley β-D-glucan	665 ± 8
Carboxymethylcellulose-Na (CMC-Na)	450 ± 10
Lichenan	393 ± 8
Hydroxyethyl cellulose (HEC)	333 ± 3
Phosphoric Acid Swollen Cellulose (PASC)	86.6 ± 2.6
Cellulose powder	0.32 ± 0.02
Avicel	0.21 ± 0.01
Tamarind xyloglucan	343 ± 1
Konjac glucomannan	285 ± 12
Carob galactomannan	16 ± 1
Beechwood xylan	21.7 ± 4.5
Birchwood xylan	9.7 ± 3.3

All the experiments were carried in triplicates (n=3) and mean ± SD for each experiment is shown here.

The multifunctionality is less commonly observed among endoglucanases. This characteristic of *Rf*GH5_4 is noteworthy here which showed efficient hydrolysis of both cellulosic as well as hemicellulosic substrates. It has been reported earlier that, the aromatic residues like tryptophan near the active-site protruding outwards from the surface of the protein have been responsible for the substrate selectivity and multifunctionality of GH5_4 glucanases (Glasgow *et al.*, 2020). The efficient hydrolysis of cellulosic substrates like CMC-Na and PASC and various hemicellulosic polymers such as barley β-D-glucan and lichenan by *Rf*GH5_4 makes it a potential cellulase candidate in the valorisation of lignocellulosic biomass. Overall, the multifunctional nature of *Rf*GH5_4 widens its canvas in the application sectors like

bioenergy, synthesis of prebiotics, modification of cellulosic surfaces, paper, pulp, food and pharmaceutical industry as also reported earlier (Rashmi and Siddalingamurthy, 2018).

3.3.2 Biochemical properties of RfGH5_4

The optimization of pH and temperature for the enzyme assay of *Rf*GH5_4 was performed by using CMC-Na as the substrate. Maximum activity of *Rf*GH5_4 was recorded at the temperature of 55℃ and pH 5.5 (**Fig. 3.4a and b**, respectively). *Rf*GH5_4 retained most of its activity in the temperature range of 5℃ – 45℃ for 1 h of incubation (**Fig. 3.4c**). It retained 75% activity at 45℃ after 1 h of incubation. However, the enzyme was less stable at its optimum reaction temperature of 55℃, at which the enzyme was almost inactive after 1 h of incubation. *Rf*GH5_4 displayed stability in the pH range, 5-8 of 20 mM sodium phosphate and MES buffers (**Fig. 3.4d**). Thus, despite being optimally active at higher temperature, *Rf*GH5_4 may be suitable for cellulosic bioethanol production by carrying out SSF at the temperature of 30℃, displaying significant activity of 149 U/mg. Therefore, *Rf*GH5_4 could help in enhancing the efficiency of single-step fermentation processes such as SSF.

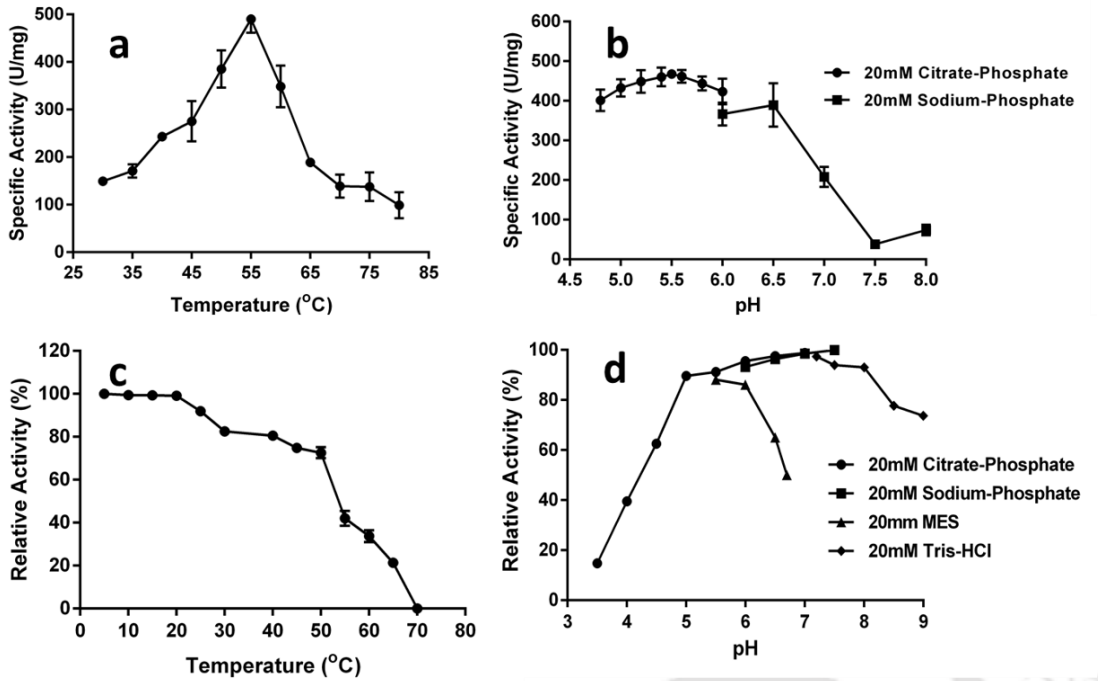

Fig. 3.4 (a) Optimum temperature, (b) optimum pH, (c) thermostability and (d) pH-stability of *Rf*GH5_4. All experiments were carried out in triplicate sets. The data shown here are the mean of three independent experiments. Error bars indicate standard deviation (±SD) for each point of datum.

3.3.3 Kinetic parameters of RfGH5_4

The Henri-Michaelis-Menten kinetic profiles of *Rf*GH5_4 against various substrates are shown in **Fig. 3.5** (**a**, barley β-D-glucan, **b**, lichenan, **c,** CMC-Na, **d**, HEC, **e**, tamarind xyloglucan and **f**, konjac glucomannan) and its kinetic parameters (V_{max}, K_m, k_{cat} and k_{cat}/K_m) against these substrates are summarized in **Table 3.2.** *Rf*GH5_4 showed remarkably high V_{max} and low K_m against soluble cellulosic substrates. *Rf*GH5_4 displayed V_{max} and K_m of 689 U/mg and 0.54 mg/mL, respectively against β-D-glucan (**Table 3.2**). Similarly, it showed V_{max} of 525 U/mg and K_m of 2.47 mg/mL against CMC-Na (**Table 3.2**). It displayed the V_{max} of 381 U/mg with K_m of 0.89 mg/mL against tamarind xyloglucan (**Table 3.2**). Interestingly, *Rf*GH5_4 also displayed a V_{max} of 316 U/mg and K_m of 0.41 mg/mL against konjac glucomannan (**Table 3.2**).

The V_{max} of 525 U/mg displayed by *Rf*GH5_4 against CMC-Na was significantly higher than V_{max}, 160.6 U/mg of endoglucanase *CS10* cloned from the metagenome of gut microflora of the black soldier fly, *Hermetia illucens* (Lee *et al.,* 2014) and V_{max} of 32.1 U/mg of mutant endoglucanase, CMCase-UV2 from *Bacillus amyloliquefaciens* SS35 (Singh *et al.,* 2020). However, an endoglucanase, *CtCel5E* (GH5) from *Clostridium thermocellum*, showed V_{max} of 736.2 U/mg against CMC-Na (Yuan *et al.,* 2015). The fungal cellulases dominate the industrial cellulase market in the present state. The most studied and industrially used organism for cellulase production, *Trichoderma reesei* displayed V_{max} of 351 U/mg for endoglucanase, *EGIII* (*Cel12A*) against CMC-Na (Macarron *et al.,* 1993; Nakazawa *et al.,* 2008), which is much lower and that of *Rf*GH5_4 against CMC-Na. Another endoglucacanase, *EGII* (a.k.a. *Cel5A*, a family GH5 member) from *T. reesei* also showed lower V_{max} of 220.5 U/mg (Samanta *et al.,* 2012). The enzyme, *Rf*GH5_4 displayed much higher V_{max} (381 U/mg) against tamarind xyloglucan, than the endoglucanase, *XEG5* of *Paenibacillus* sp. *XEG5* showing V_{max} of 18.4 U/mg (Yaoi *et al.,* 2005).

The catalytic efficiencies (k_{cat}/K_m) of *Rf*GH5_4 for β-D-glucan, CMC-Na, konjac glucomannan and tamarind xyloglucan were 875.7, 146, 529 and 294 $mL.mg^{-1}s^{-1}$, respectively (**Table 3.2**). The catalytic efficiency, k_{cat}/K_m of 12.4 $mL.mg^{-1}s^{-1}$ of endoglucanase *CtCel5E* from *Clostridium thermocellum* against CMC-Na was significantly lower than 146 $mL.mg^{-1}s^{-1}$ of *Rf*GH5_4. The processive endoglucanases, EG5C (Wu *et al.,* 2018) and EG5C-1 D70Q/S235W (Lv *et al.,* 2021) from *Bacillus subtilis* BS-5 also showed very low k_{cat}/K_m of 38.0 and 37.3 $mL.mg^{-1}s^{-1}$, respectively, against CMC-Na.

Endoglucanase, *EGIII (Cel12A) from Trichoderma reesei*, displays the highest catalytic efficiency (k_{cat}/K_m) of 7.3 mL.mg^{-1}s^{-1} among all *T. reesei* endoglucanases. An attempt was made to increase the catalytic efficiency of *EGIII* resulting in a *2R4* mutant however, that gave enhanced pH and thermostability but with a decrease in the catalytic efficiency to 4.8 mL.mg^{-1}s^{-1} (Nakazawa *et al.*, 2009). Another endoglucacanse, *EGII* (a.k.a. *Cel5A*, a family GH5 member) from *T. reesei* showed k_{cat}/K_m of 92.77 mL.mg^{-1}s^{-1} (Samanta *et al.*, 2012), which was much lower than 146 mL.mg^{-1}s^{-1} of *Rf*GH5_4 (Table 3.2, present study). It could be concluded that k_{cat}/K_m of *Rf*GH5_4 is much higher than those of endoglucanases like *EG III* of *T. reesei* and *CtCel5E* of *C. thermocellum*. Recently, a multifunctional glucomannanase, *6XSU* of subfamily GH5_4 from *R. flavefaciens* was reported with the k_{cat}/K_m of 5.66 (mL.mg^{-1}s^{-1}) for glucomannan and 1.55 (mL.mg^{-1}s^{-1}) for xyloglucan (Glasgow *et al.*, 2020) which are also significantly lower that those of *Rf*GH5_4 (Table 3.2).

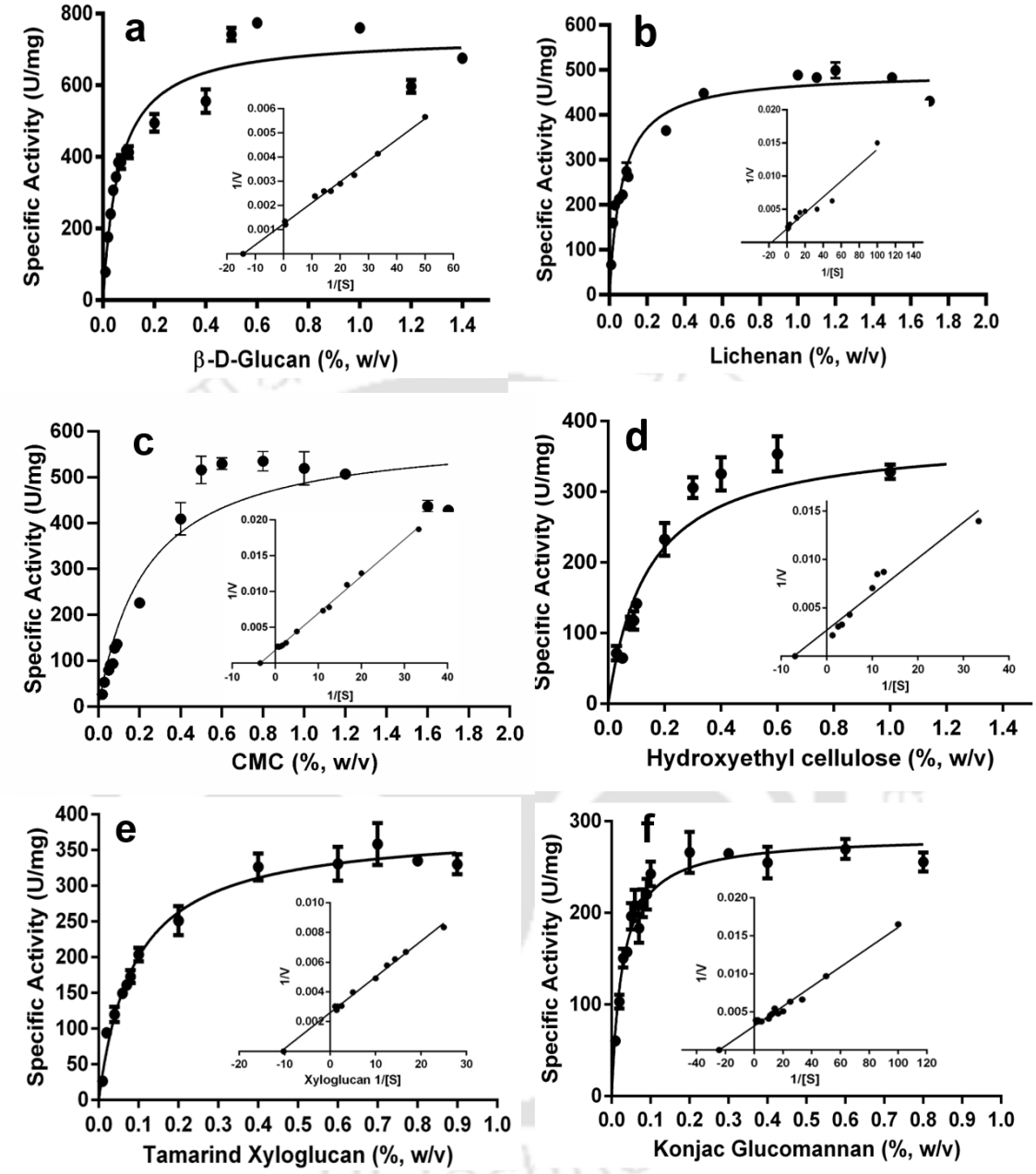

Fig. 3.5 Michaelis-Menten kinetics of *Rf*GH5_4 against (a) barley β-D-glucan, (b) lichenan, (c) CMC-Na, (d) HEC, (e) tamarind xyloglucan and (f) konjac glucomannan was studied at optimum pH of 5.5 and 55°C temperature (in 20 mM citrate-phosphate buffer) at final enzyme concentration of 5 µg/mL. The inset of every diagram shows the Lineweaver-Burk plot of *Rf*GH5_4 against respective substrates. All experiments were carried out in triplicate. The data shown here are the mean of three independent experiments. Error bars indicate standard deviation (±SD) for each data point.

Table 3.2. Kinetic properties of *Rf*GH5_4.

Substrate*	V_{max} (U/mg)	K_m (mg/mL)	k_{cat} (s^{-1})	k_{cat}/K_m ($mL.mg^{-1}s^{-1}$)
β-D-glucan	689 ± 18	0.54 ± 0.04	473 ± 18	875.7 ± 33.9
Lichenan	495 ± 9	0.58 ± 0.05	339.7 ± 7.4	585.7 ± 12.8
CMC-Na	525 ± 14	2.47 ± 0.14	360.3 ± 11.7	146 ± 5
HEC	371 ± 27	1.43 ± 0.30	254.6 ± 23.9	178 ± 16
Tamarind Xyloglucan	381 ± 12	0.89 ± 0.01	261.4 ± 10.1	294 ± 11
Konjac Glucomannan	316 ± 5	0.41 ± 0.02	216.8 ± 5.2	529 ± 12

**The assays were carried out in triplicates (n=3) with 5 µg/mL (final concentration) of Rf*GH5_4 *at an optimum pH 5.5 (20 mM citrate phosphate buffer) and 55℃ for 2 min. Mean ± SD for each experiment is shown here.*

The kinetic parameter and the catalytic efficiency of some earlier studied and commercially available endoglucanases, including this study, are described in **Table 3.3**. Moreover, the appreciable catalytic efficiency of *Rf*GH5_4 across a wide range of cellulosic and hemicellulosic substrates is noteworthy. *Rf*GH5_4 stands out as an efficient multifunctional endoglucanase among the other known endoglucanases. The comparison underlines the potential of *Rf*GH5_4 over the other available endoglucanases for industrial applications like lignocellulosic bioethanol and other value-added products.

Table 3.3. Comparison of *Rf*GH5_4 with certain other recombinant and commercial endoglucanases.

Enzyme	Substrate	Organism	Optimum	Kinetics	Reference
Endoglucanase, *Rf*GH5_4	CMC-Na	*Ruminococcus flavefaciens* FD-1 v3	pH 5.5, 55.5 °C, 41.18 kDa	*SA* – 450 K_m – 2.47 k_{cat}/K_m – 145.9	This Study
Endoglucanase, *Egst*	CMC-Na, Medium Viscosity	*Scytalidium thermophilum*	pH 5.5, 60 °C, 66 kDa	*SA* – 246 K_m – 10.5 k_{cat}/K_m – 45.2	Meleiro *et al.*, 2018
β-1,4-endoglucanase, *rMt-egl*	CMC-Na	*Myceliopthora thermophila BJA*	pH 10, 50 °C, 47 kDa	*SA* – 119.4 K_m – 5 k_{cat}/K_m – 204	Phadtare *et al.*, 2017
Bifunctional Cellulase, *CtCel5E*, (GH5)	CMC	*Clostridium thermocellum*	pH 5.0, 50 °C	*SA* – 736 K_m – 2.1 k_{cat}/K_m – 12.40	Yuan *et al.*, 2015
β-1,4-endoglucanase, *TaEGII*	CMC-Na	*Trichoderma atroviride*	pH 5.0, 60 °C, 44.23 kDa	*SA* – 119.4 K_m – 0.014 k_{cat}/K_m – 6.50	Huang *et al.*, 2017
Endoglucanase, *EGIII (Cel12A)*	CMC-Na	*Trichoderma reesei*	pH 5.0, 30 °C, 48 kDa	*SA* – 351 K_m – 5 k_{cat}/K_m – 7.33	Macarron *et al.*, 1993
Spezyme#3 (β-1,4-endoglucanase)	CMC-Na	*Trichoderma reesei*	pH 4.0, 50 °C	*SA* – 25	Genencor Intl., USA
Cellulase AP30K (β-1,4-endoglucanase)	CMC-Na	*Trichoderma reesei*	pH 4.5, 60 °C	*SA* – 21	Amano Enzyme, USA
E-CELTH (β-1,4-endoglucanase, GH6)	CMC-Na	*Thermobifida halotolerans*	pH 8.5, 80 °C	*SA* – 16	Megazyme Ltd., Ireland
ACCELLERASE 1500 (β-1,4 endoglucanase)	CMC-Na	*Trichoderma reesei*	pH 4.8, 50 °C	*SA* – 2.8	DuPont De Nemours, Inc. USA
C2730-Celluclast (β-1,4-endoglucanase)	Cellulose	*Trichoderma reesei*	pH 6, 52 °C	*SA* – 0.7	Sigma-Aldrich Corp./Novozyme Corp., USA
C1184-Cellulase (β-1,4-endoglucanase)	Cellulose	*Aspergillus niger*	pH 5, 37 °C	*SA* – 0.3	Sigma-Aldrich Corp./Novozyme Corp., USA

SA = Specific Activity (U/mg), K_m = mg/mL, k_{cat}/K_m = mL.mg^{-1}s^{-1}.

3.3.4 Effect of metal ions and other additives on RfGH5_4 activity

The influence of various metal ions and additives on the activity of *Rf*GH5_4 was explored using CMC-Na as the substrate. The activity of *Rf*GH5_4 increased in the

presence of 10 mM of K^+ or Li^+ ions. The enzyme activity of *Rf*GH5_4 was increased by 8.3, 32.2 and 50% in presence of 2, 5 and 10 mM of K^+ ions, whereas Li^+ ions could increase its activity by 6% upto 10 mM of concentrations (**Table 3.4**). A family GH5 endoglucanase, *CelRH5* from the rhizosphere, also showed a similar type of increase in enzyme activity in the presence of Li^+ or K^+ ions (Wierzbicka-Woś *et al.,* 2019). The cations, K^+ and Li^+ can alter the active-site structure of endoglucanase, thereby increasing the enzyme activity (Kui *et al.,* 2010). The K^+ ions may be performing the role of mediator between the substrate and the enzyme's active-site, the characteristic of Type II enzyme activator thereby making conformational changes in *Rf*GH5_4 as also reported earlier (Gohara and Di Cera 2016). Thus, the increment in the enzyme activity of *Rf*GH5_4 in presence of K^+ ions could be due to their participation at the active-site, thereby stabilizing the intermolecular interactions. The use of K^+ ions during synergistic saccharification of lignocellulosic biomasses to enhance the enzyme activity of *Rf*GH5_4 could be further explored. The activity of *Rf*GH5_4 was retained near to 100% at 2, 5 and 10 mM of Na^+ ions which suggested the suitability of sodium phosphate for its storage. The Mg^{2+} ions at 5 mM concentrations increased the enzyme activity of *Rf*GH5_4 by 5% but the enzyme activity in the presence of 10 mM of the Mg^{2+} ions was decreased by 33%. Surprisingly, Ca^{2+} ions at 2, 5 and 10 mM concentrations drastically lowered the *Rf*GH5_4 enzyme activity by 27.7, 68.2 and 29%, respectively. Similarly, Ni^{2+}, Fe^{2+}, Co^{2+}, Zn^{2+}, Mn^{2+} or Cu^{2+} ions adversely affected the enzyme activity of *Rf*GH5_4, where 80% to 100% loss of the enzyme activity was observed (**Table. 3.4**). A similar abrupt decrease in the enzyme activity by these metal ions was also reported for endoglucanase, *Ba-EGA* from *Bacillus* sp. AC-1 (Li *et al.,* 2006) and *CelRH5* from rhizosphere (Wierzbicka-Woś *et al.,* 2019).

Interestingly, *Rf*GH5_4 enzyme activity increased by nearly 10% in the presence of 2, 5 or 10 mM of EDTA. Similarly, 10 mM of EGTA displayed an increment of 32% in the enzyme activity of *Rf*GH5_4, whereas it was increased by about 7% at 2 and 5 mM concentrations (**Table 3.4**). The increase in the enzyme activity by EDTA or EGTA indicated the absence of Ca^{2+} or Mg^{2+} ions in the *Rf*GH5_4 structure as also reported earlier (Naika and Tiku, 2011). This increment in the enzyme activity could be attributed to a flexible conformational change in the secondary structure elements caused by EDTA or EGTA resulting in increased enzyme activity and overall stability of *Rf*GH5_4 which was also observed for an endoglucanase, *AaEnd* from fungus, *Aspergillus aculeatus* (Naika and Tiku, 2011).

*Rf*GH5_4 retained nearly 98% to 99% of enzyme activity in presence of 1% (v/v) DMSO or various conentrations (1, 2, 4, 8 or 10 %, v/v) of glycerol, respectively. The stability of *Rf*GH5_4 in presence glycerol indicates its possible use in the formulation of base which may be explored to enhance its storage stability. 10 mM guanidine hydrochloride drastically decreased the *Rf*GH5_4 enzyme activity by 23%, whereas it was reduced by 15% (w/v) in the presence of 100 mM urea (**Table 3.4**). The enzyme activity of *Rf*GH5_4 increased by 8.7% in presence of 0.5% (v/v) Triton X-100, whereas 1% (v/v) Tween 80 decreased its activity by 17% (**Table 3.4**). *Rf*GH5_4 lost 100% of enzyme activity in the presence of 0.1% SDS, indicating the sensitivity towards anionic detergents. Endoglucanase *CS10* from *Hermetia illucens* was reported with a similar type of decrease in the enzyme activity in the presence of guanidine hydrochloride, urea and SDS (Lee *et al.*, 2014).

Table 3.4. Effect of metal ions and additives on *Rf*GH5_4 enzyme activity.

Metal ion (10 mM) / Additive	Relative Enzyme Activity (%)		
	2 mM	5 mM	10 mM
Control			100.0 ±9
K^+	108.3 ±2.2	132.2 ±3.5	150.0 ±2
Li^+	105.2 ±1.3	106.3 ±2.7	106.1 ±2.8
Na^+	98.7 ±1.6	95.8 ±1.7	92.4 ±4.5
Mg^{2+}	100.5 ±1.5	105.0 ±1.0	67.2 ±3.7
Ca^{2+}	72.3 ±3.3	31.8 ±4.7	29.0 ±4
Ni^{2+}	96.0 ±1.0	68.6 ±4.4	14.1 ±2.4
Co^{2+}	93.1 ±0.6	47.9 ±4.6	13.5 ±5.9
Mn^{2+}	22..9 ±9.1	14.6 ±4.6	0.9 ±0.6
Fe^{2+}	8.6 ±4.6	0	0
Zn^{2+}	10.5 ±4.3	0	0
Cu^{2+}	4.5 ±2.8	0	0
Ethylene Diamine Tetraacetic Acid (EDTA 10 mM)	108.3 ±3.7	110.0 ±9.2	110.2 ±2.9
Ethylene Glycol Tetraacetic Acid (EGTA (10 mM)	107.0 ±3.6	106.0 ±12	132.4 ±2.2
DMSO (1%, v/v)			98.0 ±3.0
Glycerol (1%, v/v)			106.4 ±10.4
Glycerol (2%, v/v)			106.7 ±6.6
Glycerol (4%, v/v)			106.5 ±11.5
Glycerol (8%, v/v)			98.2 ±3.2
Glycerol (10%, v/v)			99.0 ±6.0
Guanidine hydrochloride (10 mM)	124.5 ±11.5	110.2±13.8	77.0 ±3.0
Urea (100 mM)	100.5±5.3	114.7±8.5	85.0 ±5.0
Triton X-100 (0.5%, v/v)			108.7 ±8.7
Tween 80 (1%, v/v)			83.0 ±11.0
Sodium Dodecyl Sulphate, SDS, 0.1%, w/v			0

All the experiments were carried in triplicates (n=3) and mean ± SD for each experiment is shown here. Specific activity of Control was 341 U/mg and considered as 100%.

3.3.5 Protein melting analysis of RfGH5_4

The complete melting of *Rf*GH5_4 was observed at 70°C (T_m). The addition of 10 mM EDTA slightly increased the T_m of *Rf*GH5_4 by 2°C (72°C), showing the absence of any metal ion in the enzyme structure (**Fig. 3.6**). This result corroborated with the result from the previous section, where incubation of *Rf*GH5_4 with 10 mM

EDTA increased the enzyme activity by 10%, indicating the absence of any metal ion in the structure. A decrease in the T_m of *Rf*GH5_4 by 2°C was observed by 10 mM Ca^{2+} ions. This was also in agreement with results of previous **Section 3.3.4**, where 10 mM of Ca^{2+} ions lowered the enzyme activity to 29%, thereby displaying the role of Ca^{2+} ions in destabilizing the *Rf*GH5_4 structure. 10 mM of K^+ ions did not affect the T_m of *Rf*GH5_4 (70°C). However, 10 mM of K^+ ions displayed increase in the enzyme activity of *Rf*GH5_4 as mentioned in the previous **Section 3.3.4**. It is possible that the K^+ cations play a role in the catalysis of *Rf*GH5_4 affecting its conformation rather than its stability as also observed for an endoglucanase from a plant root ericoid mycorrhizal fungus *Leohumicola* sp. (Kui *et al.*, 2010; Adeoyo *et al.*, 2017).

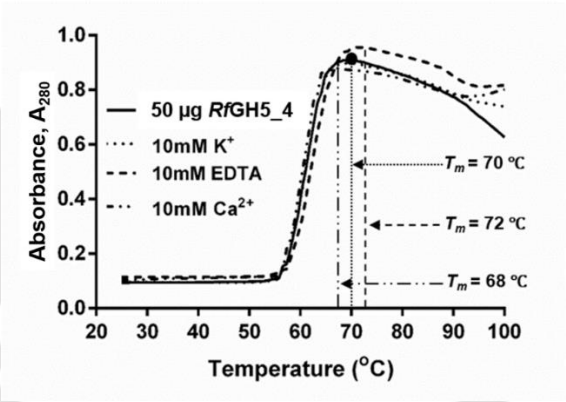

Fig. 3.6 Protein melting analysis of *Rf*GH5_4. Total 50 µg of *Rf*GH5_4 was incubated independently as well as with 10 mM of EDTA, K^+ or Ca^{2+} ions in 1 mL of 20 mM MES buffer (pH 5.5). The reaction mixture was subjected to varying temperature within a range, 25°C – 100°C. The absorbance at 280 nm (A_{280}) was recorded.

3.3.6 Analysis of hydrolysis mechanism and multifunctionality of RfGH5_4 by TLC

A chromatogram developed with the CMC-Na hydrolysates produced by *Rf*GH5_4 displayed cellotriose, cellotetraose and other higher oligosaccharides after 2 min of reaction in a 24 h hydrolysis (**Fig. 3.7a**).

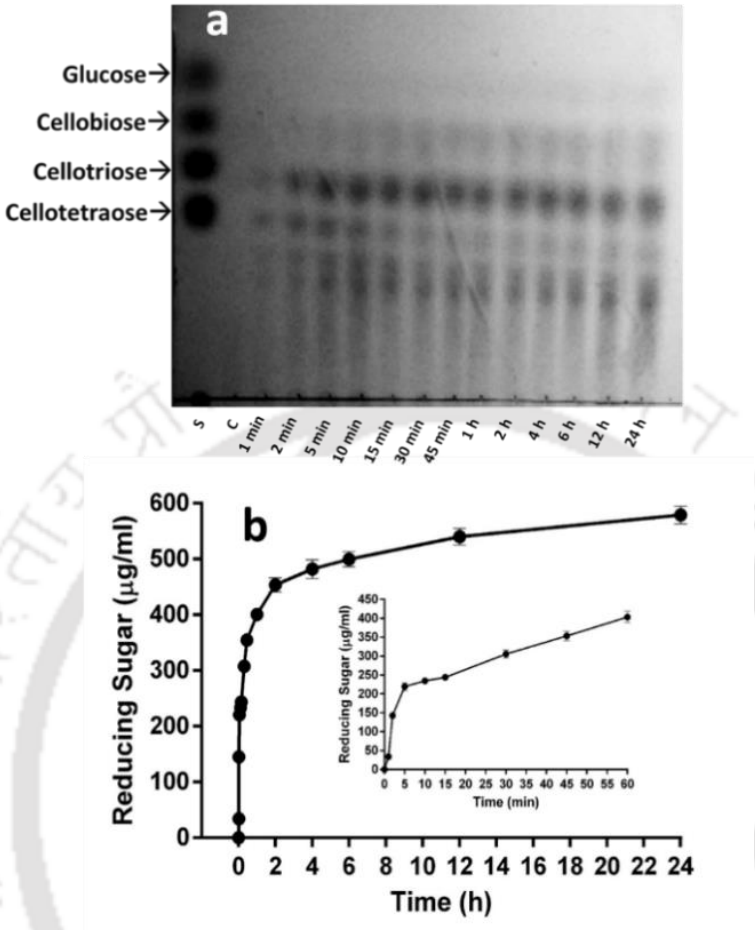

Fig. 3.7 Time-dependant TLC analysis of hydrolyzed products of CMC-Na by *Rf*GH5_4 was carried out. (a) Chromatogram of TLC of the hydrolysate of CMC-Na by *Rf*GH5_4 for different time intervals. Lane S represent Standards (Glucose, Cellobiose, Cellotriose, Cellotetraose) and lane 2 is Control (C), the substrate CMC-Na without treatment with enzyme. Other lanes are time-dependent reaction products of enzyme, *Rf*GH5_4 with CMC-Na. (b) Plot of reducing sugar concentration with time. Inset shows the reducing sugar released in the first 60 min. The assay was performed with the 50 µg/mL final enzyme concentration of *Rf*GH5_4 at 30°C and pH 5.5 (20 mM citrate-phosphate) for 1 min to 24 h.

However, after 1 h, cellotriose was a predominant oligosaccharide alongwith the other higher oligosaccharides (>DP4), confirming the *endo*-acting mode of hydrolysis by *Rf*GH5_4. Quantification of reducing sugars revealed that the hydrolysis process achieved saturation after 12 h of the reaction (**Fig. 3.7b**).

3.3.7 TLC and MALDI-TOF MS analyses of RfGH5_4 hydrolysed polysaccharides

The hydrolysis of various substrates by *Rf*GH5_4 was analysed and confirmed by TLC. The substrates used were barley β-D-glucan, CMC-Na, lichenan, tamarind xyloglucan, HEC, konjac glucomannan, beechwood xylan, birchwood xylan and carob galactomannan. The release of oligosaccharides of various degrees of polymerization by *Rf*GH5_4 was observed in these substrate hydrolysates. Cellobiose, cellotriose, cellotetraose and other higher cellooligosaccharides could be detected in the barley β-D-glucan, CMC-Na and HEC hydrolysates (**Fig. 3.8a**, Lanes 2, 3 and 6, respectively). Cellotetraose was the primary product in lichenan and konjac glucomannan hydrolysis (**Fig. 3.8a**, Lanes 4 and 7, respectively). The hydrolysis of xyloglucan showed multiple spots on the TLC plate which were attributed to different xyloglucan oligosaccharides (XyGs) (**Fig. 3.8a**, Lane 5). The TLC analysis also confirmed the hydrolysis of beechwood xylan, birchwood xylan and carob galactomannan by *Rf*GH5_4 (**Fig. 3.8a**, Lanes 8, 9 and 10, respectively). The cellooligosaccharides higher than cellotetraose were observed in the *Rf*GH5_4 hydrolysates of Avicel and cellulose powder (**Fig. 3.8b**, Lane 2 and 3, respectively). However, PASC was hydrolysed to oligosaccharides smaller than cellotetraose like cellotriose and cellobiose (**Fig. 3.8b**, Lane 4). The degree of crystallinity of PASC was reported to be 0.04 (Zhang *et al.*, 2006) which makes the cellulose chains freely accessible to the endoglucanase, unlike avicel, for which the degree of crystallinity is 0.6 (Bala and Singh, 2019). The higher the degree of crystallinity of a cellulosic substrate, lower is the accessibility of an endoglucanase for the cellulose chain and vice-versa.

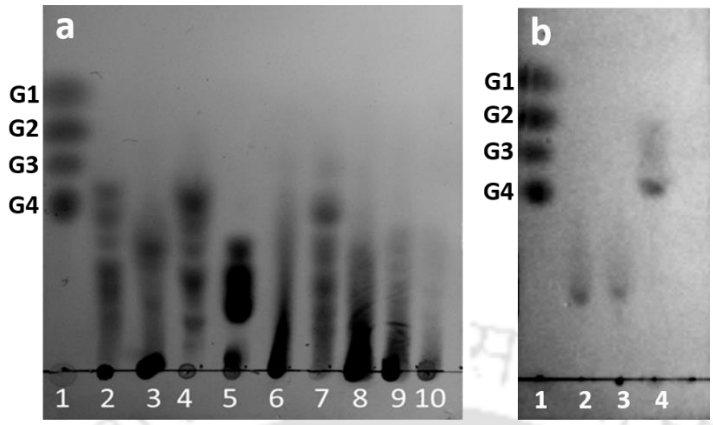

Fig. 3.8 The TLC analysis of *Rf*GH5_4 generated hydrolysates (a) of different carbohydrate polymers was also performed. Lanes, 1- Standards: G1-glucose, G2-cellobiose, G3-cellotriose and G4-cellotetraose. 2- barley β-D-glucan, 3- CMC-Na, 4- lichenan, 5- tamarind xyloglucan, 6- HEC, 7- konjac glucomannan, 8- beechwood xylan, 9- birchwood xylan, 10- carob galactomannan. Hydrolysates of avicel, cellulose powder and PASC are shown in Fig. (b). Lanes, 1- Standards: G1- glucose, G2- cellobiose, G3-cellotriose and G4- cellotetraose, 2- avicel, 3-cellulose powder, 4- PASC. The assay was performed with 50 µg/ml working concentration of *Rf*GH5_4 at 30 ℃ in 20 mM citrate-phosphate, pH 5.5 and. For avicel, cellulose powder and PASC, a rotation of 200 rpm was provided for 2 h.

MALDI-TOF analysis of PASC hydrolysate revealed the presence of majorly cellobiose (DP2) and cellotriose (DP3) in **Fig. 3.9a**, indicating the processive hydrolysis of amorphous cellulose by *Rf*GH5_4 as also described and confirmed in the **Section 3.3.8**, later in this Chapter. CMC-Na hydrolysate of *Rf*GH5_4 also showed cellooligosaccharides ranging between DP2 and DP6 as shown for PASC. The MALDI-TOF MS (DP, *m/z* ratio) analysis revealed that *Rf*GH5_4 randomly hydrolyses β-D-glucan and lichenan to give cellooligosaccharides ranging from DP2 to DP12, thus further confirming *Rf*GH5_4 as an endoglucanase (**Fig. 3.9b**). D-Glucose was not detected in any of these hydrolysates upon the MALDI-TOF analysis.

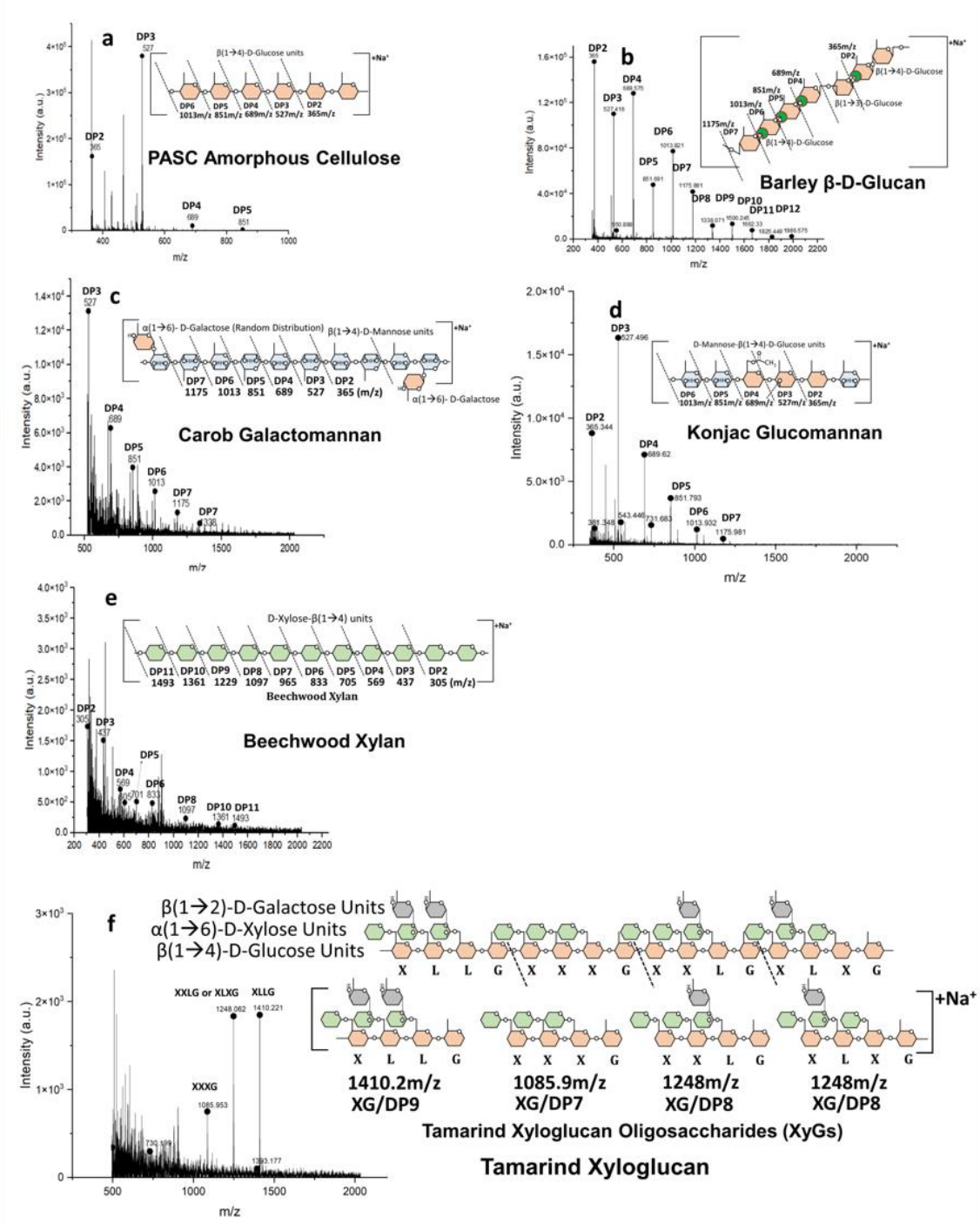

Fig. 3.9 MALDI-TOF analyses of hydrolysates of various cellulosic and hemicellulosic substrates by *Rf*GH5_4. (a) PASC amorphous cellulose, (b) Barley β-D-Glucan, (c) carob galactomannan, (d) konjac glucomannan, (e) beechwood xylan and (f) xyloglucan.

The presence of oligosaccharides ranging from DP3 to DP7 in the hydrolysate of *Rf*GH5 treated carob galactomannan as analysed by MALDI-TOF MS, confirmed its

activity on the main chain of galactomannan randomly cleaving the β-1,4 linked mannose units thus revealing it as an endomannanase (**Fig. 3.9c**). Hydrolysate of konjac glucomannan from *Rf*GH5_4 (**Fig. 3.9d**) showed oligosaccharides of DP2 (*m/z* ratio, 365), DP3 (527), DP4 (689), DP5 (851.2), DP6 (1013.2) and DP7 (1175.2) thus revealing the glucomannanase function of *Rf*GH5_4. *Rf*GH5_4 also hydrolysed xylan (**Fig. 3.9e**) as can be stated from the presence of xylooligosaccharides ranging from DP2 to DP11 in the beechwood xylan hydrolysates. Interestingly, xyloglucan oligosaccharides obtained after hydrolysis of tamarind xyloglucan by *Rf*GH5_4 were, XXXG (1085.9), XLXG (1248), XXLG (1248) and XLLG (1410.2), showing it to be a xyloglucanase (**Fig. 3.9f**). Subfamily GH5_4 endoglucanases from family GH5 uniquely hydrolyse xyloglucan as can be also observed for *Rf*GH5_4. Moreover, *Rf*GH5_4 multifunctional endoglucanase could also be the member of EC 3.2.1.151 along with being a member of EC 3.2.1.4, because it also efficiently hydrolyses β-(1,4) main chain linkage of xyloglucan alongwith that of cellulosic substrates. Xyloglucan specific endoglucanase from *Paenibacillus* sp. strain KM21 having EC 3.2.1.151 was reported previously in GH5_4 subfamily (Aspeborg *et al.,* 2012). Thus, the MALDI-TOF MS analyses of different substrate hydrolysates confirmed the multifunctionality and broad substrate specificity of *Rf*GH5_4 endoglucanase. Various degrees of higher oligosaccharides (DP2-DP12) detected by MALDI-TOF MS in the hydrolysates of *Rf*GH5_4 treated substrates further confirmed its *endo*-acting mode of polysaccharide hydrolysis.

3.3.8 Processivity of RfGH5_4 on cellulose

3.3.8.1 HPLC chromatogram of cellooligosaccharide standards

The cellooligosaccharides (D-glucose-G1, Cellobiose-G2, Cellotriose-G3 and Cellotetraose-G4) standards were also run for HPLC analysis under the same experimental conditions mentioned in **Section 3.2.10**. The chromatograms generated for G1, G2, G3 and G4 **(Fig. 3.10)** at their respective retention time were used for comparative analysis of cellooligosaccharides in the *Rf*GH5_4 hydrolysates as described in the below.

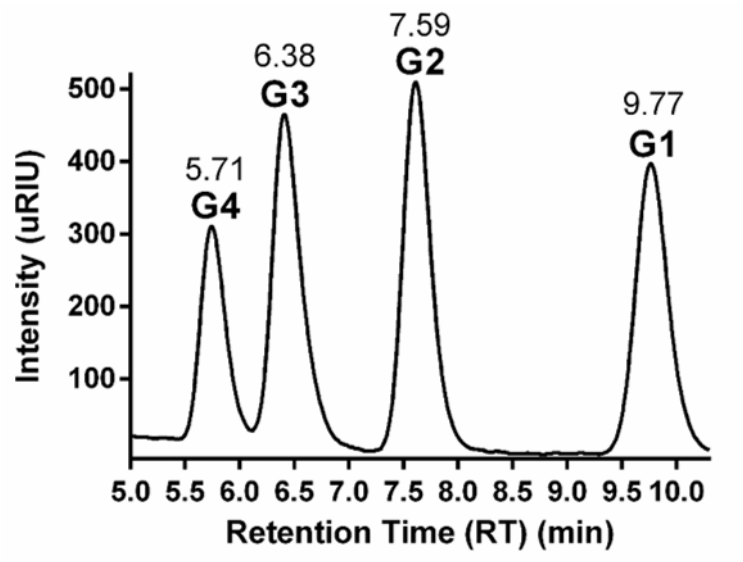

Fig. 3.10 Cellooligosaccharide standards were run in HPLC system for the comparative analysis of *Rf*GH5_4 hydrolysates (G1-Glucose, G2-Cellobiose, G3-Cellotriose, G4-Cellotetraose). The retention time of respective cellooligosaccharides is shown on the top of their respective chromatogram.

3.3.8.2 TLC, HPLC and Processivity Index profile of PASC's RfGH5_4 hydrolysate

The mode of cellulose hydrolysis by *Rf*GH5_4 was further elucidated. The processivity of *Rf*GH5_4 was studied against PASC at different time intervals up to 24 h, as shown in the TLC analysis (**Fig. 3.11a**). *Rf*GH5_4 hydrolysed the cellulosic

substrate, PASC and gave cellotetraose during first 30 min of the reaction. However, consistent release of cellotriose and cellobiose was observed after 1 h of the reaction (**Fig. 3.11a**). This is the processive behaviour of *Rf*GH5_4 endoglucanase, where it consistently cleaved PASC into cellotriose and cellobiose and this mode is called processivity as reported earlier (Zheng and Ding, 2013). The HPLC chromatogram of PASC hydrolysates from different time intervals also showed that the release of cellooligosaccharides from PASC was confined to cellotriose (G3) and cellobiose (G2) after 1 h of the reaction conforming with the G3 and G2 standards as shown in **Fig. 3.10**, thereby, further confirming the processivity of *Rf*GH5_4 (**Fig. 3.11b**). The Processivity Index, PI of the hydrolysis reaction of *Rf*GH5_4 on PASC (calculated as per the formula given in **Section 3.2.10)** was found to be consistently increasing from 0.1 to 4.9 at 1 min to 24 h, respectively (**Fig. 3.11c**), further confirming the processive behaviour of *Rf*GH5_4 as also shown in **Table 3.5**.

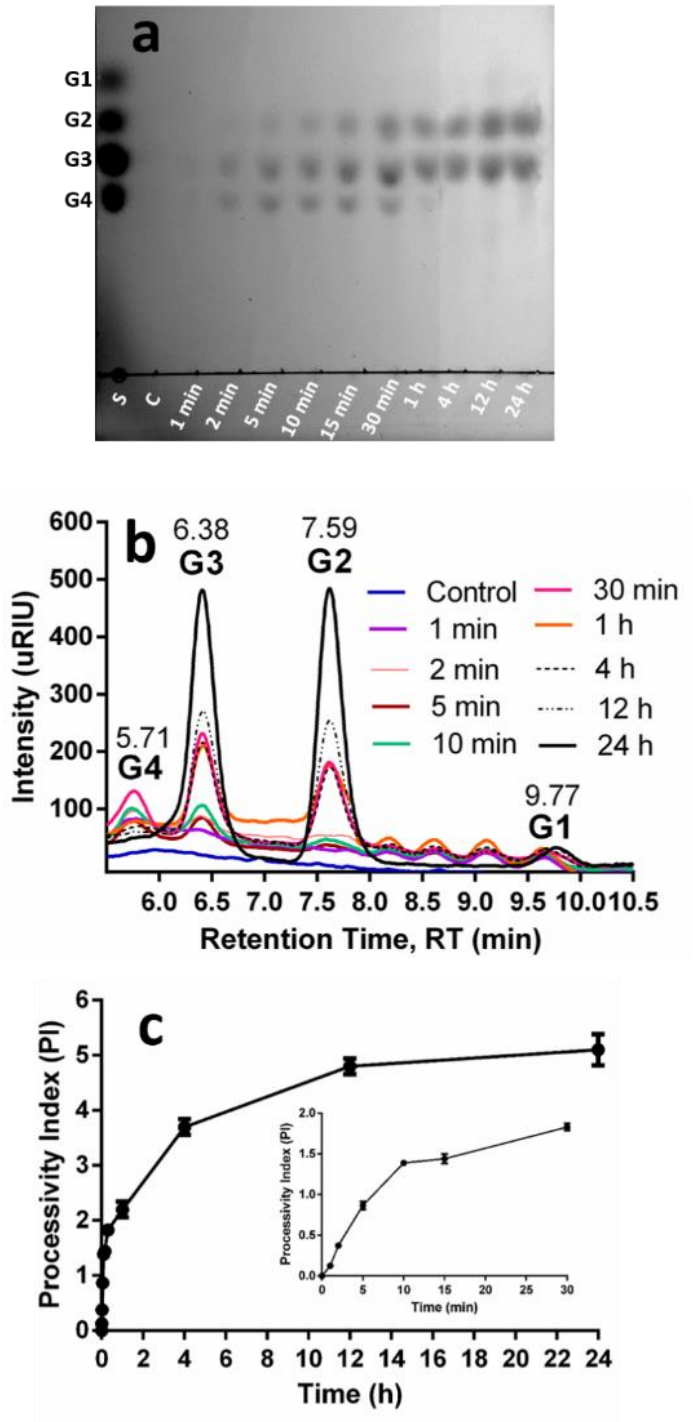

Fig. 3.11 (a) Time-dependent TLC analysis of hydrolyzed products of PASC by *Rf*GH5_4, Lane S- Standards (G1, Glucose, G2, Cellobiose, G3, Cellotriose, G4, Cellotetraose) and lane 2 is Control (C) PASC without enzyme. Other lanes are time-dependent enzyme reaction products of PASC by *Rf*GH5_4. (b) HPLC analysis of *Rf*GH5_4 treated PASC hydrolysate at different time intervals, (c) Processivity Index (PI) analysis of *Rf*GH5_4 with the time-dependent hydrolysis of PASC.

Table 3.5. Processivity index of *Rf*GH5_4.

Reaction	Reducing Sugar (µg/mL) [R]	Reducing Ends in Insoluble Fraction (µg/mL) [I]	Processivity Index (PI) [R / I] *
Control	0.0	53.0	0.0
1 min	8.7	61.7	0.1
2 min	29.3	66.3	0.4
5 min	64.7	68.3	0.9
10 min	108.0	79.0	1.4
15 min	120.0	86.7	1.4
30 min	158.3	90.0	1.8
1 h	178.3	86.0	2.1
4 h	318.3	88.7	3.6
12 h	460.0	97.0	4.7
24 h	501.7	103.0	4.9

**Persistent increase in PI with time indicates processivity of Rf*GH5_4.

In the PI analysis, the accumulation of soluble reducing sugars usually increases with time, whereas the reducing ends of cellulose in the insoluble un-hydrolysed substrate fractions remain constant. The processive endoglucanases show an increasing PI over a time period of incubation as also observed for *Rf*GH5_4 (**Fig. 3.11c**). Zheng et al. (Zheng and Ding, 2013) also reported the initial release of cellotetraose followed by cellotriose and cellobiose from PASC by an GH5 endoglucanase, *EG1* from fungus *Volvariella volvacea.* As observed for *Rf*GH5_4, the processive endoglucanases, EG5C and EG5C-1 from *Bacillus subtilis* also showed the release of cellobiose and cellotriose from PASC and Avicel (Wu *et al.*, 2018). Processivity is rarely observed in endoglucanases from family GH5. Therefore, the processive endoglucanases from family GH5 should be explored for their applications in renewable energy sector. The drawback of using non-processive endoglucanases in biomass conversion is that they produce only higher oligosaccharides unlike *Rf*GH5_4. Therefore, processive endoglucanase, *Rf*GH5_4 could be used in conjunction with other cellulases for

bioethanol production as the enzyme releases significant amount of cellobiose through its processivity.

3.4 Conclusion

This study established *Rf*GH5_4 as an efficient and multifunctional endoglucanase, because it is able to hydrolyse a wide range of both cellulosic and hemicellulosic polysaccharides. The maximum enzyme activity of *Rf*GH5_4 was observed against barley β-D-glucan (665 U/mg) followed by CMC-Na (450 U/mg), lichenan (393 U/mg), xyloglucan (343 U/mg), HEC (333 U/mg) and konjac glucomannan (285 U/mg). It gave maximum activity at pH 5.5 and 55°C and was stable in pH range, 5.0-8.0 and between, 5°C-45°C. *Rf*GH5_4 showed notably high V_{max} and low K_m against β-D-glucan (689 U/mg and 0.54 mg/mL) and CMC-Na (525 U/mg and K_m of 0.58 mg/mL). The catalytic efficiency of *Rf*GH5_4 against CMC-Na (146 $mL.mg^{-1}s^{-1}$) and konjac glucomannan (529 $mL.mg^{-1}s^{-1}$) was significantly superior to earlier reported endoglucanases. The melting temperature (T_m) of *Rf*GH5_4 was 70°C, that was unaffected by EDTA or EGTA, indicating the absence of inherent divalent metal ions in the *Rf*GH5_4 structure. Conversely, ingress 10 mM K^+ ions showed significant increment in the enzyme activity of *Rf*GH5_4 by 50%, but showed no effect on its thermostability as analyzed by protein melting analysis. This increased enzyme activity by K^+ ions, a Type II enzyme activator, could be due to the facilitation of interactions between substrate and the active-site of *Rf*GH5_4. The time-dependent TLC analysis of CMC-Na hydrolystaes by *Rf*GH5_4 showed the presence of cellotriose, cellotetraose and other higher oligosaccharides which demonstrated its *endo*-acting mode. TLC and MALDI-TOF MS analyses of hydrolysates of *Rf*GH5_4 treated various polysaccharides showed oligosaccharides (DP2-DP12) further confirming the *endo*-mode and multi-ligand activity of the enzyme. TLC analysis showed that *Rf*GH5_4 could hydrolyze crystalline cellulose, Avicel although to a lesser

extent than the soluble and amorphous forms of cellulose. *Rf*GH5_4 hydrolyzed tamarind xyloglucan with a significantly high V_{max} of 381 U/mg which is the signature enzyme activity of subfamily GH5_4 enzymes. The hydrolysate of *Rf*GH5_4 treated xyloglucan contained XyG-oligosaccharides namely, XXXG, XLXG, XXLG and XLLG as analysed by TLC and MALDI-TOF MS, further confirmed its xyloglucanase activity. The HPLC, TLC analyses showed that *Rf*GH5_4 hydrolysed amorphous cellulose, PASC in a processive manner, where it generated cellobiose along with cellotriose thereby reducing the need of cellobiohydrolase. The PI of *Rf*GH5_4 hydrolysate of PASC increased from 0.1 to 4.9 in 24h, further confirming the enzymes's processivity. The efficiency, multifunctionality and suitability for SSF puts *Rf*GH5_4 in the category of cellulases important for renewable energy sector. Based on the unique properties, *Rf*GH5_4 is prospected as a potent cellulase for lignocellulosic bioethanol production. The multifunctionality of *Rf*GH5_4 could be further explored for generation of various carbohydrate oligosaccharides useful in feed, food, prebiotics and health sector.

References

Adeoyo OR, Pletschke BI, Dames JF (2017) Improved endoglucanase production and mycelial biomass of some ericoid fungi. *AMB Express* 7:1–8

Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biololgy* 12:1–16

Bala A, Singh B (2019) Cellulolytic and xylanolytic enzymes of thermophiles for the production of renewable biofuels. *Renewable Energy* 136:1231–1244

Barros ME, Thomson JA (1987) Cloning and expression in *Escherichia coli* of a cellulase gene from *Ruminococcus flavefaciens*. *Journal of Bacteriology* 169:1760–1762

Cheng R, Cheng L, Wang L, Fu R, Sun X, Li J, Wang S, Zhang J (2019) Characterization of an alkali-stable xyloglucanase/mixed-linkage β-glucanase Pgl5A from *Paenibacillus* sp. S09. *International Journal of Biological Macromolecules* 140:1158–1166

Foong FC, Doi RH (1992) Characterization and comparison of Clostridium cellulovorans endoglucanases-xylanases EngB and EngD hyperexpressed in *Escherichia coli*. *Journal of Bacteriology* 174:1403–1409

Fox JD, Robyt JF (1991) Miniaturization of three carbohydrate analyses using a microsample plate reader. *Annals of Biochemistry* 195:93–96

Gavande P V, Basak A, Sen S, Lepcha K, Murmu N, Rai V, Mazumdar D, Saha SP, Das V, Ghosh S (2021) Functional characterization of thermotolerant microbial consortium for lignocellulolytic enzymes with central role of Firmicutes in rice straw depolymerization. *Scientific Report* 11:3032

Glasgow EM, Kemna EI, Bingman CA, Ing N, Deng K, Bianchetti CM, Takasuka TE, Northen TR, Fox BG (2020) A structural and kinetic survey of GH5_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis. *Journal of Biological Chemistry* 295:17752–17769

Gohara DW, Di Cera E (2016) Molecular mechanisms of enzyme activation by monovalent cations. *Journal of Biological Chemistry* 291:20840–20848

Huang X, Li Q, Chen X, Fan J, Xu X, Sun X, Li D, Zhao H (2017) Expression and characteristics of an endoglucanase from *Trichoderma atroviride* (TaEGII) in *Saccharomyces cerevisiae*. *Applied Biochemistry and Biotechnology* 182:1158–1170

Johnson KA, Goody RS (2011) The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry* 50:8264–8269

Kui H, Luo H, Shi P, Bai Y, Yuan T, Wang Y, Yang P, Dong S, Yao B (2010) Gene cloning, expression, and characterization of a thermostable xylanase from *Nesterenkonia xinjiangensis* CCTCC AA001025. *Applied Biochemistry Biotechnology* 162:953–965

Kumar K, Singal S, Goyal A (2019) Role of carbohydrate binding module (CBM3c) of GH9 β-1, 4 endoglucanase (Cel9W) from *Hungateiclostridium thermocellum* ATCC 27405 in catalysis. *Carbohydrate Research* 484:107782

Lee C-M, Lee Y-S, Seo S-H, Yoon S-H, Kim S-J, Hahn B-S, Sim J-S, Koo B-S (2014) Screening and characterization of a novel cellulase gene from the gut microflora of *Hermetia illucens* using metagenomic library. *Journal of Microbiology and Biotechnology* 24:1196–1206

Li Y-H, Ding M, Wang J, Xu G, Zhao F (2006) A novel thermoacidophilic endoglucanase, Ba-EGA, from a new cellulose-degrading bacterium, *Bacillus* sp. AC-1. *Applied Microbiology and Biotechnology* 70:430–436

Lima AOS, Quecine MC, Fungaro MHP, Andreote FD, Maccheroni W, Araújo WL, Silva-Filho MC, Pizzirani-Kleiner AA, Azevedo JL (2005) Molecular characterization of a β-1, 4-endoglucanase from an endophytic *Bacillus pumilus* strain. *Applied Microbiology and Biotechnology* 68:57–65

Lineweaver H, Burk D (1934) The determination of enzyme dissociation constants. *Journal of American Chemical Society* 56:658–666

Lv K, Shao W, Pedroso MM, Peng J, Wu B, Li J, He B, Schenk G (2021) Enhancing the catalytic activity of a GH5 processive endoglucanase from *Bacillus subtilis* BS-5 by site-directed mutagenesis. *International Journal of Biological Macromoles* 168:442–452

Macarron R, Acebal C, Castillon MP, Dominguez JM, De la Mata I, Pettersson G, Tomme P, Claeyssens M (1993) Mode of action of endoglucanase III from *Trichoderma reesei*. *Biochemical Journal* 289:867–873

Meleiro LP, Carli S, Fonseca-Maldonado R, da Silva Torricillas M, Zimbardi ALRL, Ward RJ, Jorge JA, Furriel RPM (2018) Overexpression of a cellobiose-glucose-halotolerant endoglucanase from *Scytalidium thermophilum*. *Applied Biochemistry and Biotechnology* 185:316–333

Naika GS, Tiku PK (2011) Influence of Ethylenediaminetetraacetic Acid (EDTA) on the Structural Stability of Endoglucanase from *Aspergillus aculeatus*. *Journal of Agricultural and Food Chemistry* 59:7341–7345.

Nakazawa H, Okada K, Kobayashi R, Kubota T, Onodera T, Ochiai N, Omata N, Ogasawara W, Okada H, Morikawa Y (2008) Characterization of the catalytic domains of *Trichoderma reesei* endoglucanase I, II, and III, expressed in *Escherichia coli*. *Applied Microbiology and Biotechnology* 81:681–689

Nakazawa H, Okada K, Onodera T, Ogasawara W, Okada H, Morikawa Y (2009) Directed evolution of endoglucanase III (Cel12A) from *Trichoderma reesei*. *Applied Microbiology and Biotechnology* 83:649–657

Phadtare P, Joshi S, Satyanarayana T (2017) Recombinant thermo-alkali-stable endoglucanase of *Myceliopthora thermophila* BJA (rMt-egl): biochemical characteristics and applicability in enzymatic saccharification of agro-residues. *International Journal Biological Macromolecules* 104:107–116

Raabo BE, Terkildsen TC (1960) On the enzymatic determination of blood glucose. *Scandinavian Journal of Clinical and Laboratory Investigation* 12:402–407

Rashmi R, Siddalingamurthy KR (2018) Microbial xyloglucanases: A comprehensive review. *Biocatalysis and Biotransformation* 36:280–295

Samanta S, Basu A, Halder UC, Sen SK (2012) Characterization of *Trichoderma reesei* endoglucanase II expressed heterologously in *Pichia pastoris* for better biofinishing and biostoning. *Journal of Microbiology* 50:518–525

Sharma K, Antunes IL, Rajulapati V, Goyal A (2018) Molecular characterization of a first endo-acting β-1, 4-xylanase of family 10 glycoside hydrolase (PsGH10A) from *Pseudopedobacter saltans* comb. nov. *Process Biochemistry* 70:79–89

Singh S, Dhillon A, Goyal A (2020) Enhanced catalytic efficiency of *Bacillus amyloliquefaciens* SS35 endoglucanase by ultraviolet directed evolution and mutation analysis. *Renewable Energy* 151:1124–1133.

Somogyi M (1952) Determination of reducing sugars by Nelson-Somogyi method. *Journal of Biological Chemistry* 200:245

Somogyi M (1945) A new reagent for the determination of sugars. *Journal of Biological Chemistry* 160:61–68

Somogyi N, Nelson N (1944) A photometric adaptation of the Somogyi method for the determination of glucose. *Journal of Biological Chemistry* 153:375–380

Son J, Cha S (2014) MALDI mass spectrometric analysis of nonderivatized steroids using cyclodextrin-supported 2, 5-dihydroxybenzoic acid as matrix. *Bulletin of the Korean Chemical Society* 35:1409–1412

Wierzbicka-Woś A, Henneberger R, Batista-García RA, Martínez-Ávila L, Jackson SA, Kennedy J, Dobson ADW (2019) Biochemical characterization of a novel monospecific endo-β-1, 4-glucanase belonging to GH family 5 from a rhizosphere metagenomic library. *Frontiers in Microbiology* 10:1342

Wu B, Zheng S, Pedroso MM, Guddat LW, Chang S, He B, Schenk G (2018) Processivity and enzymatic mechanism of a multifunctional family 5 endoglucanase from Bacillus subtilis BS-5 with potential applications in the saccharification of cellulosic substrates. *Biotechnology for Biofuels* 11:1–15

Yaoi K, Nakai T, Kameda Y, Hiyoshi A, Mitsuishi Y (2005) Cloning and characterization of two xyloglucanases from *Paenibacillus* sp. strain KM21. *Applied and Environmental Microbiology* 71:7670–7678

Yin Y-R, Li T, Sang P, Yang R-F, Liu H-Y, Xiao M, Li S, Yang L-Q, Li W-J (2020) Characterization of an alkali-tolerant, thermostable, and multifunctional GH5 family endoglucanase from *Thermoactinospora rubra* YIM 77501 T for prebiotic production. *Biomass Conversion and Biorefinery* 1–10

Yuan S-F, Wu T-H, Lee H-L, Hsieh H-Y, Lin W-L, Yang B, Chang C-K, Li Q, Gao J, Huang C-H (2015) Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*. *Journal of Biological Chemistry* 290:5739–5748

Zhang Y-HP, Cui J, Lynd LR, Kuang LR (2006) A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* 7:644–648

Zheng F, Ding S (2013) Processivity and enzymatic mode of a glycoside hydrolase family 5 endoglucanase from *Volvariella volvacea*. *Applied and Environmental Microbiology* 79:989–996

Zheng F, Vermaas J V, Zheng J, Wang Y, Tu T, Wang X, Xie X, Yao B, Beckham GT, Luo H (2019) Activity and thermostability of GH5 endoglucanase chimeras from mesophilic and thermophilic parents. *Applied Environmental Microbiology* 85:e02079-18

Chapter 4

Structure analysis of multifunctionality and mechanism of processivity of endoglucanase, *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 v3 by *in silico* modeling, molecular dynamics simulation and small angle X-ray scattering

4.1 Introduction

Endoglucanase is a primary member of the cellulase enzyme system which synergistically deconstructs the cellulose polymers having β-1,4 linked D-glucose monomers. They are grouped under 15 different Glycoside Hydrolase (GH) families in the Carbohydrate Enzyme (CAZy, http://www.cazy.org/Glycoside-Hydrolases.html) database (Cantarel *et al.,* 2009). GH5 is one of the largest families among GH, which currently contains more than 27,129 GH sequences, of which 616 enzymes are characterized and around 107 structures solved by X-ray crystallography as of July 2023. GH5 family primarily represents endo-β-1,4-glucanase. However, many members of GH5 possess broad specificity showing other enzyme activities like mixed-β-1,3;1,4-glucanase, xylanase, galactanase and xyloglucanase (Aspeborg *et al.*, 2012).

As a result, GH5 family is further divided into various subfamilies, from GH5_1 to GH5_57. The subfamily GH5_4 describes the most versatile among GH5 subfamilies, whose signature enzyme activity is marked as endoglucanase but is also capable of hydrolysing the xyloglucan (Aspeborg *et al.,* 2012). Thus, the members of subfamily GH5_4 are multifunctional endoglucanases showing more than one type of glycoside hydrolysis.

The subfamily GH5_4 mainly contains bacterial endoglucanases which show canonical $(\beta/\alpha)_8$ TIM-barrel structure, the characteristic of GH5 family (Segato *et al.,* 2014). Interestingly, some of the endoglucanases from subfamily GH5_4 show processivity on cellulose, where they consistently release similar product in the form of certain cellooligosaccharides. The end product released by such processive endoglucanase is always a cellobiose along with cellooligosaccharides having Degree of Polymerization (DP) DP3 or DP4 as seen for GH5 endoglucanases, EG1 from fungus *Volvariella volvacea* (Zheng and Ding, 2013) and EG5C-1 from bacterium *Bacillus subtilis* (Wu *et al.,* 2018; Lv *et al.,* 2021). Generally, the family GH9 endoglucanases are known for the processivity, where a Carbohydrate Binding Module, CBM3c is essentially required for its catalytic activity as demonstrated by GH9 exocellulase, E4 from a soil bacterium *Thermomonospora fusca* (Sakon *et al.,* 1997; Kumar *et al.,* 2019). Thus, the processive multifunctional endoglucanases can curtail the requirement of various lignocellulolytic enzymes including cellobiohydrolase for biomass saccharification making the process more cost-effective (Fan *et al.,* 2021). The processivity of family GH5 endoglucanases possessing broad substrates specificity thus, becomes an interesting phenomenon in cellulose hydrolysis, where no CBM is essential to show the processivity.

The structural basis and mechanism of processivity and multifunctionality of subfamily GH5_4 endoglucanases are not well understood. Various endoglucanases of subfamily GH5_4 from different organisms including *Hungateiclostridium thermocellum* and *Ruminococcus flavefaciens* were recently investigated for their structural insights (Glasgow *et al.*, 2020). Here, the aromatic residues near the active-site were potentially accounted for substrate recognition and polyspecificity of GH5_4 endoglucanases. Hitherto, structural basis of processivity and multifunctionality of subfamily GH5_4 endoglucanases is still to be established.

As described in Chapter 1, the inhabitant bacteria of herbivorous rumen have accustomed themselves for the cellulosic pressure since millions of years. This cellulosic pressure has evolved efficient machineries in the rumen loving bacterial genomes, cellulosome being the most notable representative. Bacterium *Ruminococcus flavefaciens* FD-1 v3 is a Gram-positive rumen loving firmicute which harbours a cohort of efficient cellulase genes within its genome (Bryant *et al.*, 1958). In the last decade, *R. flavefaciens* FD-1 v3 genome was sequenced which was found to be the reservoir of several efficient glycoside hydrolases especially, cellulases (Rincon *et al.*, 2010). As shown in Chapter 3, the biochemical and functional characterization of *Rf*GH5_4 from *R. flavefaciens* revealed that the endoglucanase *Rf*GH5_4 is a multifunctional and catalytically efficient processive enzyme (Gavande *et al.*, 2022). However, the structural features and the catalytic mechanism involved in the multifunctionality and processivity of *Rf*GH5_4 are needed to be unveiled. In this study, the structural and functional basis of multifunctionality and processivity of *Rf*GH5_4 endoglucanase was investigated. Therefore, the current study deals with the structural aspects of *Rf*GH5_4 by using *in silico* techniques like homology modeling, molecular

docking, molecular dynamics (MD) simulations as well as using the experimental tools such as circular dichroism (CD), structure in solution by Small Angle X-ray Scattering (SAXS) and Dynamic Light Scattering (DLS).

4.2 Materials and methods

4.2.1 Amino acid sequence analysis, evolutionary investigation

The amino acid sequence of *Rf*GH5_4 from the genome of *R. flavefaciens* was retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/WP_009984467) as discussed in Chapter 2. To unveil the percent identity of *Rf*GH5_4 with its homologous protein sequences, *Rf*GH5_4 amino acid sequence was subjected to NCBI-BLAST analysis against the PDB database (Boratyn *et al.,* 2013). Multiple sequence alignment (MSA) of *Rf*GH5_4 amino acid sequence with above mentioned PDB homologues was performed using the Clustal Omega server of EMBL-EBI (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers and Higgins, 2018) and analysed by the ESPript 3.0 server (http://espript.ibcp.fr/ESPript/ESPript/) (Robert and Gouet, 2014). Phylogenetic analysis of endoglucanase *Rf*GH5_4 was carried out by Neighbour-Joining method using MEGA11 software (Tamura *et al.,* 2021) for determining its evolutionary relationship with the GH5 family glycoside hydrolases.

4.2.2 Secondary structure analyses of RfGH5_4

The secondary structure composition of the enzyme, *Rf*GH5_4 was predicted by using online server, PSIPRED (McGuffin *et al.,* 2000) and SOPMA (Geourjon and Deleage 1995) tool. It was further confirmed by circular dichroism (CD) by subjecting *Rf*GH5_4 (2 µM) at 25°C to far UV region between 190 to 240 nm in 20 mM sodium phosphate buffer, pH 7.0 with 1 nm of bandwidth and scanning rate of 50 nm/min using a CD spectrophotometer (Jasco J-815, Japan). The CD data were drawn in the form of difference in molar extinction coefficient ($\Delta\varepsilon$, Deciliter mol^{-1} cm^{-1}) as a function of wavelength (Kelly *et al.,* 2005). The final percentage of α-helices and β-sheets

calculated by CD analysis were estimated by by K2D3 software (http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/) as per the formula given below (Kelly *et al.*, 2005; Kumar *et al.*, 2021).

$$mre = \frac{\theta \times mrw}{10 \times c \times Path\ length}$$

Where,

θ = Observed ellipticity (milli degree), recorded against respective wavelength,

mrw = Molar residual weight of the protein :

$$\frac{\text{Molecular weight of } Rf\text{GH5\_4}}{\text{Total Number of amino acids} - 1}$$

c = Concentration of the protein in mg/mL,

Path length = 0.1 cm for the CD cuvette.

The differential extinction coefficient, Δε was calculated by using following formula as reported earlier (Louis-Jeune *et al.*, 2012) in which the above calculated value of *mre* was used.

$$\Delta\varepsilon = \frac{mre}{3298}$$

4.2.3 Homology modeling

The amino acid sequence of cloned *Rf*GH5_4 protein containing His_6-tag at N-terminal was submitted to different online homology modelling servers namely SwissModel (Waterhouse *et al.*, 2018), RaptorX (Källberg *et al.*, 2014), I-TASSER (Yang and Zhang, 2015), Phyre2 (Kelley *et al.*, 2015) and Modeller 9.21 (Webb and Sali, 2017) software. The 3D structures were obtained from all the above-mentioned servers in the form of PDB files.

4.2.4 Refinement, energy minimization and quality assessment

The obtained PDB structures were subjected to energy minimization using the online tool, YASARA (http://www.yasara.org/minimizationserver.php) (Land and Humble, 2018). The energy minimized PDB structures of *Rf*GH5_4 generated from different servers were analysed for their quality using UCLA (USA) SAVES v6.0 server (Sharma *et al.*, 2021). SAVES utilizes various tools like Verify3D, ERRAT and PROCHECK-Ramachandran Plot to evaluate the fitness of a protein model. ProSA server (https://prosa.services.came.sbg.ac.at/prosa.php) (Wiederstein and Sippl, 2007) was also employed to check the statistical Z-score deviation of the modeled structure by comparing it with the Z-score of known NMR and X-ray based high resolution crystallographic structures of the different proteins. Based on the SAVES server and Ramachandran plot analysis, the 3D-model generated by RaptorX server was found to best satisfy the quality assessment criteria (**Table 4.1**). The 2D topology of *Rf*GH5_4 was drawn by using PDBSum database (http://www.ebi.ac.uk/thorn-ton-srv/databases/pdbsum/Generate.html).

Table 4.1. Quality assessment of *Rf*GH5_4 models by SAVES v6.0.

Server	Verify -3D (%)	ERRAT (%)	PROCHECK Ramachandran Plot (Percentage, % and Residue Number)				
			Most Favoured %	*Additional Allowed %*	*Generously Allowed %*	*Total Allowed %*	*Dis-allowed %*
I-TASSER	94.30 PASS	97.08	72.3 (227)	20.4 (64)	6.3 (20)	99.0	1.0 (3)
Phyre2	94.33 PASS	66.76	83.4 (262)	12.7 (40)	2.6 (8)	98.7	1.3 (4)
SWISS Model	90.14 PASS	91.36	87.2 (269)	11.3 (35)	0.5 (2)	99.0	1.0
Raptor X[a]	95.91 PASS	88.03	87.4 (285)	11.7 (38)	0.9 (3)	100	0.00
Modeller	90.74 PASS	78.21	84.7 (276)	12 (39)	0.8 (3)	97.5	2.5 (8)

[a]*The highlighted model was found most favourable and was used for downstream experiments.*

The total number of amino acids are shown in parenthesis.

4.2.5 Molecular dynamics simulations of RfGH5_4

The stability and conformational dynamics of the modeled 3D structure of the endoglucanase, *Rf*GH5_4 was further assessed by molecular dynamics (MD) simulations. The model of *Rf*GH5_4 generated by RaptorX server was simulated using GROMACS v5.1.4 software (Berendsen *et al.*, 1995) and the PARAM-ISHAN High Performance Computing (HPC) supercomputer facility available at Indian Institute of Technology Guwahati (IITG), India. Gromos53a6 all atom force field was utilized to evaluate the protein forces of *Rf*GH5_4 structure. The 3D modeled structure of *Rf*GH5_4 was situated in a cube of Single Point Charge (SPC) along with 29,088 water molecules as solvent. The protein was neutralized electrostatically with the help of 8 counter ions of Na^+. The whole system was equilibrated following the two-phase equilibration process. The first equilibration phase of isothermal-isochoric NVT ensemble was run for 500 ps with the iteration time of 2 fs at stable temperature of

300K. The second equilibration phase consisted of an isothermal-isobaric NPT assemblage which was also run for 500 ps with the iteration time of 2 fs and 1 bar pressure (Hess *et al.,* 2008). The modeled *Rf*GH5_4 structure was simulated for 100 ns and its stability and compactness was assessed as a function of time. The alteration in the structural conformations was evaluated by calculating Root Mean Square Deviation (RMSD, nm) of the *Rf*GH5_4 structure and Root Mean Square Fluctuation (RMSF) of the amino acids. RMSD was analyzed by least square fitting method. The *'gmx_mpi rmsd' and 'gmx_mpi rmsf'* commands were used in GROMACS software to calculate the RMSD and RMSF of *Rf*GH5_4, respectively. The Solvent Accessible Surface Area (SASA) and radius of gyration (R_g) of *Rf*GH5_4 was analysed by running *'gmx_mpi sasa'* and *'gmx_mpi gyrate'* commands, respectively. Thus, the affirmed and fit 3D structure of *Rf*GH5_4 was studied for its interactions with carbohydrate.

4.2.6 Protein-ligand interactions and active-site analyses of RfGH5_4

The interactions of catalytic residues present at the active-site of *Rf*GH5_4 with various cellulosic ligands was assessed by performing the molecular docking studies. Protein-ligand interactions were studied by using SwissDock server (Grosdidier *et al.,* 2011). Cellooligosaccharide ligands of various degree of polymerization (DP, DP2-DP10) were prepared by using GLYCAM server. These ligands were namely, cellobiose, cellotriose, cellotetraose, cellopentaose, cellohexaose and cellodecaose, extensively branched xyloglucan-oligosaccharide-XLLG (D-β-Glu-(α-1-6-D-Xyl)-β-1,4-D-Glu-(α-1-6-D-Xyl-β-1,2-D-Gal)-β-1,4-D-Glu-(α-1-6-D-Xyl-β-1,2-D-Gal)-β-1,4-D-Glu), manno(galacto)hexaose and gluco-mannohexaose. The protein-ligand complexes with lowest binding energy and having the ligand best fitting at the recognized active-site were used to draw the 2D hydrophilic-hydrophobic interaction diagram using

LigPlot^{+} tool (Laskowski and Swindells, 2011). The ligand interacting hydrophilic and hydrophobic residues were subsequently noted down and visualized using freely available version of visualizing software such as PyMOL v2.5.2 and ChimeraX v1.6.

4.2.7 Molecular dynamics simulation of RfGH5_4-Cellopentaose complex

The interaction of *Rf*GH5_4 with cellulosic ligands, where *Rf*GH5_4-Cellopentaose complex from **Section 4.2.6** was chosen for protein-ligand interaction studies, observing its remarkably low binding energy, greater stability and fitness. Moreover, cellopentaose is an oligosaccharide with sufficiently high DP (DP5) which completely fits inside the active-site of *Rf*GH5_4. The protein forces were determined using GROMOS96 43A1 all atom force field, whereas PRODRG server was used to build the ligand topology for cellopentaose to make it compatible for GROMACS. The *Rf*GH5_4-Cellopentaose complex was set up in a single point charge (SPC) box containing 19,959 water molecules as solvent. The charge on the complex was neutralized by the addition of 12 Na^{+} ions. The primary equilibration of entire system was carried out in an NVT assemblage (stable number of particles, volume and temperature) for 500 ps. The secondary equilibration in an NPT assemblage was carried out for 500 ps, where number of particles, pressure and temperature were constant. For both the equilibrations, the iteration time was 2 fs. The NPT equilibrated system was subjected to final MD simulation with the run time of 100 ns and integration time of 2 fs. The results of MD simulation of the protein-ligand complex were visualized by PyMOL 2.5.2 software (Schrodinger) and the 2D interaction diagram showing its hydrogen bonding interactions was developed by LigPlot^{+} tool (Laskowski and Swindells, 2011). The conformational properties of *Rf*GH5_4-Cellopentaose complex such as stability, RMSD, RMSF, R_g, SASA were evaluated as mentioned in **Section**

4.2.5 and compared with those of only *Rf*GH5_4 MD simulation results.

4.2.8 Solution structure analysis of RfGH5_4 by SAXS

The conformational adaptation of endoglucanase, *Rf*GH5_4 in solution form was analysed by Small Angle X-ray scattering technique on a local facility (SAXSpace, Anton Paar) BM29-BioSAXS at two different concentrations of 2.5 mg/mL and 5.0 mg/mL. The purified protein, *Rf*GH5_4 was centrifuged at 12000g and 4℃ for 15 min to remove the pellet containing possible protein aggregates. The protein, *Rf*GH5_4 and the buffer solution (50 mM sodium phosphate, pH 7.4) were filtered using 0.25 µm membrane syringe filter. The protein concentration was measured by taking absorbance at 280 nm on UV-Visible Spectrophotometer with the help of extinction coefficient (78630 $M^{-1}.cm^{-1}$) for *Rf*GH5_4. The scattering data for *Rf*GH5_4 and buffer solution were collected at wavelength (λ), 1.54 Å and 10℃ as reported earlier (Ahmed *et al.*, 2021; Sharma *et al.*, 2021). The sample to CMOS methane detector distance was kept 3.17 cm to fulfil the scattering vector (Q) range, $Q = 4 * \pi * \frac{Sin\theta}{\lambda}$ of 0.12-7.0 nm^{-1} where Q = Scattering vector, λ = wavelength and 2θ = Scattering angle (Kikhney and Svergun, 2015). The temperature of 10℃ was maintained during data collection using the Peltier Temperature controller. Two subsequent scattering frames of 30 minutes for each concentration of *Rf*GH5_4 and buffer were collected after stabilizing the beam. The scattering profile of buffer was subtracted from protein sample profiles to get absolute scattering pattern of the protein. The processing of SAXS data was performed by using ATSAS 3.0.3 software suit of EMBL laboratory (Franke *et al.*, 2017; Higasi *et al.*, 2021) PrimusQt tool of ATSAS software suit was employed to generate the Guinier plot to know the globular radius of gyration (R_g) or rod shape (Rc) of protein in both the concentrations of *Rf*GH5_4 (Guinier *et al.*, 1955). The maximum diameter

(D_{max}) and function of distance distribution, P(R) plot was deduced by using indirect Fourier transformation method by GNOM tool (Svergun, 1992). The estimation of molecular mass of *Rf*GH5_4 from SAXS data was carried out by using SAXSMoW software (Fischer *et al.,* 2010). ATSAS DAMMIF generated the twenty *ab initio* models of *Rf*GH5_4 which were further processed for averaging, clustering, filtering and refining using DAMAVER and DAMMIN to get the final shape (Volkov and Svergun, 2003). The averaged and filtered *ab initio* model *Rf*GH5_4 was superposed with the homology model using PyMOL v2.5.2 and UCSF Chimera v1.16 software. The fitting of *ab initio* model data and homology model of *Rf*GH5_4 was performed by using CRYSOL program (Svergun *et al.*, 1995).

4.2.9 Dynamic Light Scattering analysis of RfGH5_4

The behaviour of *Rf*GH5_4 in aqueous environment was studied by Dynamic Light Scattering (DLS) (Particle Analyzer, Litesizer 500, Anton Paar, Graz, Austria). Two different concentrations (1 and 3 mg/mL) of purified enzyme followed by centrifugation (13000***g***, 10 min, 4℃) and filtration through 0.25 µm membrane (Pall Corporation, USA) were subjected to DLS analysis. The measurements were made with with 40 mW diode laser of wavelength 658 nm at back-scatter angle (175º) at a constant temperature of 25℃ maintained by the Peltier-based temperature controller. The Kalliope software was used for particle size analysis after an average of 1000 processed runs by recording the apparent hydrodynamic diameter (D_h) (Ahmed *et al.,* 2021; Gavande *et al.,* 2023). The charge present on the enzyme in solution form, 400 µL (0.3 mg/mL) of filtered sample of *Rf*GH5_4 was analysed by DLS in Omega cuvette capable of carrying voltage and the zeta potential (ζ) of each sample was recorded in mV.

4.3 Results and Discussion

4.3.1 Sequence and Phylogenetic analysis of RfGH5_4

*Rf*GH5_4 is an endoglucanase, that is one of the catalytic modules of the cellulosomal complex, *Rf*GH5$_{1/2}$ from *R. flavefaciens* FD-1 v3 as discussed in **Section 2.3.1** of Chapter 2. The recombinant endoglucanase, *Rf*GH5_4 contained 367 amino acids, of which, 348 amino acids from wildtype protein, whereas 19 residues were part of the vector sequence having His$_6$-tag at N-terminal. The molecular mass of *Rf*GH5_4 was estimated to be 41.18 kDa with the help of ExPasy server (https://web.expasy.org/protparam/) which corroborated with molecular size of approximately, 41 kDa analysed by SDS-PAGE (Gavande *et al.*, 2022) as shown in Chapter 2, **Section 2.3.2**.

The PDB-BLAST analysis of *Rf*GH5_4 showed a maximum score of 270, 90% query coverage and 43.38% sequence similarity with *Endoglucanase A* of *Ruminiclostridium cellulolyticum* H10 (PDB ID: 1EDG) (**Table 4.2**). It also had remarkable homology for other bacterial endoglucanases such as *Endoglucanase D* (PDB-ID: 3NDY), *Endoglucanase-Cel5C* (PDB ID: 4NF7), *Endoglucanase E* (PDB ID :4IM4), β-1,3-1,4-glucanase (PDB ID: 4X0V) and a ruminal fungal endoglucanase (PDB ID: 3AYR) from *Piromyces rhizinflatus*. *Rf*GH5_4 sequence was aligned with its homologs as concluded from the PDB_BLAST analysis. The multiple sequence alignment (MSA) revealed the conserved and semi-conserved and non-conserved residues present in the sequence of *Rf*GH5_4. The conserved residues are shown in dark red background, whereas semi-conserved residues are depicted in red colour (**Fig. 4.1a**). A total of 14 helices (12α + 2Ƞ), 10 β-strands and 5 turns (TT) were noticed in the *Rf*GH5_4. It was evident from the MSA that Trp58, Arg80, His122, Asn167,

Glu168, Trp179, Ala243, Tyr245, Tyr248, Glu292, Trp325 and His340 are the conserved residues.

Table 4.2. PDB-BLAST analysis of *Rf*GH5_4.

Organism	Enzyme	PDB ID	Max. Score	Query Cover (%)	Identity (%)	E-Value	PDB Resolution (Aº)
Ruminiclostridium cellulolyticum H10	Endoglucanase A	1EDG	270	90	43.38	8e-88	1.6
Clostridium cellulovorans	Endoglucanase D	3NDY	268	91	44.28	3e-87	2.0
Butyrivibrio proteoclasticus B316	Endo-1,4-beta-glucanase Cel5C	4NF7	266	90	41.44	1e-86	2.1
Hungateiclostridium thermocellum	Endoglucanase E	4IM4	261	91	43.57	1e-84	2.4
Piromyces rhizinflatus	Endoglucanase	3AYR	260	89	41.42	7e-84	2.0
Caldicellulosiruptor sp. F32	β-1,3-1,4-glucanase	4X0V	258	94	41.96	5e-83	2.8

Glu168 and Glu292 were observed to be catalytic residues though MSA (**Fig. 4.1b**). These residues are also found to be conserved throughout the GH5 family (Ducros *et al.*, 1995; Bianchetti *et al.*, 2013).

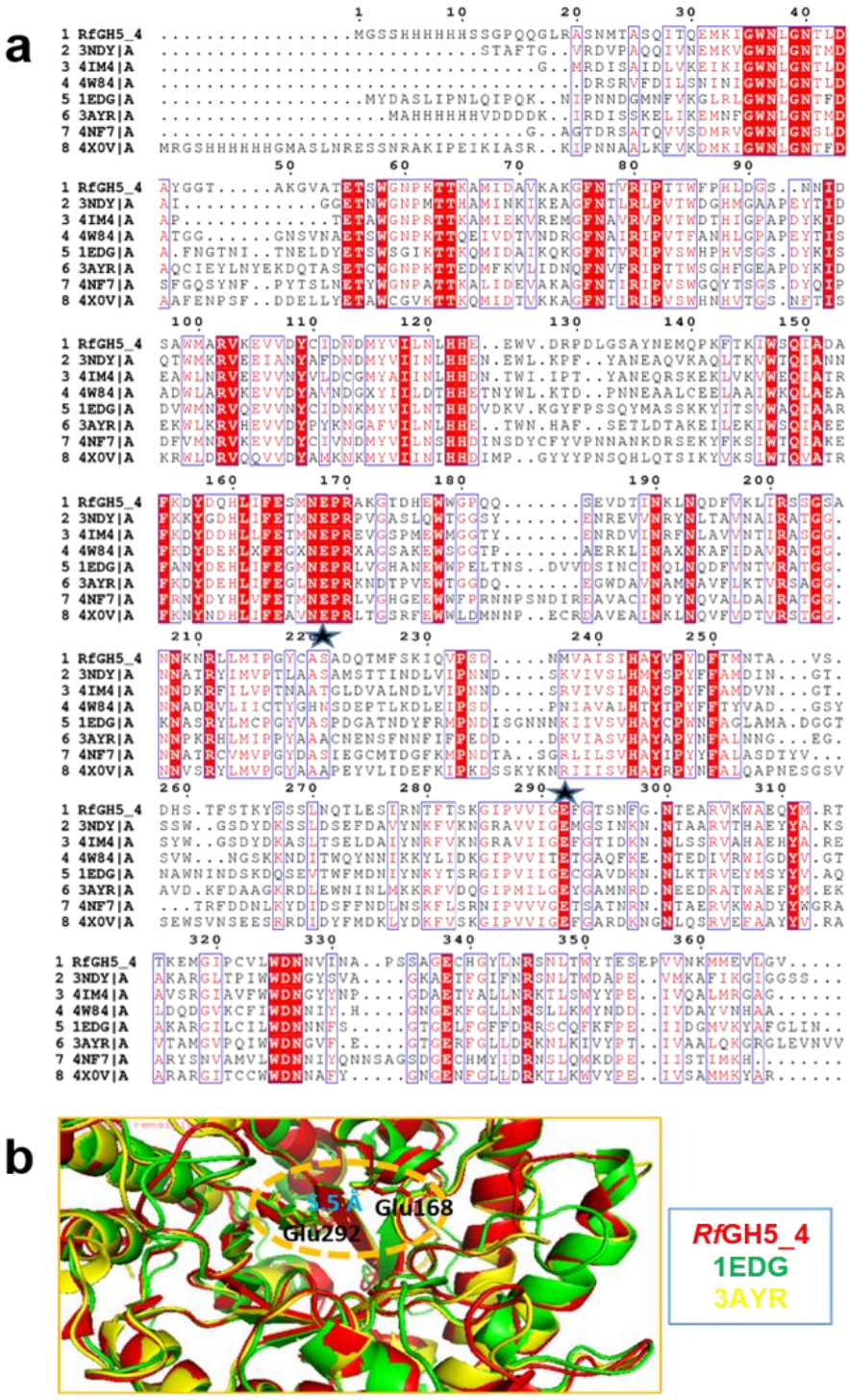

Fig. 4.1 Conserved residues and active-site mapping of *Rf*GH5_4. (a) Multiple sequence alignment of *Rf*GH5_4 with its PDB homologs, (b) Superposition of *Rf*GH5_4 with its PDB homologs 1EDG and 3AYR from bacterium *Ruminiclostridium cellulolyticum* H10 and *Piromyces rhizinflatus,* respectively, showing catalytic residues Glu168 and Glu292, 5.5 Å apart confirming the retaining type of catalytic mechanism.

The phylogenetic analysis of *Rf*GH5_4 with family GH5 endoglucanases showed that it shares common evolutionary ancestors with various bacterial as well as fungal endoglucanases (**Fig. 4.2**). The phylogenetic analysis further revealed that *Endoglucanase A* of *Ruminiclostridium cellulolyticum* H10 (PDB ID: 1EDG) is the nearest relative of *Rf*GH5_4 followed by an endoglucanase from *R. champanellensis.*

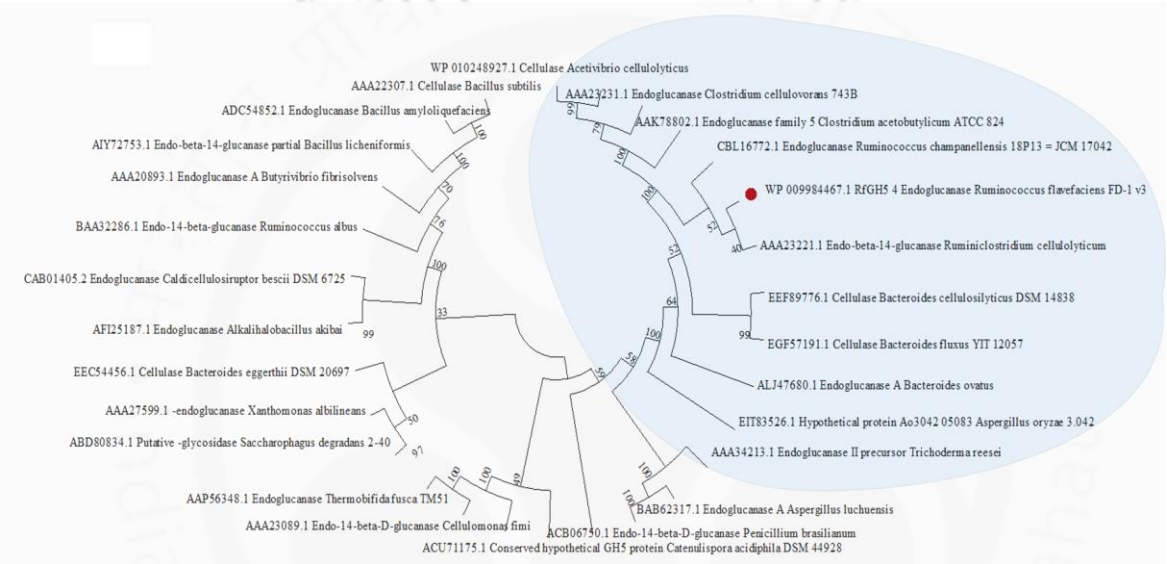

Fig. 4.2 Phylogenetic relationship of *Rf*GH5_4 with family GH5 endoglucanases in CAZy database.

4.3.2 Secondary structure determination of RfGH5_4

The percentage distribution of secondary structures in the modeled structure of *Rf*GH5_4 was assessed by using PSIPRED (**Fig. 4.3a**) and SOPMA (Geourjon and Deleage, 1995) servers available online which predicted the presence of around 36% of α-helices and 15% of β-strands as well as about 48% turns and coils (**Table 4.3**). The CD curve of *Rf*GH5_4 showing absorption due to α and β secondary structures was generated in Far-UV region as shown in (**Fig. 4.3b**). The CD analysis of *Rf*GH5_4 by using K2D3 software (Louis-Jeune *et al.*, 2012) gave 40.83% α-helices, 13.84% β-strands and 45.33% random coils and turns. The results of CD corroborated with the *in*

silico predicted secondary structure elements of *Rf*GH5_4 by PSIPRED and SOPMA.

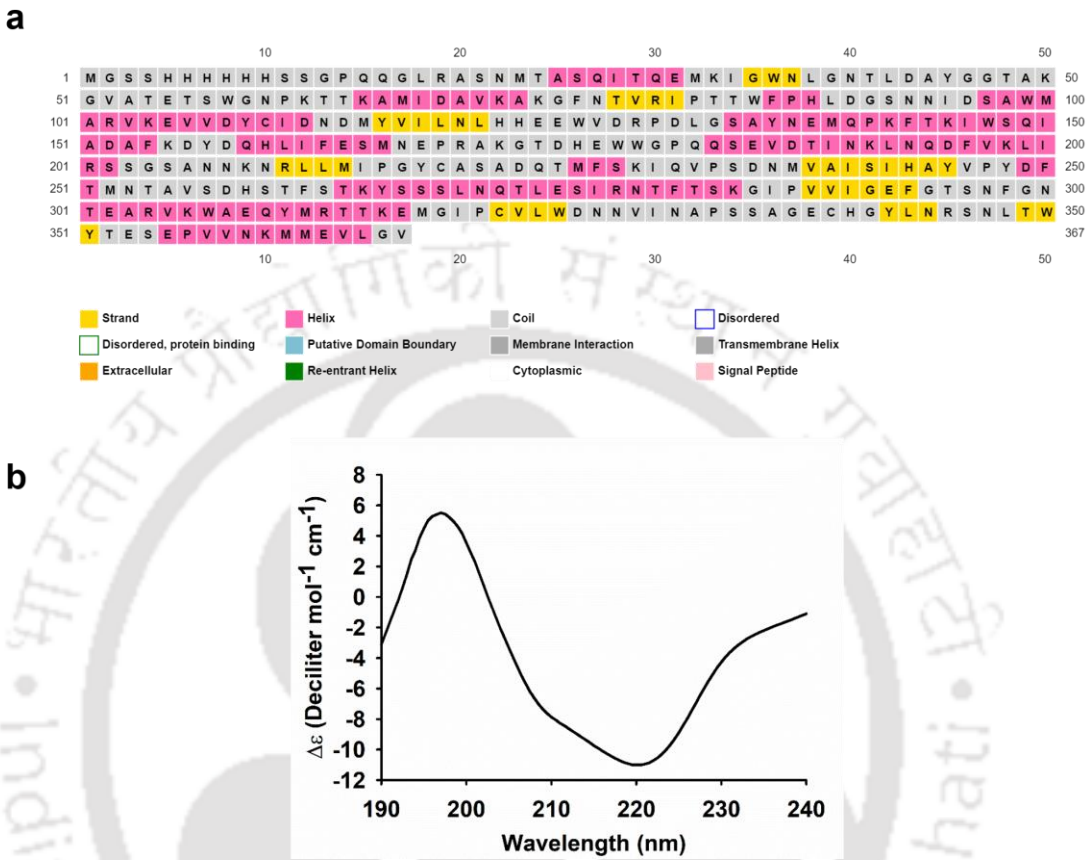

Fig. 4.3 Secondary structure analyses of *Rf*GH5_4 by using (a) PSIPRED software and (b) CD spectrum.

Table 4.3. Secondary structure analysis of *Rf*GH5_4.

Secondary structure element	PSIPRED (%)	SOPMA (%)	CD (%)
α-Helix	37.88	36.00	40.83
β-Strand	11.17	15.26	13.84
Turns/ Random coil	50.95	48.74	45.33

4.3.3 Homology modeling and structure quality assessment

The core of *Rf*GH5_4 structure is formed with β_1-β_8 sheets arranged cylindrically, whereas α_1- α_8 helices encircle the β-sheet apparatus. The α-helix, β-sheets and the loops were numbered as per the 2D topology file (**Fig. 4.4a**) generated by PDBSum server. *Rf*GH5_4 showed a total of 8 β-sheet and 10 α helices (**Fig. 4.4b**). All the 8 β-sheets

and 8 out of 10 α helices are involved in the classical $(\beta/\alpha)_8$ TIM-barrel motif formation at the core of *Rf*GH5_4 (**Fig. 4.4c**). RaptorX server generated the best fitted 3D homology model of *Rf*GH5_4 by using endoglucanase, 1EDG as the closest homolog. The catalytic residues Glu168 and Glu292 are located in the loop 1 (L1) and loop 7-8 (L7/L8) regions (**Fig. 4.4d**). L1-L8 have been marked in the surface view here in **Fig. 4.4d**. The tertiary structure of GH5 family enzymes is shaped into $(\beta/\alpha)_8$ –TIM barrel fold as can be observed for 1EDG endoglucanase from *Ruminiclostridium cellulolyticum* H10 (Ducros *et al.*, 1995). The $(\beta/\alpha)_8$ TIM-barrel fold of *Rf*GH5_4 was in agreement with an earlier reported multifunctional glucomannanase, (PDB ID: 6XSU) from *R. flavefaciens* (Glasgow *et al.*, 2020). The Ramachandran plot of energy minimized *Rf*GH5_4 structure generated by RaptorX showed 87.4% residues in most favoured region followed by 11.7% residues in additional allowed and 0.9% residues in generously allowed region (**Fig. 4.5a**). Therefore, total 100% residues of *Rf*GH5_4 were present in allowed region and no residue was detected in the disallowed region (**Table 4.1**). It was conferred from the Ramachandran plot (Hollingsworth and Karplus, 2010) that the *Rf*GH5_4 structure followed a correct backbone with permissible phi (φ) and psi (ψ) angles. ProSA analysis also showed that *Rf*GH5_4 modeled structure is free from any errors and falls within the X-ray-NMR protein crystal zone having Z-score of -8.66 (**Fig. 4.5b**). ProSA compares the modeled structure parameter with those of solved structures of various proteins using the X-ray crystallography and Nuclear Magnetic Resonance (NMR).Verify-3D tool showed 95.96 % of total amino acids of *Rf*GH5_4 above the average 3D-1D score (≥ 0.2) which established the correctness of the model (**Fig. 4.5c**). Verify-3D awarded a green Pass to the modeled structure of *Rf*GH5_4 (**Table 4.1**). The overall quality threshold for *Rf*GH5_4 by ERRAT server

was 88% which further confirmed the good quality of the structure (**Table 4.1**). Thus, various assessments of the modeled *Rf*GH5_4 structure marked the *Rf*GH5_4 structure to be satisfactory and thus it was used for further analyses.

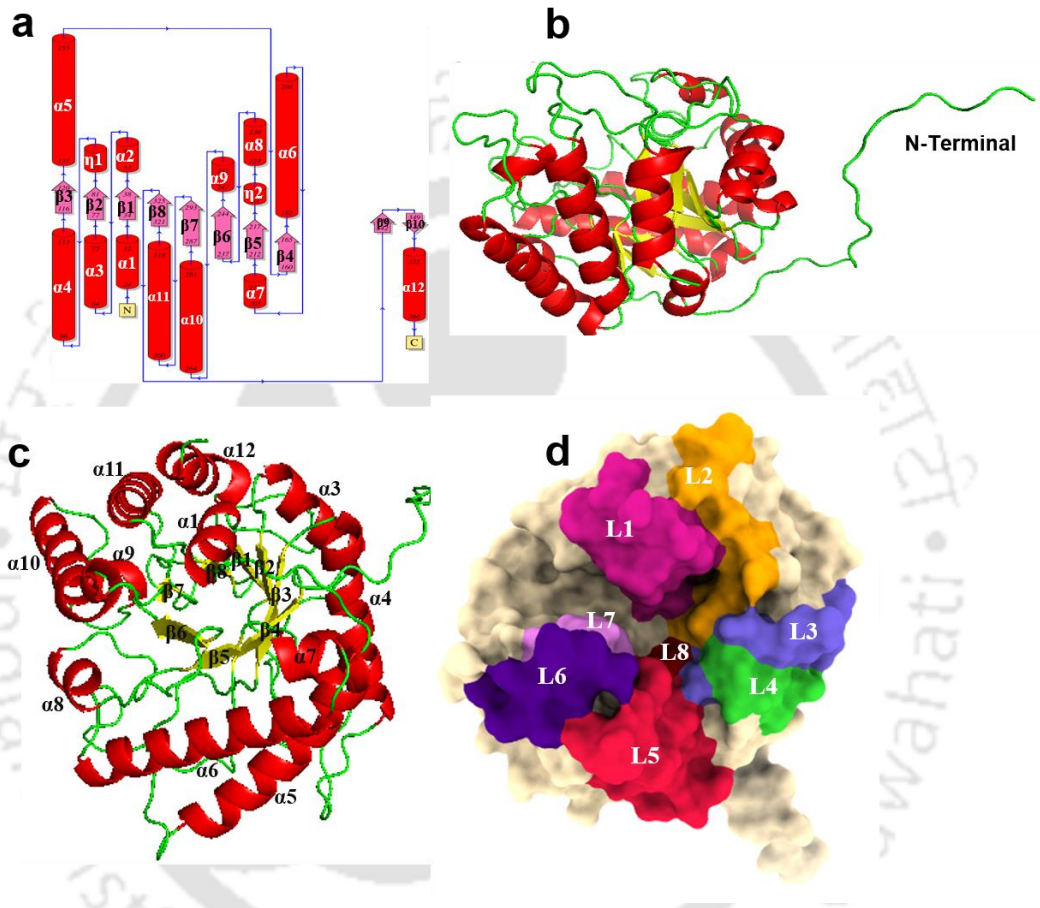

Fig. 4.4 Homology modeled *Rf*GH5_4. (a) 2D topology generated by PDBSum sever, (b) α-helix and β-strand, forming barrel structure-side view, (c) $(\beta/\alpha)_8$-TIM barrel fold of *Rf*GH5_4, (d) Loop mapping in the surrounding region of the active-site of *Rf*GH5_4.

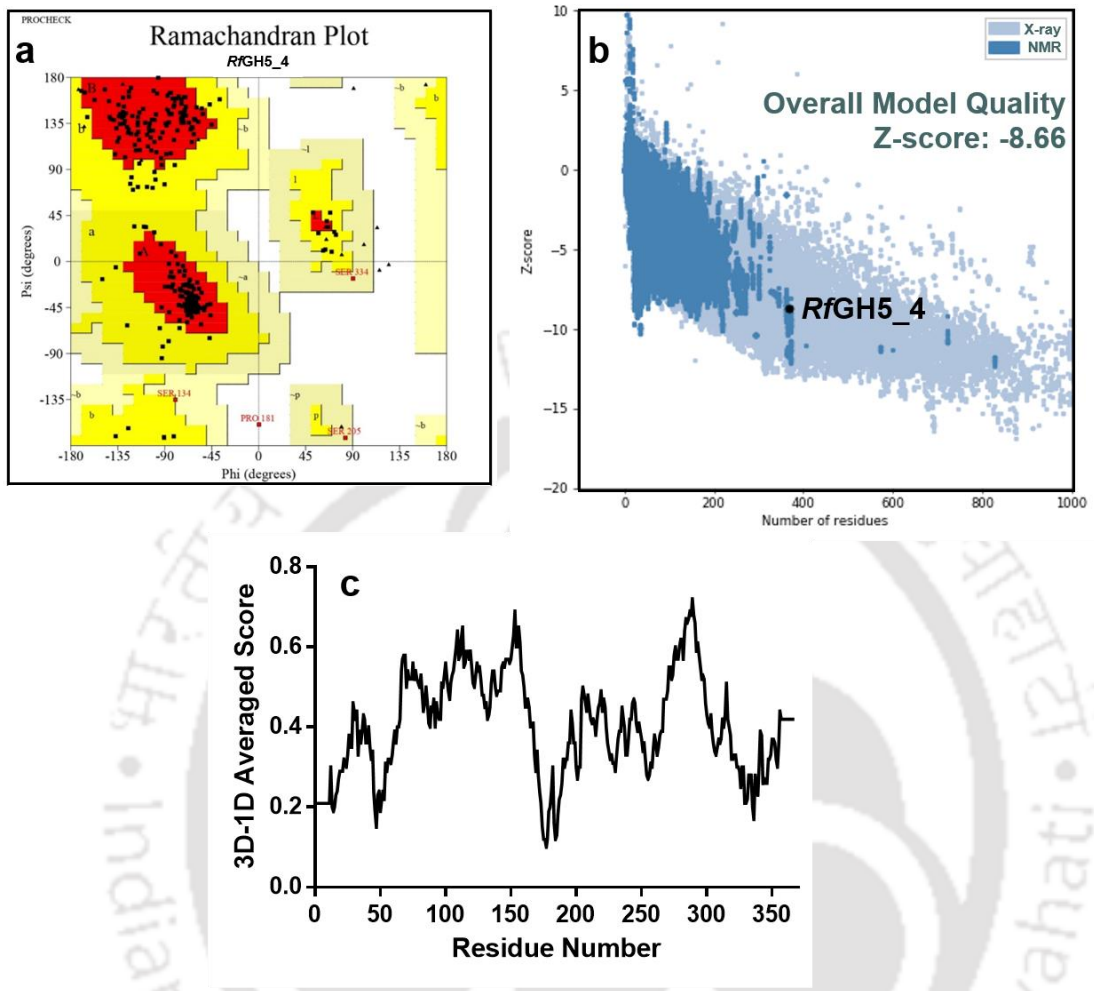

Fig. 4.5 Quality assessment of modeled *Rf*GH5_4. (a) Ramachandran plot of *Rf*GH5_4, (b) PROSA analysis showing Z-score within XRD/NMR region, (c) Verify-3D analysis showing 3D-1D averaged score >0.2.

4.3.4 Active-site analysis and mechanism of hydrolysis

Glu168 and Glu292 were analysed as the catalytic residues (**Fig. 4.1b**) when *Rf*GH5_4 structure was superposed with its homologs from PDB namely, 1EDG from bacterium *Ruminiclostridium cellulolyticum* H10 and 3AYR from fungus *Piromyces rhizinflatus*. Superposition also revealed that Glu168 is the general acid, whereas Glu292 acts as a catalytic base for *Rf*GH5_4.

Glycoside hydrolases follow two schemes for catalysis which are namely, inverting- and retaining-type mechanisms. The conformation of anomeric carbon is retained in the

retaining-type, whereas it gets altered in the inverting-type mechanism. The crucial requirement to choose one of the schemes is the distance between the carboxylic group of the two catalytic residues. The average distance between the carboxyl oxygen (δ-O) of catalytic base and catalytic acid should be 5.5 Å for retaining-type, whereas the distance of 10.5 Å between them, follows the inverting mechanism (Zechel and Withers, 2000). Glu168 and Glu292 were found 5.5 Å apart from each other as seen by PyMol 2.0 (**Fig. 4.1b**) which confirmed the retaining-type of catalytic mechanism followed by *Rf*GH5_4. GH5 family members follow the retaining-type of reaction mechanism which is also known as the double displacement mechanism of hydrolysis (Lombard *et al.,* 2013). The endoglucanases (PDB ID: 4IM4, 6MQ4, resolution 2.4 Å and 1.4 Å, respectively) from *Acetivibrio thermocellus* (formerly, *Hungateiclostridium thermocellum or Clostridium thermocellum*) and multifunctional glucomannanase (PDB ID: 6XSU) from *Ruminococcus flavefaciens* also showed the retaining-type of catalytic mechanism (Glasgow *et al.,* 2020). The open groove shape of the active-site of *Rf*GH5_4 (**Fig. 4.4d**) supports the endo-acting mode on cellulose. The open groove at the active-site helps in accommodating the longer cellulosic chains which is then hydrolysed randomly in an endo-acting mode to cellooligosaccharides as also reported earlier (Segato *et al.,* 2014).

4.3.5 Molecular dynamics simulation of modeled RfGH5_4 structure

The 3D structure of *Rf*GH5_4 was evaluated for stability and compactness through molecular dynamics simulation in an NVT and NPT environment. RMSD analysis revealed that the structural folds of *Rf*GH5_4 are consistently stable from 35 ns to 100 ns for which the average RMS Deviation was found to be 0.71 nm (**Fig 4.6a**). After initial fluctuations till 12 ns, radius of gyration (R_g) of *Rf*GH5_4 remained notably

stable at 1.85 nm throughout the 100 ns simulation (**Fig 4.6b**). Constant R_g confirmed the stability and global compactness of secondary structure conformations of *Rf*GH5_4 which reveals the robustness of the overall 3D structure. The backbone of *Rf*GH5_4 structure was further evaluated through RMSF (root mean square fluctuation) to find the changes occurred in C_α coordinates of amino acids due to any atomic displacement. RMSF of *Rf*GH5_4 amino acid residues belonging to flexible secondary structures, especially of loops, was observed to be more than 0.2 nm (**Fig. 4.6c**). The fluctuating residues such as amino acids 46-59 are part of loops L3 and L4, whereas residues 121-137 form Loop L2 (**Fig. 4.6c,** Red circles). Loop L1 is made up of residues 163-185 which also contains the catalytic Glu168 acting as a general acid. However, the position of Glu168 was notably stable in MD simulation which showed RMSF of only 0.12 nm which significantly is lower than the recommended threshold of 0.2 (**Fig. 4.6c,** E168 Green circles). The general base Glu292 is part of Loop L7 and RMSF of 0.10 nm which shows its stability (**Fig. 4.6c,** E292 Green circles). Thus, only loops of *Rf*GH5_4 was recorded to be variable but the catalytic residues were stable. Also, the residues present in various loops and at protein ends were found with fluctuation. Moreover, the position of catalytic residues Glu168 and Glu292 which are present on the loops was notably stable as seen in core of the superposed structure of *Rf*GH5_4 (**Fig. 4.6d**). The SASA of *Rf*GH5_4 lowered down from 180 nm^2 to 168 nm^2 in initial 20 ns and was notably stable till 100 ns (discussed later in this Chapter, in **Fig. 4.12d**). The stable SASA over a period of 100 ns thus entrusted fitness of the structure devoid of any deformity and the availability of its catalytic substrates to the substrate. The superposition of MD simulated structure of *Rf*GH5_4 with initial model structure showed the unchanged positions of catalytic residues further ensuring its compactness

(**Fig. 4.6c**). The stability of *Rf*GH5_4 at 300K i.e., 27℃ in MD simulation (**Fig. 4.6a and b**) is in agreement with its experimentally determined thermostability (**Fig. 3.4c**) in the temperature range of 5-40℃ as mentioned in Chapter 3, Section 3.3.2.

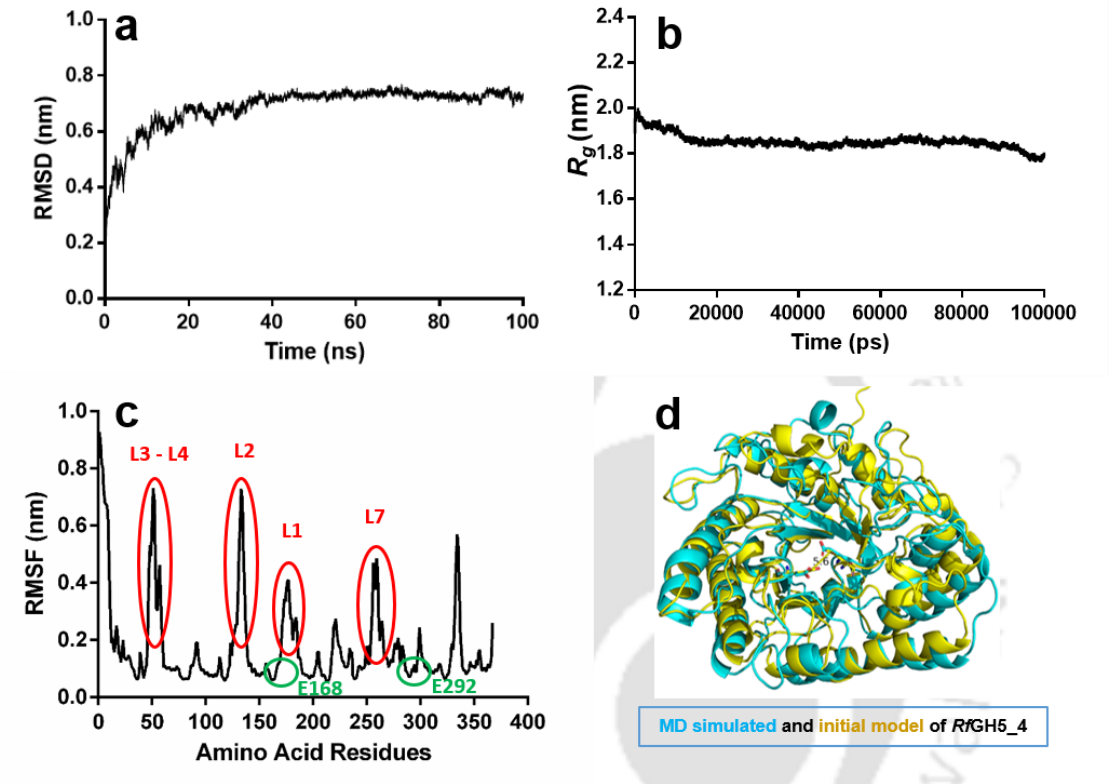

Fig. 4.6 MD simulation of RaptorX model of *Rf*GH5_4 for 100 ns. (a) RMSD, (b) R_g, (c) RMSF and (d) Superposition of MD simulated and initial model of *Rf*GH5_4.

4.3.6 Molecular docking analyses of RfGH5_4 with various cellulosic ligands

The interaction of cellooligosaccharides at the catalytic cleft of *Rf*GH5_4 was examined by using SwissDock server. The cellulosic ligands used for the experiment were cellobiose, cellotriose, cellotetraose, cellopentaose, cellohexaose and cellodecaose. The conserved residues as revealed from the MSA analysis, form the active site pocket of *Rf*GH5_4 and are part of sub-sites when interacting with the ligands. The binding energy (ΔG, kcal/mol) of the respective protein-

cellooligosaccharide complex is listed in the **Table 4.4**. Cellodecaose showed the most favourable protein-ligand interactions with release of a maximum ΔG of -12.64 kcal/mol. The increment in the release of ΔG was observed as the Degree of Polymerization (DP) of the ligand increased from DP2 to DP10. This remarks the affinity of *Rf*GH5_4 for higher order cellulosic ligands and thus, its capability of cellulose deconstruction. This is anticipated as being a GH5 endoglucanase, it carries an open groove active-site capable of accommodating cellooligosaccharides. Cellohexaose, cellopentaose, cellotetraose, cellotriose and cellobiose also showed a remarkable ligand fit at the active-site cleft for which the binding energy, ΔG was noted to be -11.70, -10.40, -9.57, -7.65, and -7.64 kcal/mol, respectively (**Fig. 4.7A-E**). The docked complexes *Rf*GH5_4 with these cellooligosaccharides were further analysed through LigPlot^{+} software to reveal the residues involved in hydrogen bond, hydrophobic and Van der Waals interaction with the respective ligand molecule (**Fig. 4.7a-e**).

The protein-ligand complexes shown here illustrate the positions of ligands at the active site. As seen for *Rf*GH5_4-cellopentaose complex (*Rf*GH5_4-*Cp*), the residues Tyr248, Tyr245, Asp223, Ser221, Glu168 and Asn40 form hydrogen bond with the cellopentaose (**Fig. 4.7d**), whereas the residues His223 and Trp225 are involved in hydrophobic interactions (**Fig. 4.7d**). The residues involved in hydrophilic and hydrophobic interactions with various ligands are enlisted in the **Table 4.4**. The *Rf*GH5_4-*Cp* complex showed a good fit at active-site and was subjected for MD simulation for 100 ns to study the protein-ligand interactions in the complex as discussed later in **Section 4.3.9**.

Table 4.4. Protein-Carbohydrate interactions of *Rf*GH5_4.

Ligand	Binding Energy ΔG (kcal/mol)	Residues involved in Polar Interactions[a]	Residues involved in Hydrophobic Interactions[b]
Cellobiose	-7.64	Glu292, Asn167, Glu338, Asn40, Trp325	His122, His123, Try245, Met252, His340, Asn327
Cellotriose	-7.65	Glu168, Tyr245, Thr273, Ser221, Ala122, Asp223,	Met252, Tyr248, Val246, Trp178, Ala244
Cellotetraose	-9.57	Asp223, Cys219, Trp178, Trp245, His243, Glu292, His123, Arg170	Tyr248, Ser221, Ala220, Glu168, Asn167, Met252
Cellopentaose	-10.40	Asn40, Tyr248, Tyr245, Asp223, Ser221, Glu168	Asn327, Trp58, His340, Trp325, Glu338, Met252, Trp178, Val246, His243, Pro247, Thr273
Cellohexaose	-11.70	Trp58, Glu338, Asp223, Tyr245, His243, Glu168, Tyr245	Tyr45, Met252, Val246, Trp178
Cellodecaose	-12.64	Asn40, Gln224, Asp223, Asn327	Tyr45, Trp58, His123, Glu168, Arg170, Trp178, Trp179, His243, Tyr245, Tyr248, Met252, Gln272, Glu292, Glu338

a- Blue and b- Orange in protein-ligand interactions by molecular docking in Fig. 4.7.

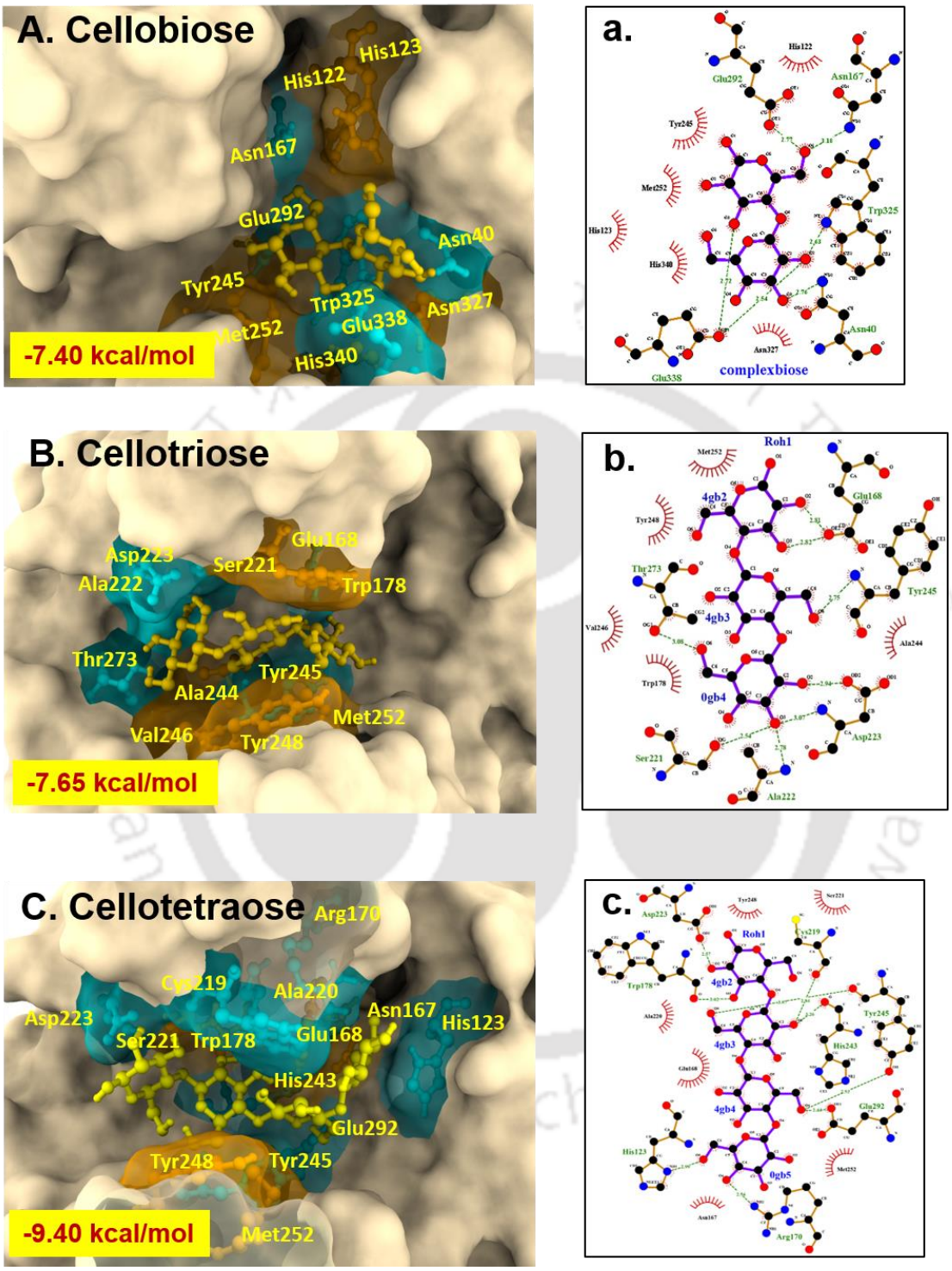

Fig. 4.7 Interactions of *Rf*GH5_4 with cellooligosaccharides, (A-a) cellobiose, (B-b) cellotriose, (C-c) cellotetraose. The 3D figures were generated with UCIF ChimeraX v1.6. Blue- hydrophilic and orange-hydrophobic interactions. 2D plots were derived by using LigPlot^{+} software.

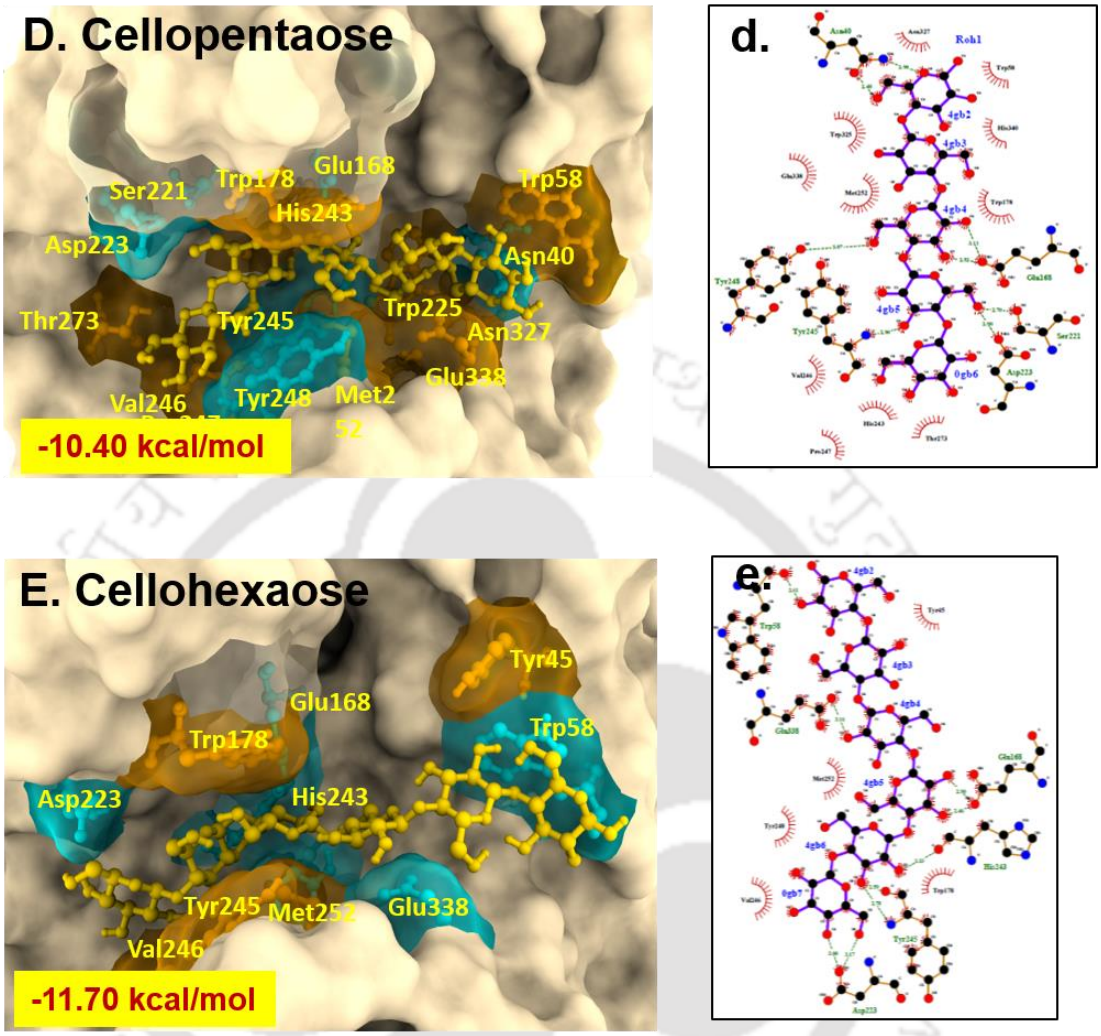

Fig. 4.7 (Contd.) Interactions of *Rf*GH5_4 with cellooligosaccharides, (D-d) cellopentaose. (E-e) cellohexaose. The 3D figures were generated with UCIF ChimeraX v1.6. Blue- hydrophilic and orange-hydrophobic interactions. 2D plots were derived by using LigPlot^{+} software.

4.3.7 Structural and functional basis of processivity of RfGH5_4

The processivity of endoglucanases is defined as the consistent release of a defined group of oligosaccharides from the cellulose chain before the enzyme falls off the polymer (Sakon *et al.*, 1997). The endoglucanase (EC 3.2.1.4), exoglucanase/cellobiohydrolase (EC 3.2.1.91) and β-glucosidase (3.2.1.21) comprise the complete cellulase enzyme system. β-Glucosidase hydrolyses cellobiose into two D-glucose units. Exoglucanases release cellobiose from either reducing or non-reducing

end of the cellulose chain and are naturally processive cellulases as they possess tunnel-shape active-site cleft as demonstrated for cellobiohydrolase (CBH-II) from fungus *Trichoderma reesei* (Rouvinen *et al.,* 1990). Endoglucanases on the other hand, carry out random hydrolysis of cellulose into cellooligosaccharides. However, some endoglucanases possess the capability to processively hydrolyse cellulose. The consistent release of cellotetraose, cellotriose and cellobiose by endoglucanase, *Rf*GH5_4 in the first 30 min of hydrolysis of Phosphoric Acid Swollen Cellulose (PASC) was earlier demonstrated through TLC (**Fig. 3.11a**) and HPLC (**Fig. 3.11b**) analyses (Gavande *et al.*, 2022) as discussed in Chapter 3. The processivity is generally observed in the family GH9 endoglucanases, which relies on CBM3c for its processive mode of catalysis (Sakon *et al.*, 1997). So, it was pertinent to know what is the structural basis of processivity of this family GH5 endoglucanase, *Rf*GH5_4 during cellulose hydrolysis. The *Rf*GH5_4 interactions with cellotetraose by molecular docking studies revealed that it fits in the catalytic cleft from subsites +2 to -2 (**Fig. 4.8a and b**). The loop L2 at the active-site of the enzyme possibly acts as a barrier for reducing end of the cellulose chain. The cellotetraose is thus accessible to the catalytic residues E168 and E292 to get cleaved into two cellobiose moieties (**Fig. 4.8a**) as cellotetraose disappeared after 30 min in TLC and HPLC experiments (Chapter 3, **Fig. 3.11a and b**). The processive behaviour displayed by a family GH9 endoglucanase, *Ht*GH9 from *Acetivibrio thermocellus (*formerly, *Hungateiclostridium thermocellum)* also had a loop barrier at the active-site (Kumar *et al.,* 2021). The processivity of GH5 endoglucanases on cellulose often possess similarities to that of GH9 endoglucanases where the product released ranges from cellotetraose (DP4) to cellobiose (DP2) (Wu and Wu, 2020). Cellotriose and cellobiose were also detected as the major products, along with the

cellotetraose during 24 h of hydrolysis of PASC by *Rf*GH5_4 (Chapter 3, **Fig. 3.11a and b**). Here, loop L2, L3 and L4 possibly act as barrier to bend the cellulose polymer outward of the open groove so that cellotriose part of cellulose chain can sit from subsite +3 to +1 and get released or cellobiose can sit from +2 to +1 and get cleaved (**Fig. 4.9a**). Similar results on generation of cellotriose and cellobiose from PASC by a GH5 endoglucanase, *EngE* (PDB ID: 6UI3) from *Clostridium cellulovorans* in 2 h reaction was reported earlier (Glasgow *et al.,* 2020). Interestingly, docking studies of *Rf*GH5_4 with cellopentaose showed that it sits from subsite +2 to -3 at the active-site. However, cellopentaose was not produced or detected in TLC or HPLC analyses. It may be possible that cellopentaose has been hydrolysed instantaneously into cellotriose and cellobiose (**Fig. 4.9a**). The residues Trp58 and Trp178 of *Rf*GH5_4 were observed to interact with cellopentaose (**Fig. 4.7D**) and cellohexaose (**Fig. 4.7E**). Similarly, the amino acid residue, Trp197 at the active-site was demonstrated to be important for the processivity of a family GH5 processive endoglucanase, *CHU_2103* from *Cytophaga hutchinsonii* (Zhang *et al.,* 2014).

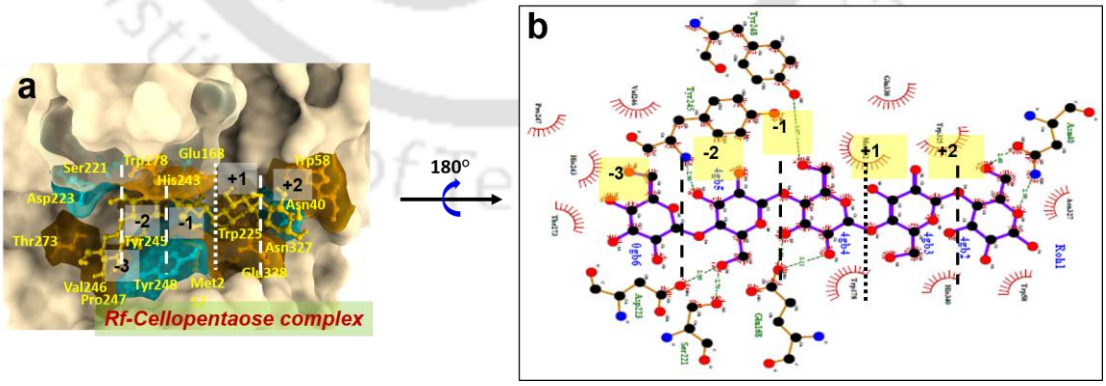

Fig. 4.8 Processivity of *Rf*GH5_4 on cellulose. (a) Cellopentaose located at subsite +2 to -3 of *Rf*GH5_4 active-site, (b) Protein-ligand 2D plot *Rf-Cp* complex with +2 to -3 subsites.

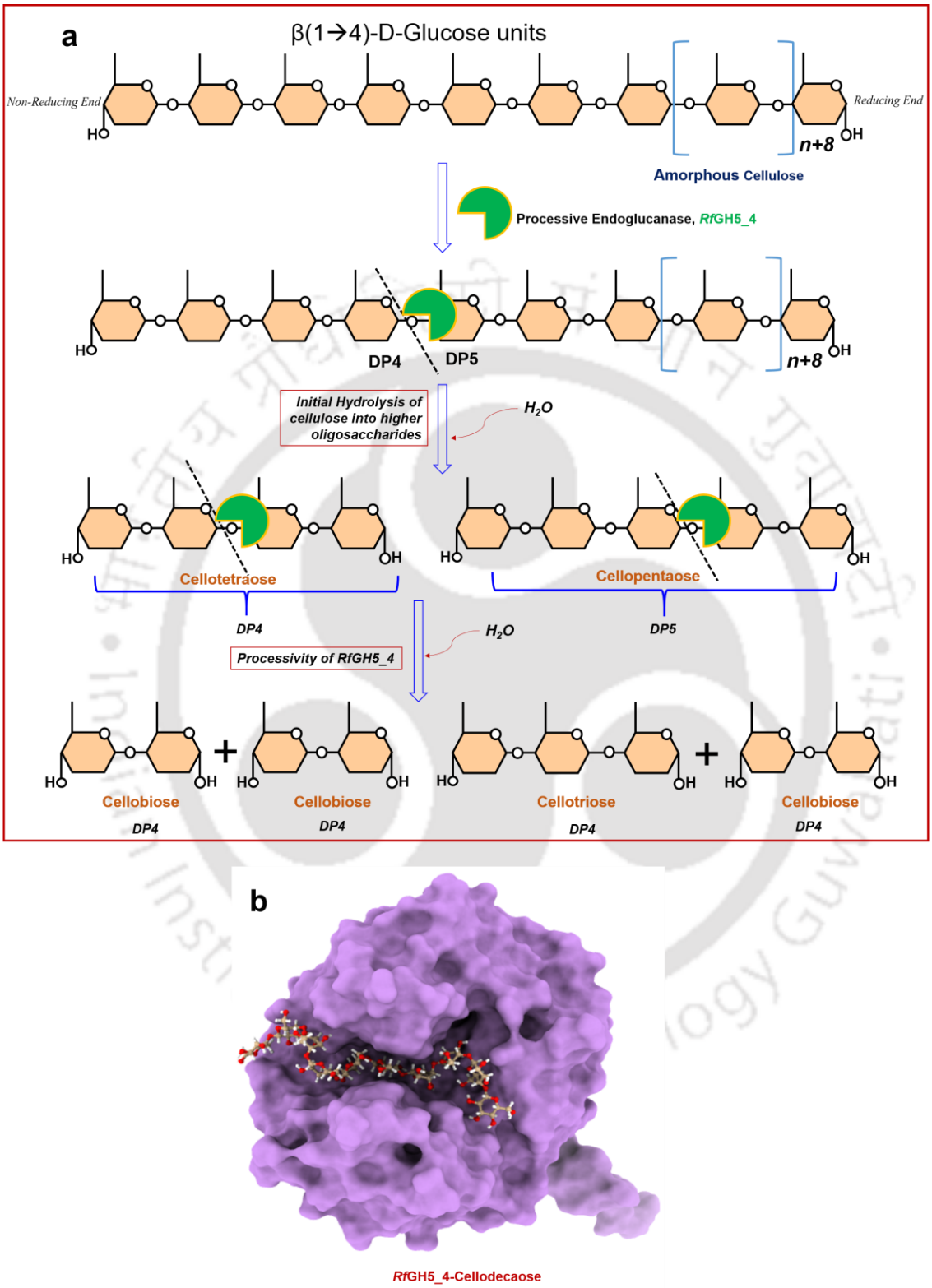

Fig. 4.9 Processivity of *Rf*GH5_4 on cellulose. (a) Mechanism of processivity of *Rf*GH5_4 elucidated from structural and functional analysis, (b) Docked *Rf*GH5_4-Cellodecaose complex.

The loop L4 containing Trp58, acting as a barrier plays an important role in the processivity of *Rf*GH5_4 as also seen for Trp197 in *CHU_2103* from *Cytophaga hutchinsonii*. The mutational studies would further confirm the importance of amino acid, Trp58 in the processivity of *Rf*GH5_4. Overall, the aromatic residues Trp58, Trp178, Trp179, Tyr245 and Tyr248 at active-site of *Rf*GH5_4 possibly guide the binding of cellulosic ligands towards the catalytic residues for processive hydrolysis as also reported earlier for endoglucanases, *XEG5A* and *XEG5B* (Dos Santos *et al.*, 2015). Moreover, another processive endoglucanase, EG1 of family GH5 from *Volvariella volvacea* required carbohydrate binding module (CBM1) for its processivity (Zheng and Ding, 2013). However, the recombinant *Rf*GH5_4 endoglucanase does not require any assisting CBM for its processivity. Therefore, the processivity of endoglucanase, *Rf*GH5_4 was observed to be distinct from that of an exoglucanase (cellobiohydrolase) as well as other family GH5 endoglucanases. Thus, *Rf*GH5_4 displayed unique diversity by performing processivity that is capable of cleaving cellulose into cellobiose, cellotriose and cellotetraose and also, cellotetraose into two cellobiose as well as cellopentaose into cellotriose and cellobiose without the assistance of any CBM. Moreover, *Rf*GH5_4 being an endoglucanase could accommodate cellodecaose (DP10) at its active-site as observed in the molecular docking analysis of *Rf*GH5_4-Cellodecaose (ΔG, -12.64 kcal/mol) interactions (**Fig. 4.9b**).

The accommodation of higher DP oligosaccharides further confirmed the endo acting mode of *Rf*GH5_4 displaying its function as a random endoglucanase acting on cellulosic substrates, while also hydrolysing the amorphous cellulose, PASC processively. *Rf*GH5_4 produced disaccharide and trisaccharide as well as oligosaccharides of DP higher than DP4 (DP6-12) from CMC-Na, avicel and cellulose

powder as analysed by MALDI-TOF analysis of hydrolysates and TLC (Gavande *et al.*, 2022) as reported in Chapter 3, **Section 3.3.7**. Moreover, the cellopentaose moiety was slightly detectable in the MALDI-TOF analysis of PASC hydrolysate (Chapter 3, **Fig. 3.9a**) which further supports the sitting of cellopentaose at the active-site of *Rf*GH5_4 and its subsequent hydrolysis into cellobiose and cellotriose. Therefore, the capability of simultaneously performing random as well as processive endo-hydrolysis of cellulose imparts the credibility to *Rf*GH5_4 to explore its application for bioethanol production. The processivity of *Rf*GH5_4 could prove cost-effective as it will reduce the dependency on cellobiohydrolase during saccharification for conversion of cellooligosaccharides into cellobiose which is finally hydrolysed by β-glucosidase to glucose.

4.3.8 Structural and functional basis of multifunctionality of RfGH5_4

The molecular docking of *Rf*GH5_4 with various ligands shed light on the structural basis of its multi-substrate specificity. As discussed in Chapter 3, TLC (**Fig. 3.8**) and MALDI-TOF (**Fig. 3.9**) analyses demonstrated that *Rf*GH5_4 efficiently hydrolysed β-1,4 linkages of cellulose (PASC, **Fig. 3.9a**), β-1,3;1,4 mixed linkages of barley β-D-glucan (**Fig. 3.9b**), β-1,4 linkages of carob galactomannan (**Fig. 3.9c**), β-1,4 linkages of D-glucose/D-mannose in konjac glucomannan (**Fig. 3.9d**), xylan from beechwood (**Fig. 3.9e**) and also the extensively branched hemicellulosic biopolymer, like tamarind xyloglucan (**Fig. 3.9f**). Xyloglucan hydrolysis is the characteristic of endoglucanases from subfamily GH5_4 (Cantarel *et al.*, 2009). The highly branched oligosaccharide, XLLG from xyloglucan (XyG), was found to be accommodated in the active-site groove of *Rf*GH5_4 showing the binding energy of -13.2 kcal/mol. The XLLG sits after the -1 subsite of *Rf*GH5_4 (**Fig. 4.10a-b**) where sufficient space is

available to fit such a large ligand. Moreover, the aromatic residues, Trp179 and Asp223 were involved in hydrogen bonding with D-xylose and D-galactose, the side chain residues of XLLG (**Fig. 4.10c-d**), stabilizing the protein-carbohydrate complex. Similar kind of affinity and accommodation of carob galactomannohexaose (**Fig. 4.10e**) and konjac glucomannohexaose (KG6, **Fig. 4.9f** and **4.11a,b,c**) in the catalytic cleft of *Rf*GH5_4 was observed. The open groove of the active-site of *Rf*GH5_4 can accommodate these large ligands, thus imparting the characteristic multifunctionality. Therefore, the molecular docking studies of *Rf*GH5_4 corroborated the experimental results shown in Chapter 3, **Section 3.3.7** for MALDI-TOF and TLC analyses showing its multifunctionality against broad range of cellulosic and hemicellulosic substrates (Gavande *et al.*, 2022).

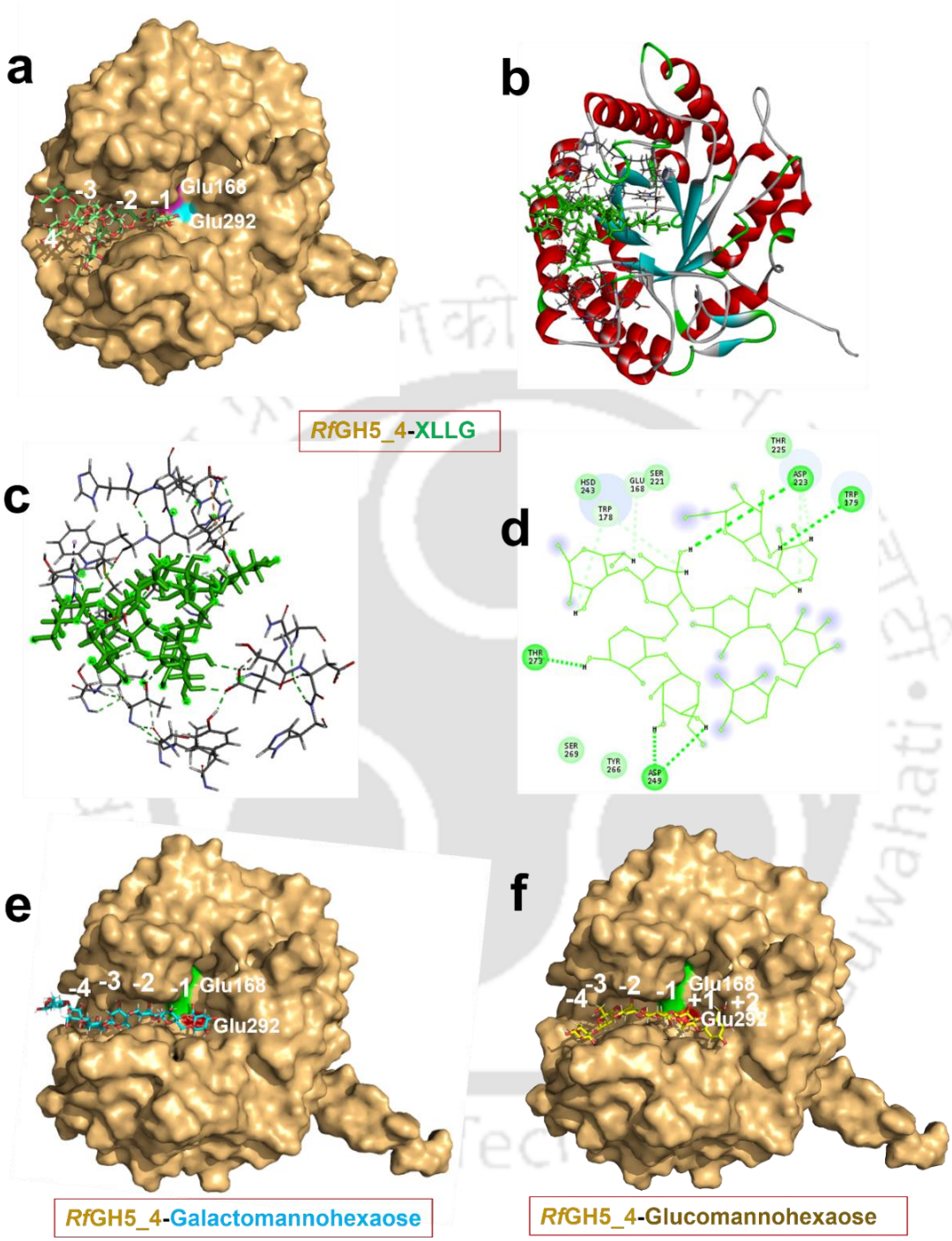

Fig. 4.10 Multifunctionality of *Rf*GH5_4 demonstrated with Xyloglucan oligosaccharide- XLLG, galactomannohexaose and glucomannohexaose. (a) XLLG sitting at the active site of *Rf*GH5_4, (b) XLLG location at *Rf*GH5_4 model structure, (c) Amino acids present in active-site pocket interacting with XLLG, (d) 2D map of active-site residue and XLLG interactions (Green dotted lines represent Hydrogen bonding), (e) *Rf*GH5_4-Galactomannohexaose complex (f) *Rf*GH5_4-Glucomannohexaose (KG6) complex.

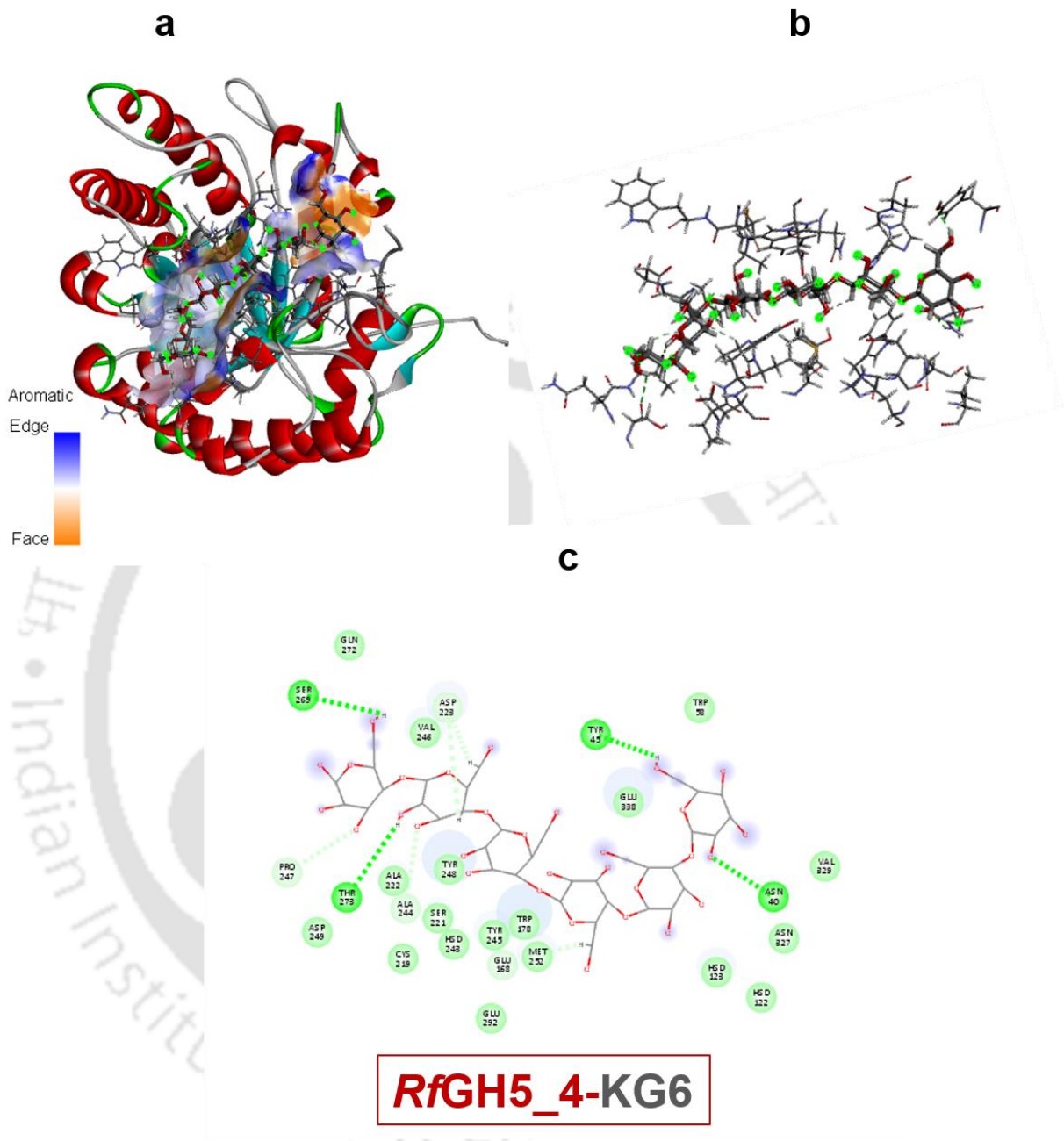

Fig. 4.11 Multifunctionality of *Rf*GH5_4 endoglucanase with glucomannohexaose ligand of konjac glucomannan (KG6). Glucomannan was observed to get comfortably accommodated at the active-site of *Rf*GH5_4 as seen through (a) the interactions of *Rf*GH5_4 amino acids with KG6 at its active-site (b) 3D model *Rf*GH5_4-KG6 interactions (c) 2D plot of *Rf*GH5_4-KG6 interactions. Green line are hydrogen bonds while blue lines show hydrophobic interactions. The plots were generated using Biovia Discovery Studio visualizer (DSV 2021).

4.3.9 Stability analysis of RfGH5_4-Cellopentaose complex by MD simulation

Endoglucanase, *Rf*GH5_4-cellopentaose (*Rf-Cp*) complex was subjected to MD simulation for 100 ns and compared with previously simulated *Rf*GH5_4 through various parameters. The average RMSD of *Rf-Cp* complex was 0.55 nm which is 0.16 nm lesser than the RMSD (0.71 nm) of independently simulated *Rf*GH5_4 (**Fig. 4.12a**). This indicated the stability of *Rf*GH5_4 with cellopentaose ligand. The average R_g of *Rf-Cp* complex (1.93 nm) differed from that of unbound *Rf*GH5_4 (2.03 nm) protein (**Fig. 4.12b**). Due to the bound cellopentaose, the *Rf*GH5_4 protein structure became more compact possibly lowering the R_g by 0.1 nm. The amino acid residues in *Rf-Cp* complex, as a part of α-helices and β-strands of secondary structure were found stable as analysed from RMSF analysis (**Fig. 4.12c**). The fluctuation was also less in the loop regions as compared with the individually simulated *Rf*GH5_4. Only the loops L1, L3 and C-terminal loops were recorded with significant flexibility (**Fig. 4.12c**). However, the low values of RMSF 0.09 and 0.07 nm of C_α atoms of catalytic residues, E168 and E292), respectively in the *Rf-Cp* complex (**Fig. 4.12c**), showed the increased stability of the residues due to the presence of ligand. The residues at N-terminal and C-terminal having higher RMSF between 0.3 to 0.8 nm showed higher flexibility (**Fig. 4.12c**). Flexible loops possibly help in accommodating long chain cellulose and hemicellulose polymers like xyloglucan as also reported previously (Nath *et al.*, 2020). The SASA of *Rf-Cp* complex was stable at 150 nm^2 which is ~15 nm^2 lesser than SASA of *Rf*GH5_4 (**Fig. 4.12d**). This smaller decrease in SASA was anticipated as the solvent accessibility to catalytic cleft was slightly abandoned owing to the presence of the ligand, cellopentaose. Stable SASA indicated that the substrate is accessible to the catalytic residues at the active-site.

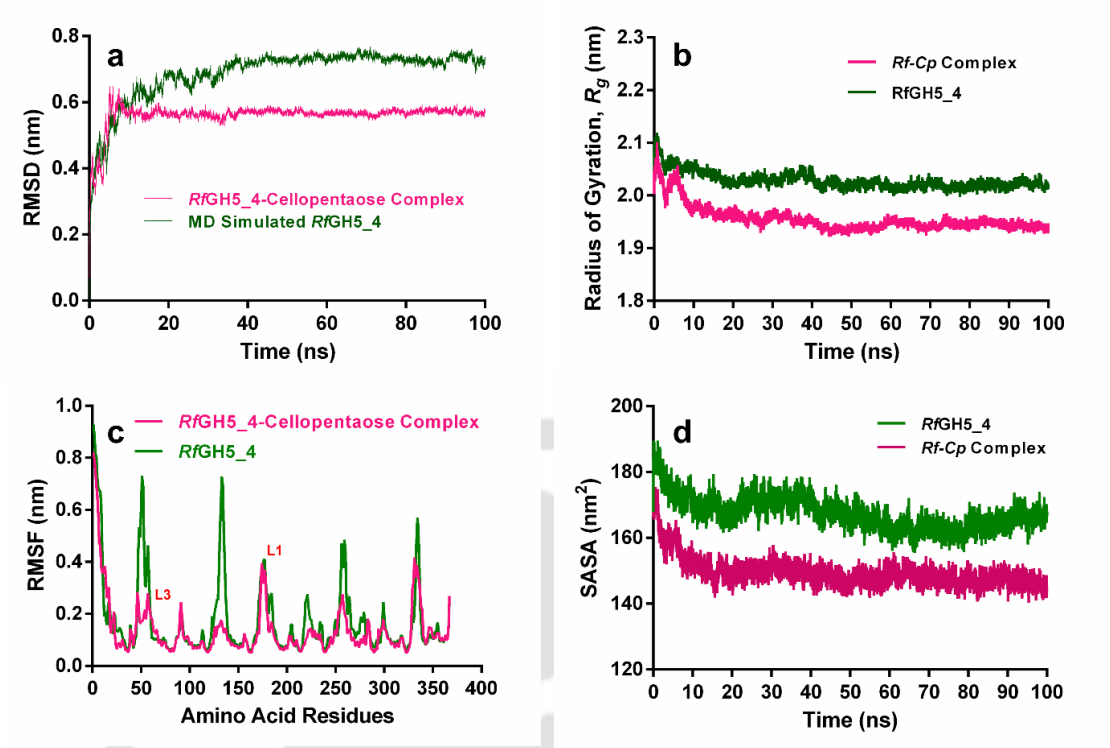

Fig. 4.12 Molecular dynamics simulation of *Rf-Cp* complex compared with that of *Rf*GH5_4 showing a) RMSD plot, b) R_g plot, c) RMSF plot and d) Solvent Accessible Surface Area (SASA) plot.

4.3.10 Solution structure of RfGH5_4 using SAXS

Endoglucanase, *Rf*GH5_4 was found stable in the temperature range of 5-40°C in our earlier report (Gavande et al. 2022). So, the behaviour, conformation and dispersity of *Rf*GH5_4 in solution (50 mM Sodium phosphate buffer, pH 7.0) at 10°C was analysed by SAXS. A good SAXS pattern for *Rf*GH5_4 was observed at 2.5 mg/mL (**Fig. 4.13a**). However, *Rf*GH5_4 was presumably aggregated at a higher concentration, 5 mg/mL as shown by ATSAS PrimusQt software 3.0.3. Thus, the SAXS data of *Rf*GH5_4 obtained at 2.5 mg/mL were used for modeling and conformational analysis (**Table 4.5**). The radius of gyration (R_g) of *Rf*GH5_4 was recorded to be 2.73 ± 0.53 nm. The low q region in Guinier plot (**Fig. 4.13a**) was observed to be linear, which ensured the monodispersed behavio4ur of *Rf*GH5_4 in solution form at 2.5 mg/mL.

Similar monodispersity of endoglucanase *Ht*GH9 of *H. thermocellum* in aqueous solution using SAXS was reported (Kumar *et al.*, 2021). The scattering profile of *Rf*GH5_4 was Fourier transformed in to P(R) plot (**Fig. 4.13b**). The diameter (D_{max}, 6.0 nm) and R_g (2.61 nm) obtained from P(R) plot (**Fig. 4.13b**) corroborated with R_g from Guinier plot (2.73 nm, **Fig. 4.13a**) and thus asserted its globular nature. The Kratky plot gave the global compactness and flexibility of *Rf*GH5_4. Bell shaped peak observed in the low q region of Kratky plot suggested the compact and proper folding of the protein structure (**Fig. 4.13c**). The molecular mass of *Rf*GH5_4 by SAXSMoW was found to be ~44.7 kDa which is near to the theoretical molecular weight of 41.18 kDa of *Rf*GH5_4. The *ab initio* model of *Rf*GH5_4 was generated at the resolution of 0.6 nm using DAMFILT facility of ATSAS DAMMIF software (Volkov and Svergun, 2003; Sharma *et al.*, 2021). A single domain *Rf*GH5_4 with a tail for the N-terminal His_6-tag was observed. The *ab initio* model (**Fig. 4.13d,e**) was found superposed with the homology modeled 3D structure of *Rf*GH5_4 (**Fig. 4.13f**). However, the variable region due to the tail of His_6-tag (**Fig. 4.13g**) could be seen in the superposed structure of *ab initio* and homology model of *Rf*GH5_4 as also reported earlier for recombinant xylanase, *Ps*GH10A from *Pseudopedobacter saltans* comb. nov. (Sharma *et al.*, 2018). The *ab initio* model of *Rf*GH5_4 was observed to be having a rattle-toy shape. A multifunctional glucomannanase of family GH5, *6XSU* from *R. flavefaciens* (PDB ID: 6XSU) also showed excellent fit with the SAXS generated *ab anitio* model of *Rf*GH5_4 which was observed to be rattle-toy shaped (**Fig. 4.14**). Moreover, a notable open-groove like shape (**Fig. 4.13e**) near the active-site region in *ab initio* model of the solution structure of *Rf*GH5_4 was detected, which was also seen in its homology modeled structure.

Table 4.5. SAXS data analysis parameters of *Rf*GH5_4.

Data-collection parameter	*Rf*GH5_4
Instrument	SAXSPace 2.0
Wavelength (Å)	1.54 Å
Q range (nm^{-1})	-0.8-0.8
Exposure time (min)	2x30
Temperature (°C)	10
Protein Concentration mg/mL	2.5
Structural parameter	
Q range (nm^{-1}) used for R_g analysis	0.04-0.12
I(0) au from Guinier	19325.10±1698.38
R_g nm from Guinier	2.73±0.53
R_g nm from P(R)	2.61
D_{max} (nm)	6.0
Porod volume (nm^3)	437.61
Molecular Mass Determination	
Theoretical molecular mass (kDa)	41.18
Molecular mass from SAXSMoW (kDa)	44.7
Modeling Parameters	
Normalized spatial discrepancy (NSD)	0.807
χ^2 Value from DAMMIF	0.769
Resolution (nm)	0.6
S range (nm^{-1}) used for CRYSOL	0.05-0.1
χ^2 Value from CRYSOL	0.76
Software Utilised	
Data analysis	ATSAS Primus
P(R) function	ATSAS GNOM
Ab initio model	ATSAS DAMMIF
Validation and average	ATSAS DAMAVER DAMFILT
SAXS data and structure model fitting	CRYSOL
Superposition	SUPCOMB
3D representation	PyMOL 2.5.2

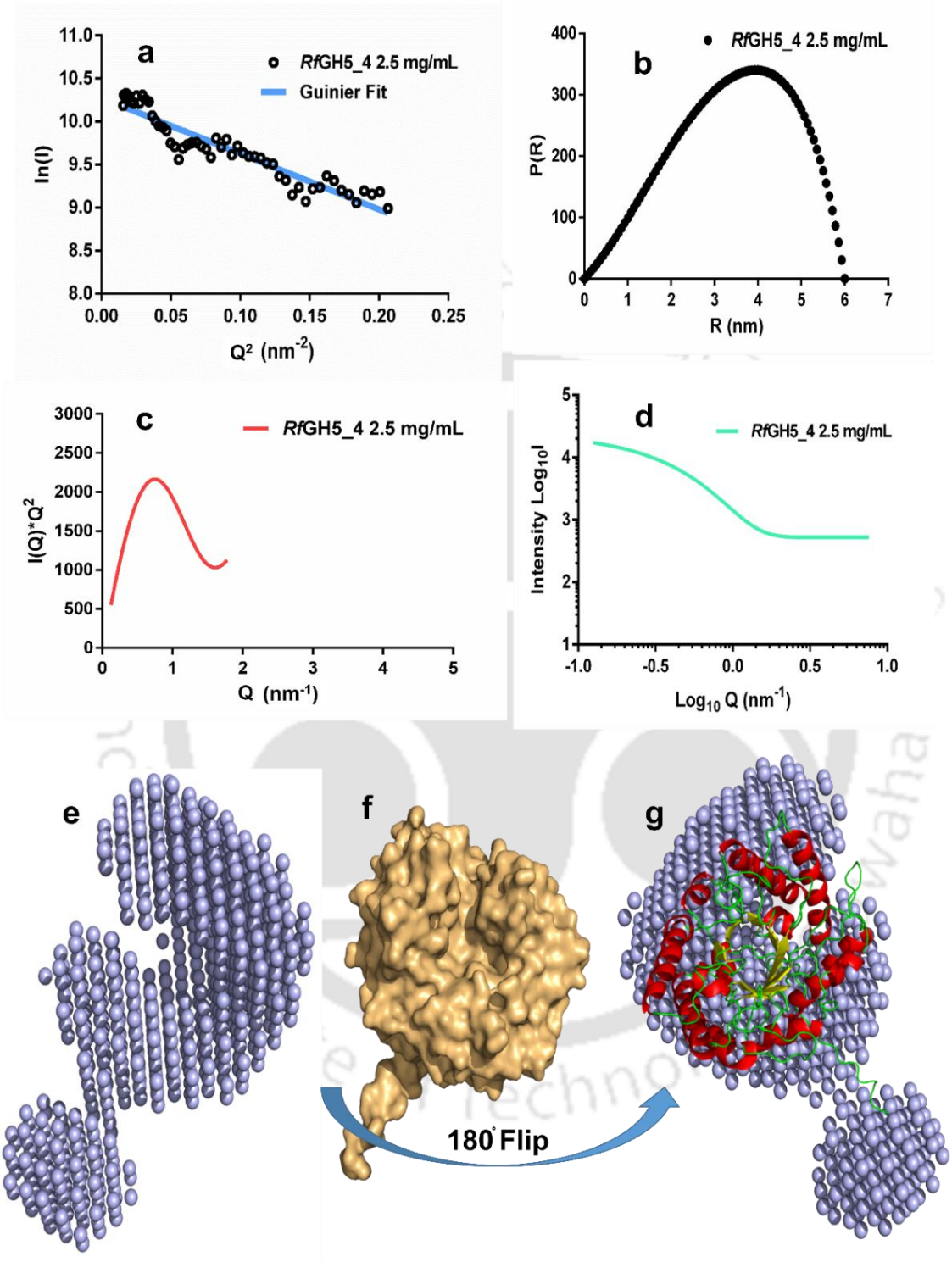

Fig. 4.13 Solution structure behaviour of *Rf*GH5_4 at 2.5 mg/mL. (a) Guinier plot showing SAXS scattering pattern, (b) P(R) plot, (c) Kratkey plot, (d) Scattering profile of *Rf*GH5_4 at 2.5 mg/mL concentration, (e) SAXS DAMFILT *ab-initio* structure of *Rf*GH5_4 (f) *Rf*GH5_4 Model-Surface view, (g) Superposition of SAXS DAMFILT *ab initio* envelope and homology modeled structure of *Rf*GH5_4.

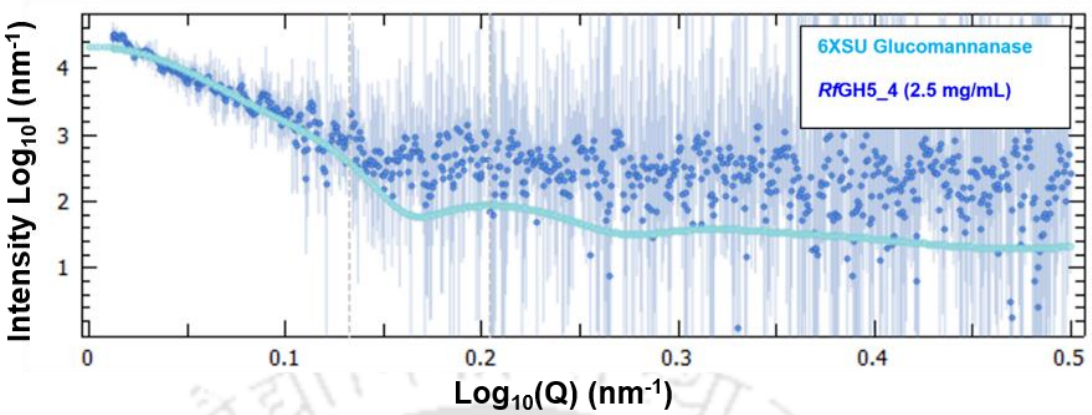

Fig. 4.14 Superpostion of a family GH5 multifunctional glucomannanase (PDB ID: 6XSU) from *R. flavefaciens* with SAXS profile of *Rf*GH5_4 using ATSAS CRYSOL. Cyan is 6XSU and Blue dotted SAXS scattering profile is for *Rf*GH5_4.

4.3.11 Hydrodynamic diameter and zeta potential of RfGH5_4 by DLS

DLS analysis of *Rf*GH5_4 at two different concentrations (1 and 3 mg/mL) showed the hydrodynamic diameter (D_h) of 6.0 ±1.79 nm suggesting the monomeric form of the protein (**Fig. 4.15a**). The hydrodynamic radius, R_h (3.0 nm) derived from D_h also matched with the R_g, elucidated the SAXS analysis (2.73 nm). The average zeta potential (ξ) of *Rf*GH5_4 was revealed to be notably higher (-16.7 ±2.1 mV) in the 50 mM sodium phosphate buffer at pH 7.5 which is possibly responsible for its higher stability (**Fig. 4.15b**). As zeta potential is the charge density on surface of the protein, a protein carrying charge denotes its solubility and hence resistance towards aggregation (Salgin *et al.*, 2012). Thus, a higher zeta potential denotes the long-term stability of the *Rf*GH5_4, thereby making it a suitable candidate for various industrial bioprocesses and bioethanol production.

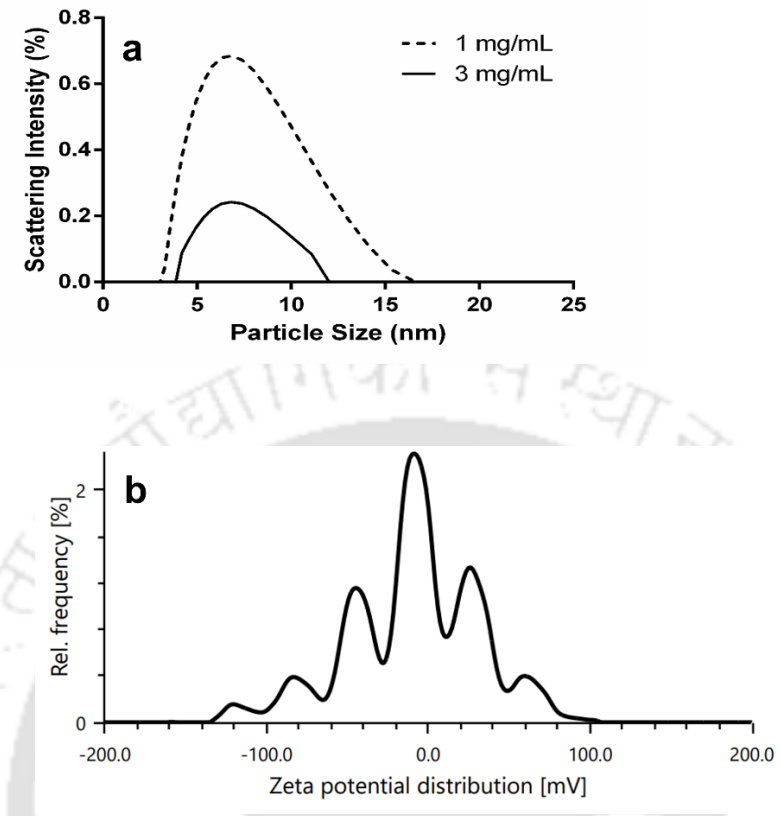

Fig. 4.15 (a) Hydrodynamic diameter of *Rf*GH5_4 by using DLS analysis and (b) Graphical representation of total Zeta potential distribution of *Rf*GH5_4.

4.4 Conclusion

The homology modeling analysis showed the $(\beta/\alpha)_8$-TIM barrel structure of endoglucanase, *Rf*GH5_4. The multiple sequence alignment of *Rf*GH5_4 showed Glu168 and Glu292 as the catalytic residues. Trp58, Arg80, His122, Asn167, Glu168, Trp179, Ala243, Tyr245, Tyr248, Glu292, Trp325 and His340 are the conserved residues. The CD analysis for secondary structure of *Rf*GH5_4 displayed 40.83% α-helices, 13.84% β-strands and 45% random coils which also corroborated with the results of prediction by PSIPRED and SOPMA. Molecular dynamic simulation of *Rf*GH5_4 structure up to 100 ns at 27°C revealed that it retained a stable conformation as an independent molecule as well as a *Rf*-Cellopentaose complex. The RMSD of *Rf*GH5_4-cellopentaose complex (0.55 nm) was lower than that of *Rf*GH5_4 (0.71 nm) which suggested its compatibility and compactness with the cellulosic ligands. The molecular docking studies of *Rf*GH5_4 with cellulosic ligands showed strong affinity with cellotetraose, cellopentaose, cellohexaose and cellodecaose. Moreover, *Rf*GH5_4 showed broad ligand specificity, wherein its active-site also accommodated various hemicellulosic ligands such as branched xyloglucan (XLLG) and glucomannan oligosaccharides. The open groove of active-site of *Rf*GH5_4 was responsible for the multifunctionality. The barrier-like conformation of loop L2, L3 and L4 containing amino acid Trp58 at +2 subsite with cellopentaose and cellotetraose imparted the processive behaviour to *Rf*GH5_4 by which it can consistently hydrolyse cellooligosaccharides into cellotriose and cellobiose. SAXS analysis showed monodispersed state of *Rf*GH5_4 at 2.5 mg/mL concentration and the rattle-toy shape in the solution form. Zeta potential on the surface of *Rf*GH5_4 at pH 7.5 was -16 mV which indicated its higher solubility in water and thus stable for longer duration.

Overall, the structural elucidation of *Rf*GH5_4 through homology modelling, molecular docking and MD simulations revealed various properties of its protein-substrate interactions. Moreover, the study of molecular basis of multifunctionality and processivity at the active-site of *Rf*GH5_4 can pave the way in further improving its catalytic diversity, thermostability and hydrolytic efficiency against various lignocellulosic substrates. This may lead to the application of this multifunctional endoglucanase, *Rf*GH5_4 in the renewable energy industry as well as in food sector.

References

Ahmed J, Kumar K, Sharma K, Fontes CMGA, Goyal A (2021) Computational and SAXS-based structure insights of pectin acetyl esterase (Ct Pae12B) of family 12 carbohydrate esterase from *Clostridium thermocellum* ATCC 27405. *Journal of Biomolecular Structure Dynamics* 40:8437–8454

Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology* 12:1–16

Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics and Communications* 91:43–56

Bianchetti CM, Brumm P, Smith RW, Dyer K, Hura GL, Rutkoski TJ, Phillips Jr GN (2013) Structure, dynamics, and specificity of endoglucanase D from *Clostridium cellulovorans*. *Journal of Molecular Biology* 425:4267–4285

Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezhuk Y (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Research* 41:W29--W33

Bryant MP, Small N, Bouma C, Robinson IM (1958) Characteristics of ruminal anaerobic cellulolytic cocci and *Cillobacterium cellulosolvens* n. sp. *Journal of Bacteriology* 76:529–537

Cantarel BL, Coutinho PM, 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

Dos Santos CR, Cordeiro RL, Wong DWS, Murakami MT (2015) Structural Basis for Xyloglucan Specificity and α-d-Xyl p (1→ 6)-d-Glc p Recognition at the− 1 Subsite within the GH5 Family. *Biochemistry* 54:1930–1942

Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe H-P, Belaich J-P, Davies GJ, Haser R (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* 3:939–949

Fan MZ, Wang W, Cheng L, Chen J, Fan W, Wang M (2021) Metagenomic discovery and characterization of multi-functional and monomodular processive endoglucanases as biocatalysts. *Applied Sciences* 11:5150

Fischer H, Oliveira Neto M de, Napolitano HB, Polikarpov I, Craievich AF (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography* 43:101–109

Franke D, Petoukhov M V, Konarev P V, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography* 50:1212–1225

Gavande PV, Kumar K, Ahmed J, Goyal A (2023) Multifunctionality and mechanism of processivity of family GH5 endoglucanase, RfGH5_4 from *Ruminococcus flavefaciens* on lignocellulosic polymers. *International Journal of Biological Macromolecules* 224:1395–1411

Gavande PV, Nath P, Kumar K, Ahmed N, Fontes CMGA, Goyal A (2022) Highly efficient, processive and multifunctional recombinant endoglucanase RfGH5_4 from *Ruminococcus flavefaciens* FD-1 v3 for recycling lignocellulosic plant biomasses. *International Journal of Biological Macromolecules* 209:801–813

Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11:681–684

Glasgow EM, Kemna EI, Bingman CA, Ing N, Deng K, Bianchetti CM, Takasuka TE, Northen TR, Fox BG (2020) A structural and kinetic survey of GH5_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis. *Journal of Biological Chemistry* 295:17752–17769

Grosdidier A, Zoete V, Michielin O (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research* 39:W270–W277

Guinier A, Fournet G, Yudowitch KL (1955) *Small-angle scattering of X-rays Wiley, New York*

Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Compution* 4:435–447

Higasi PMR, Velasco JA, Pellegrini VOA, de Araújo EA, França BA, Keller MB, Labate CA, Blossom BM, Segato F, Polikarpov I (2021) Light-stimulated T. thermophilus two-domain LPMO9H: Low-resolution SAXS model and synergy with cellulases. *Carbohydrate Polymers* 260:117814

Hollingsworth SA, Karplus PA (2010) A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins *Biomolecular Concepts* 1:271-283

Källberg M, Margaryan G, Wang S, Ma J, Xu J (2014) RaptorX server: a resource for template-based protein structure modeling. *In: Protein structure prediction. Springer*, pp 17–27

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10:845–858

Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1751:119–139

Kikhney AG, Svergun DI (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters* 589:2570–2577

Kumar K, Singal S, Goyal A (2019) Role of carbohydrate binding module (CBM3c) of

GH9 β-1, 4 endoglucanase (Cel9W) from *Hungateiclostridium thermocellum* ATCC 27405 in catalysis. *Carbohydrate Research* 484:107782

Kumar K, Singh S, Sharma K, Goyal A (2021) Computational modeling and small-angle X-ray scattering based structure analysis and identifying ligand cleavage mechanism by processive endocellulase of family 9 glycoside hydrolase (HtGH9) from *Hungateiclostridium thermocellum* ATCC 27405. *Journal of Molecular Graphics and Modelling* 103:107808

Land H, Humble MS (2018) YASARA: a tool to obtain structural guidance in biocatalytic investigations. *In: Protein Engineering. Springer*, pp 43–67

Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand–protein interaction diagrams for drug discovery *ACS Journal of Chemical Information and Modeling* 51:2778-2786

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42:D490-5

Louis-Jeune C, Andrade-Navarro MA, Perez-Iratxeta C (2012) Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Structure, Function and Bioinformatics* 80:374–381

Lv K, Shao W, Pedroso MM, Peng J, Wu B, Li J, He B, Schenk G (2021) Enhancing the catalytic activity of a GH5 processive endoglucanase from *Bacillus subtilis* BS-5 by site-directed mutagenesis. *International Journal of Biological Macromolecules* 168:442–452

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16:404–405

Nath P, Sharma K, Kumar K, Goyal A (2020) Combined SAXS and computational approaches for structure determination and binding characteristics of Chimera (CtGH1-L1-CtGH5-F194A) generated by assembling β-glucosidase (CtGH1) and a mutant endoglucanase (CtGH5-F194A) from *Clostridium thermocellum*. *International Journal of Biological Macromolecules* 148:364–377

Rincon MT, Dassa B, Flint HJ, Travis AJ, Jindou S, Borovok I, Lamed R, Bayer EA, Henrissat B, Coutinho PM (2010) Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS One* 5:e12476

Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 42:W320--W324

Rouvinen J, Bergfors T, Teeri T, Knowles JK, Jones TA (1990) Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* 249:380–386

Sakon J, Irwin D, Wilson DB, Karplus PA (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology* 4:810–818

Salgin S, Salgin U, Bahadir S (2012) Zeta potentials and isoelectric points of

biomolecules: the effects of ion types and ionic strengths. *International Journal of Electrochemical Sciences* 7:12404–12414

Schrodinger LLC The AxPyMOL molecular graphics plugin for microsoft powerpoint

Segato F, Damásio ARL, de Lucas RC, Squina FM, Prade RA (2014) Genomics Review of Holocellulose Deconstruction by Aspergilli. *Microbiology and Molecular Biology Reviews* 78:588.

Sharma K, Antunes IL, Rajulapati V, Goyal A (2018) Low-resolution SAXS and comparative modeling based structure analysis of endo-β-1, 4-xylanase a family 10 glycoside hydrolase from *Pseudopedobacter saltans* comb. nov. *International Journal of Biological Macromolecules* 112:1104–1114

Sharma K, Fontes CMGA, Najmudin S, Goyal A (2021) Small angle X-ray scattering based structure, modeling and molecular dynamics analyses of family 43 glycoside hydrolase α-L-arabinofuranosidase from *Clostridium thermocellum*. *Journal of Biomolecular Structure Dynamics* 39:209–218

Sievers F, Higgins DG (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Science* 27:135–145

Svergun D, Barberato C, Koch MHJ (1995) CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* 28:768–773

Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* 25:495–503

Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution* 38:3022–3027

Volkov V V, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* 36:860–864

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* 46:W296--W303

Webb B, Sali A (2017) Protein structure modeling with MODELLER. *In: Functional Genomics. Springer*, pp 39–54

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

Wu B, Zheng S, Pedroso MM, Guddat LW, Chang S, He B, Schenk G (2018) Processivity and enzymatic mechanism of a multifunctional family 5 endoglucanase from *Bacillus subtilis* BS-5 with potential applications in the saccharification of cellulosic substrates. *Biotechnology for Biofuels* 11:1–15

Wu S, Wu S (2020) Processivity and the mechanisms of processive endoglucanases. *Applied Biochemistry and Biotechnology* 190:448–463

Yang J, Zhang Y (2015) Protein structure and function prediction using I-TASSER.

Current Protocols in Bioinformatics 52:5–8

Zechel DL, Withers SG (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. Acc Chem Res 33:11–18

Zhang C, Wang Y, Li Z, Zhou X, Zhang W, Zhao Y, Lu X (2014) Characterization of a multi-function processive endoglucanase CHU_2103 from *Cytophaga hutchinsonii*. *Appllied Microbiology and Biotechnology* 98:6679–6687

Zheng F, Ding S (2013) Processivity and enzymatic mode of a glycoside hydrolase family 5 endoglucanase from *Volvariella volvacea*. *Applied and Environmental Microbiology* 79:989–996

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 5

Deconstruction of agricultural lignocellulosic biomass by multifunctional endoglucanase, *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 v3

5.1 Introduction

Renewable energy (RE) is the keyword in the 21st century, the age of energy crisis. Bioethanol (a.k.a. ethyl alcohol) is considered one of the top RE solutions in this energy crisis era (Lange, 2007). In the antiquity, ethanol was majorly produced by mankind for consumption and cosmetic purposes from grapes, barley, jaggery or potatoes since thousands of years (Rasmussen, 2014; Niphadkar *et al.*, 2018). The starchy material of plant is fermented into ethanol with the help of various fungal species through anaerobic respiration (Matsumoto *et al.*, 1985; Lin and Tanaka, 2006). The Arab, Chinese, Indian and Greeks are accounted to distil the alcoholic beverages in the antiquity as a part of the drinks (al-Hassan, 2002). The word alcohol comes from the Arabic word, *al-kuhl* which means alcoholic sprit. However, the use of ethanol as a fuel is purely a modern age concept evolved during 19th century after its first ever

chemical synthesis from ethylene by Michael Faraday in 1825 (Faraday, 1833). The idea that blending of ethanol with gasoline (petroleum) can reduce the emission of carbon monoxide from vehicles revolutionized the biofuel industry (Solomon *et al.,* 2007). In fact, Henry Ford manufactured his initial motor models in 1908 compatible for both gasoline and alcohol (Kovarik, 1998). However, as the petroleum became the motor fuel of choice, the interest in ethanol got dampened. Interestingly, the exploration of ethanol as fuel resurrected in 1970s, initiated by Brazil, in response to uncertainties over the crude oil supply and prices.

In the past in 20th century, corn and starch was the basic source of bioethanol (or ethanol), which was termed as 1st generation (or 1G) bioethanol (Rosillo-Calle and Cortez, 1998). But in the context of food security, corn or starch based bioethanol production was criticized (Giampietro *et al.,* 1997; Mohanty and Swain, 2019). In the light of these concerns regarding 1G bioethanol, lignocellulose, the structural component of plants made up of polysaccharides as well as lignin, was considered for bioethanol production which is known as the 2nd generation (or 2G) bioethanol (Berndes *et al.,* 2001). Nowadays, algal biomass based bioethanol production is also being studied which is referred to as the 3rd generation (or 3G) bioethanol (Saini *et al.,* 2015; Jambo *et al.,* 2016). The exploration of genetically modified organisms and resources for bioethanol production comes under the umbrella of 4th generation (or 4G) bioethanol (Kumar *et al.,* 2020). Interestingly, the concept of lignocellulose based 2G bioethanol emerged in late 50's of the last century (Aditiya *et al.,* 2016). According to a report by International Energy Association (IEA), the major percentage (>80%) of our worldwide energy requirements comes from hydrocarbon based fossil fuels like petroleum, which are extracted from the crust of earth

(https://www.iea.org/reports/worlD-energy-balances-overview/world). The fossil fuels are being consumed excessively and are expected to be depleted by the year 2050 (Abas *et al.,* 2015; Helm, 2016). Lignocellulose is the most abundant storehouse of renewable organic carbon which consists of cellulose, hemicellulose, lignin and pectin as discussed in the **Chapter 1, Section 1.1.** The polymeric cellulose comprises β-1,4 linked D-glucose monomers with the average degree of polymerization (DP) upto ~300. Cellulose is the most abundant organic polymer on Earth and thus most sought for the source of D-Glucose. Monomeric D-glucose can be converted into bioethanol (EtOH) through fermentation process using microorganisms like fungus *Saccharomyces cerevisiae* and bacterium *Zymomonas mobilis* (Gonçalves *et al.,* 2016).

The major bottleneck in converting the lignocellulosic D-glucose into bioethanol is the crystalline and recalcitrance nature of the cellulose (Berndes *et al.,* 2001). The tightly bound lignin to cellulose and hemicellulose in the plant cell wall makes the access of cellulose degrading enzymes difficult (Zhao *et al.,* 2012a). However, with the advent of modern technologies, the pre-treatment of lignocellulose by various efficient methods greatly reduce the recalcitrance of plant biomass by removing lignin and hemicellulose content. Various types of chemical methods such as alkali, acid, physical methods (such as microwave, steam explosion), physico-chemical methods (such as liquid-hot water (LHW), steam explosion (SE), ammonium fiber explosion (AFEX), and soaking in aqueous ammonia (SAA) and biological pre-treatment have been employed on lignocellulose biomass (Zhao *et al.,* 2012b).

The conversion of cellulose polymer into D-glucose requires an armoury of three cellulase enzymes *viz.* endoglucanase, cellobiohydrolase and β-glucosidase as described in **Chapter 1, Section 1.2**. With the reduced recalcitrance of lignocellulose

by a suitable pre-treatment, the cellulose degrading enzymes, for example, endoglucanase can get access to the glycosidic linkages of cellulose, releasing various cellooligosaccharides of various degrees of polymerization (DPs) and D-glucose. These cellooligosaccharides are converted to cellobiose by cellobiohydrolase, which is finally converted to two D-glucose units by β-glucosidase. Therefore, the catalytic efficiency of endoglucanase is vital for the overall cost-effective bioconversion process (Eriksson *et al.,* 2002). The cohort of cellulases secreted by the important cellulolytic microbes, for example, from fungus *Trichoderma reesei* contains all of these cellulolytic enzymes (Henrissat *et al.,* 1985). However, to develop an artificial cocktail of cellulase system, endoglucanase, cellobiohydrolase and β-glucosidase are needed to be mixed in a particular ratio to achieve efficient saccharification. Thus, synergy among these three components of a cellulase system is pivotal for efficient biomass conversion into D-glucose for its fermentation to produce bioethanol (Azhar *et al.,* 2017).

A multi-substrate specific endoglucanase, *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 v3 was established as an efficient cellulolytic enzyme as demonstrated in **Chapter 3**. A cellobiohydrolase, *Ct*CBH5A (Zverlov *et al.,* 2002; Nedumaran *et al.,* 2020) and β-glucosidase, *Ct*GH1 (Sharma *et al.,* 2019) both from *Clostridium thermocellum* were reported earlier for their compatibility with an endoglucanase, *Ba*GH5-UV2 from *Bacillus amyloliquefaciens* (Singh *et al.,* 2020a) in the saccharification of *Sorghum durra* stalk (Singh *et al.,* 2020b). In the present study, the endoglucanase, *Rf*GH5_4 was explored for its application in the saccharification of various lignocellulosic agricultural residual biomasses. Moreover, the synergy of endoglucanase, *Rf*GH5_4 with cellobiohydrolase, *Ct*CBH5A and β-glucosidase, *Ct*GH1 for saccharification of pre-treated lignocellulosic biomass and Whatman Filter

Paper No. 1 to give D-glucose was also investigated under various enzyme combinations of these enzymes. *Rf*GH5_4 is stable at 30°C for 24 h (as described in **Chapter 3, Section 3.3.3),** which is the optimum temperature for carrying out simultaneous saccharification and fermentation, SSF as reported earlier (Olofsson *et al.*, 2008). Therefore, the ethanol tolerance of *Rf*GH5_4 was also explored for its possible application in SSF.

5.2 Materials and methods

5.2.1 Chemicals and reagents

Luria Bertani (LB) bacterial growth medium, kanamycin, sodium hydrogen phosphate (di- and monobasic), sodium hydroxide (NaOH), potassium bromide (KBr), cellulose powder and cellobiose were purchased from HiMedia Laboratories Pvt. Ltd., India. Glucose oxidase-Peroxidase (GOD-POD) kit for D-glucose estimation was purchased from Coral Clinical Systems (Tulip Diagnostics Pvt. Ltd., India). Syringe membrane filter of pore size 0.25 µm were procured from Axiva (Axiva Sichem Pvt. Ltd., India). Sodium salt of carboxymethyl cellulose (CMC-Na, low viscosity), Isopropyl-β-D thiogalactopyranoside (IPTG), Avicel, D-glucose and a commercial cellulase mixture (700 U/mg) of *Trichoderma reesei* were purchased from Sigma-Aldrich Co. LLC., USA. Whatman Filter Paper No. 1 (Whatman™ 1, WFP1), was procured from Cytiva-Global Life Sciences Solutions, USA. Xylan from beechwood was purchased from Sisco Research Laboratories (SRL) Pvt. Ltd., India. Alkaline lignin was purchased from Tokyo Chemical Industry (TCI) Co. Ltd., Japan. The thin layer chromatography plate (TLC Silica gel 60 F_{254}, 20 × 20 cm) was procured from Merck KGaA, Darmstadt, Germany.

5.2.2 Selection of biomass for saccharification

The purpose of the study was to explore the capability of endoglucanase, *Rf*GH5_4 in lignocellulosic biomass saccharification. So, various agricultural residual biomasses namely Cotton Main Stalk, CMS (*Gossypium hirsutum* L.), Cotton Small Branches (CSB), Maize Leaves, MZL (*Zea mays*) were collected from an agriculture farm, Nandra Haveli, Jalgaon, Maharashtra, India. Sugarcane Bagasse, SBG (*Saccharum officinarum*) was collected from Local Market in Guwahati, Assam.

Sorghum Durra stalk, SDR (*Sorghum durra*), Finger Millet Stalk, FMS (*Eleusine coracana*) were made available from Thenur, Madurai, Tamil Nadu, India. The feedstocks were individually washed and dried completely under the natural sunlight. The dried biomasses were powdered, refined to <1 mm particle size by using wire mesh and stored at 25℃ until further use.

5.2.3 Processing and pretreatment of biomass

The selected biomasses were alkali pretreated as per their previously reported respective pre-treatment methods. The lignocellulosic biomasses (5%, w/v, 50 mL) of cotton main stalk (CMS) and cotton small branches (CSB) were pre-treated in 250 mL Erlenmeyer flask with 2% (w/v) NaOH followed by autoclaving at 121℃ under 15 psi pressure for 20 min (Baig and Dharmadhikari, 2014). Sugarcane Bagasse, SBG (Talha et al. 2016), Sorghum Durra stalk, SDR (Jamaldheen *et al.*, 2018), Finger Millet Stalk, FMS (Jamaldheen *et al.*, 2019) and Maize Leaves, MZL (Madadi *et al.*, 2017) were pre-treated with 1% (w/v) NaOH followed by autoclaving at 121℃ under 15 psi pressure for 20 min. The pre-treated biomass was separated from the lignin containing hydrolysate, filtered using multifolded muslin cloth, neutralised to pH 7.0 using water followed by the drying at 70℃ for 12 h, the pH was confirmed to be 7.0 using a pH indicating strip. The respective biomasses were further washed with 500 mL of distilled water. The respective pre-treated solid neutralized biomasses were dried in hot air oven at 70℃ for 24 h and used for further experiments.

5.2.4 Characterization of raw and pretreated biomasses

The raw and pre-treated biomasses (CSB, CMS, SBG, FMS, SDR and MZL) were characterized by powder X-ray diffraction (pXRD), Field Emission Scanning

Electron Microscope (FE-SEM) based imaging and Fourier Transform-Infrared spectroscopy (FT-IR) analysis.

5.2.4.1 Degree of crystallinity of raw, pretreated and saccharified biomasses by pXRD

The difference between the raw and pre-treated biomass samples was assessed using the Crystallinity Index (*CrI*) derived by pXRD. The *CrI* of raw biomasses (CMS, CSB, SBG, SDR, FMS and MZL), pretreated biomasses (ptCMS, ptCSB, ptSBG, ptSDR, ptFMS and ptMZL) and *Rf*GH5_4 post-saccharified samples of each pre-treated lignocellulosic biomasses was assessed by X-ray diffractometer (Bruker, D8 Advance, Germany). One gram of each biomass sample was subjected to X-ray scan over a range of 2θ = 2 to 50° with the step size of 0.02°. The *CrI* (Segal *et al.*, 1959; Jamaldheen *et al.*, 2018) was calculated as follows:

$$Crystallinity\ Index\ (CrI, \%) = \frac{I_{crystalline} - I_{amorphous}}{I_{crystalline}} \times 100$$

Where, $I_{crystalline}$ is the intensity at 2θ = 22° and $I_{amorphous}$ is the intensity at 2θ = 18°.

5.2.4.2 FE-SEM analysis of raw and pre-treated biomass

The difference between surface texture of pre-treated and raw biomass was observed by FE-SEM (Zeiss, Sigma, Germany). A small mass (pinch) of each dried raw (untreated) and pre-treated biomass (ptCMS, ptCSB, ptSBG, ptSDR, ptFMS and ptSBG) sample was separately spreaded over a carbon tape and placed on the surface of stub of an electron microscope. All the samples were gold plated and put in a vacuum chamber for imaging at 5K magnification.

5.2.4.3 FT-IR spectroscopy of raw, pre-treated and enzyme saccharified biomasses

The FT-IR analysis of raw, pre-treated and *Rf*GH5_4 post-saccharified residual lignocellulosic biomasses namely, CMS, CSB, SBG, SDR, FMS and MZL biomasses was performed. Two milligram of powdered biomass was homogenized alongwith potassium bromide (KBr) in the ratio of 1:100 (w/w). The disk of KBr containing respective sample was formed individually using hydraulic pellet press (KP, Kimyaya Engineers, India) under 15 ton pressure. The FT-IR spectra of untreated, pre-treated and post-saccharified biomass were recorded as a function of % transmittance (% T) the range of IR spectrum between 4000-450 cm^{-1} using FTIR spectrophotometer (Spectrum, Two, PerkinElmer, Waltham, MA, USA). FT-IR spectra of Avicel (Sigma), cellulose powder (HiMedia), xylan from beechwood (SRL), alkaline lignin (TCI), Phosphoric Acid Swollen Cellulose (PASC, prepared from microcrystalline Avicel as shown in **Chapter 3, Section 3.2.3.1**) were also recorded as positive control for the comparative analysis.

5.2.5 Production of cellulases: endoglucanase (RfGH5_4), cellobiohydrolase (CtCBH5A) and β-glucosidase (CtGH1)

The gene encoding recombinant endoglucanase, *Rf*GH5_4 was cloned and expressed from *R. flavefaciens* as described in **Chapter 2, Section 2.2.17**. The recombinant cellobiohydrolase, *Ct*CBH5A was a kind gift from Prof. Carlos M.G.A. Fontes, NZYTech Ltd., Lisbon, Portugal. The gene encoding *Ct*CBH5A cloned in pHTP1 vector was transformed using *E. coli* BL21(DE3) cells using 50 µg/mL kanamycin. β-Glucosidase, *Ct*GH1 was cloned from *Clostridium thermocellum* (a.k.a *Acetivibrio thermocellus*) in our Carbohydrate Enzyme Biotechnology Laboratory (CEBL), IIT Guwahati as reported earlier (Sharma *et al.*, 2019). Individual glycerol

stock of *E. coli* BL21 (DE3) cells (1%, v/v) harbouring recombinant plasmids of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 were independently grown at 37℃ until the OD_{600} reached 0.4-0.6 in 400 mL LB medium using 50 µg/mL kanamycin. These cells were induced by 1 mM IPTG and further incubation at 24℃ for 16 h for enzyme expression. The *E. coli* BL21(DE3) cells for all the three expressed proteins namely, *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 were independently centrifuged at 6000***g*** and 4℃ for 20 min. The supernatant was discarded and the pellets were stored at 4℃ for further use.

It was reported earlier that *Ct*CBH5A and *Ct*GH1 are stable for more than 48 h at 40℃, retaining around 75% of their enzyme activity (Singh *et al.*, 2020b). *Rf*GH5_4 was also found stable between pH 5.5 to 7.5 of 20 or 50 mM citrate-phosphate as well as sodium phosphate buffer, as described in **Chapter 3, Section 3.3.2.** Moreover, the optimum pH of both *Ct*CBH5A and *Ct*GH1 is also within the range of 6.0-6.5 (Sharma *et al.*, 2019). So, a common buffer, 50 mM sodium phosphate, pH 6.5 was selected for preparation of cell extract of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 from the induced cell pellets. The cell pellets of 400 mL culture of all the three enzymes were seperately suspended in 10 mL of 50 mM sodium phosphate buffer, pH 6.5 and sonicated at 4 s ON and 8 s OFF cycle for 20 min. The sonicated cells were centrifuged at 12000***g*** (Thermo Scientific, Model-Legend XTR, USA) and 4℃ for 50 min. Sodium azide, 0.005% (w/v) was added to the obtained cell extracts to avoid fungal contamination. The cell extract of each enzyme was filtered aseptically through 0.25 µm membrane using the syringe filter and stored at 4℃ for their downstream application in the saccharification. Protein concentration of cell extracts was estimated by UV method and Bradford method using Bovine Serum Albumin (BSA) as standard as mentioned in **Chapter 2, Section 2.2.20.** The cell extracts of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 were

directly used for the saccharification process, instead of purifying them, to make the process cost-effective from the industrial point of view. However, purified enzyme (as described in **Chapter 2, Section 2.2.19)** was used for preliminary studies on the deconstruction of different lignocellulosic biomass hydrolysis by *Rf*GH5_4 described later in **Section 5.2.8**.

5.2.6 Enzyme assay of RfGH5_4, CtCBH5A and CtGH1

The optimum reaction pH of *Rf*GH5_4 is 5.5 of 50 mM citrate phosphate buffer and temperature 55℃, as described in **Chapter 3, Section 3.3.2**. The optimum pH and temperature for both enzymes, *Ct*CBH5A (Nath et al. 2021) and *Ct*GH1 is 6.5 of 50 mM sodium phosphate buffer and 65℃ (Sharma et al. 2019). The enzyme activity of cell extract (10 µL) of endoglucanase, *Rf*GH5_4 was determined in a 100 µL reaction mixture using 1% (w/v) CMC-Na at 55℃ and pH 5.5 of 50 mM citrate phosphate buffer (and also at pH 6.5 of 50 mM sodium phosphate buffer, as it is the common pH for all the three enzymes) for 2 min of reaction duration. The enzyme activity of cell extract of cellobiohydrolase, *Ct*CBH5A was also determined in a similar way as that of *Rf*GH5_4 by incubating it with 1% (w/v) CMC-Na for 5 min at 65°C and the reaction pH was 6.5 of 50 mM sodium phosphate buffer. The enzyme activity of both *Rf*GH5_4 and *Ct*CBH5A was determined by measuring the released reducing sugar using Nelson-Somogyi method (Nelson, 1944; Somogyi, 1945) as described in **Chapter 3, Section 3.2.2.1** against the respective cell extract as Blank (to normalize the possible reducing sugar contribution from cell extracts). For enzyme assay of β-glucosidase, *Ct*GH1, in a 100 µL final reaction volume, 10 µL of cell extract was incubated with 1.1% (w/v) cellobiose (90 µL) at 65℃ and pH 6.5 of 50 mM sodium-phosphate buffer for 5 min as reported earlier (Sharma et al. 2019). The released D-glucose was measured by taking

absorbance at A_{505} against cell extract of *Ct*GH1 as Blank using GOD-POD method (Raabo and Terkildsen, 1960), as discussed in **Chapter 3, Section 3.2.2.4**. One unit (U) of enzyme activity for all the three enzymes was defined as 1 µmole of D-Glucose equivalents released by the respective enzyme per min (µmol/min). The enzyme activity was expressed in U/mL (µmol/min/mL).

5.2.7 TLC, TRS and MALDI-TOF analyses of saccharified biomasses by purified RfGH5_4

The role of *Rf*GH5_4 in deconstruction of various pre-treated lignocellulosic biomass was explored through preliminary saccharification experiments. For saccharification, 20 mg of each dried pre-treated biomass (2%, w/v) namely CMS, CSB, SBG, FMS, SDR and MZL were individually suspended in 600 µL of 50 mM citrate phosphate buffer, pH 5.5. Antifungal reagent sodium azide, 10 µL (0.005%, w/v) was also added in the reaction to prevent any contamination. The pre-treated biomasses were independently saccharified by using 390 µL (50 µg/mL) or 87 PASC U/$g_{Biomass}$ of purified *Rf*GH5_4 endoglucanase (as described in **Chapter 2, Section 2.3.5**) at 30°C, 180 rpm for 48 h, making the final reaction volume of 1 mL. The concentration of *Rf*GH5_4 in reaction mixture was 19.5 µg/mL. The Total Reducing Sugar (TRS) was estimated by Nelson (1944) and Somogyi (1945) methods as described in **Section 5.2.6**, whereas the analysis of hydrolysed products of various biomasses by TLC was performed by using the mobile phase of n-butanol:acetic acid:water (2:1:1) and MALDI-TOF MS by using dihydroxybenzene matrix as described in **Chapter 3, Section 3.2.8 and 3.2.9**.

5.2.8 Synergy of RfGH5_4, CtCBH5A and CtGH1 on WFP1 and ptSDR

The synergy among three enzymes, *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 was explored against WFP1 and ptSDR. Total 200 $U/g_{Biomass}$ of enzyme cell extract in the ratio of 50:50:100 (*Rf*GH5_4 + *Ct*CBH5A + *Ct*GH1) was taken thereby incubating with 2% (w/v) Whatman Filter Paper No. 1 (WFP1) in a 1 mL reaction at 40°C and pH 6.5 of 50 mM sodium phosphate buffer for 48 h. Alongside, the independent reaction of each enzyme *Rf*GH5_4, *Ct*CBH5A (50 U/g_{WFP1}), *Rf*GH5_4 + *Ct*CBH5A (100 U/g_{WFP1}), *Ct*GH1 (100 U/g_{WFP1}), *Ct*CBH5A + *Ct*GH1 (150 U/g_{WFP1}), *Rf*GH5_4 + *Ct*GH1 (150 U/g_{WFP1}, 50:100) was also set up. A blank of WFP1 without enzyme was run as negative control. Similarly, the synergistic saccharification of 2% (w/v) pre-treated SDR was also investigated by using total 200 U/g of pre-treated SDR Biomass ($U/g_{Biomass}$, 50:50:100) of cellulase system, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 in a 1 mL reaction at 40°C, 180 rpm and pH 6.5 of 50 mM sodium phosphate buffer for 48 h. A blank of pre-treated SDR without enzyme was run as negative control. A separate reaction of *Ct*GH1 (100 U/g) with cellobiose was performed as a positive control to check the release of D-glucose. For comparative analysis of the saccharification, a commercial cellulase (50 $U/g_{Biomass}$) from *Trichoderma reesei* (*Tr*CellulaseSigma) was used to saccharify WFP1 and ptSDR at 40°C, 180 rpm, pH 4.8 of 50 mM citrate phosphate buffer for 48 h.

The TRS in the saccharification of WFP1 and pre-treated SDR was estimated by Nelson-Somogyi method (Nelson, 1944; Somogyi, 1945) as discussed in **Chapter 3, Section 3.2.2.** Also, the yield of D-glucose in WFP1 and pre-treated SDR saccharification was estimated using GOD-POD reagent as per the protocol given in **Chapter 3, Section 3.2.2.4**. The hydrolysed samples of WFP1 and pre-treated SDR were further analysed by TLC as described in **Chapter 3, Section 3.2.8** except the

samples were not precipitated using three volume of ethanol. Alternatively, 2 µL hydrolysate, directly from the saccharification reaction was loaded on to the TLC plate, to see the release of D-glucose in order to observe the synergy among the enzymes. The experiments were performed in triplicate and mean ±SD are shown in the graphs. The statistical analysis of significance among each reaction was performed by One-way ANOVA using the GraphPad Prism v9. The *p*-value <0.001 was considered significant using *Rf*GH5_4 saccharification reaction as the comparative control in ANOVA.

5.2.9 Ethanol tolerance of RfGH5_4

The cell extract of *Rf*GH5_4 (0.5 mL, 4 U/mL) was incubated with different concentrations of ethanol (0, 2, 4, 6, 7, 10, 20, 30 and 50%, v/v) in a final volume of 1 mL in 50 mM sodium phosphate buffer pH 6.0 for different time intervals followed by measuring the enzyme activity as described in **Section 5.2.6,** by determining the released D-glucose equivalents. The reaction mixture, 100 µL contained 90 µL of 1.1% (w/v) CMC-Na and 10 µL of pre-incubated cell extract of *Rf*GH5_4 with various concentrations of ethanol for respective time intervals of 0, 12, 24, 48, 72 and 96 h. It was assayed for the residual enzyme activity at optimum reaction pH 5.5 of 50 mM citrate phosphate buffer by incubating at 55°C for 2 min followed by of reducing sugar estimation using Nelson (1944) and Somogyi (1945) methods. The enzyme activity of cell extract *Rf*GH5_4 in U/mL without ethanol (0%) at 0^{th} h was considered as Control with 100% activity. The relative (residual) activity of *Rf*GH5_4 in presence of various ethanol concentrations was recorded with respect to that of Control (0^{th} h without ethanol). All reactions were performed in a set of triplicate and their mean (±SD) was plotted using GraphPadPrism v6.0 software.

5.3 Results and Discussion

5.3.1 Selection, processing and pre-treatment of raw biomasses

The raw agricultural residues of cotton (CMS, CSB), sugarcane bagasse (SBG), sorghum (SDR), finger millet (FMS) and maize leaves (MZL) are shown sequentially in **Fig. 5.1 (I, A-F)**. The dried lignocellulosic biomasses selected for saccharification purpose were mechanically processed into powder form **Fig. 5.1 (II)** followed by alkali pre-treatment by 1% (w/v) NaOH for SBG, SDR, FMS, MZL and 2% (w/v) NaOH for CMS and CSB (**Fig. 5.1 (III)**. The visual analysis and comparison of pre-treated biomasses (CMS**, Fig. 5.1 (III, a_2)** and CSB, **Fig. 5.1 (III, b_2)** SBG, **Fig. 5.1 (III, c_2)**, SDR, **Fig. 5.1 (III, d_2)**, FMS, **Fig. 5.1 (III, e_2),** MZL, **Fig. 5.1 (III, f_2)** with that of their respective untreated biomasses **Fig. 5.1 (II, a1-f1)** showed notable change in the texture of pre-treated ones indicating the effect of pre-treatment **Fig. 5.1 (III)**.

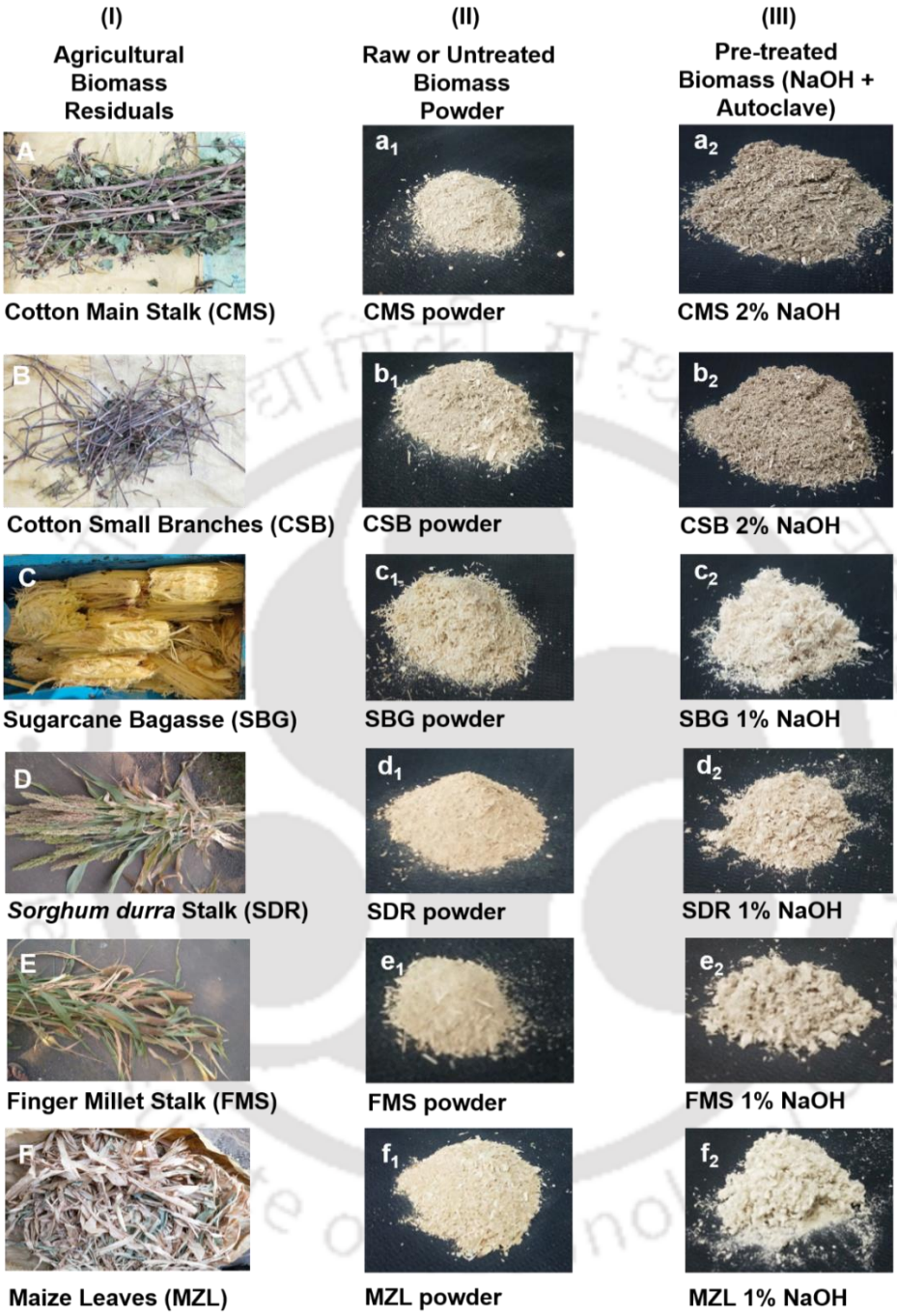

Fig. 5.1 Harvested agricultural residues biomasses (I-A-F,), processed raw biomasses (II-a_1-f_1) and alkali pre-treated biomasses (III-a_2-f_2). A-a_1-a_2: Cotton (CMS), B-b_1-b_2:Cotton (CSB), C-c_1-c_2:Sugarcane bagasse (SBG), D-d_1-d_2: Sorghum (SDR), E-e_1-e: Finger millet (FMS) and F-f_1-f2: Maize leaves (MZL). Alkali pre-treatment of these lignocellulosic biomasses was carried out using 1%, w/v NaOH for SBG, SDR, FMS, MZL and 2%, w/v NaOH for CMS and CSB, followed by autoclaving at 121℃ for 20 min.

5.3.2 Characterization of raw and pretreated biomasses

5.3.2.1 Total carbohydrate content of pretreted biomasses

The comparative analysis of total carbohydrate content of raw and NaOH pre-treated lignocellulosic biomasses (CMS, CSB, SBG, SDR, FMS and MZL) was collected from the literature and compiled in the **Table 5.1**. Total cellulose content of all the NaOH pre-treated biomasses is significantly higher as compared to their raw biomasses as can be seen from the **Table 5.1**. Interestingly, the total percentage of lignin in the alkali pre-treated biomasses is seen reduced after alkali pre-treatment which signifies the efficient lignin removal in post-pretreated biomasses. This comparative carbohydrate content analysis inferred that the alkali pre-treated lignocellulosic biomasses used in the present study had enriched cellulose content. As can be inferred from **Table 5.1,** ptSDR shows the highest cellulose content and least lignin after the alkali pre-treatment.

Table 5.1 Survey of carbohydrate-lignin content in the selected biomasses.

Biomass Type	Holocellulose (%)	Cellulose (%)	Hemicellulose (%)	Lignin	Reference
CMS/CSB (Raw)	56.2	43.7	12.5	28.6	(Shahzad *et al.*, 2019)
ptCMS/ptCSB	63.7	52.5	11.2	19.8	
SBG (Raw)	68.3	42.5	25.8	22.1	(Talha *et al.*, 2016)
ptSBG	70.1	47.4	22.8	12.3	(Nath *et al.*, 2021)
SDR (Raw)	55.0	31.8	23.2	5.7	(Jamaldheen *et al*, 2018)
ptSDR	76.4	54.7	21.7	2.3	(Nedumaran *et al.*, 2020)
FMS (Raw)	69.3	35.8	33.5	5.1	(Jamaldheen *et al.*, 2019)
ptFMS	67.2	51.8	15.4	3.1	(Jamaldheen *et al.*, 2019)
MZL (Raw)	58.93	36.17	22.76	19.98	(Cai *et al.*, 2016)
ptMZL	73.86	48.11	25.75	14.44	(Chen *et al.*, 2013)

ptCMS and ptCSB: 2% NaOH+Autoclaving;
ptSBG, ptSDR, ptFMS and ptMZL: 1% NaOH+Autoclaving.

5.3.2.2 Comparative pXRD based crystallinity index analysis of the biomasses

The crystallinity indices, *CrI* of un-treated and pre-treated biomasses were measured by pXRD to assess the exposure of cellulose content after pre-treatment with 1% or 2% (w/v) NaOH followed by autoclaving. The *CrI* of all the pre-treated biomasses was increased when compared with respective un-treated samples as seen in **Fig. 5.2A-F** and **Table 5.2**. The maximum increase in *CrI* after pre-treatment was observed for CSB (16.3) and SDR (15.8). The *CrI* of pre-treated CMS increased to 46.2% from 34.7% of untreated CMS (**Fig. 5.2A)**, whereas it increased to 41.7% for pre-treated CSB from 25.4% of untreated CSB (**Fig. 5.2B)**. This indicated the effectiveness of alkali pre-treatment of 2% (w/v) NaOH+autoclaving on CMS and CSB. Similarly, the *CrI* of pre-treated SBG increased to 45.2 from 40.6 of untreated SBG (**Fig. 5.2C)**. The *CrI* of pre-treated SDR significantly increased to 44.2% from 28.4% of untreated SDR (**Fig. 5.2D**). The *CrI* of pre-treated FMS showed significant increment to 38.7% from 27.7% of untreated FMS (**Fig. 5.2E**). MZL *CrI* also increased to 38.5% from 23.4 of untreated MZL (**Fig. 5.2F**). Similar increase in the *CrI* of same pre-treated SDR used in the present study by 1% (w/v) NaOH pre-treatment with autoclaving was reported earlier (Jamaldheen *et al.,* 2018) thus signifying its qualification for saccharification. As per an earlier report, the same raw SDR contained 55.0% holocellulose and 5.7% of lignin which increased to 76.4% holocelllose and lignin conent reduced to 2.3%, after the pre-treatment with 1% NaOH followed by autoclaving (Nedumaran *et al.,* 2020). Interestingly, Jamaldheen *et al.,* 2018 showed presence of D-xylose and L-arabinose in the pre-treated hydrolysate thereby indicating its reduced hemicellulose content. The increased *CrI* of pre-treated SDR showed that the lignin and hemicellulose content are significantly reduced after the alkali pre-

treatment thereby exposing the cellulosic fraction. The enhanced cellulose content was further validated by the notable increment in the *CrI* of *Rf*GH5_4 treated biomass samples which incorporated the cellulose contribution from exposure of newly formed amorphous cellulose regions due to enzymatic action (**Table 5.2**). It could be concluded that increased crystallinity of pre-treated CMS, CSB, SBG, SDR, FMS and MZL indicated higher exposure of cellulose content as compared to respective untreated biomasses. The specific increase in *CrI* (%) of all pre-treated biomasses with respect to their respective untreated samples is also shown in **Table 5.2** highlighting their efficient pre-treatment by NaOH.

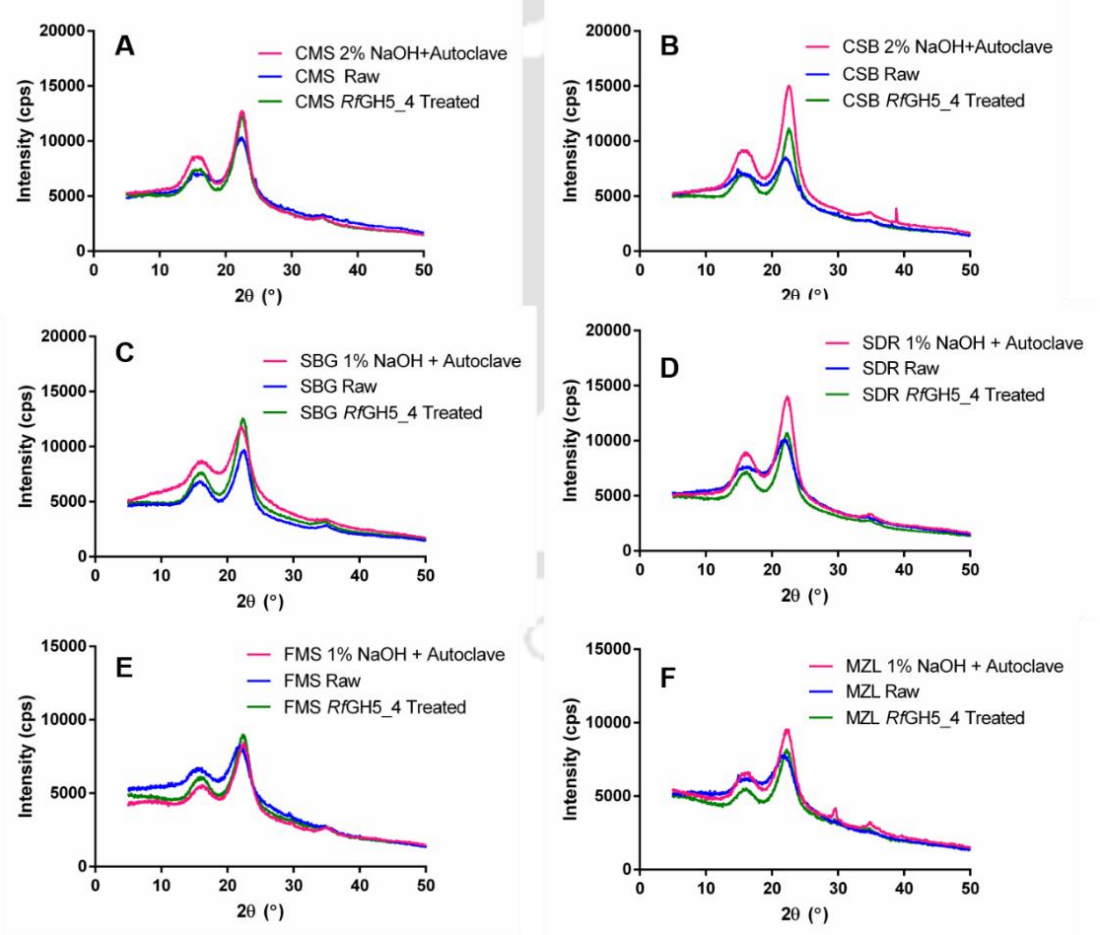

Fig. 5.2 pXRD analysis of untreated (Raw), pretreated and residuals from *Rf*GH5_4 saccharified biomasses, A) Cotton CMS, B) Cotton CSB, C) Sugarcane bagasse (SBG), D) Sorghum stalk (SDR), E) Finger millet (FMS) and F) Maize leaves (MZL).

Table 5.2 *CrI* of raw, pretreated and *Rf*GH5_4-saccharified biomass.

Biomass Type	*CrI* (%)	Increase in *CrI* as Compared to Raw [c] (%)	Increase in *CrI* of pretreated biomass after *Rf*GH5_4 saccharification [d] (%)
CMS (Raw)[a]	34.7		
ptCMS (Pretreated)[b]	46.2	11.5	
ptCMS *Rf*GH5_4-treated[e]	50.9		4.7
CSB (Raw)[a]	25.4		
ptCSB (Pretreated)[b]	41.7	16.3	
ptCSB *Rf*GH5_4-treated	43.8		2.0
SBG (Raw)[a]	40.6		
ptSBG (Pretreated)[b]	45.2	4.6	
ptSBG *Rf*GH5_4-treated	49.9		4.7
SDR (Raw)[a]	28.4		
ptSDR (Pretreated)[b]	44.2	15.8	
ptSDR *Rf*GH5_4-treated	45.1		0.9
FMS (Raw)[a]	27.7		
ptFMS (Pretreated)[b]	38.7	11	
ptFMS *Rf*GH5_4-treated	43.8		5.2
MZL (Raw)[a]	23.4		
ptMZL (Pretreated)[b]	38.5	15.1	
ptMZL *Rf*GH5_4-treated	42.1		3.6

The calculation for c and d was performed using formula; c = b-a and d = e-b. ptCMS, ptCSB, ptSBG, ptSDR, ptFMS and ptMZL *stands for pre-treated biomasses.*

5.3.2.3 Comparative FE-SEM analysis of untreated and pre-treated biomasses

Field Emission Scanning Electron Microscopic (FE-SEM) analyses of untreated and pre-treated lignocellulosic biomasses showed that there is an increase in surface roughness and thus accessibility of the pre-treated biomasses as compared with un-treated biomasses, thereby indicating disruption of lignocellulosic structure (**Fig. 5.3)**.

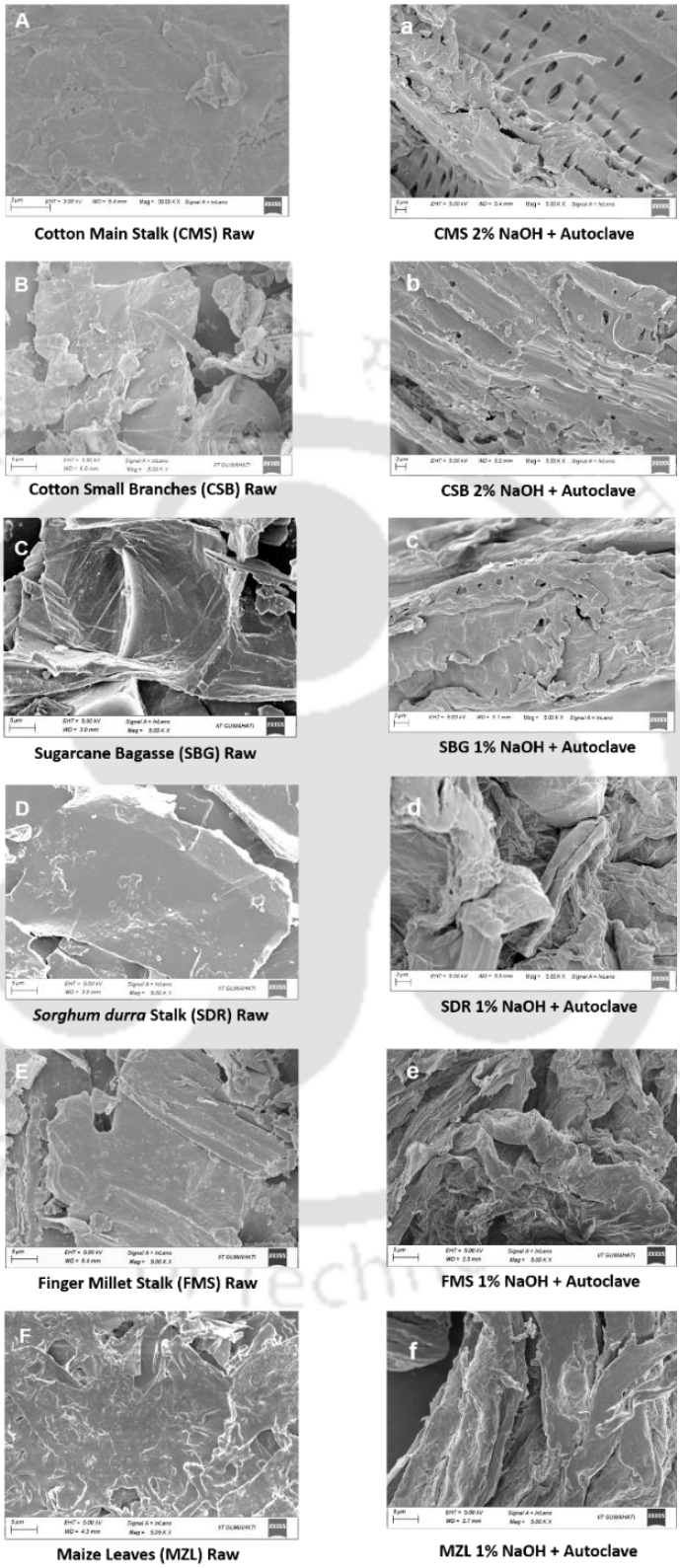

Fig. 5.3 FE-SEM analysis of untreated/raw lignocellulosic biomasses (A-F) and their respective NaOH pre-treated (a-f) samples. A-a: CMS, B-b: CSB, C-c: SBG, D-d: SDR, E-e: FMS, F-f: MZL. All images were taken at 5KX magnification.

The surface of pre-treated biomasses was broken suggesting loosening of the cellulosic polymers present in lignocellulose. It can be seen from the **Fig. 5.3** that surface topology of the pretreated samples is rough and uneven as compared with untreated biomass for all the six lignocellulosic biomasses namely, CMS (**Fig. 5.3A)**, CSB (**Fig. 5.3B)**, SBG (**Fig. 5.3C)**, SDR (**Fig. 5.3D)**, FMS (**Fig. 5.3E)** and MZL (**Fig. 5.3F)**. The *CrI* of ptSDR marginally increased after saccharification by *Rf*GH5_4 suggesting that the alkali pre-treatment sufficiently reduced its recalcitrance and prepared it for hydrolysis by *Rf*GH5_4. Therefore, alkali pre-treated ptSDR was inferred to be the most suitable biomass for *Rf*GH5_4

5.3.2.4 FT-IR spectroscopy of raw, pre-treated biomasses and commercial substrates

The comparative FT-IR analysis of untreated and pre-treated CMS, CSB, SBG, SDR, FMS and MZL biomasses showed that the peaks for lignin (2937, 1263, 2875 cm^{-1}) is significantly reduced in pre-treated biomass samples as seen for ptCMS (**Fig. 5.4a)**, ptCSB (**Fig. 5.4b)**, ptSBG (**Fig. 5.4c)**, ptSDR (**Fig. 5.4d**), ptFMS (**Fig. 5.4e)** and ptMZL (**Fig. 5.4f)**. The signature peaks recorded as Control for lignin (2937 cm^{-1}), cellulose (3335 cm^{-1}, Avicel/cellulose powder), Beechwood xylan (1446 cm^{-1}) (**Fig. 5.4g)** were compared with the raw and pre-treated lignocellulosic biomass samples and were found in agreement as mentioned earlier (Sills and Gossett, 2012). The sharp peak of crystalline cellulose at 3335 cm^{-1} was also broadened and distorted in the pre-treated samples as compared with untreated biomass samples. The deformation in the peak at 1735 cm^{-1} in the pretreated biomasses clearly indicated the dissolution of majority of hemicellulose after alkali pre-treatment thus increasing the cellulose content as also reported earlier (Gonultas and Candan, 2018). It confirmed the effective alkali pretreatment of the biomass samples indicating removal of lignin and hemicellulose

content which was also validated by increase in the *CrI* of pre-treated biomasses indicating enhanced cellulose content after alkali pre-treatement, as discussed in the earlier **Section 5.3.2.2**. Interestingly, the peak for amorphous cellulose (893 cm^{-1}) became prominently visible in all the pre-treated biomasses as compared to untreated samples. This peak of amorphous cellulose of all biomasses (pretreated and *Rf*GH5_4 treated/post-saccharified) was in agreement with that of pure amorphous cellulose, PASC (893 cm^{-1}) (**Fig. 5.4g**). These results further confirmed that the alkali pre-treatment of lignocellulosic biomasses enhanced the total content of amorphous cellulose. Thus, the alkali pre-treated lignocellulosic biomasses were found suitable for enzymatic saccharification in the downstream processing. Similar type of FT-IR based characterization of pretreated lignocellulose was also reported earlier for hardwood as well as softwood to confirm the reduction of lignin and hemicellulose content (Sills and Gossett, 2012).

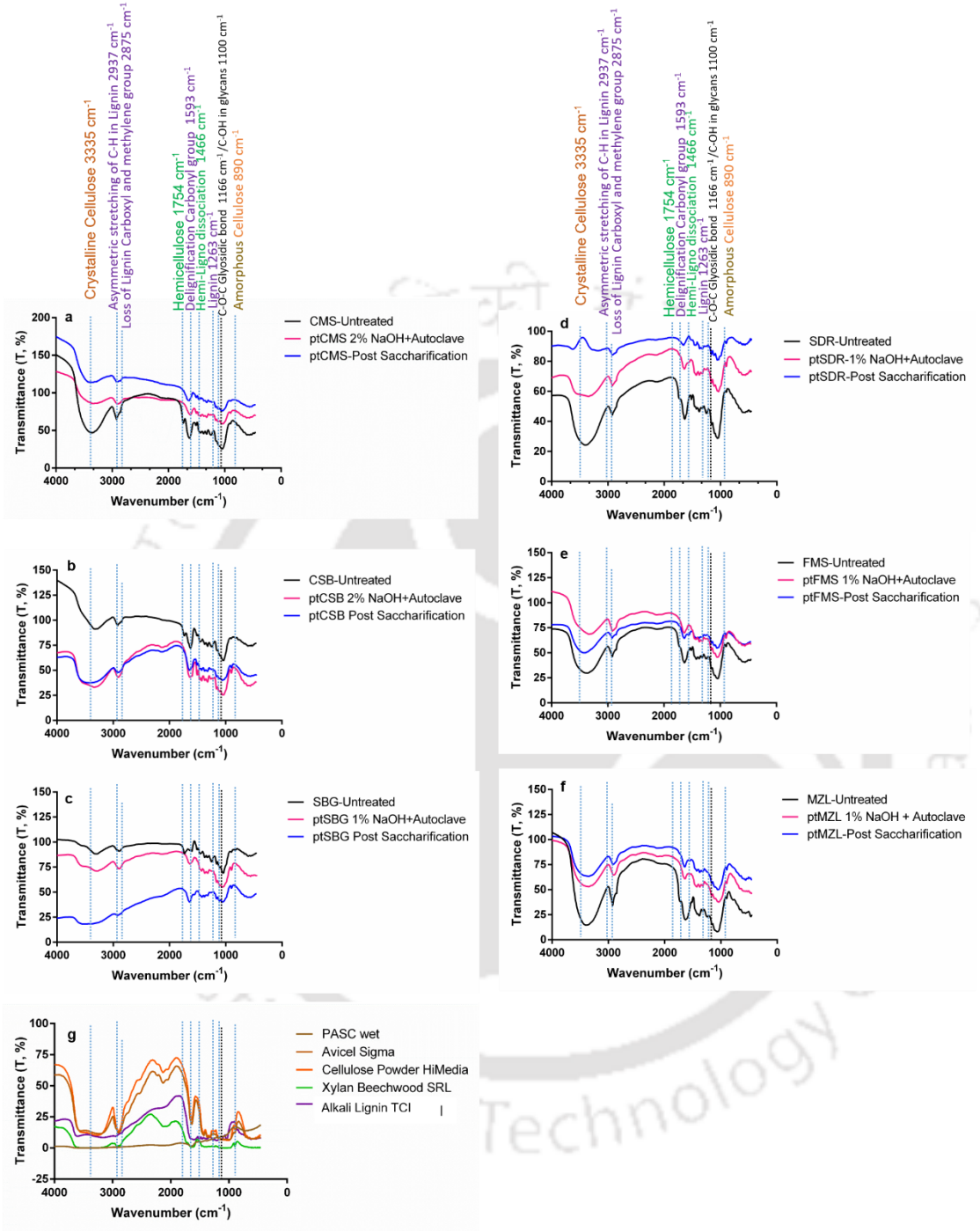

Fig. 5.4 FT-IR spectral analysis of untreated lignocellulosic biomasses and their respective NaOH pre-treated samples. a: CMS, b: CSB, c: SBG, d: SDR, e: FMS, f: MZL. FT-IR spectra of Avicel and PASC, Beechwood xylan, alkali lignin (g) was recorded as positive control. The transmittance, T (%) shown is qualitative for the detection of required functional groups of lignocellulose and does not intend quantitative comparative analysis with respect to a specific sequence for curves.

5.3.3 Production of RfGH5_4, CtCBH5A and CtGH1 enzymes

Ten mililiter of cell extract after sonication was obtained in 50 mM sodium-phosphate buffer, pH 6.5 for each of the protein *viz.* *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 from 400 mL culture of their independently expressed *E. coli* BL21 (DE3) cells. The presence of desired protein in the respective cell extract was confirmed by SDS-PAGE using 12% (w/v) acrylamide gel for *Rf*GH5_4 (**Fig. 5.5A**), *Ct*CBH5A (**Fig. 5.5B**) and *Ct*GH1 (**Fig. 5.5C**). The prominent bands of approximately 41 kDa for *Rf*GH5_4, 62 kDa for *Ct*CBH5A and 52 kDa for *Ct*CH1 were observed through the SDS-PAGE analysis of cell extracts (**Fig. 5.5**). The total protein concentration of cell extract of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 after filtration was 1.2, 5.7 mg/mL and 3.0 mg/mL, respectively.

5.3.4 Enzyme assays of RfGH5_4, CtCBH5A and CtGH1 enzymes

The enzyme activity of cell extract of endoglucanase, *Rf*GH5_4 against 1% (w/v) CMC-Na at 55℃ was 4.05 U/mL in pH 5.5 of 50 mM citrate phosphate buffer and 3.12 U/mL at pH 6.5 of 50 mM sodium phosphate buffer. The cell extract of Cellobiohydrolase, *Ct*CBH5A yielded 1.48 U/mL of enzyme activity at 65℃ in the optimum pH 6.5 of 50 mM sodium phosphate buffer. The enzyme activity of cell extract of β-glucosidase, *Ct*GH1 was found 8.92 U/mL at 65℃, at the optimum pH 6.5 of 50 mM sodium phosphate buffer. The compiled details on enzyme activity of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 and their respective reaction parameters are shown in **Table 5.3**.

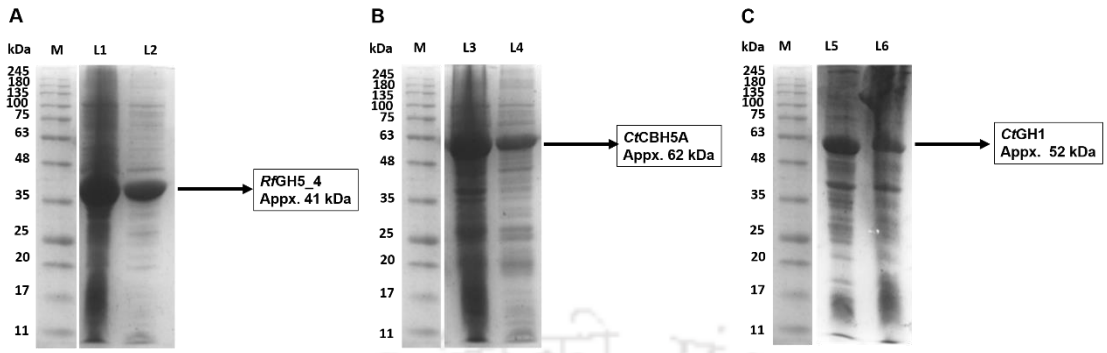

Fig. 5.5 SDS-PAGE analysis using 12% (w/v) gel for expressed proteins of (A) *Rf*GH5_4, (B) *Ct*CBH5A and (C) *Ct*GH1 and their respective cell extracts (CE). For all A, B, C panels, Lane M- protein marker (HiMedia Laboratories Pvt. Ltd., India), L1- expressed *E. coli* BL21 (DE3) cells with *Rf*GH5_4, L2- CE of *Rf*GH5_4, L3- expressed *E. coli* BL21 (DE3) cells with *Ct*CBH5A, L4: CE of *Ct*CBH5A, L5- expressed *E. coli* BL21 (DE3) cells with *Ct*GH1, L6- CE of *Ct*GH1.

Table 5.3 Enzyme activity of cell extracts of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1.

Cell Extract	Reaction Conditions	Enzyme Activity (U/mL)
*Endoglucanase, Rf*GH5_4 50 mM citrate phosphate buffer pH 5.5, 50 mM sodium phosphate buffer, pH 6.5,	1% (w/v) CMC-Na 55°C, 2 min Method: Nelson-Somogyi Assay at A_{500}.	4.05 ±0.19 3.12 ±0.13
*Cellobiohydrolase, Ct*CBH5A 50 mM sodium phosphate buffer, pH 6.5,	1% (w/v) CMC-Na 65°C, 5 min Method: Nelson-Somogyi Assay at A_{500}.	1.48 ±0.13
*β-glucosidase, Ct*GH1 50 mM sodium phosphate buffer, pH 6.5,	1% (w/v) Cellobiose 65°C, 5 min Method: GOD-POD Assay at A_{505}	8.92 ±0.69

5.3.5 Saccharification of pre-treated biomass by purified RfGH5_4

The highest TRS ($mg/g_{of\ biomass}$) was achieved from the hydrolysis of SDR (72) followed by FMS (62), SBG (38), CSB (27), CMS (19) and MZL (8.5) (**Fig. 5.6a**) by *Rf*GH5_4 (87 PASC $U/g_{Biomass}$ or 525 U/mg enzyme activity on CMC-Na) treatmet. The TLC analysis hydrolysates of six different types of alkali pre-treated biomasses *viz.*

CMS, CSB, SBG, SDR, FMS, MZL showed that the purified *Rf*GH5_4 resulted in various degree of cellooligosaccharides after 48 h of saccharification (**Fig. 5.6b**). The degree of polymerization (DP) of cellooligosaccharides produced by *Rf*GH5_4 was between DP2-DP11 as analysed by TLC (**Fig. 5.6b**) and MALDI-TOF MS for ptCMS (**Fig. 5.7a**), ptCSB (**Fig. 5.7b**), ptSBG (**Fig. 5.7c**), ptSDR (**Fig. 5.7d**), ptFMS (**Fig. 5.7e**) and ptMZL (**Fig. 5.7f**). The analysis of DP of respective cellooligosaccharides detected by MALDI-TOF MS corroborated with earlier reports (Silva et al. 2020). This demonstrated that *Rf*GH5_4 is capable of saccharifying the lignocellulosic biomass and can be also used along with cellobiohydrolase and β-glucosidase for its hydrolysis to D-glucose.

The TRS yields, 72 and 38 mg/g generated by purified *Rf*GH5_4 from alkali pre-treated SDR and SBG respectively were significantly higher than the earlier reported TRS yields SDR and SBG (34.2 mg/g for SDR, (Jamaldheen et al. 2018) and 5.9 mg/g for SBG, (Nath *et al.*, 2021). Jamaldheen *et al.* (2018) used purified endoglucanase (*Ct*Cel8A, 80 U/mg) and purified β-glucosidase (*Ct*Bgl1A, 33 U/mg) of *C. thermocellum* for SDR saccharification. Nath *et al.*, (2019) employed cellulolytic chimera (*Ct*GH1-L1-*Ct*GH5-F194A, 240 U/g) comprising β-glucosidase (*Ct*GH1) and endoglucanase (*Ct*GH5F194A) along with purified cellobiohydrolase, *Ct*CBH5A (360 U/g) from *C. thermocellum* for saccharification of SBG. However, the TRS yield from *Rf*GH5_4 saccharified CMS (19 mg/g) and CSB (27 mg/g) in 48 h was significantly lower than the 490 mg/g value reported earlier (Baig and Dharmadhikari, 2014) which could be because they used 5%, w/v alkali pre-treated cotton stalk that was saccharified by 100 $U/g_{ptBiomass}$ of commercial cellulase enzyme mixture at 50°C for 72 h.

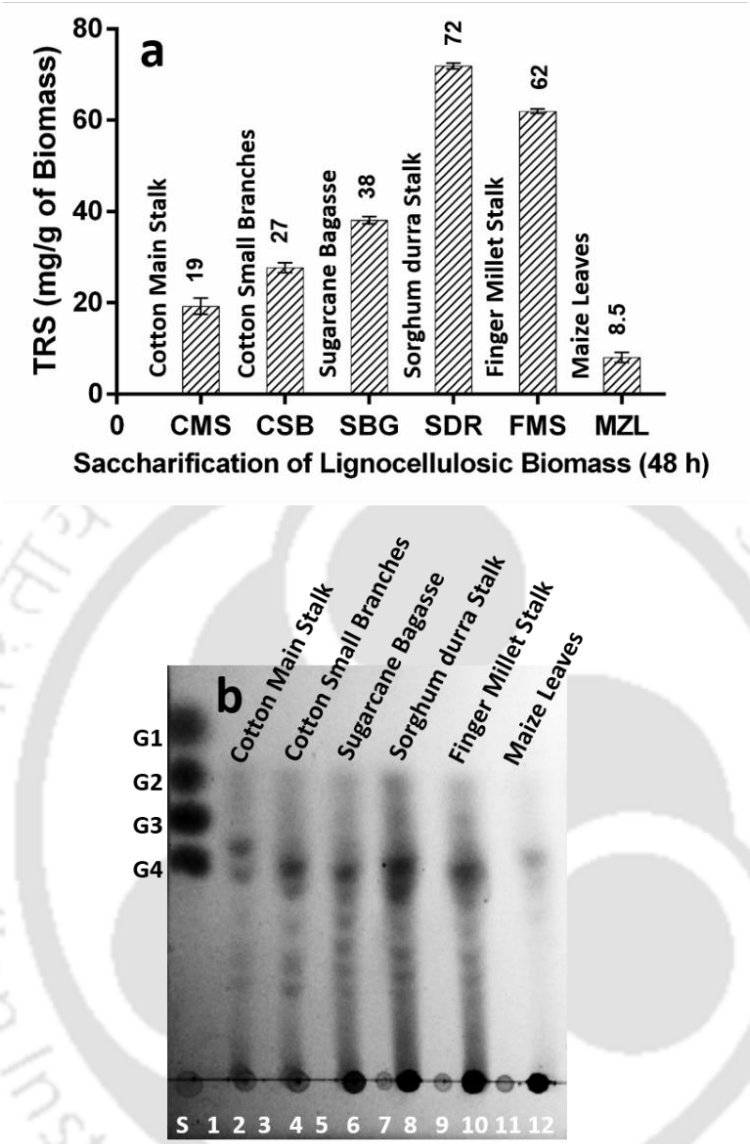

Fig. 5.6 (a) TRS yield in mg/g of pre-treated biomass, released by purified *Rf*GH5_4 from the various lignocellulosic biomasses was calculated. The saccharification of biomasses was carried out at 30℃, pH 5.5 of 20 mM citrate-phosphate buffer and 180 rpm for 48 h. The data represented here are the mean of three experimental replicates with error bars showing ±SD (b) The TLC analysis of hydrolysates of pre-treated biomasses generated by *Rf*GH5_4 after 48 h saccharification. Lanes S- Standard (G1, Glucose, G2, Cellobiose, G3, Cellotriose, G4, Cellotetraose), 1- Blank of CMS, 2- CMS, 3- Blank of CSB, 4- CSB, 5- Blank of SBG, 6- SBG, 7- Blank of SDR, 8- SDR, 9- Blank of FMS, 10- FMS, 11- Blank of MZL, 12-MZL.

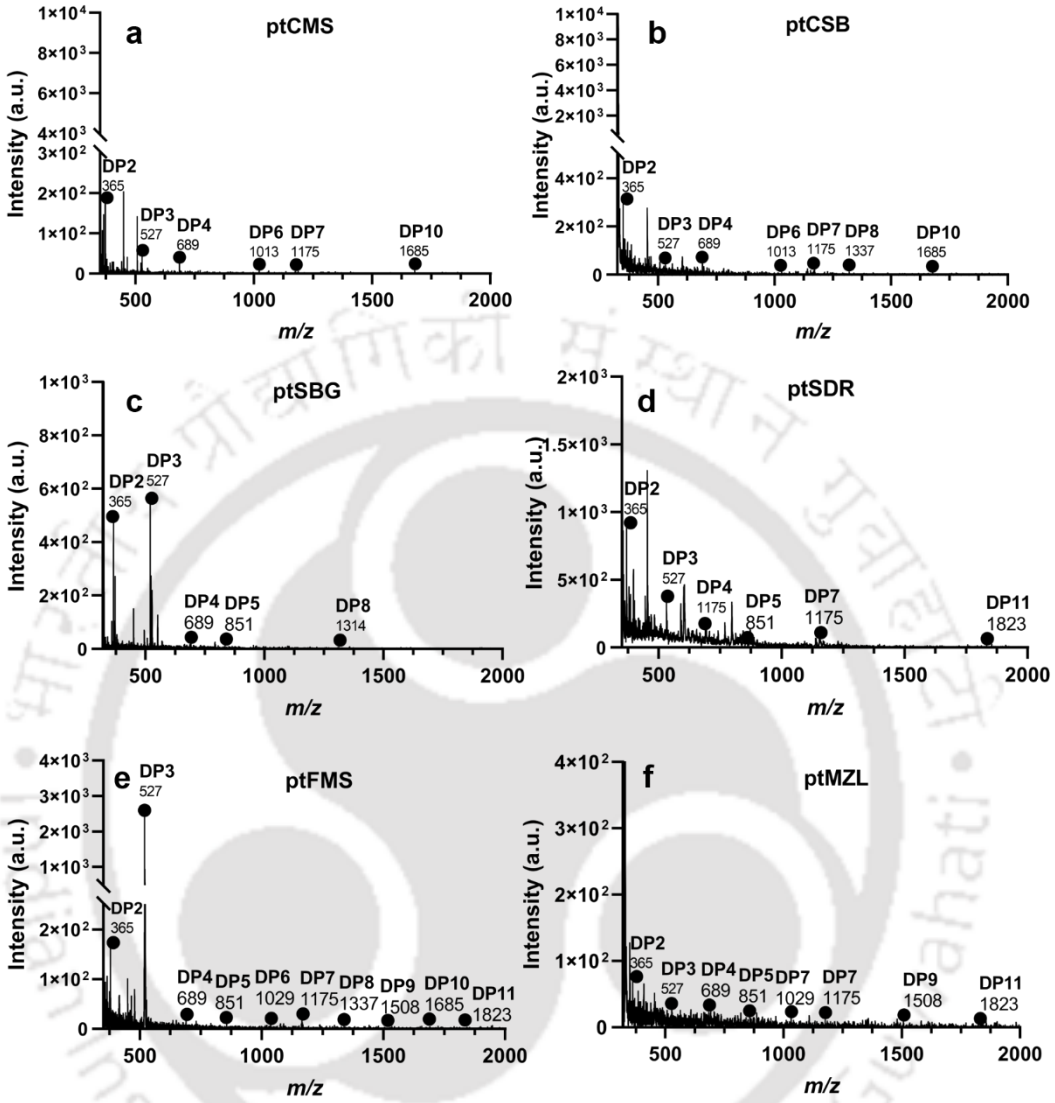

Fig. 5.7 MALDI-TOF MS analysis of hydrolysates of *Rf*GH5_4 from lignocellulosic biomasses (a) ptCMS, (b) ptCSB, (c) ptSBG, (d) ptSDR, (e) ptFMS, (f) ptMZL. The *m/z* ratio of mother peak $[M]^+$ cellooligosaccharides of respective DP are labelled with addition of atomic mass of $[Na]^+$ or $[K]^+$ ion.

The TRS obtained in the pre-treated ptCMS or ptCSB was less because of lower biomass loading, 2% (w/v) was used and for shorter period (48h) using an in-house produced single enzyme, endoglucanase (*Rf*GH5_4) at lower temperature, 30°C. Moreover, all the TRS yields from the aforementioned other reports were from the synergistic action of the three cellulases, *viz*. endoglucanase, cellobiohydrolase and β-glucosidase, whereas the TRS from saccharification of SDR, SBG, CMS, and CSB

reported in present study is only by a single enzyme, endoglucanase, *Rf*GH5_4. Sorghum stalk and finger millet stalk were found to be the best lignocellulosic substrates for *Rf*GH5_4. Annual residual biomass of sorghum is estimated around 11 metric tons in India (Jamaldheen *et al.,* 2018). Finger millet is a drought tolerant rain-fed crop in India (Jamaldheen *et al.,* 2019). It is noteworthy that *Rf*GH5_4 hydrolyzed the cotton biomasses, CSB and CMS quite effectively which are abundantly available as the agricultural waste in western indian regions of Maharashtra and Gujarat as well as in Telangana and Karnataka. Cotton is an important cash crop worldwide having roughly 32-million-hectare area under cotton cultivation of which India accounts for about 10 million-hectares (Baig and Dharmadhikari, 2014). Farmers usually burn the biomass like cotton stalk after cotton plucking is complete. Cultivation of cotton generates abundant residual cotton biomass which could be used as a rich source of lignocellulose for bioethanol production. Thus, residual cotton stalk could be a good biomass resource for lignocellulosic bioethanol production. The TRS yield from all biomasses can be further increased after optimising the pre-treatment of individual biomass and further optimising the saccharification conditions and parameters. Efficient hydrolysis of lignocellulosic biomasses could be achieved by synergistic action of endoglucanase-*Rf*GH5_4 with cellobiohydrolase and β-glucosidase for release of the end product, D-glucose for bioethanol production.

*Rf*GH5_4 hydrolysed the aforementioned biomasses into cellobiose (DP2) alongwith cellooligosaccharides of higher DP (DP3-DP11) as observed in TLC (**Fig. 5.6b**) and MALDI (**Fig. 5.7**) analysis after 48 h of saccharification. Therefore, *Rf*GH5_4 endoglucanase present in sufficient concentration may contribute to generate cellobiose that can be utilised by cellobiohydrolase during saccharification to give D-

glucose. *Rf*GH5_4 could also be immobilized using metal ion assisted recyclable pH-responsive polymer like Eudragit S100 to hydrolyse lignocellulose in bioethanol industry (Lv *et al.*, 2021). Moreover, *Rf*GH5_4 is capable of efficiently hydrolysing hemicellulosic polysaccharides as described in **Chapter 3, Section 3.3.2** which is an additional benefit to hydrolyse hemicellulases to loosen the cellulose present in lignocellulosic biomasses as per the requirement. Overall, this proof-of-concept study of lignocellulose saccharification by *Rf*GH5_4 confirmed that it has the ability to serve the lignocellulose conversion industry as a multifunctional yet a highly efficient endoglucanase.

5.3.6 Analysis of synergy among RfGH5_4, CtCBH5A and CtGH1 on WFP1

The saccharification of WFP1 was performed using cell extract of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 for which the crude enzyme activity is listed in **Table 5.3**. The hydrolysis of 2% (w/v) WFP1 by cocktail of all the three enzymes *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 (50:50:100 U/g) released 26.6 mg/g_{ptWFP1} of TRS (**Fig. 5.8 a**) and 7.5 mg/g_{ptWFP1} of D-glucose (**Fig. 5.8b**). This TRS and D-glucose was notably greater than that of independent hydrolysis by individual (50 U/g) *Rf*GH5_4 (1.5 and 0.4 mg/g, respectively, **Fig. 5.8 a**) and *Ct*CBH5A (8.4 and 0.7 mg/g, respectively, **Fig. 5.8 b**), thus highlighting the synergy among the three enzymes *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1. No TRS and D-glucose was released in the saccharification reaction of only β-glucosidase, *Ct*GH1 (100 U/g_{WFP1}) which is anticipated as it does not act on cellulose polymer except cellobiose and smaller cellooligosaccharides (~DP3, DP4). The optimization of ratio of individual enzymes *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 in a cocktail may further enhance the D-glucose yield from WFP1. The release of D-glucose by *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 and *Rf*GH5_4+*Ct*CBH5A was further confirmed by

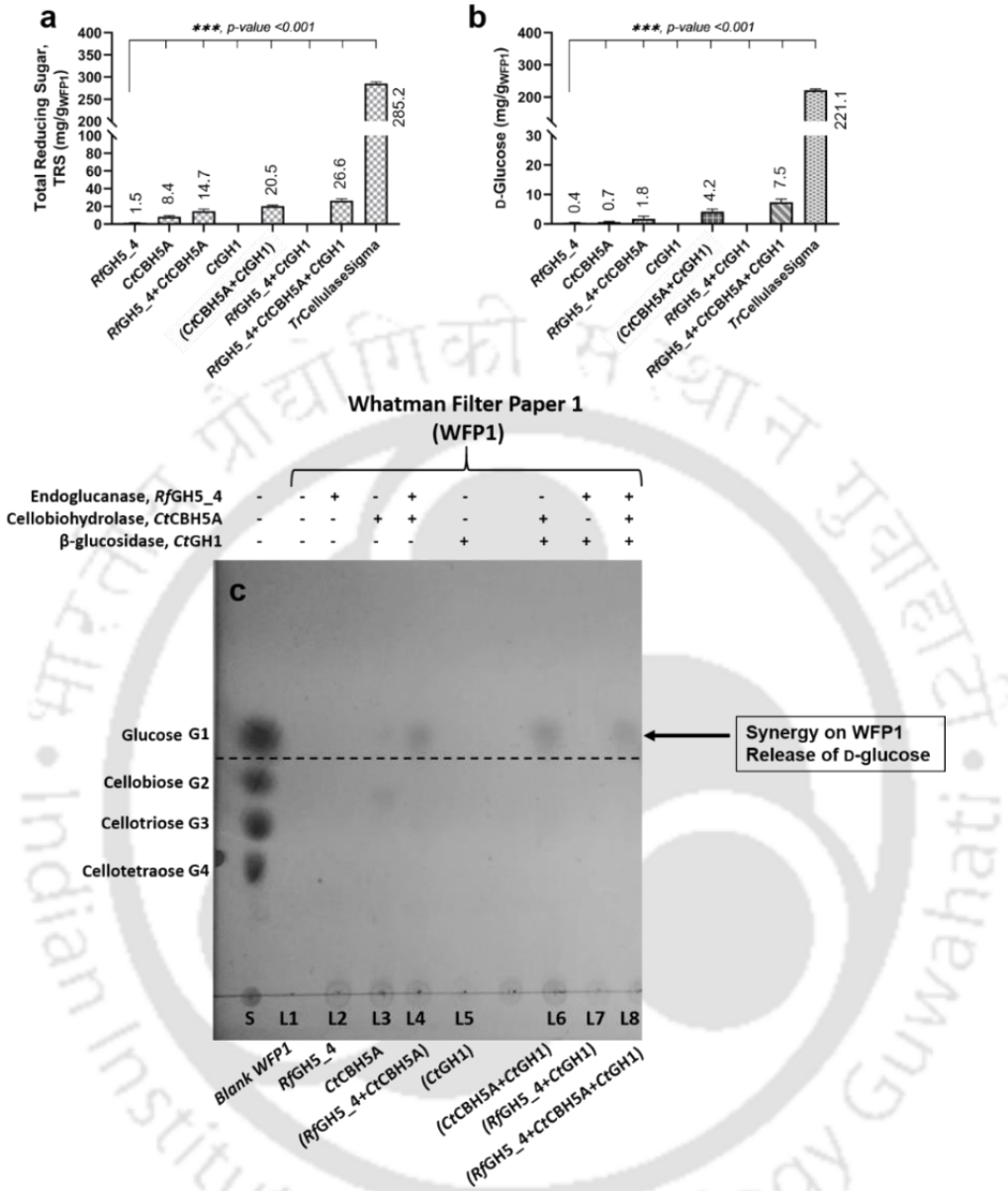

Fig. 5.8 Analysis of synergistic action of cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 and the different combinations of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 on saccharification of WFP1. (a) TRS and (b) D-glucose from WFP1 was observed. The data represent Mean±SD for each experiment with statistical evaluation by One-way ANOVA. The *p*-value <0.001 was considered significant. (c) TLC analysis of WFP1 treated by *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 in various combinations. Lane S: standard glucose (G1), cellobiose (G2), cellotriose (G3), cellotetraose (G4), L1-WFP1 Blank, L2- *Rf*GH5_4 (50 U/g), L3-*Ct*CBH5A (50 U/g), L4-*Rf*GH5_4+*Ct*CBH5A (50:50 U/g), L5-*Ct*GH1 (100 U/g), L6-*Ct*CBH5A+*Ct*GH1 (50:100 U/g), L7- *Rf*GH5_4+*Ct*GH1 (50:100 U/g), L8-*Rf*GH5_4+*Ct*CBH5A+*Ct*GH1. Total 200 U/g_{WFP1} of *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 (50:50:100 U/g) was incubated with 2% (w/v) WFP1 in 1 mL reaction volume at 40℃, pH 6.5 of 50 mM sodium phosphate buffer, 180 rpm for 48 h.

TLC analysis (**Fig. 5.8c**). The TLC analysis showed that the cocktails of *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 (50:50:100 U/g), *Rf*GH5_4+*Ct*CBH5A (50:50 U/g) and *Ct*CBH5A+*Ct*GH1 (50:100 U/g) are able to produce D-glucose, whereas the cocktail *Rf*GH5_4+*Ct*GH1 (50:100 U/g) could not, indicating essential requirement of *Ct*CBH5A enzyme for efficient synergistic hydrolysis of WFP1. Thus, the cellulolytic cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 was found suitable for the saccharification of WFP1 to give D-glucose.

5.3.7 Saccharification of ptSDR by RfGH5_4, CtCBH5A and CtGH1

The saccharification of 2% (w/v) lignocellulosic biomass ptSDR was performed using cell extract of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 with their crude enzyme activity as listed in **Table 5.3**. In the preliminary saccharification experiments, performed with various delignified biomasses, the maximum TRS released by purified *Rf*GH5_4 was from ptSDR (**Fig. 5.6a**). Therefore, ptSDR was selected for saccharification by *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 and to check synergy among them. A total of 200 $U/g_{Biomass}$ (50:50:100 U/g) of cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 released 62.0 mg/g_{ptSDR} of TRS (**Fig. 5.9a**) and 28.8 mg/g_{ptSDR} of D-glucose (**Fig. 5.9b**) from pre-treated SDR at 40°C and pH 6.5 of 50 mM sodium phosphate buffer after 48h of saccharification. A statistically optimized crude recombinant cellulase cocktail of endoglucanase, *EGII*, exoglucanase, *CBHII* and β-glucosidase, *BGLI* of *Trichoderma reesei* resulted in release of 41.4 mg/g of D-glucose from acid (3%, v/v H_2SO4) pre-treated bagasse after 48 h of saccharification (Soleimani and Ranaei-Siadat, 2017). The cocktail, *Ct*CBH5A+*Ct*GH1 (50:100 U/g) gave 36.5 mg/g_{ptSDR} of TRS and 19.6 mg/g_{ptSDR} of D-glucose indicating the synergy between the two, although lesser than the cocktail of three, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1. The cocktail of *Rf*GH5_4+*Ct*CBH5A

could produce, 27.7 mg/g_{ptSDR} TRS (**Fig. 5.9b**) without D-glucose release, clearly indicating the need of *Ct*GH1. Similarly, the cocktail of *Rf*GH5_4+*Ct*GH1 (150 U/g, 50:100) was able to produce 14.3 mg/g_{ptSDR} of TRS but no D-glucose (**Fig. 5.9b**) was detected indicating the essential requirement of *Ct*CBH5A. The saccharification reaction of *Rf*GH5_4 and *Ct*CBH5A separately resulted 15.5 mg/g_{ptSDR} and 25.6 mg/g_{ptSDR} of TRS, respectively. However, negligible amount of D-glucose was produced from ptSDR (**Fig. 5.9b**) when it was incubated with individual *Rf*GH5_4 (0.3 mg/g_{ptSDR}) or *Ct*CBH5A (3.3 mg/g_{ptSDR}) enzyme.

The released D-glucose from ptSDR by cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 was detected by TLC analysis (**Fig. 5.8c**), demonstrating the synergy among three enzymes in the saccharification of ptSDR. The positive control having 100 U/g of *Ct*GH1 incubated with cellobiose with similar reaction conditions also showed its conversion into D-glucose (**Fig. 5.9c**, Lane L9 and L10). This confirmed the role of *Ct*GH1 in saccharification of ptSDR by cocktails, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 and *Ct*CBH5A+*Ct*GH1, releasing the glucose from cellobiose. Thus, the release of D-glucose by the newly formulated cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 demonstrated the synergistic conversion of lignocellulosic biomass, ptSDR, making it applicable for bioethanol production.

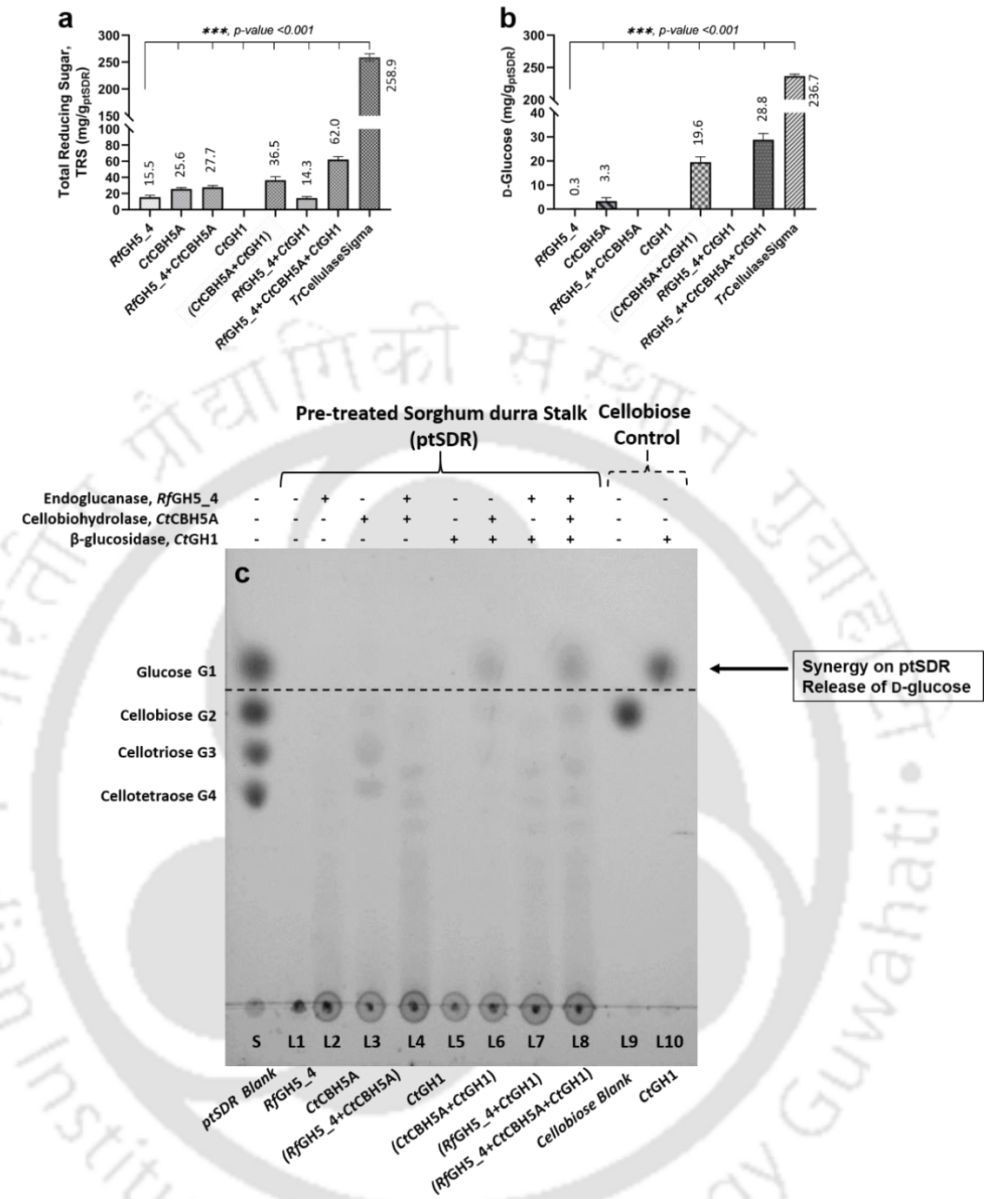

Fig. 5.9 Analysis of synergistic saccharification of ptSDR by cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 and other combinations of *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1. (a) TRS and (b) D-glucose released in from ptSDR. The data represent Mean±SD for each experiment with statistical evaluation by One-way ANOVA. (c) TLC analysis of synergy among *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1 in various combinations. Lane S: standard glucose (G1), cellobiose (G2), cellotriose (G3), cellotetraose (G4), L1-ptSDR Blank, L2- *Rf*GH5_4 (50 U/g), L3-*Ct*CBH5A (50 U/g), L4-*Rf*GH5_4+*Ct*CBH5A (50:50 U/g), L5-*Ct*GH1 (100 U/g), L6-*Ct*CBH5A+*Ct*GH1 (50:100 U/g), L7-*Rf*GH5_4+*Ct*GH1 (50:100 U/g), L8- *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1. Total 200 U/g_{ptSDR} of *Rf*GH5_4+*Ct*CBH5A+ *Ct*GH1 (50:50:100) were incubated with 2% (w/v) ptSDR biomass in 1 mL reaction at 40℃, pH 6.5 of 50 mM sodium phosphate buffer, 180 rpm for 48h. L9- cellobiose Blank, L10- Positive Control *Ct*GH1 (100 U/g) with cellobiose.

5.3.8 Saccharification of ptSDR and WFP1 by commercial cellulase

The saccharification of 2% (w/v) ptSDR or WFP1 by commercial cellulase, *Tr*CellulaseSigma (50 $U/g_{Biomass}$, a.k.a. Celluclast®1.5L when sold by Novozymes Corp., Denmark) released 258.9 mg/g_{ptSDR} (**Fig. 5.9a**) and 285.2 mg/g_{ptWFP1} (**Fig. 5.8a)** of TRS, respectively, after 48h of saccharification. Similarly, 236.7 mg/g_{ptSDR} and 221.1 mg/g_{WFP1} of D-glucose was generated by *Tr*CellulaseSigma from ptSDR (**Fig. 5.9b**) and WFP1 (**Fig. 5.8b)**, respectively. The amount of TRS and D-glucose released by *Tr*CellulaseSigma are significantly higher than those released by cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 from ptSDR (62.0 mg/g_{ptSDR} and 28.8 mg/g_{ptSDR}, respectively) and WFP1 (26.6 mg/g_{WFP1} and 7.5 mg/g_{WFP1}, respectively) described in the previous **Sections 5.3.6** and **5.3.7**. The enhanced efficiency of saccharification by *Tr*CellulaseSigma could be due to the presence several and different types of purified endoglucanases (*END-I* to *END-V*), exoglucanases (*CBH1* and *CBHII*) β-glucosidases (*BGLI* and *BGLII*) in *Tr*CellulaseSigma as reported earlier (Kumar and Murthy, 2017).

5.3.9 Ethanol tolerance of RfGH5_4 for prospected SSF application

Endoglucanase, *Rf*GH5_4 was found remarkably stable 30°C as discussed in **Chapter 3, Section 3.3.3** which is the appropriate temperature for single-step fermentation process, SSF. *Rf*GH5_4 need to withstand the alcohol concentration for SSF. Therefore, the ethanol tolerance of *Rf*GH5_4 cell extract was studied. The ethanol tolerance analysis of *Rf*GH5-4 at 30°C for 96 h showed that *Rf*GH5_4 is able to tolerate around 20% of ethanol, where it could retain more than 80% of its activity after 96h of incubation (**Fig. 5.10**). The highest concentration of ethanol achieved during standard SSF was reported to be around 7% (Larsen *et al.*, 2008). Interestingly, *Rf*GH5_4 retained 95% activity at 7% ethanol even after 96 h of incubation (**Fig. 5.10**). *Rf*GH5_4

lost activity only at higher conentrations such as 30% and 50% of ethanol within 10 h. An endoglucanase, *Ta*Cel12 from fungus *Trichoderma asperellum* ND-1 retained 80% of enzyme activity in 20% (v/v) of ethanol after 1 h of incubation at 25℃ (Zheng *et al.*, 2024), whereas *Rf*GH5_4 retained 80% of enzyme activity at 30℃ even after 96 h of incubation. Moreover, the cellulosic mixture of endoglucanases from *Clostridium cellulolyticum* (Anieto, 2015) could sustain upto 25% of ethanol at 37℃ incubation for 48 h. These results showed that *Rf*GH5_4 can sustain the higher alcohol concentration thereby enhancing its suitability for SSF.

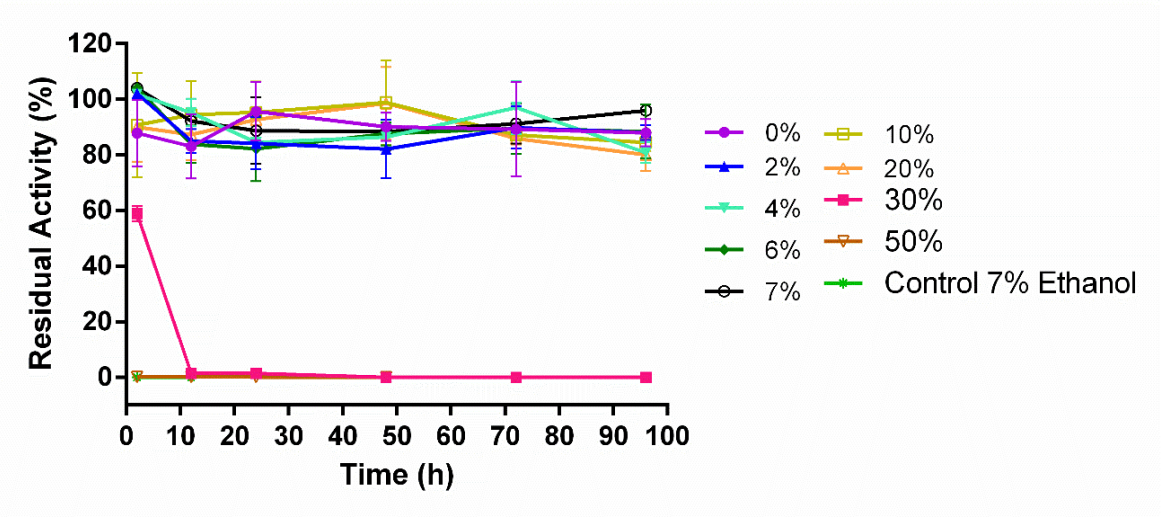

Fig. 5.10 Tolerance of *Rf*GH5_4 at various concentrations of (0, 2, 4, 6, 7,10, 20, 30 and 50%) ethanol. Cell extract of *Rf*GH5_4 (0.5 mL, 4 U/mL) was incubated in the ethanol containing 50 mM sodium phosphate buffer, pH 6.0 at 30℃ for 96 h. All the experiments were carried out in triplicate and data shown is the Mean ± SD.

5.4 Conclusion

Alkali NaOH based pre-treatment of lignocellulosic biomasses (Cotton Main Stalk, CMS, Cotton Small Branches, CSB), Sugarcane Bagasse, SBG), *Sorghum durra* stalk, SDR), Finger Millet Stalk, FMS) and Maize Leaves (MZL) was found to be effective for removal lignin and hemicellulose. This was inferred from the notable increment in the crystallinity index (*CrI*) of pre-treated biomasses as compared with the untreated ones with maximum increase in ptSDR. FE-SEM analyses of alkali pre-treated biomasses showed rough surface and texture against smooth surface of untreated ones further indicating the increased surface area and effectiveness of the alkali pre-treatment. *Rf*GH5_4 was used for the deconstruction of delignified lignocellulose, where it yielded maximum TRS (mg/g) from ptSDR (72) followed by ptFMS (62) and ptSBG (38). The TLC and MALDI-TOF MS analyses of hydrolysates from *Rf*GH5_4 saccharified pre-treated biomasses showed release of mixture of cellooligosaccharides (DP2-DP11) demonstrating its capability to hydrolyse lignocellulose material. The cocktail of three enzymes *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 at ratio 50:50:100 U/g showed synergy against pure cellulosic substrate, WFP1 and delignified lignocellulose, ptSDR where this cocktail yielded 26.6 mg/g_{WFP1}, 62.0 mg/g_{ptSDR} of TRS and 7.5 mg/g_{WFP1}, 28.8 mg/g_{ptSDR} of D-glucose, respectively. The release of D-glucose from WFP1 and ptSDR by treatment with *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 visualised by TLC, confirmed the synergy among the three enzymes *Rf*GH5_4, *Ct*CBH5A and *Ct*GH1. These results showed that the new cellulase cocktail, *Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 is capable of efficiently hydrolysing delignified biomass like ptSDR. *Rf*GH5_4 could tolerate upto 20% (v/v) of ethanol till 96 h at 30℃ further confirmed its suitability for SSF. Therefore, th endoglucanase, *Rf*GH5_4 and the newly formulated cellulase cocktail,

*Rf*GH5_4+*Ct*CBH5A+*Ct*GH1 has a potential for serving as cellulase toolbox for bioethanol production. Further statistical optimization of ratio of individual enzyme in cellulase cocktail and optimization of saccharification conditions could enhance the TRS and D-glucose yield. Moreover, the notable synergy observed between *Rf*GH5_4 and *Ct*CBH5A for cellulose conversion may be explored further and a bifunctional chimeric cellulase, *Rf*GH5_4-*Ct*CBH5A could be constructed to render the saccharification process time saving and cost-effective.

References

Abas N, Kalair A, Khan N (2015) Review of fossil fuels and future energy technologies. *Futures* 69:31–49

Aditiya HB, Mahlia TMI, Chong WT, Nur H, Sebayang AH (2016) Second generation bioethanol production: A critical review. *Renewwable and Sustainable Energy Reviews* 66:631–653

al-Hassan AY (2002) Alcohol and the Distillation of Wine in Arabic Sources from the Eighth Century Onuards. *History of Science and Technology in Islam* http//www Hist com/notes/notes7 html# _edn6

Anieto U (2015) Investigation of stress tolerance of endoglucanases of the cellulosomes of *Clostridium cellulolyticum* to ethanol. *African Journal of Biotechnology* 14:2095–2102

Azhar SHM, Abdulla R, Jambo SA, Marbawi H, Gansau JA, Faik AAM, Rodrigues KF (2017) Yeasts in sustainable bioethanol production: A review. *Biochemical and Biophysical Reports* 10:52–61

Baig MZ, DHArMADHikAri SmM (2014) Bioethanol production from enzymatically hydrolysed cotton stalk: one approach towards sustainable energy development. *Current World Environemnt* 9:940–946

Berndes G, Azar C, Kåberger T, Abrahamson D (2001) The feasibility of large-scale lignocellulose-based bioenergy production. *Biomass and Bioenergy* 20:371–383

Cai D, Li P, Luo Z, Qin P, Chen C, Wang Y, Wang Z, Tan T (2016) Effect of dilute alkaline pretreatment on the conversion of different parts of corn stalk to fermentable sugars and its application in acetone–butanol–ethanol fermentation. *Bioresource Technology* 211:117–124

Chen Y, Stevens MA, Zhu Y, Holmes J, Xu H (2013) Understanding of alkaline pretreatment parameters for corn stover enzymatic saccharification. *Biotechnology for Biofuels* 6:1–10

Eriksson T, Karlsson J, Tjerneld F (2002) A model explaining declining rate in hydrolysis of lignocellulose substrates with cellobiohydrolase I (Cel7A) and endoglucanase I (Cel7B) of *Trichoderma reesei*. *Applied Biochemistry and Biotechnology* 101:41–60

Faraday M (1833) On new compounds of carbon and hydrogen, and on certain other products obtained during the decomposition of oil by heat. In: Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London. *The Royal Society London*, pp 248–249

Giampietro M, Ulgiati S, Pimentel D (1997) Feasibility of large-scale biofuel production. *Bioscience* 47:587–600

Gonçalves FA, Ruiz HA, dos Santos ES, Teixeira JA, de Macedo GR (2016) Bioethanol production by *Saccharomyces cerevisiae*, *Pichia stipitis* and *Zymomonas mobilis* from delignified coconut fibre mature and lignin extraction according to biorefinery concept. *Renewable Energy* 94:353–365

Gonultas O, Candan Z (2018) Chemical characterization and FT-IR spectroscopy of thermally compressed eucalyptus wood panels. *Maderas Ciencia y Tecnologia* 20:431–442

Helm D (2016) The future of fossil fuels—is it the end? *Oxford Review of Economic Policy* 32:191–205

Henrissat B, Driguez H, Viet C, Schülein M (1985) Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Biotechnology* 3:722–726

Jamaldheen SB, Sharma K, Rani A, Moholkar VS, Goyal A (2018) Comparative analysis of pretreatment methods on sorghum (*Sorghum durra*) stalk agrowaste for holocellulose content. *Preparative Biochemistry and Biotechnology* 48:457–464

Jamaldheen SB, Thakur A, Moholkar VS, Goyal A (2019) Enzymatic hydrolysis of hemicellulose from pretreated Finger millet (*Eleusine coracana*) straw by recombinant endo-1, 4-β-xylanase and exo-1, 4-β-xylosidase. *International Journal of Biological Macromolecules* 135:1098–1106

Jambo SA, Abdulla R, Azhar SHM, Marbawi H, Gansau JA, Ravindra P (2016) A review on third generation bioethanol feedstock. *Renewable and Sustainable Energy Reviews* 65:756–769

Kovarik B (1998) Henry Ford, Charles F. Kettering and the fuel of the future. *Automotive History Review* 32:7–27

Kumar B, Bhardwaj N, Agrawal K, Verma P (2020) Bioethanol production: generation-based comparative status measurements. *Biofuel Production Technologies: Critical Analysis for Sustainability* 155–201

Kumar D, Murthy GS (2017) Development and validation of a stochastic molecular model of cellulose hydrolysis by action of multiple cellulase enzymes. *Bioresource and Bioprocessing* 4:1–17

Lange J (2007) Lignocellulose conversion: an introduction to chemistry, process and economics. *Biofuels, Bioproducts and Biorefining Innovation for a Sustainable Economy* 1:39–48

Larsen J, Østergaard Petersen M, Thirup L, Wen Li H, Krogh Iversen F (2008) The IBUS process–lignocellulosic bioethanol close to a commercial reality. *Chemical Engineering & Technology* 31:765–772

Lin Y, Tanaka S (2006) Ethanol fermentation from biomass resources: current state and prospects. *Applied Microbiology and Biotechnology* 69:627–642

Lv K, Yu Z, Pedroso MM, Wu B, Gao Z, He B, Schenk G (2021) Metal affinity immobilization of the processive endoglucanase EG5C-1 from *Bacillus subtilis* on a recyclable pH-responsive polymer. ACS Sustain Chem Eng 9:7948–7959

Madadi M, Tu Y, Abbas A (2017) Recent status on enzymatic saccharification of lignocellulosic biomass for bioethanol production. *Electronic Journal of Biology* 13:135–143

Matsumoto N, Yoshizumi H, Miyata S, Inoue S (1985) Development of the noncooking and low temperature cooking systems for alcoholic fermentation of grains. *Journal of Agricultural Chemical Society Japan*

Mohanty SK, Swain MR (2019) Bioethanol production from corn and wheat: food, fuel, and future. *In*: *Bioethanol production from food crops.* Elsevier, pp 45–59

Nath P, Maibam PD, Singh S, Rajulapati V, Goyal A (2021) Sequential pretreatment of sugarcane bagasse by alkali and organosolv for improved delignification and cellulose saccharification by chimera and cellobiohydrolase for bioethanol production. *3 Biotech* 11:1–16

Nedumaran M, Singh S, Jamaldheen SB, Nath P, Moholkar VS, Goyal A (2020) Assessment of combination of pretreatment of Sorghum durra stalk and production of chimeric enzyme (β-glucosidase and endo β-1, 4 glucanase, *Ct*GH1-L1-Ct GH5-F194A) and cellobiohydrolase (*Ct*CBH5A) for saccharification to produce bioethanol. *Preparative Biochemistry and Biotechnology* 50:883–896

Niphadkar S, Bagade P, Ahmed S (2018) Bioethanol production: insight into past, present and future perspectives. *Biofuels* 9:229–238

Olofsson K, Bertilsson M, Lidén G (2008) A short review on SSF–an interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnology for Biofuels* 1:1–14

Raabo BE, Terkildsen TC (1960) On the enzymatic determination of blood glucose. *Scandinavian Journal of Clinical and Laboratory Investigation* 12:402–407

Rasmussen SC (2014) The quest for Aqua Vitae: The history and chemistry of alcohol from antiquity to the Middle Ages. *Springer Science & Business*

Rosillo-Calle F, Cortez LAB (1998) Towards ProAlcool II—a review of the Brazilian bioethanol programme. *Biomass and Bioenergy* 14:115–124

Saini JK, Saini R, Tewari L (2015) Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech* 5:337–353

Segal L, Creely JJ, Martin Jr AE, Conrad CM (1959) An empirical method for estimating the degree of crystallinity of native cellulose using the X-ray diffractometer. *Textile Research Journal* 29:786–794

Shahzad K, Sohail M, Hamid A (2019) Green ethanol production from cotton stalk. *In*: *IOP Conference Series: Earth and Environmental Science. IOP Publishing*, p 12025

Sharma K, Thakur A, Kumar R, Goyal A (2019) Structure and biochemical characterization of glucose tolerant β-1, 4 glucosidase (*Ht*Bgl) of family 1 glycoside hydrolase from *Hungateiclostridium thermocellum*. *Carbohydrate Research* 483:107750

Sills DL, Gossett JM (2012) Using FTIR to predict saccharification from enzymatic hydrolysis of alkali-pretreated biomasses. *Biotechnology and Bioengineering* 109:353–362

Silva C de OG, Teixeira TS, Rodrigues KB, Souza AA, Monclaro AV, Mendes TD, de Aquino Ribeiro JA, de Siqueira FG, de Lima Fávaro LC, Abdelnur PV (2020) Combination of MALDI-TOF MS and UHPLC-ESI-MS for the characterization of lytic polysaccharide monooxygenase activity. *Analytical Methods* 12:149–161

Singh S, Dhillon A, Goyal A (2020a) Enhanced catalytic efficiency of *Bacillus amyloliquefaciens* SS35 endoglucanase by ultraviolet directed evolution and mutation analysis. *Renewable Energy* 151:1124–1133. https://doi.org/https://doi.org/10.1016/j.renene.2019.11.105

Singh S, Rajulapati V, Jamaldheen SB, Moholkar VS, Goyal A (2020b) Statistically designed cellulase mixture for saccharification of pretreated *Sorghum durra* stalk. *Industrial Crops and Products* 154:112678

Soleimani S, Ranaei-Siadat S-O (2017) Preparation and optimization of cellulase cocktail to improve the bioethanol process. *Biofuels* 8:291–296

Solomon BD, Barnes JR, Halvorsen KE (2007) Grain and cellulosic ethanol: History, economics, and energy policy. *Biomass and Bioenergy* 31:416–425

Somogyi M (1944) 164, 61 and 74 (1945); Nelson, N. *Journal of Biological Chemistry* 153:375

Somogyi M (1945) A new reagent for the determination of sugars. *Journal of Biological Chemistry* 160:61–68

Somogyi N, Nelson N (1944) A photometric adaptation of the Somogyi method for the determination of glucose. *Journal of Biological Chemistry* 153:375–380

Talha Z, Ding W, Mehryar E, Hassan M, Bi J (2016) Alkaline pretreatment of sugarcane bagasse and filter mud codigested to improve biomethane production. *BioMed Research International* 2016

Zhao X, Zhang L, Liu D (2012a) Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. *Biofuels, Bioproducts and Biorefining* 6:465–482

Zhao X, Zhang L, Liu D (2012b) Biomass recalcitrance. Part II: Fundamentals of different pre-treatments to increase the enzymatic digestibility of lignocellulose. *Biofuels, Bioproducts and Biorefining* 6:561–579

Zheng F, Basit A, Wang J, Zhuang H, Chen J, Zhang J (2024) Characterization of a novel acidophilic, ethanol tolerant and halophilic GH12 β-1, 4-endoglucanase from *Trichoderma asperellum* ND-1 and its synergistic hydrolysis of lignocellulosic biomass. *International Journal of Biological Macromolecules* 254:127650

Zverlov V V, Velikodvorskaya GA, Schwarz WH (2002) A newly described cellulosomal cellobiohydrolase, CelO, from *Clostridium thermocellum*: investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. *Microbiology* 148:247–255

List of Publications

From Thesis

1. **Parmeshwar Vitthal Gavande,** K. Kumar, J. Ahmed, A. Goyal, Highly efficient, processive and multifunctional recombinant endoglucanase *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 v3 for recycling lignocellulosic plant biomasses, Int. J. Biol. Macromol. 209 **(2022a)** 801–813 https://doi.org/10.1016/j.ijbiomac.2022.04.059 **4th April 2022. (JIF 8.0)**
2. **Parmeshwar Vitthal Gavande,** P. Nath, K. Kumar, N. Ahmed, C.M.G.A. Fontes, A. Goyal, Multifunctionality and mechanism of processivity of family GH5 endoglucanase, *Rf*GH5_4 from *Ruminococcus flavefaciens* on lignocellulosic polymers, Int. J. Biol. Macromol. 224 **(2022b)** 1395–1411. https://doi.org/10.1016/j.ijbiomac.2022.10.227 **26th Oct 2022. (JIF 8.0).**
3. **Parmeshwar Vitthal Gavande** and Arun Goyal, Development of a novel synergistic cellulase cocktail containing endoglucanase, *Rf*GH_4 from *Ruminococcus flavefaciens* FD-1 v3, cellobiohydrolase, *Ct*CBH5A and β-glucosidase, *Ct*GH1 from *Clostridium thermocellum* for deconstruction of delignified lignocellulosic biomass of *Sorghum durra* stalk. ***(Manuscript Under Preparation)***

Book:

1. Book Edited by **Parmeshwar Vitthal Gavande**, Virang Shukla, Rahul Agrawal and Anand Patel, Futuristic Trends in Renewable and Sustainable Energy, IIP Series. ***(In Press)***

Publications from other collaborations

1. **Parmeshwar Vitthal Gavande,** Shyam Ji, Carlos M.G.A. Fontes and Arun Goyal, Interesting Bifunctional exo-1,4-β-xylosidase and endoxylanase activities by glycosidase, *Bo*Xyl43A of family GH43 from bacterium *Bacteroides ovatus* against natural xylan ***(Under Revision)***
2. Ji Shyam, **Parmeshwar Vitthal Gavande**, Choudhury Bipasha and Arun Goyal. Computational design and structure dynamics analysis of bifunctional chimera of endoxylanase from *Clostridium thermocellum* and xylosidase from *Bacteroides ovatus*. 3 Biotech. 2023 Feb;13(2):59.
3. **Parmeshwar Vitthal Gavande** and Arun Goyal Molecular dynamics based structural insights of the first putative endoglucanase, *Ps*GH5A of glycoside hydrolase family 5 from *Pseudopedobacter saltans.* J Mol. Model 2023, Jan, 29(6):1-6.
4. Shubha Singh, Jebin Ahmed, **Parmeshwar Vitthal Gavande,** Carlos M.G.A Fontes and Arun Goyal, Structural and functional insights into the glycoside hydrolase family 30 xylanase of the rumen bacterium *Ruminococcus flavefaciens*, J. Mol. Struc. 1272 (2022) 134155, 1-18.
5. **Parmeshwar Vitthal Gavande**, Basak A., Sen S., Lepcha K., Murmu N., Rai, V., Saha, S., Mazumdar, D., Das, V. and Ghosh, S. Functional characterization of thermotolerant microbial consortium for lignocellulolytic enzymes with central role of Firmicutes in rice straw depolymerisation. Sci. Rep. (2021) 11(1): 1-13.

Book Chapter

1. Akshita Kanwar, **Parmeshwar Vitthal Gavande** and Arun Goyal, Chapter 3: Plant Based Biofuels *In* Emerging Sustainable Technologies for Biofuel Production, *Ed.* Deepanwita Deka, Springer. (2023) ***(In Press)***

2. **Parmeshwar Vitthal Gavande** and Arun Goyal, ***Chapter 1:*** **Carbohydrate-Active Enzymes: Overview,** *In Volume:* Glycoside Hydrolase: Biochemistry, Biophysics and Biotechnology, Ed. Arun Goyal and Kedar Sharma, Elsevier *Foundations and Frontiers of Biocatalysis* series (2023).
3. **Parmeshwar Vitthal Gavande** and Arun Goyal, ***Chapter 6:*** **Endo-β-1,3-glucanase,** *In Volume:* Glycoside Hydrolase: Biochemistry, Biophysics and Biotechnology, Ed. Arun Goyal and Kedar Sharma, Elsevier *Foundations and Frontiers of Biocatalysis* series (2023).
4. **Parmeshwar Vitthal Gavande** and Arun Goyal, ***Chapter 3:*** **Endo-β-1,4-glucanase,** *In Volume:* Glycoside Hydrolase: Biochemistry, Biophysics and Biotechnology, Ed. Arun Goyal and Kedar Sharma, Elsevier *Foundations and Frontiers of Biocatalysis* series (2023).
5. Kaustubh C. Khaire, S. Patel, **Parmeshwar Vitthal Gavande**, V.A. Moholkar, A. Goyal, **Extremophilic biofilms: Exploring the prospects**, In *Introduction to Biofilm Engineering*, American Chemical Society (2019). Pp.141-157.

Conferences

Oral Presentations

1. **Parmeshwar Vitthal Gavande*** and Arun Goyal, Elucidation of structural basis of multifunctionality of endoglucanase *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 v3 for efficient saccharification of lignocellulose, International Conference on Biotechnology for Better Tomorrow, **29-30th Oct, 2022,** Udayana University, **Bali, Indonesia**. ***(Best Offline Oral Presentation Award).***
2. **Parmeshwar Vitthal Gavande*,** Priyanka Nath, Krishan Kumar, Nazneen Ahmed, Carlos M.G.A. Fontes and Arun Goyal. An efficient cellulase enzyme enriches the toolbox of biomass conversion and bioethanol production. Research and Industrial Conclave 2022: An amalgamation of Academia, Industry & Start-up **(RIC), January, 20–23, 2022,** Indian Institute of Technology Guwahati, **Guwahati, Assam, India**. **(*2nd Prize for Best Presentation Award 2022*).**
3. **Parmeshwar Vitthal Gavande,** Priyanka Nath, Krishan Kumar, Nazneen Ahmed, Carlos M.G.A. Fontes and Arun Goyal Processive yet multifunctional recombinant endoglucanase, *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 efficiently deconstructs lignocellulosic biomasses of cotton, sugarcane bagasse, sorghum and finger millet. 18th BRSI-International Conference on Biotechnology for Resource Efficiency, Energy, Environment, Chemicals and Health **(BRSI-BRE3CH 2021) December 1 – 4, 2021,** CSIR-Indian Institute of Petroleum (IIP), **Dehradun, Uttarakhand**, **India**. (***Offline Oral Presentations*****)**
4. **Parmeshwar Vitthal Gavande,** Priyanka Nath, Krishan Kumar, Nazneen Ahmed, Carlos M. G. A. Fontes and Arun Goyal. A multifunctional endoglucanase (*Rf*GH5_4) cloned from *Ruminococcus flavefaciens* FD-1 v3 hydrolyses β-1,4 and mixed β-1,3- β-1,4-linked carbohydrate polymers. BRSI Virtual International Conference on Bioengineering Solutions for Healthcare, Food, Energy, and Environment (BSHFEE-2021), **9th-10th April, 2021**, Indian Institute of Technology Jodhpur, (IITJ), **Jodhpur, Rajasthan, India**. (***Online Flash oral talk and presentation*****)**

Poster Presentations

1. **Parmeshwar Vitthal Gavande** and Arun Goyal (2023). A promising newly constituted bacterial cellulase system for synergistic saccharification of agricultural residual lignocellulosic biomass, having endoglucanase-*Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 v3, cellobiohydrolase-*Ct*CBH5A and β-glucosidase-*Ct*GH1 from *Clostridium thermocellum*. Biotech Research Society, India's International Conference on New Horizons in Biotechnology (BRSI-NHBT2023), November 26-29, 2023, & CSIR-National Institute for Interdisciplinary Science and Technology (NIIST), **Trivandrum, Kerala, India.**

2. **Parmeshwar Vitthal Gavande,** Shyam Ji, Vânia Cardoso, Carlos M.G.A. Fontes and Arun Goyal (2023) The single-domain family GH43 of *Bacteroides ovatus* displays high exo-β-1,4-xylosidase activity against natural and purified xylans. Research and Industrial Conclave 2023, May 14-16, 2023, IIT Guwahati, **Guwahati, Assam, India.**
3. **Parmeshwar Vitthal Gavande*** and Arun Goyal, Molecular dynamics based structural insights of the first putative endoglucanase (PsGH5FL) of glycoside hydrolase family 5 from *Pseudopedobacter saltans* at Biotech Research Society India's International Conference on Biotechnology for Sustainable Bioresources and Bioeconomy (BRSI-BSBB-22), Dec 7-10 2022 Indian Institute of Technology Guwahati, **Guwahati, Assam**. ***(Best Flash Presentation & Poster Award)***
4. **Parmeshwar Vitthal Gavande*** and Arun Goyal, Generation of industrially and medically important cellulosic as well as hemicellulosic oligosaccharides by a multifunctional endoglucanase (*Rf*GH5_4) from *Ruminococcus flavefaciens*, ACCTI-CARRBO-36, **5-7th Dec, 2022,** Indian Institute of Technology Bombay, **Mumbai, Maharashtra.**
5. **Parmeshwar Vitthal Gavande*,** Priyanka Nath, Krishan Kumar, Nazneen Ahmed, Carlos M.G.A. Fontes and Arun Goyal Highly efficient, processive and multifunctional cellulase from *Ruminococcus flavefaciens* FD-1 for recycling plant biomasses to lignocellulosic bioethanol. 35th International conference on advances in chemistry and biology of carbohydrates **(ACCTI-CARBO XXXV), December 4-5, 2021,** Forest Research Institute & Association of Carbohydrate Chemist and Technologist India, **Dehradun, Uttarakhand**.
6. **Parmeshwar Vitthal Gavande,** Priyanka Nath, Nazneen Ahmed, Carlos M. G. A. Fontes and Arun Goyal. Cloning, expression, purification, biochemical, functional and structure characterization of an endoglucanase of family GH5_4 (*Rf*GH5_4) from *Ruminococcus flavefaciens* FD-1 v3. 88th Annual Meeting of Society of Biological Chemists (India, SBCI) and Conference on Advances at the Interface of Biology and Chemistry, **1st -3rd November, 2019**, Bhabha Atomic Research Centre (BARC), **Mumbai, Maharashtra**.

Workshop attended

1. Introduction to Small Angle X-ray Scattering, SAXS by Anton-Paar, Austria and Department of Chemistry, IIT Guwahati, **Guwahati, Assam,** Feb 26, 2020.

Awards and Honors

1. ***DBT-CTEP International Travel Grant to Bali, Indonesia*** to participate in ICBBT-2022**.**
2. ***Best Flash*** Presentation & Poster Award by BRSI-BSBB, 7-10th Dec, 2022, IIT Guwahati.
3. ***Best Offline Oral Presentation Award*** by ICBBT-2022 held from 29-30th Oct, 2022 at Udayana University, **Bali, Indonesia**.
4. ***Press Release*** by IIT Guwahati for the research work done on *Rf*GH5_4, 28th Dec, 2022 which was covered by various national and international media outlets such as Times of India, India Today, News18, The Better India, Sentinel Assam, etc.
5. *InSc* ***Young Researcher Award*** June, 2022 by Institute of Scholars, Bangalore.
6. *2nd* Prize for ***Best Oral Presentation*** by Research and Industrial Conclave, RIC-2022, IIT Guwahati.

Service/Contribution to Department of Biosciences and Bioengineering (BSBE)

1. Teaching Assistantship with Prof. Arun Goyal for BT608 (Microbial Technology) Jan-May 2023.
2. Teaching Assistantship with Prof. Arun Goyal for BT412 (Enzymology) Jan-May 2022.
3. Teaching Assistantship with Prof. Arun Goyal for BT631 (Protein Structure, Function and Crystallography) July-November 2021.
4. Teaching Assistantship with Prof. Arun Goyal and Prof. G. Kaur for BT305 (Introductory Biotechnology), Jan-April 2020.
5. Teaching Assistantship with Prof. Arun Goyal for BT631 (Protein Structure, Function and Crystallography) July-November 2019.
6. I was given an opportunity to carve the name for the BSBE Annual Departmental Newsletter, now known as "***Awalokan***".
7. Cultural Performance for *Nazm* and *Ghazal* in the BSBE Departmental Retreat: The Biotech Express 2021, 2020 and 2019.

Performed Registration duty for PhD admission test July 2019, 2022.

VITAE

The author was born on November 9, 1994 in Nandra Haveli, Jamner, Jalgaon, Maharashtra (India). He passed the Secondary School Certificate Examnination (10th Class in 2010 from Shri Sayajirao Maharaj Madhyamik Vidyalaya, Pune and Higher Secondary School Certificate Examination (10+2) in 2012 from Bharat English School, Shivajinagar, Pune; both the exams were conducted by Maharashtra State Secondary and Higher Secondary Examination Board, Pune, Maharashtra. School Certificate. He completed B.Sc. in Biotechnology in 2015 from Padmashree Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri, Pune which is affiliated to Savitribai Phule Pune University, Pune. He completed his M.Sc. in Biotechnology in 2017 from Department of Biotechnology, University of North Bengal, Siliguri, Darjeeling, West Bengal through a Department of Biotechnology (DBT), Government of India sponsored educational program. He worked as a Junior Research Fellow under at Department of Science and Technology (DST), Govt. of India sponsored Project till March 2018.

Mr. Parmeshwar joined the Ph.D. program in July 2018 at Department of Biosciences and Bioengineering (BSBE), Indian Institute of Technology Guwahati (IIT Guwahati), Guwahati, Assam, India under the Supervision of Prof. Arun Goyal. He received the the Institute GATE Fellowship, for whole Ph.D. tenure under the scheme run by Ministry of Education (Erstwhile, Ministry of Human Resource Development a.k.a. MHRD), Government of India, New Delhi. He delivered the open seminar (Ph.D. Synopsis) on January 9, 2024, presented his Ph.D Thesis work before Doctoral Committee and his performance was satifactiory. He submitted the Ph.D. thesis to Indian Institute of Technology Guwahati, Guwahati in January 2024 in partial fulfilment of the requirements for the award of the Degree, Doctor of Philosophy in Biocsciences and Bioengineering.