Structure and functional analysis of a recombinant endoglucanase (*At*GH9C-CBM3A-CBM3B) from *Acetivibrio thermocellus* ATCC 27405 and its application in synthesis of nanocellulose-based hydrogel for dye removal

PhD Thesis

by

ARDHENDU MANDAL

DECEMBER 2025

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI 781039, ASSAM, INDIA

Structure and functional analysis of a recombinant endoglucanase (*At*GH9C-CBM3A-CBM3B) from *Acetivibrio thermocellus* ATCC 27405 and its application in synthesis of nanocellulose-based hydrogel for dye removal

A thesis submitted for the partial fulfilment
of requirements for the award

of

DOCTOR OF PHILOSOPHY

by

ARDHENDU MANDAL

Under supervision of

Professor Arun Goyal

December 2025

BIOSCIENCES AND BIOENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI 781039, ASSAM, INDIA

Dedicated to my Teachers

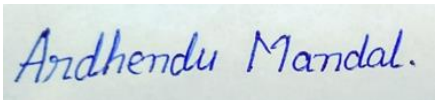
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

STATEMENT

I do hereby declare that the content embodied in this thesis entitled as **"Structure and functional analysis of a recombinant endoglucanase (*At*GH9C-CBM3A-CBM3B) from *Acetivibrio thermocellus* ATCC 27405 and its application in synthesis of nanocellulose-based hydrogel for dye removal"** 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.



Dec, 2025

Ardhendu Mandal
(206106004)

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

CERTIFICATE

It is certified that the work described in this thesis entitled "**Structure and functional analysis of a recombinant endoglucanase (*At*GH9C-CBM3A-CBM3B) from *Acetivibrio thermocellus* ATCC 27405 and its application in synthesis of nanocellulose-based hydrogel for dye removal**" by Ardhendu Mandal (Roll No. 206106004) for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision at the Department of Biosciences & Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India and this work has not been submitted elsewhere for a degree.

Dr. Arun Goyal *(MSc, MTech, PhD)*
(FAMI, FBRS, FABAP, FNABS, FNAAS, FIFIB, FMBI, FACCTI, FSEES)
(Thesis Supervisor)
Professor (HAG) and Former Head,
Department of Biosciences & Bioengineering
Indian Institute of Technology Guwahati
Guwahati, 781 039, India

ACKNOWLEDGEMENTS

The Ph.D. journey has been transformative, marked by immense growth, learning and self-discovery. While this thesis signifies the culmination of years of rigorous academic pursuit, it also stands as a testament to the collective efforts of those who have shaped my research and personal development. To all who contributed in meaningful ways, your impact on this study and my journey is deeply cherished. I am also profoundly grateful to everyone who played a role, directly or indirectly, in making this thesis possible. Their support has not only facilitated the completion of this academic endeavor but has also made it an unforgettable experience.

I would first like to express my sincere gratitude to my supervisor, Prof. Arun Goyal, Department of Biosciences and Bioengineering, IIT Guwahati, for his invaluable guidance, unwavering support and constant encouragement throughout this research journey. His critical feedback, scientific insights and mentorship have been instrumental in shaping this work and my growth as a researcher. I am deeply grateful for his patience, leadership and belief in my potential, which motivated me to push my limits and strive for excellence. His passion for scientific inquiry has been truly inspiring and the analytical thinking he instilled in me extends beyond this thesis. Our insightful discussions on research and his emphasis on soft skills have significantly contributed to both my academic and personal development. Thank you, sir, for the privilege of being your student, this journey would not have been possible without your guidance!

I am equally grateful to my doctoral committee members - Prof. Debasish Das, Prof. Manish Kumar and Dr. Souptick Chanda - for their invaluable guidance throughout my research journey. Their expert suggestions and constructive feedback during all my APS (Annual Progress Seminars) were crucial in shaping my work and steering me toward successful thesis completion. Their critical insights significantly enhanced the quality of my research.

I gratefully acknowledge the Department of Biosciences & Bioengineering and the Central Instrumentation Facility (CIF), IIT Guwahati, for their essential

infrastructural support and access to advanced research instruments. My sincere thanks also go to IIT Guwahati for providing computational resources through the Param-Kamrupa Supercomputing Facility.I extend my deep appreciation to Dr. Ravishankar Ramachandran, Principal Scientist, and his team at CSIR-Central Drug Research Institute (CDRI), Lucknow, for their generous assistance with SAXS facility access, which was invaluable to my research.

I sincerely thank the present and former Heads of the Department of Biosciences & Bioengineering, IIT Guwahati – Prof. Utpal Bora, Prof. Rakhi Chaturvedi, and Prof. Latha Rangan – for their leadership and for ensuring access to essential research facilities. I am also deeply grateful to all faculty members, administrative staff, and technical personnel of the department for their invaluable support and assistance throughout my research journey. I extend my sincere gratitude to the Ministry of Education (MoE), Government of India, for the financial support provided during this research work.

I am deeply grateful to all members of the Carbohydrate Enzyme Biotechnology Laboratory (CEBL) for their invaluable support throughout my research journey. Special thanks to my seniors - Dr. Kaustubh Khaire, Dr. Maibam Premeshworii Devi, Dr. Jebin Ahmed, Dr. Parmeshwar Gavande and Robin - for their expert guidance and insightful suggestions. I sincerely appreciate my colleague Madhulika and juniors Vishwanath, Shreya, Akshita, Aishwarya, Bipasha, Sushruta, Akshay, Ashwani, Tiyasa, Deblina, Annu and Nilakshi for their constant assistance in laboratory work. The collaborative spirit and supportive environment you all created made this research journey both productive and enjoyable. This work would not have been possible without their collective contributions.

I have been fortunate to meet extraordinary people during this journey who enriched both my research and personal growth. My deepest gratitude goes to my dear friends Aishwarya, Harish, Kuhelika, Kalyan, Pabir, Sayanti, Shubhangini and Simra for their constant encouragement and unwavering support.

A special acknowledgement goes to C. Raghav – our late-night discussions about science, politics and life have profoundly shaped my perspectives. Thank you for always stepping in to help with my work and for giving me strength during challenging times. Your support kept me moving forward.

I also extend my sincere appreciation to Gourav Bhaiya, Siddhartha Bhaiya, Pratik Da, and Gayatri Di for their invaluable guidance and assistance throughout this journey. I would also like to express my sincere gratitude to my fellow members of the IIT Guwahati football community. Your camaraderie and the joyful moments we shared on the field served as the perfect antidote to stress throughout my PhD journey.

The completion of my PhD journey would not have been possible without the boundless love, unwavering faith, and endless sacrifices of my beloved mother. To her, I owe every achievement – it is my deepest wish to make her proud.

I am eternally grateful to my high school science teacher, Mr. R.K. Ghosh, whose inspirational guidance first ignited my passion for science. His steadfast encouragement and character-building challenges (which I now cherish as blessings) have shaped both my academic path and personal growth.

My profound thanks to my entire family and my dearest friend Subho – your unconditional support, quiet prayers, and constant belief in me have been my unshakable foundation through every challenge. This accomplishment belongs as much to you as it does to me.

Finally, I offer my deepest gratitude to the Almighty for the countless blessings that sustained me throughout this remarkable journey. I am eternally thankful for the divine grace that granted me both the physical endurance and mental fortitude to complete this work.

Ardhendu Mandal
Dec, 2025

SYNOPSIS

Introduction

Plants serve as the principal autotrophic producers within ecological food webs, sustaining the energy demands of nearly all organisms across diverse habitats through the intricate biochemical process of photosynthesis. They accumulate energy in the form of carbohydrates (sugars), which are stored in various tissues, including leaves, seeds, and roots. Carbohydrates are defined as organic compounds characterized by polyhydroxy aldehydes or ketones, or derivatives thereof, formed via fundamental chemical modifications such as hydrolysis, oxidation, or reduction. These biomolecules constitute up to 90% of the primary cell wall, playing critical roles in structural integrity, cellular protection, and signaling processes. The major carbohydrate constituents of plant cell walls include cellulose, hemicellulose, and pectin. Cellulose consists of β-1,4-linked glucan chains that assemble into crystalline microfibrils through hydrogen bonding. Hemicelluloses comprise heterogeneous polysaccharides, such as xyloglucans, arabinoxylans, arabinogalactans, mannans, xylans, and mixed-linkage glucans, with xylan and mannan being the most prevalent. Pectin, the most structurally intricate polysaccharide, is predominantly localized in the primary cell wall and middle lamella, contributing to cell adhesion and matrix cohesion.

Cellulose is a linear polymer composed of β-1,4-linked D-glucose units. The glucose molecules are connected by glycosidic bonds, forming long, unbranched chains. These chains interact through extensive hydrogen bonding, creating rigid and crystalline

microfibrils. The hierarchical organization of cellulose includes microfibrils with alternating crystalline regions, which are highly ordered, and amorphous regions, which are disordered. These microfibrils bundle together to form the structural framework of plant cell walls, where they are further reinforced by hemicellulose and lignin. Cellulose exists in several structural forms with distinct properties. Native cellulose (Cellulose I) occurs naturally in plant cell walls (cotton, wood) and appears in two crystalline variants: I_α (triclinic, dominant in algae/bacteria) and I_β (monoclinic, prevalent in higher plants). Through chemical processing (mercerization/dissolution), cellulose converts to the more stable Cellulose II form (rayon, lyocell), characterized by antiparallel chain packing. Amorphous cellulose, produced via acid hydrolysis or mechanical grinding, lacks crystalline structure, enhancing its reactivity for industrial applications.

Plant biomass deconstruction requires a diverse array of carbohydrate-active enzymes (CAZymes), which hydrolyze, transfer or modify carbohydrate molecules. The CAZy database systematically classifies carbohydrate-active enzymes into five principal functional categories: Glycoside Hydrolases (GHs), Glycosyl Transferases (GTs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs), and Auxiliary Activities (AAs). GHs are particularly crucial, with endoglucanases (distributed across GH5, GH9, GH12, GH45 and GH74) initiating cellulose breakdown by randomly cleaving internal β-1,4-glycosidic bonds, producing oligosaccharides with new chain ends. This endo-acting activity generates new chain ends, facilitating subsequent hydrolysis by exoglucanases (cellobiohydrolases) and β-glucosidases. The synergistic

action of these enzymes ensures efficient cellulose depolymerization into fermentable sugars. Endoglucanases employ general acid-base catalysis, utilizing two carboxylate residues (Glu/Asp) as proton donor and nucleophile/base to hydrolyze glycosidic bonds via either retaining or inverting mechanisms. These enzymes display diverse structural architectures, including $(\alpha/\beta)_8$ barrels and β-jelly roll folds, with GH9 members typically adopting $(\alpha/\alpha)_6$-barrel configurations that optimize amorphous cellulose degradation. Most endoglucanases feature a modular organization, combining: i) a Catalytic module for bond cleavage and ii) Carbohydrate-binding modules (CBMs) that enhance substrate affinity and targeting. CBMs, classified into families by sequence specificity, critically influence enzyme processivity on crystalline cellulose. The CBM3 family, common in bacterial systems, mediates crystalline cellulose recognition through planar aromatic residue arrays.

Structural understanding of multi-modular endoglucanases remains limited due to scarce full-length structural data, hindering insights into module synergy and linker optimization. The mechanisms of substrate interaction by multi-modular endoglucanases in complex lignocellulosic matrices are poorly characterized. While CBMs enhance catalytic efficiency on crystalline substrates by up to 10-fold, their quantitative effects on kinetic parameters across enzyme families require systematic study. Processivity mechanisms in CBM-containing enzymes also need clarification. Industrial applications face challenges including product inhibition, thermostability-activity tradeoffs and pH sensitivity. Moreover, the application of endoglucanase-

derived nanocellulose has gained attention for use in hydrogels, dye adsorption, and composite material development. Currently, no three-module endoglucanase has been structurally or functionally characterized. Addressing this gap requires identifying thermostable, high-activity variants for biochemical and structural analysis coupled with protein engineering to optimize performance. Such efforts will elucidate modular synergy and advance industrial biocatalyst development for industrial applications.

Acetivibrio thermocellus is a Gram-positive, thermophilic, anaerobic bacterium characterized by a rod-shaped morphology and the presence of cell-surface cellulosomes. The organism is known to synthesize a sophisticated cellulosomal system, comprising both cell-free and cell-bound cellulosome complexes. While multiple glycoside hydrolases from this organism have been biochemically characterized, the structural and functional properties of a three-module endoglucanase remain unexplored. A gene (GenBank accession number ABN53959.1) encoding a glycoside hydrolase family 9 (GH9) enzyme from *Acetivibrio thermocellus* was identified and designated as *At*GH9C-CBM3A-CBM3B. *At*GH9C-CBM3A-CBM3B is proposed to be an endoglucanase of family 9 of glycoside hydrolases belonging to the cellulosomal complex from *Acetivibrio thermocellus*. This study reports the cloning, purification, and comprehensive biochemical characterization of full-length multi-modular GH9 family endoglucanase, *At*GH9C-CBM3A-CBM3B and its truncated derivatives (*At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B) from *Acetivibrio thermocellus*. The structural and functional properties of the endoglucanase *At*GH9C-CBM3A-CBM3B

were investigated using *in silico* analyses and small-angle X-ray scattering (SAXS). Additionally, the synergistic action of endoglucanase *At*GH9C-CBM3A-CBM3B and cellobiohydrolase *At*CBH5A was assessed for nanocellulose production from sugarcane trash cellulose, followed by the fabrication of nanocellulose-based hydrogels for dye removal applications. The molecular architecture of full-length endoglucanase, *At*GH9C-CBM3A-CBM3B (gene locus Cthe_2760), currently under investigation, is shown below (**Fig. 1**).

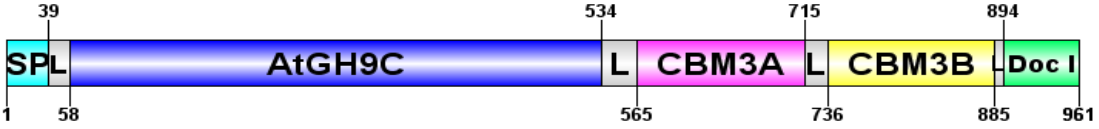


Fig. 1. Structural organization of the full-length endoglucanase, *At*GH9C-CBM3A-CBM3B, encoded by locus Cthe_2760 (GenBank ABN53959.1) from *Acetivibrio thermocellus* ATCC 27405, illustrating the GH9 catalytic module (blue) flanked by CBM3A (pink) and CBM3B (yellow) modules.

Present study

The **present study**, entitled "**Structural and functional analysis of a recombinant endoglucanase (*At*GH9C-CBM3A-CBM3B) from *Acetivibrio thermocellus* ATCC 27405 and its application in synthesis of nanocellulose-based hydrogel for dye removal**" is organized into five distinct chapters.

Chapter 1 provides a general introduction and literature review on plant cell wall polysaccharides and their enzymatic degradation, focusing on GH9 endoglucanases. Plant cell walls consist primarily of cellulose, hemicellulose, and pectin. Cellulose, a

linear β-1,4-glucan polymer, forms crystalline and amorphous microfibrils, influencing enzymatic accessibility. Hemicellulose and pectin further contribute to structural complexity. Synthesized by membrane-bound cellulose synthase complexes, cellulose predominantly exists as the native cellulose I allomorph. Plant biomass degradation is mediated by carbohydrate-active enzymes (CAZymes), including glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activity (AA) enzymes. GHs hydrolyze glycosidic bonds via retaining or inverting mechanisms, often using aspartate/glutamate catalytic residues. Many GHs feature modular architectures, combining catalytic domains with carbohydrate-binding modules (CBMs) like CBM3, which targets crystalline cellulose. GH9 endoglucanases, characterized by an $(\alpha/\alpha)_6$-barrel fold, cleave internal β-1,4-glycosidic bonds, preferentially targeting amorphous cellulose. Bacterial GH9 enzymes often include CBMs that enhance substrate binding, catalytic efficiency, and stability. The cellulosome of *Acetivibrio thermocellus* exemplifies an efficient multi-enzyme system for biomass deconstruction, with GH9 enzymes playing a key role in cellulose breakdown. This study investigates the *A. thermocellus* endoglucanase *At*GH9C-CBM3A-CBM3B, examining the functional contributions of its CBMs to enzymatic activity. This multi-domain enzyme holds promise for engineering advanced cellulases, with applications in biofuels and nanocellulose production.

Chapter 2 focuses on cloning and hyper-expression of modular endoglucanase, *At*GH9C-CBM3A-CBM3B, along with its truncated derivatives from *Acetivibrio*

thermocellus ATCC 27405, to facilitate their structural and functional characterization. The gene of interest, identified by the UniProt ID A3DJ30 and GenBank accession number ABN53959.1. BLASTp analysis of gene locus A3DJ30_ACET2 identified an N-terminal signal peptide (1-39) followed by the catalytic *At*GH9C module (58-534) with flanking linkers (40-57, 535-564). The C-terminal region contains two CBMs (CBM3A:565-715 and CBM3B:736-885) separated by a linker (716-735), concluding with a short linker (886-893) and Doc I-like domain (894-961) (**Fig.1**). The individual functional contributions of these domains were assessed by PCR amplification of the DNA sequences encoding *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B from *Acetivibrio thermocellus* genomic DNA using domain-specific primers containing *NheI* and *XhoI* restriction sites. The amplified fragments were subsequently cloned into the pET-28a(+) expression vector for heterologous protein production. PCR amplification yielded expected product sizes: ~2.5 kb for the full-length gene (*At*GH9C-CBM3A-CBM3B), 2.0 kb for *At*GH9C-CBM3A, 1.4 kb for *At*GH9C, and ~0.5 kb each for CBM3A and CBM3B. The inserts were ligated into linearized vector and transformed into *E. coli* DH5α andh CBM3B showing highest transformation efficiency. The recombinant plasmids were isolated and confirmed by restriction digestion, confirming the successful cloning of the gene inserts into pET-28a(+). *NheI/XhoI* digestion confirmed all constructs, displaying correct fragment sizes. For heterologous protein production, confirmed recombinant plasmids were transformed into *E. coli* BL-21 (DE3) cells. IPTG induction (0.25-1.0 mM)

revealed optimal protein expression at 1.0 mM for all constructs and confirmed by SDS-PAGE. The overexpressed recombinant proteins were detected with approximate molecular masses consistent with the predicted sizes: ~100 kDa for the full-length *At*GH9C-CBM3A-CBM3B, ~73 kDa for *At*GH9C-CBM3A, ~54 kDa for *At*GH9C, ~16 kDa for CBM3A, and ~18 kDa for CBM3B.

Chapter 3 focuses on a novel modular β-1,4-endoglucanase, *At*GH9C-CBM3A-CBM3B and investigates the roles of its associated CBM3A and CBM3B, in enzymatic function and stability. The full-length recombinant enzyme *At*GH9C-CBM3A-CBM3B and its truncated versions: *At*GH9C-CBM3A (lacking CBM3B), *At*GH9C (catalytic module only), and isolated CBMs were expressed in *E. coli* cells. The purified constructs exhibited molecular masses matching theoretical values: 97.2 kDa (*At*GH9C-CBM3A-CBM3B), 72.3 kDa (*At*GH9C-CBM3A), 54.4 kDa (*At*GH9C), 16.4 kDa (CBM3A), and 17.9 kDa (CBM3B), as confirmed by SDS-PAGE. The full-length, *At*GH9C-CBM3A-CBM3B showed the highest purification fold (19.8), followed by *At*GH9C-CBM3A (15.4) and *At*GH9C (2.3). The full-length enzyme exhibited strong activity against β-1,4-linked substrates, with highest activity on CMC-Na (58.8±2.6 U/mg), followed by lichenan (44.5 ± 2.0 U/mg) and β-D-glucan (36.2 ± 1.7 U/mg). Minimal activity was observed on crystalline Avicel, while β-1,3/1,6-linked polysaccharides (laminarin, curdlan) were not hydrolyzed. Removal of either CBM3A or CBM3B drastically reduced activity, with the isolated catalytic module (*At*GH9C) showing only minimal activity, demonstrating the crucial role of CBMs in substrate

interaction. Biochemical characterization revealed optimal activity at pH 7.5 and 55°C, with more than 80% activity maintained between pH 6.0-9.0 and 4-60°C, indicating excellent stability. Kinetic analysis showed highest catalytic efficiency for CMC-Na (K_m = 1.5 mg/mL, V_{max} = 61 U/mg, K_{cat}/K_m = 65.5 mL·mg^{-1}·s^{-1}). Mg^{2+} enhanced activity by 29%, while Zn^{2+}, Cu^{2+} and Fe^{2+} were inhibitory. Chelators (EDTA/EGTA) reduced activity more than 70%, confirming catalytic importance of Mg^{2+} ion. CD spectroscopy revealed a T_m of 65°C, increasing to 70°C with Mg^{2+}, demonstrating its stabilizing effect. TLC analysis confirmed endo-activity, producing cellotriose/tetraose from CMC-Na within minutes, yielding 355 µg/mL reducing sugars after 12h. CBM functional studies showed CBM3A restored 47% of full-length activity (CMC-Na) at 1:3 molar ratio, while CBM3B restored only nearly 13%. Combined CBMs restored nearly 50% activity and enabled weak Avicel hydrolysis, suggesting cooperative binding. The results suggest CBM3A directly enhances catalysis, while CBM3B, likely due to its affinity for crystalline cellulose, contributes marginally unless present in higher concentration. Thermostability analysis revealed that only the presence of CBM3A significantly enhanced the thermal resilience of *At*GH9C, with activity increasing from 2.4 U/mg (*At*GH9C alone) to 24.9 U/mg (*At*GH9C + CBM3A) at 50°C. CBM3B offered minimal thermal protection, while the combined CBM3A+CBM3B mixture showed similar thermal protection as CBM3A alone. These findings affirm that physical and possibly spatial proximity of CBM3A to the catalytic domain is critical for enzyme stabilization and function. In summary, *At*GH9C-CBM3A-CBM3B represents

a thermostable GH9 endoglucanase whose full activity requires both CBMs in their native configuration. Its robust activity across wide pH and temperature ranges makes it promising for various industrial applications.

Chapter 4 details about the computational and experimental solution structures of *At*GH9C-CBM3A-CBM3B, revealing its modular architecture and dynamics. The homology model reveals a multi-domain organization comprising an $(\alpha/\alpha)_6$-barrel catalytic module (*At*GH9C) flanked by two antiparallel β-sandwich CBMs (CBM3A/CBM3B). Multiple sequence alignment and structural superposition analyses identified a conserved catalytic triad consisting of Asp98, Asp101, and Glu489. The spatial arrangement of catalytic bases Asp98 and Asp101 positioned approximately 8.5 Å from the catalytic acid Glu489. This configuration is characteristic of inverting glycoside hydrolases. Circular dichroism spectroscopy validated the secondary structure composition, quantifying 25.2% α-helix, 18.4% β-sheet, and 56.4% random coil, which correlated well with computational predictions from multiple servers. Molecular dynamics simulations for 200 ns and 300 K provided valuable insights into the structural dynamics of enzyme. The RMSD analysis revealed an average value of 1.8 nm after 140 ns of simulation, indicating a stable conformational state. RMSF analysis showed that loop regions, particularly L3 and L4, were more flexible, and likely contribute to substrate gating, while the catalytic triad remained rigid (RMSF $\leq$ 0.15 nm). R_g showed a consistent value of 4.2 nm from 140-200 ns, confirming global compactness. Notably, RMSF analysis identified high flexibility in CBM3B regions, suggesting this module

may have reduced functional importance in substrate binding compared to CBM3A. Molecular docking studies demonstrated strongest binding affinity for cellotetraose (ΔG = –5.05 kcal/mol), indicating the active-site of enzyme groove optimally accommodates tetrasaccharide units. Binding was mediated by hydrogen bonding and hydrophobic interactions involving His173, Trp176, Trp378, and Tyr485. The active-site cleft could accommodate up to four glucose units (subsites –3 to +1), suggesting a processive endoglucanase mechanism, similar to Cel9M, with cellotetraose as the primary product. MD simulations of the enzyme-substrate complex confirmed ligand-induced stabilization (RMSD reduced to 1.5 nm), with lower flexibility in catalytic residues and substrate-binding loops (L1–L4). CBM3A showed greater stabilization than CBM3B. R_g remained stable (3.3 nm), with reduced SASA (405 nm²) and increased intramolecular H-bonds (610), confirming stabilization of the active-site cleft and CBMs. Processivity analysis on phosphoric acid-swollen cellulose classified *At*GH9C-CBM3A-CBM3B as a pEG4-type processive endoglucanase. Loop 3 (residues 342–379) likely acts as a barrier at the non-reducing end, promoting cellotetraose release. SAXS analysis of *At*GH9C-CBM3A-CBM3B in phosphate buffer (pH 7.5, 5 mg/mL) showed an elongated, monodisperse structure (R_g = 5.10 nm, D_{max} = 16.2 nm), with Guinier analysis confirming solution homogeneity (R_g = 4.50 ± 0.33 nm, R_c=1.05 ± 0.93 nm). Kratky plots revealed compact folding (persistence length = 15.2 nm), while *ab initio* modeling generated a fist-and-elbow shaped monomer (94.3 kDa, NSD<1.80±0.14) with spatially distinct CBM3B. Complementary DLS data showed

concentration-dependent hydrodynamic radii (R_h = 4.6-5.8 nm) and strong negative surface charge (-18.6 to -24.2 mV), accounting for solution stability. These results validate the multi-domain architecture of the enzyme and propose a framework for targeted mutagenesis of catalytic residues and loops to optimize activity or tailor products for industrial use.

Chapter 5 describes the enzymatic synthesis of nanocellulose *via* synergistic hydrolysis by endoglucanase (*At*GH9C-CBM3A-CBM3B) and cellobiohydrolase (*At*CBH5A) and its application in dye adsorption using nanocellulose-carboxymethyl cellulose (CMC) composite hydrogels. The enzymatic synthesis of nanocellulose presents an environmentally sustainable approach, overcoming the limitations of conventional energy-intensive and low-yield chemical methods. The recombinant endoglucanase, *At*GH9C-CBM3A-CBM3B (97.2 kDa) and cellobiohydrolase, *At*CBH5A (62 kDa) were expressed and purified, showing specific activities of 58.8 ± 0.8 U/mg and 96.8 ± 1.2 U/mg, respectively, against 1.0% (w/v) CMC-Na. Cellulose extracted from sugarcane trash (SCT) through optimized alkali pretreatment was enzymatically hydrolyzed (3.0 mg enzyme/g cellulose, 6 h) to produce enzymatic nanocellulose (EN-NC). Comparative analysis with TEMPO-oxidized nanocellulose (TO-NC) revealed similar FTIR and XRD spectral patterns, while FESEM demonstrated distinct morphological characteristics. The SCT cellulose displayed micron-scale architecture (~10 μm), whereas both EN-NC and TO-NC exhibited nanoscale dimensions (10–100 nm), with EN-NC showing superior fibril uniformity (≤20 nm), indicative of enzymatic

specificity. The EN-NC-CMC-Na hydrogel was synthesized by dispersing EN-NC (0.5 g in 50 mL 2% acetic acid) into a CMC-Na solution (1.0 g in 50 mL deionized water, pH 6.0) at 35°C, reaching hydration equilibrium within 6 h. Characterization revealed a hydrogel pHpzc of 6.6 and a surface area of 4.81 m^2/g. Response surface methodology-optimized dye removal studies demonstrated exceptional performance: 88.5% Congo Red (CR) removal (48.75 mg/L) at pH 7.5 with 0.47 g hydrogel in 4.75 h, and 94.7% Methylene Blue (MB) removal (37.5 mg/L) at pH 6.0 with 0.6 g hydrogel in 3.5 h. Adsorption isotherm analysis revealed CR followed both Freundlich and Langmuir models, while MB adsorption conformed exclusively to the Freundlich model, indicating multilayer adsorption. Pseudo-second-order kinetics suggested chemisorption mechanisms involving hydrogen bonding, electrostatic interactions, and n–π stacking. The hydrogel exhibited excellent regeneration capacity using 0.1 M NaOH/50% ethanol for CR and 0.1 M HCl/HNO_3 for MB. These findings position the EN-NC-CMC-Na hydrogel as a sustainable, high-performance solution for tertiary treatment of dye-contaminated industrial wastewater.

CONTENTS

Chapter 1

General Introduction

1.1 Carbohydrates

Plants serve as fundamental biotic components of global ecosystems, functioning as primary producers that support trophic dynamics across terrestrial and aquatic habitats. This multicellular, eukaryotic organism fulfills the primary requirements of energy through the complex process of photosynthesis and weaves the invisible thread of life on Earth. Plants predominantly accumulate and store energy as carbohydrates, subsequently serving as a metabolic resource for heterotrophic organisms [1]. Carbohydrates, the most abundant polysaccharides on Earth, consist of carbon (C), hydrogen (H) and oxygen (O), with the general formula $(CH_2O)_n$, where $n \geq 3$ [2]. Classification of the carbohydrates is based on the degree of polymerization or the number of saccharide units that are present in them. A single unit of polyhydroxy ketone or aldehydes is termed a monosaccharide. The most abundant monosaccharide

in nature is glucose, which is the key substrate in cellular respiration for energy production [3]. Two to ten monosaccharides assemble to form oligosaccharides and more than 20 monosaccharide units are called polysaccharides [3]. Various cellulolytic bacteria and fungi, along with associated microbes, convert complex carbohydrates into soluble sugars like glucose and cellobiose for cellular uptake. The assimilation and utilization of the carbohydrates by cells depends on its overall structural composition, degree of polymerization, position of the anomeric carbon and nature of glycosidic linkage between the sugars present [4]. In plants and animals, carbohydrate reserves are primarily stored as starch and glycogen, respectively [5,6]. Similarly, other organisms utilize distinct polysaccharides for energy storage: laminarin in brown algae, carrageenan in red seaweeds and fructans (including inulin) in certain plants and bacteria [5]. These storage carbohydrates exhibit structural diversity tailored to their biological roles. They contribute significantly to the structural integrity and defense mechanisms of plant cells and additionally serve as signaling molecules involved in diverse physiological and biochemical pathways [5].

1.2 Structural composition of plant polysaccharides

Primary cell walls in plant are primarily composed of the polysaccharides, *viz.* cellulose, hemicellulose and pectin. The cellulose provides the structural strength, the hemicellulose involved in cross-linking microfibrils and pectin regulates the porosity and signaling [1]. In contrast, secondary cell walls are frequently fortified by the incorporation of lignin, a structurally complex and heterogeneous aromatic polymer, which confers increased mechanical strength and rigidity [7]. Almost 90 percent of the primary cell wall of the plants is primarily composed of carbohydrates and around 10% is made up of extensins, enzymes and structural glycoproteins, enabling protection,

structural support and signalling functions (**Fig. 1.1**) [7]. The secondary cell wall of plant comprises approximately, 10–30% hydrophobic polyphenylpropanoid lignin macromolecules, along with linear, unsubstituted or minimally substituted polysaccharides, including cellulose and heteroxylans [8]. The macromolecular, complex carbohydrate polysaccharides of plants are generally linear or branched chains of monosaccharides glued together by glycosidic bonds, often obtained after a condensation reaction [3].

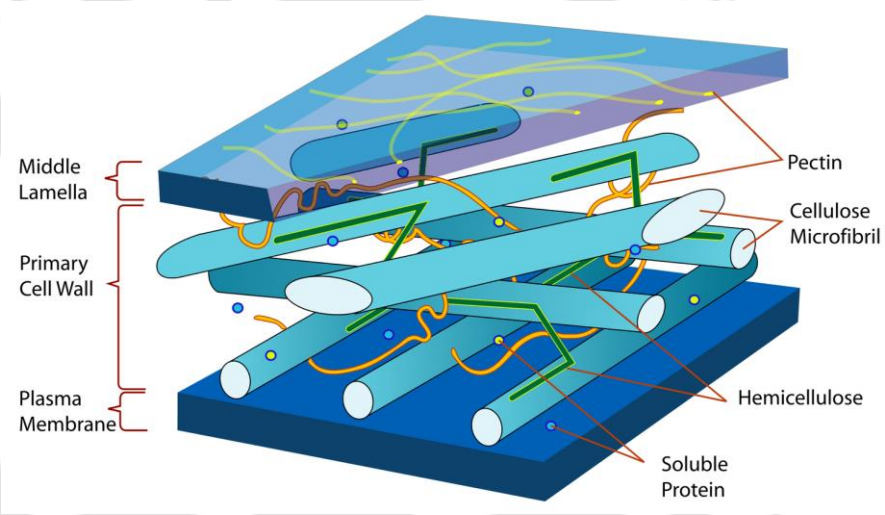


Fig. 1.1 Structural components of plant cell wall showing cellulose, hemicellulose, and pectin within the plant cell wall [7].

The primary component of the plant cell wall is cellulose, an unbranched homopolymer of β-(1→4)-glucose that adopts a predominantly linear configuration [9]. Cellulose is regarded as the most abundant polysaccharide (33–50% of terrestrial plant biomass) and organic substance on Earth and is believed to exist in two major forms: crystalline and amorphous. In its crystalline state, cellulose chains undergo self-association through intra- and intermolecular hydrogen bonding, as well as Van der Waals interactions, leading to the formation of crystalline microfibrils [10]. Fibrils are produced by plasma membrane-bound complexes (~30 nm in diameter), each

containing around 36 cellulose synthase (CESA) protein subunits. These complexes synthesize cellulose fibrils that provide mechanical strength to the plant cell wall [10]. Amorphous cellulose represent isotropically distributed, straight cellulose chains that lacks higher lacttice order and susceptible to degradation [11].

In addition to cellulose, the plant cell wall houses other different structural polysaccharides termed as hemicellulose and pectins. Hemicellulose constitutes a heterogeneous mixture of diverse monosaccharide units, including xylose, glucose, mannose, galactose, arabinose and uronic acids such as glucuronic and galacturonic acid. Xylan is a β-(1→4)-linked xylopyranose backbone polysaccharide, often substituted with side chains such as α-L-arabinofuranosyl, β-D-glucuronic acid or 4-O-methyl-D-glucuronic acid. These substitutions give rise to structural variants, including arabinoxylans, glucuronoxylans, and glucuronoarabinoxylans, which differ in their side-chain composition and functional properties [12,13]. Arabinoxylans consist of a β-(1→4)-linked D-xylopyranose backbone, substituted at O-2 and/or O-3 with α-L-arabinofuranosyl residues. Arabinogalactans are composed of a β-(1→3)-linked galactan backbone with β-(1→6)-galactan side chains. Xylans are located in the plant cell wall and plasma membrane, are involved in cell surface recognition and signalling [14]. Mannans are homopolysaccharides consisting of β-(1→4)-linked D-mannopyranose units. Based on their substitution patterns, mannans are classified into: (i) Glucomannans, which have a β-(1→4)-linked backbone of D-mannopyranosyl units with randomly interspersed β-D-glucopyranosyl residues in an approximate mannose-to-glucose ratio of 3:1; (ii) Galactomannans, which feature a β-(1→4)-linked mannose backbone substituted with galactose residues with side chains of single α-(1→6) linked D-galactopyranosyl residues; (iii) Galactoglucomannan consists of a β-(1→4)-linked

backbone of D-mannopyranose and D-glucose units, with the glucose residues further substituted by α-(1→6)-linked D-galactose [15]. Another principal class of structural polysaccharides present in plant cell walls includes pectins and modified heteropolysaccharides, which feature a backbone of α-(1→4)-linked D-galacturonic acid residues adorned with side chains of rhamnose, galactose and arabinose [1].

1.2.1 Cellulose and its derivatives

Cellulose is a linear, naturally occurring polysaccharide consisting of β(1→4)-linked D-glucose monomers with the molecular formula, $(C_6H_{10}O_5)_n$ [16]. Cellulose constitutes nearly 40% of plant carbon and provides the primary structural scaffold for plant cell walls, conferring mechanical integrity [17]. This naturally occurring biodegradable polymer predominantly exists as microfibrils in plant cell walls, woody tissues, algal matrices, and tunicate epidermal membranes. Additionally, certain bacterial species produce cellulose as nanofibrous networks without pectin/pectic contaminants, conferring structural superiority [17]. Cellulosic materials exhibit a multiscale hierarchical architecture, with structural organization extending from the nanoscale fibrillar level to macroscopic dimensions. Each superfine fibril comprises multiple cellulose chains exhibiting alternating crystalline and amorphous domains (**Fig. 1.2**). The cross-sectional dimensions of these fibrils range from 2-20 nm, varying according to their biosynthetic origin [17]. Individual cellulose chains traverse multiple crystalline and amorphous regions within fibrils, interconnected by robust β-(1→4) glycosidic linkages. Within crystalline domains, these chains demonstrate remarkable longitudinal alignment [17]. Cellulose and its derivatives have emerged as promising biocompatible polymers for biomedical applications, including advanced

wound dressings, drug delivery systems, and implantable medical devices, due to tunable physicochemical properties and exceptional mechanical performance [18].

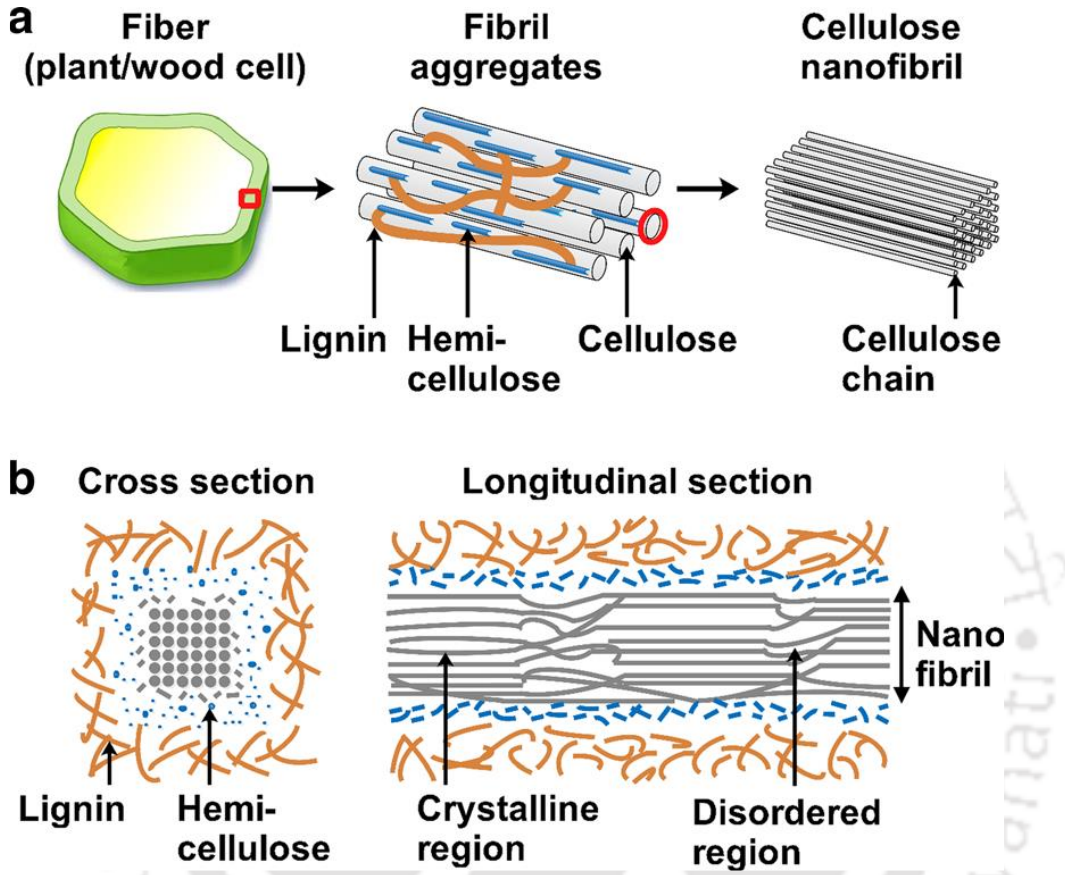


Fig. 1.2 Microstructure of cellulosic fiber: (a) Plant fiber, fibril aggregates, and cellulose-containing nanofibrils with lignin-hemicellulose matrix; (b) Cross/longitudinal sections showing crystalline/amorphous regions in nanofibrils embedded in lignin-hemicellulose [17].

1.2.1.1 Biosynthesis of cellulose

The cellulose comprises β(1→4)-linked D-glucopyranose units in $^4C_1$ chair conformation (equatorial -OH groups) with the degree of polymerization (DP) around 10,000 in native state and occur in six different polymorphs (I, II, III_I, III_{II}, IV_I and IV_{II}) [19]. Current evidence indicates cellulose synthesis occurs at or beyond the plasma membrane. Freeze-fracture microscopy reveals membrane-associated particle clusters, termed terminal complexes (TCs) or rosettes [20,21]. These terminal complexes (TCs)

are glued with microfibril termini and are hypothesized to represent cellulose synthase complexes responsible for microfibril elongation. Consequently, all chains within a microfibril must elongate synchronously, necessitating a multisubunit complex where each subunit polymerizes an individual chain. Given that microfibrils comprise 30–200 chains, this mechanism exhibits extraordinary complexity [22]. Cellulose exists in multiple crystalline polymorphs, designated as Cellulose I, II, III, and IV, which are generated through artificial processing methods such as recrystallization, acid hydrolysis and thermal treatment [22,23]. Cellulose I exists as a metastable crystalline polymorph, which is improbable to arise from the post-synthetic crystallization of cellulose chains. In contrast, cellulose II is generated through in vitro crystallization processes, resulting in the thermodynamically stable form [23,24]. The evidence suggests that crystallization occurs either in parallel with or immediately after TC-mediated glucose incorporation into nascent chains. Cellulose III and IV are thermodynamically derived polymorphs obtained through specific processing treatments of native cellulose (Cellulose I) or regenerated cellulose (Cellulose II). Thus, intermolecular hydrogen bond formation is closely linked to the entire crystallization process [23].

1.2.1.2 Types of cellulose

While all celluloses share the same basic chemical structure, they can differ significantly in their physical form, crystallinity, source and processing methods. Thus, cellulose can be classified based on its source, structure and treatment method into different polymorphs, including native cellulose, regenerated cellulose, microcrystalline cellulose, nanocellulose, bacterial cellulose, etc. The following sections describe its classification in detail.

1.2.1.2.1 Cellulose I, II, III and IV

The crystalline cellulose comprises four distinct structural allomorphs: cellulose I, II, III and IV. Crystallographic analysis revealed that cellulose features a monoclinic unit cell (a=0.61 nm, b=1.034 nm, c=0.54 nm, β=88°) with alternating glycosidic linkages (1→1)/(4→4) along the chain axis [25]. Cellulose I exhibits a parallel-chain crystalline structure stabilized by interchain hydrogen bonding and van der Waals interactions, conferring its characteristic mechanical rigidity and structural integrity [24]. Cellulose I is composed of mixture of 2 metastable polymorph triclinic (I_α) and monoclinic (I_β) and mainly occur in plant-based or tunicate-based cellulose . The I_α allomorph adopts a triclinic unit cell, while I_β exhibits a monoclinic unit cell with two parallel chains [24,26]. Cellulose II, the predominant polymorph after Cellulose I, is produced via two routes: (1) regeneration: dissolving cellulose I in a solvent, followed by aqueous dilution-induced precipitation to yield cellulose II or (2) mercerization: involves NaOH-induced swelling of native fibers, producing cellulose II upon alkali removal (**Fig. 1.3**). Cellulose III allomorphs are reversibly derived from cellulose I or II through liquid ammonia/amine treatment followed by solvent evaporation forming cellulose III_I and III_{II}, respectively [26]. Cellulose IV can be derived from either cellulose III_I or III_{II} through controlled thermal processing in glycerol at 206°C [11].

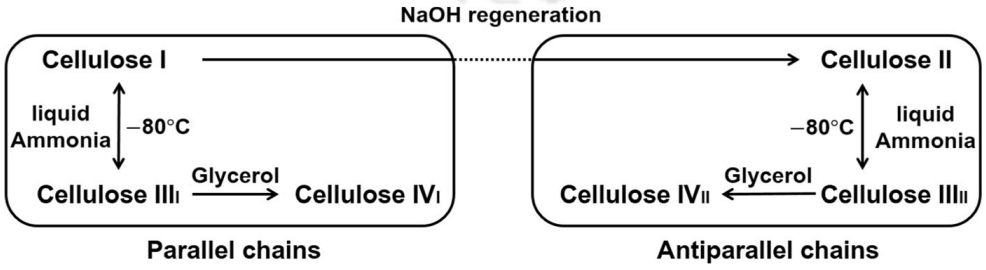


Fig. 1.3 Interconversion of cellulose crystalline phases (allomorphs I-IV) [17].

1.2.1.2.2 Bacterial cellulose

Bacterial cellulose (BC), also termed microbial cellulose, bacterial nanocellulose, or biocellulose, was first identified by Brown in 1886 as a robust gelatinous pellicle forming on vinegar fermentation broth [27]. BC is produced by terminal complexes in a near-pure state (>90%), unbound to external polymers like lignin or hemicellulose, enabling simple isolation and purification without intensive processing. It possesses unique properties, including high purity, crystallinity, mechanical strength, biocompatibility and water retention. Its non-toxic nature further enhances its suitability for biomedical applications [28]. BC is synthesized by various Gram-negative bacterial species, as well as the Gram-positive genus Sarcina, utilizing available carbon, nitrogen and oxygen sources. BC exhibits key structural properties, including high crystallinity (80-90%), extensive polymerization (7,000-16,000 glucose units), and exceptional mechanical strength comparable to steel or Kevlar, attributable to its nano-fibrillar architecture. BC is produced by acetic acid bacteria in both synthetic and natural media, and can be used for forming highly absorbent hydrogels with water content exceeding 99% in some cases [29]. Thus, BC demonstrates remarkable elasticity, exceptional strength in hydrated conditions and optimal conformability. The unique properties of BC make it particularly suitable for biomedical applications, notably in tissue engineering and pharmaceutical fields, where it serves as implants and scaffolds.

1.2.1.2.3 Algal cellulose

Marine algae, including brown (*Posidonia oceanica*), green (*Cladophora*) and red (*Gelidium elegans*) algae, represent significant cellulose sources, with their cell walls containing substantial cellulose content [30]. Red algae such as *Gelidium*

elegans contain abundant carbohydrates, primarily comprising energy-enriched cellulose and mucilaginous agar. Algae represent sustainable alternatives for cellulose production, utilizing waste-derived nutrients (e.g., wastewater, flue gas) while enabling scalable cultivation with environmental benefits. Algal-derived cellulose typically exists in association with hemicellulose, proteins and lignin, rather than in pure form [31]. Green algal-derived cellulose exhibits superior properties compared to wood cellulose (WC), Plant cellulose (PC) and BC, including high crystallinity (>70%), low moisture adsorption, mesoporous structure and large specific surface area [32]. These characteristics, combined with its biocompatibility and ease of functionalization, make algal nanocellulose particularly promising for biomedical applications like tissue engineering [33].

1.2.1.2.4 Microcrystalline cellulose

Microcrystalline cellulose (MCC) forms through selective extraction of crystalline regions from microfibrils via combinations of tailored mechanical, chemical and enzymatic treatments [34]. MCC particles exhibit rigid, rod-like morphology with highly ordered crystalline cellulose chains, initially characterized by Ranby (1949) and marketed as Avicel® [35]. Pharmaceutical-grade MCC (Avicel® PH) typically exhibits micrometer-scale particle lengths and is predominantly derived from wood cellulose. Microcrystalline cellulose (MCC) is produced via hydrochloric acid hydrolysis of alpha cellulose, which selectively depolymerizes amorphous regions to yield shorter, highly crystalline particles. This preferential hydrolysis of disordered domains results in characteristic high crystallinity of MCCs [36]. MCCs are rod-like particles measuring 10–200 μm, demonstrates excellent compatibility with active pharmaceutical ingredients (APIs) owing to its inert physiological properties, processability,

biodegradability and reliable sourcing. MCCs display a combination of biodegradability, thermal stability, superior mechanical properties (high modulus, aspect ratio) and low density, thus serving as an excellent biocomposite filler [37].

1.2.1.2.5 Nanocellulose

The nanocellulose is characterized as a cellulosic material possessing at least one nanoscale dimension. Nanocellulose is categorized into three primary types: bacterial nanocellulose (BNC), cellulose nanocrystals (CNCs) and cellulose nanofibrils (CNFs) [38]. The production methods, preparation techniques and functional properties of nano-cellulose are predominantly determined by the source material and processing conditions, which influence inherent characteristics such as crystallinity, thermal stability and morphology [39]. Nanocrystalline cellulose (NCC) and nanofibrillated cellulose (NFC) are typically produced, employing enzymatic, physical (defibrillation), or chemical extraction methods (alkaline, bleaching and acid hydrolysis) from agricultural crops, plants and waste byproducts. Bacterial nanocellulose (BNC) is biosynthesized *via* a bottom-up approach, where specific bacterial species convert glucose into exopolysaccharides that form stable hydrogels at air-liquid interfaces [40,41]. Nano-cellulose exhibits an aspect ratio of 20–500 nm length and 5–70 nm width. In nanocellulose, the percolation forms an interpenetrating fibrillar network within the polymer matrix, increasing modulus and tensile strength [42]. In pharmaceutical and drug delivery systems, BNC is the material of choice, while NFC proves more suitable for large-scale applications such as rheology modification and composite reinforcement. Its sustainable origin and exceptional properties make it a promising eco-friendly substitute for synthetic materials.

1.2.2 Hemicellulose

Hemicellulose is an amorphous, branched heteropolysaccharide with a β-(1→4)-linked backbone of glucose, galactose, mannose, arabinose or xylose found in plant cell walls, comprising 15-35% of lignocellulosic biomass [43]. The D-pyranosyl residues (e.g., glucose, mannose and xylose) in hemicellulose are interconnected via β-(1→4)-glycosidic linkages at the O4 position (oxygen of C4) in an equatorial conformation [44]. The structural complexity and relative abundance of hemicelluloses exhibit significant variation across species and cellular compartments. The composition and abundance of functional groups in hemicellulose side chains vary according to lignocellulosic biomass origin. Based on the predominant β-(1→4)-linked monosaccharides in their backbone, hemicelluloses are categorized into four major subclasses: mannans, xyloglucans, arabinans and xylans [43]. Xylan represents the predominant hemicellulose, followed by mannans, arabinans and xyloglucans, with compositional variations across species. Hemicelluloses primarily function in cell wall reinforcement through interactions with lignin and cellulose [43].

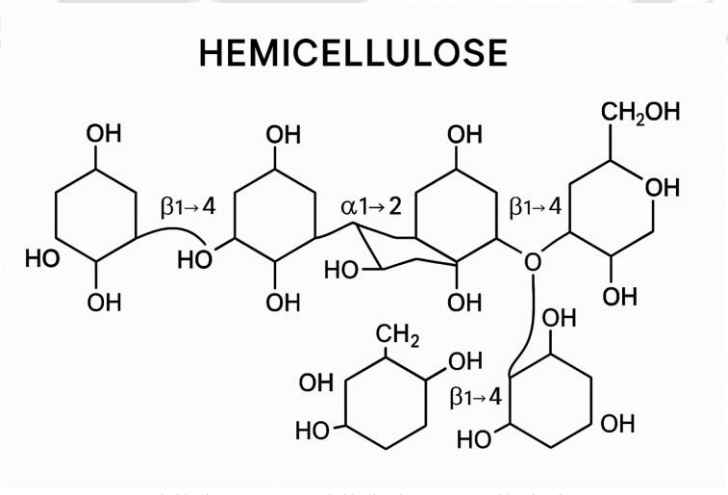


Fig. 1.4 Structural organization of hemicellulose showing β-(1→4) backbone linkages of glucose, mannose, or xylose and side-chain substitutions [43].

Hemicellulose branches exhibit a short-chain structure, contributing to their amorphous nature (**Fig. 1.4**). These polysaccharides are biosynthesized in the Golgi apparatus through the action of multiple glycosyltransferases [44,45]. While glycosyltransferases involved in xyloglucan and mannan biosynthesis are well-characterized, the enzymatic mechanisms underlying xylan and mixed-linkage β(1→3,1→4)-glucan formation remain poorly understood.

1.2.3 Pectins

Pectin, the most structurally intricate, versatile heteropolysaccharide found in plant cell walls, predominantly localizes in the primary cell wall and middle lamella, where it plays an important role in maintaining cell wall integrity [46]. Pectin constitutes an essential structural component in all higher plant cell walls, representing 0.5-4.0% of total fresh biomass. Pectins are characterized by polymeric chains of galactopyranosyluronic acid (GalpA) monomeric units, which can form calcium and magnesium pectates (**Fig. 1.5**) These divalent cations (Ca^{2+} and Mg^{2+}) establish salt bridges, resulting in insoluble gel formation [47].

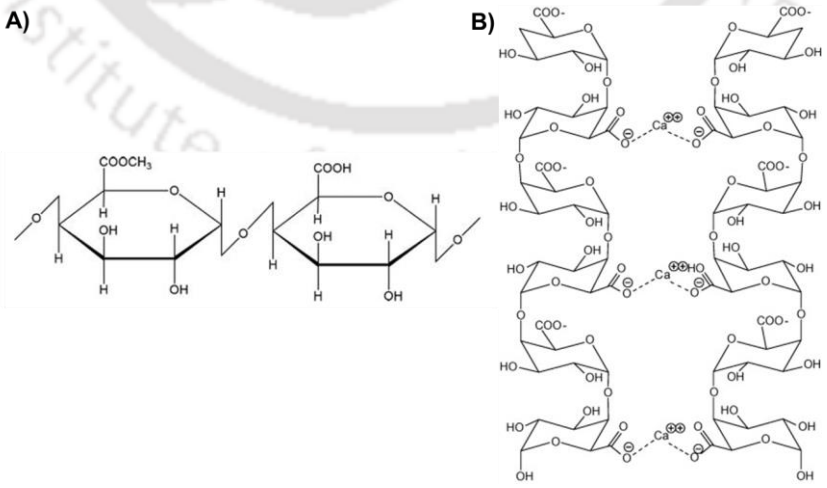


Fig. 1.5 Structural organization of pectin showing: a) α-D-galactopyranosyluronic acid, the primary repeating unit of pectin; b) Divalent cation (Ca^{2+}) coordination stabilizing pectate dimers [48].

Pectins are galacturonic acid-rich polysaccharides with two structural domains: a linear α-(1→4)-homogalacturonan (HG) ("smooth") with methyl/acetyl groups, and a branched region ("hairy") containing rhamnogalacturonan I (RG I), rhamnogalacturonan II (RG-II) and xylogalacturonan (XGA) [47]. The pectin matrix forms covalent (ester) and non-covalent (hydrogen-bonding) interactions with cellulose, hemicellulose and structural proteins including extensins and arabinogalactan proteins (AGPs). Pectin-modifying enzymes in the cell wall regulate its multifaceted biological functions, including plant morphogenesis, growth regulation, defense responses, cellular adhesion, wall structural integrity, permeability control, and fruit maturation [48]. Pectin is commercially valuable for its functional (thickening, gelling, emulsifying) and bioactive (anti-inflammatory, immunomodulatory) properties, with applications spanning food, pharmaceutical, and cosmetic industries.

1.3 Carbohydrate-active enzymes

Carbohydrates serve as essential energy substrates across nearly all life forms, undergoing continuous synthesis and degradation as dynamic metabolic components. Diverse microorganisms, particularly fungi and bacteria, have evolved specialized enzymes for plant cell wall polysaccharide degradation. Certain insects and mollusks also produce such cell wall-disrupting enzymes [1]. The biosynthesis and complete depolymerization of carbohydrates into monosaccharides and derived metabolites require coordinated action of specialized enzyme systems, collectively classified as carbohydrate-active enzymes (CAZymes) [49]. These CAZymes are cataloged in the CAZy database (www.cazy.org) and classified into families based on amino acid sequence homology, which reflects shared structural features and conserved catalytic mechanisms rather than substrate specificity [50]. The CAZy classification system

results from two important organizational features: (1) Enzymes with identical EC numbers can span multiple CAZy families and (2) CAZy families can contain enzymes of different EC numbers. The database categorizes these CAZymes into five major functional classes: Glycoside Hydrolases (GHs), Glycosyl Transferases (GTs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs), and Auxiliary Activities (AAs). The current CAZy database (2024 updated) comprises 381 enzyme families, including: 181 glycoside hydrolase (GH), 115 glycosyl transferase (GT), 46 polysaccharide lyase (PL), 21 carbohydrate esterase (CE) and 18 auxiliary activity (AA) families. Among the five CAZyme classes, glycosyltransferases (GTs) catalyze polysaccharide biosynthesis, while the remaining four classes—glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and auxiliary activities (AAs), mediate polysaccharide degradation or modification [51]. Approximately, 1-3% of an organism's genome encodes glycosyltransferases (GTs) and glycoside hydrolases (GHs).

Glycosyltransferases (GTs; EC 2.4.x.y) exclusively catalyze the transfer of sugar moieties from activated donors to specific acceptors, forming glycosidic bonds. Glycosyltransferases are categorized as retaining or inverting enzymes based on the stereochemical configuration of their substrates and products. GTs exhibit near-universal phylogenetic distribution across organisms.

Polysaccharide lyases (PLs; EC 4.2.2.x) mediate β-elimination of glycosidic bonds in uronic acid-containing polysaccharides through non-hydrolytic cleavage. Although PLs constitute merely 1.2 % of CAZymes in the CAZy database, they possess significant biotechnological and biomedical utility.

Carbohydrate esterases (CEs) facilitate the removal of O- and N-linked acyl groups from saccharides, breaking down ester modifications in different glycosidic polymers to allow further hydrolysis by glycoside hydrolases (GHs) and polysaccharide lyases (PLs). CEs account for 0.5% of identified CAZymes and these enzymes are essential for carbohydrate metabolism.

The auxiliary activities (AA) family comprises redox enzymes, including lignin-degrading enzymes and lytic polysaccharide monooxygenases (LPMOs), which oxidatively cleave lignin and crystalline polysaccharides. AAs synergize with other CAZymes for efficient lignocellulose deconstruction.

Carbohydrate-active enzymes (CAZymes) frequently exhibit a modular architecture, featuring catalytic domains linked to non-catalytic carbohydrate-binding modules (CBMs). These structurally independent CBMs consist of contiguous amino acid sequences with specific carbohydrate recognition capabilities [52]. CBMs are non-catalytic modules that enhance enzymatic activity by: (1) positioning catalytic modules near substrates, (2) stabilizing enzyme structure, and (3) improving thermostability [51]. These CBMs generally have β-sandwich in structure and known to increase the local concentration of the enzyme near its target, improving catalytic efficiency. Thus, they play a crucial role in biomass degradation, biofuel production and understanding pathogen-host interactions in microbes.

Glycoside hydrolases (GHs) are the primary mediators of enzymatic carbohydrate degradation. Glycoside hydrolases (GHs; EC 3.2.1.-) catalyze the cleavage of glycosidic linkages in carbohydrate-carbohydrate or carbohydrate-aglycone complexes, representing one of the most abundant enzyme classes in nature.

The nature, types, properties, origin and reaction mechanism are discussed in details in **section 1.3.1**.

1.3.1 Glycoside Hydrolase

Glycoside hydrolase (GH) is a widespread and diverse enzyme class that catalyzes the cleavage of glycosidic bonds in oligo-/polysaccharides or between carbohydrates and non-carbohydrate groups [53]. The International Union of Biochemistry and Molecular Biology (IUBMB) classifies glycoside hydrolases primarily by substrate specificity and occasionally by reaction mechanism, without considering the structural features. In contrast, the CAZy database categorizes these enzymes into families based on amino acid sequence homology, providing continuously updated classifications [54]. According to current classification, glycoside hydrolases comprise 181 distinct families, which are further organized into 18 superfamilies based on conserved protein structural folds. Glycoside hydrolases (GHs) cleave O-, N-, or S-linked glycosidic bonds in polysaccharides or oligosaccharides, acting either endolytically (random internal cleavage) or exolytically (terminal cleavage from the reducing end), classifying them as endo- or exo-acting enzymes, respectively [51]. Endo-acting enzymes feature an open active site conformation that accommodates non-crystalline polysaccharides, enabling random cleavage and oligosaccharide production. While the exo-acting enzymes typically feature a pocket or tunnel-shaped active site that specifically accommodates the substrate's reducing end, facilitating monosaccharide or disaccharide release [55]. GHs exhibit distinct structural diversity, featuring characteristic folds including $(\alpha/\alpha)_6$, β-helix, β-propeller, β-jelly roll, and the predominant $(\alpha/\beta)_8$ TIM barrel motif observed in most GH families within the CAZy database [53]. Glycoside hydrolase homologs frequently exhibit divergent substrate

specificities despite catalytic similarity, while structurally related enzymes may display markedly different activities. Glycoside hydrolases frequently display a modular structure, consisting of a catalytic module connected through linker peptides to one or more accessory modules known as carbohydrate-binding modules (CBMs). Accurate annotation of open reading frames (ORFs) and functional prediction of glycoside hydrolases require careful analysis of their modular architecture [55]. GHs are essential for the breakdown of structural polysaccharides such as cellulose, hemicellulose and chitin, particularly in microbial and fungal systems. Consequently, they contribute to numerous biological processes and industrial applications.

1.3.2 Mechanism of action of glycoside hydrolase

Koshland (1953) first proposed the enzymatic hydrolysis mechanism of glycosidic bonds, involving acid-base catalysis with two essential residues: a proton donor and a nucleophile/base [56]. Glycoside hydrolases typically mediate this hydrolysis *via* water addition across the bond, primarily utilizing Asp or Glu as catalytic residues, with rare exceptions. Glycoside hydrolases employ two principal types of catalytic mechanisms for substrate hydrolysis i.e. i) retaining and ii) inverting. These mechanisms, respectively mediate the retention or inversion of the anomeric configuration in carbohydrate substrates. The underlying principles of these catalytic mechanisms are detailed in subsequent sections.

1.3.2.1 Retaining mechanism

Glycoside hydrolases employing retaining mechanisms typically feature a nucleophile and catalytic acid/base residue pair separated by approximately, 5.5 Å [57]. The reaction follows a double displacement mechanism, wherein a nucleophile attacks the anomeric center immediately with general acid-catalyzed protonation of the

glycosidic oxygen. This process proceeds via an oxocarbenium ion-like transition state, generating a covalent glycosyl-enzyme intermediate while preserving anomeric carbon stereochemistry. The glycosyl-enzyme intermediate is hydrolyzed through general base-catalyzed water activation and subsequent nucleophilic attack at the anomeric center (**Fig. 1.6a**).

1.3.2.2 Inverting mechanism

Inverting glycoside hydrolases maintain greater spacing between catalytic acid/base residues to facilitate direct water nucleophilic attack on the substrate and occurs via single displacement mechanism [57]. Inverting reaction mechanisms generally involve a nucleophile and a catalytic acid/base residue pair positioned at an average distance of ~9.5 Å [57]. The general base deprotonates a water molecule, enabling nucleophilic attack at the anomeric center, while the general acid simultaneously protonates the glycosidic leaving group (**Fig. 1.6b**). Inverting mechanisms proceed through an oxocarbenium ion-like transition state, yielding net anomeric configuration inversion [57].

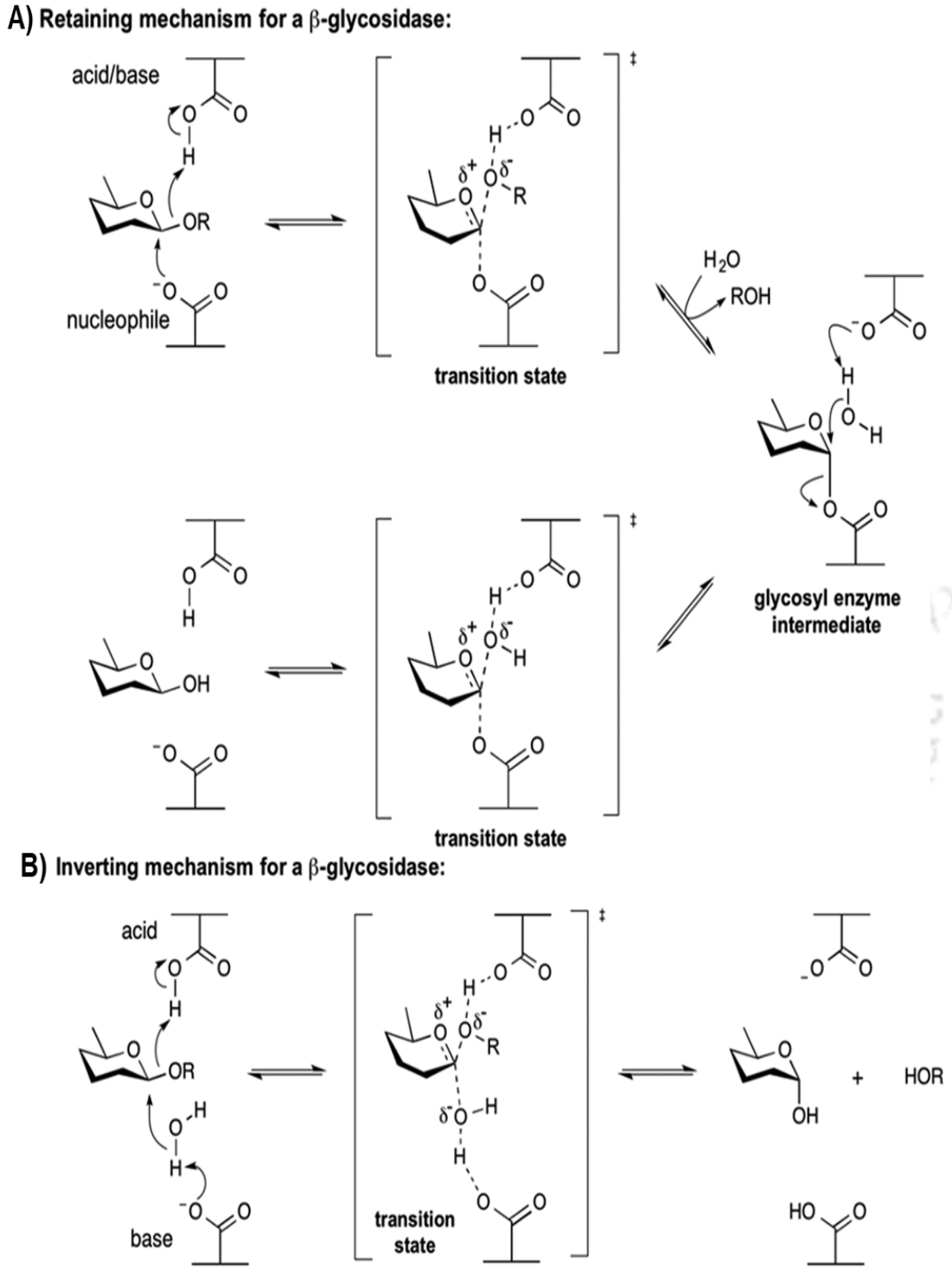


Fig. 1.6 Catalytic mechanisms of glycoside hydrolases. (a) Double-displacement (retaining) hydrolysis involving glycosyl-enzyme intermediate. (b) Single-displacement (inverting) hydrolysis *via* direct nucleophilic attack [57].

1.4 Structure of the cellulosome

Cellulolytic bacteria and fungi utilize diverse glycoside hydrolases collectively termed as cellulases to enzymatically depolymerize cellulose [1]. The cellulosome represents a supramolecular machinery in which different types of cellulases exhibit synergistic interactions to facilitate the efficient enzymatic hydrolysis of cellulose and hemicellulose [58]. The cellulosome was first characterized in *Acetivibrio thermocellus* in 1983 as a high-molecular-weight cellulolytic complex with molecular mass in range from about 40 kDa to 180 kDa [59]. The cellulosome is a modular multicomplex integrating diverse cellulases and associated enzymes through a specialized scaffolding subunit, scaffoldin, which mediates their structural organization [58]. Cellulosomes feature catalytic domains for hydrolysis and carbohydrate-binding modules (CBMs) in their modular structure. Scaffoldin mediates dual functions: cellulose binding through its carbohydrate-binding modules (CBMs) and enzymatic subunit incorporation via cohesin-dockerin interactions. Cellulosomal subunits feature modular architectures, where dockerin-bearing enzymes assemble onto cohesin-containing scaffoldin. Enzymatic subunits are covalently linked with the dockerin domain that binds non-covalently with the cohesin domains of the scaffoldin (**Fig. 1.7**) . A defining distinction between cellulosomal and noncellulosomal enzymes lies in the absence of dockerin domains in the latter [60].

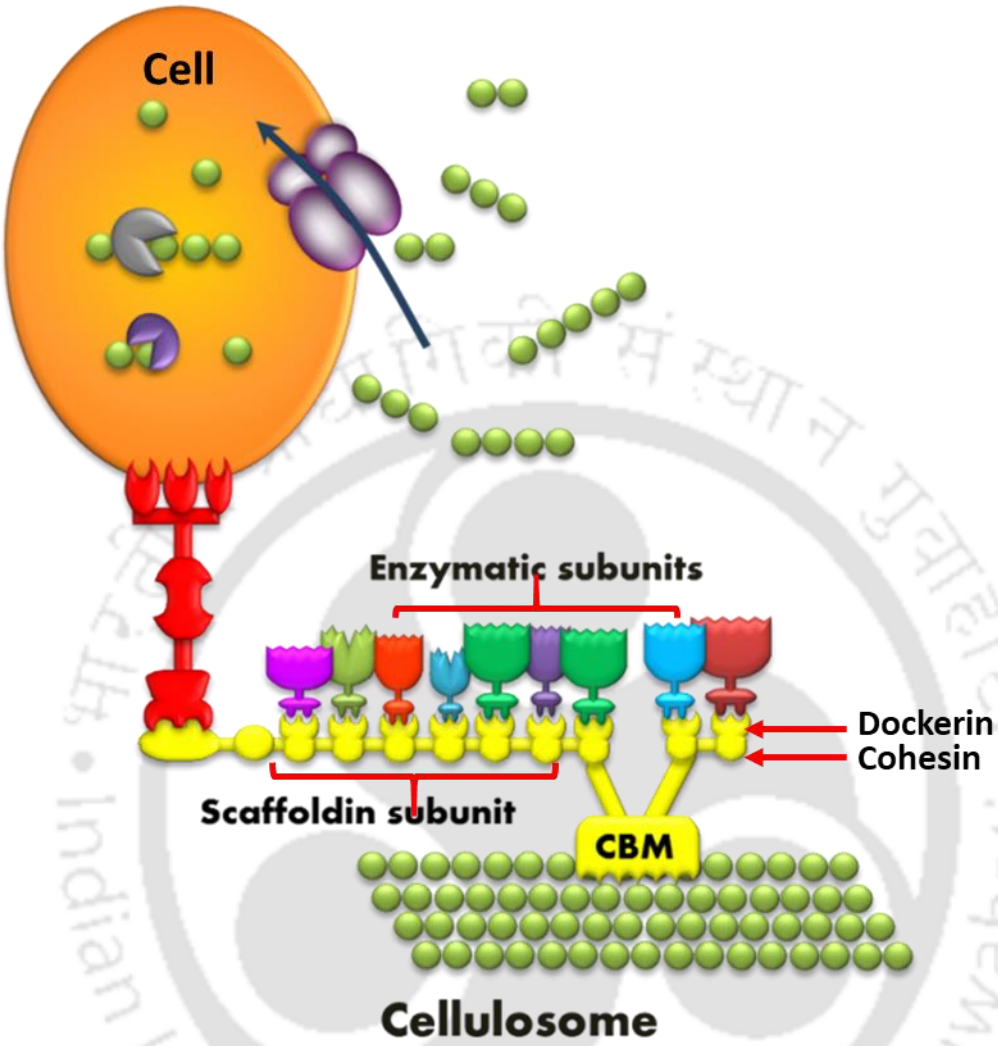


Fig. 1.7 Structural organization of a cellulosome: Enzymatic subunits bearing dockerin domains dock onto scaffoldin via cohesin-dockerin interactions, forming a multi-enzyme complex for efficient cellulose degradation [60].

1.5 Cellulolytic enzymes

Cellulolytic enzymes are glycoside hydrolases specialized for cellulosic polysaccharide degradation in plant cell walls. Cellulose hydrolysis is mediated through the coordinated activity of three key enzymes within the cellulosome complex: endoglucanase (EC 3.2.1.4), which randomly cleaves internal β-(1,4)-glycosidic bonds in amorphous cellulose regions; cellobiohydrolase, CBH (EC 3.2.1.91), which

processively liberates cellobiose units; and β-glucosidase, BGL (EC 3.2.1.21), which completes the degradation by converting oligosaccharides to cellobiose and cellobiose to glucose monomers, respectively [61] (**Fig. 1.8**). Endoglucanase (EC 3.2.1.4) catalyzes the initial depolymerization of cellulose by hydrolyzing internal β-1,4-glycosidic bonds in amorphous regions, thereby generating new reducing and non-reducing chain termini. Endoglucanases are currently classified across 15 glycoside hydrolase families (GH5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74, 124 and 148), existing either as single-domain proteins or multi-modular enzymes. Among 15 families, endoglucanases are predominantly distributed across 12 established GH families (GH5, 6, 7, 8, 9, 12, 44, 45, 48, 51, 74, 124), with sporadic activity reported in other families. No bona fide cellulose-active endoglucanases have been identified in GH families beyond GH148. Cellobiohydrolase (EC 3.2.1.91) exhibits processive exo-activity, sequentially releasing cellobiose from cellulose chain termini regardless of end polarity. CBHs primarily belong to five glycoside hydrolase families GH5, GH6, GH7, GH9 and GH48 [62]. The cellobiose produced through CBH-catalyzed hydrolysis serves as the substrate for β-glucosidases (EC 3.2.1.21), which complete cellulose degradation by converting cellobiose to glucose monomers. The catalytic activity of β-glucosidases is principally distributed among nine GH families (GH1-3, GH5, GH9, GH16, GH30, GH39 and GH116) [63].

Endoglucanases, as the primary initiators of cellulose hydrolysis, are distributed across multiple glycoside hydrolase families. Among these, GH9 represents the second most abundant family harboring endoglucanase activity. The reaction pattern and the structure analysis is described in detail in the following **section 1.6.**

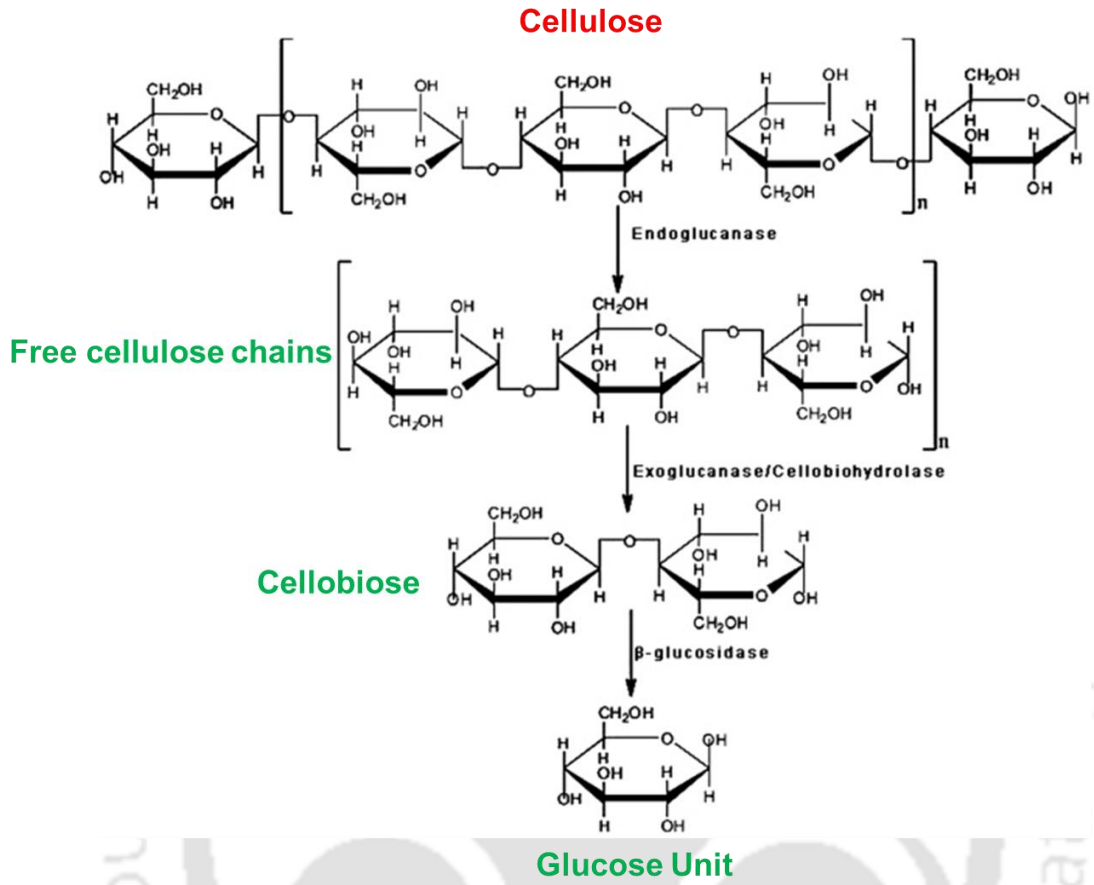


Fig. 1.8 Cellulolytic enzymes system. Endoglucanases (EGs) hydrolyze amorphous cellulose, generating free chain ends. Cellobiohydrolases (CBHs) processively degrade these chain termini, liberating cellobiose units, which are subsequently hydrolyzed by β-glucosidases (BGLs) to glucose.

1.6 Endoglucanases from GH9 family

The glycoside hydrolase family 9 (GH9), formerly designated as "Cellulase E," represents the second most largest cellulase family, predominantly consisting of endoglucanases, with a limited number exhibiting processive enzymatic activity [64]. Processive endoglucanases cleave internal β-1,4-glycosidic bonds in cellulose while remaining bound to the substrate for continuous hydrolysis, combining endo- and exo-like activity [65]. Processive endoglucanases in GH9 possess a C-terminal family 3C carbohydrate-binding module (CBM3c) and catalyze glycosidic bond cleavage via

anomeric inversion [66]. These enzymes exhibit high activity on swollen, non-crystalline cellulose but are inactive on crystalline cellulose [67]. Family 9 GH enzymes are classified into two primary subgroups: EGI, exclusively comprising bacterial cellulases (both aerobic and anaerobic) and EGII, which includes cellulases of non-bacterial origin [68]. GH9 endoglucanases are ubiquitously distributed across life domains, encompassing eubacteria, archaea, chordates and hemichordates and all known plant cellulases are classified within this family.

1.6.1 Reaction catalyzed by GH9 endoglucanases

The GH9 family endo-β-1,4-glucanases cleave internal β-1,4-glycosidic linkages in β-1,4-glucan polymers, which consist of glucose monomers connected by β-1,4 bonds [68]. Cavicchioli and Watson (1991) first reported a GH9 endoglucanase from *Fibrobacter succinogenes*, identified by functional screening of a cosmid library on CMC-containing medium [69]. The endo-β-1,4-glucanase activity is quantified by assessing the enzymatic hydrolysis of carboxymethyl cellulose and β-1,4-glucan oligomers ranging from cellobiose to cellohexose. However, true cellulolytic activity is specifically characterized by crystalline cellulose degradation. GH9 endo-β-1,4-glucanases of certain organisms, *viz. Gecarcoidea natalis, Cherax quadricarinatus, Aplysia kurodai, Euphausia pacifica* exhibit lichenase activity, hydrolyzing mixed-linkage β-glucans at rates comparable to or exceeding (up to 7-fold) those observed with carboxymethyl cellulose [70–72]. Hence, the enzymes can exhibit dual catalytic activity, functioning both as a lichenase and an endo-β-1,4-glucanase, where lichenan is an soluble glucose polymer containing mixed β-1,3 and β-1,4 glycosidic linkages. Certain GH9 endo-β-1,4-glucanases exhibit limited cellulose hydrolysis, displaying significantly higher activity toward carboxymethyl cellulose (CMC). These enzymes

typically operate optimally in the temperature range, 40-70°C and pH 4-8. The limited hydrolytic activity is attributed to preferential degradation of amorphous cellulose regions, with no detectable activity against crystalline part in certain GH9 endoglucanases. Based on k_{cat}/K_m values against preferred substrates, GH9 enzymes are principally characterized as endo-β-1,4-glucanases or lichenases with low cellulase activity.

1.6.2 Structural organization of GH9 endo-β-1,4-glucanases

The catalytic domain of GH9 endoglucanases adopts a characteristic double-layered α-helical barrel structure comprising twelve α-helices with six inner and six outer helices forming the $(\alpha/\alpha)_6$ scaffold. The enzyme features two catalytic aspartates within the conserved Asp-Ala-Asp-Gly motif and one glutamate in the Asn-Glu-Val-Ala sequence, employing an inverting mechanism that converts β- to α-anomers at the reducing end [73]. The catalytic mechanism involves two aspartates (Asp-Ala-Asp-Gly) functioning as a general base to activate a water nucleophile, which attacks the anomeric carbon with concomitant inversion of configuration, while a glutamate (Asn-Glu-Val-Ala) serves as a general acid to protonate the glycosidic oxygen [73]. Certain endoglucanases exhibit modular architecture, comprising both catalytic modules and cognate CBMs [52] as shown in **Fig. 1.9**. The active-site of endo-β-1,4-glucanase from *Thermobifida fusca* is capable of binding hexaglucoside substrate and the associated CBMs enhance crystalline cellulose hydrolysis efficiency [74]. In contrast to the bacterial counterparts, GH9 endo-β-1,4-glucanases from termites, *Nasutitermes takasagoensis* and earthworm *Eisenia fetida* lack CBMs and feature a more constrained active-site accommodating pentaglucoside substrates [75,76]. The CBM containing endo-β-1,4-glucanases exhibit an extended architecture, with mature proteins

comprising approximately, 500–900 amino acid residues and a molecular mass of ~45–100 kDa. GH9 endo-β-1,4-glucanases from invertebrates typically comprise 427–445 amino acids (~ 47–48 kDa) and contain an N-terminal 15–21 residue signal peptide, indicating their extracellular secretion into digestive fluids [64]. Homology modeling predicts that animal GH9 endo-β-1,4-glucanases share conserved structural features, including an $(\alpha/\alpha)_6$ barrel fold and catalytic residues, with an active site capable of binding pentaglucose substrate [64]. This structural configuration enables the accommodation of β-1,4-linked glucotriose/tetraose units, rationalizing the observed lichenase activity observed in GH9 endoglucanases.

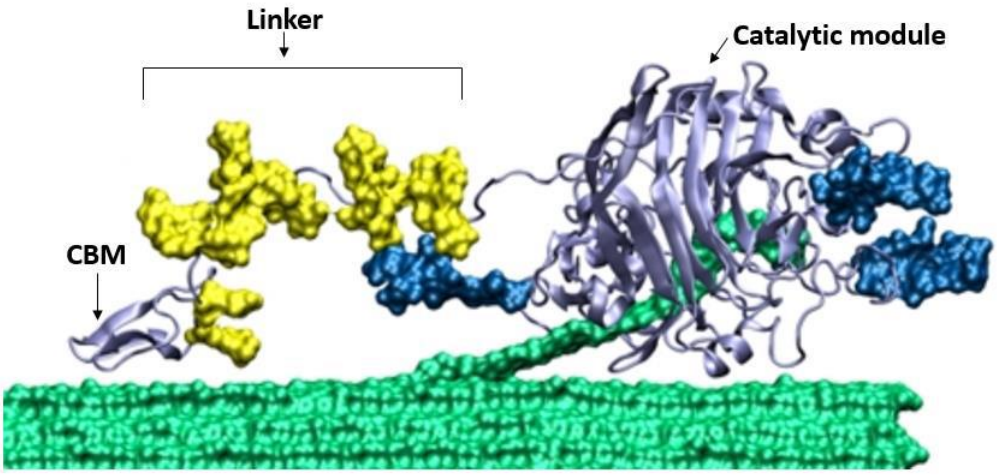


Fig. 1.9 Modular structure of endoglucanase depicting a catalytic module (CM) linked to a carbohydrate-binding module (CBM) via a flexible peptide linker [52].

1.7 Applications of the endoglucanases

Cellulases have widespread industrial applications, including textile processing, detergent formulation, pulp/paper production, animal feed additives/digestives and food processing, representing a major segment of commercial enzyme markets (**Fig. 1.10**). Among cellulases, endoglucanases significantly improve industrial processes

such as juice extraction, beer clarification, edible oil recovery from oilseeds, and nutritional improvement of baked goods and animal feedstuffs [77]. Endoglucanase-catalyzed cellulose hydrolysis constitutes an essential yet economically impactful stage in bioethanol production, accounting for approximately, 15-20% of total processing costs and significantly influencing final fuel pricing [78]. Endoglucanases enhance brewing efficiency by improving wort filtration (isolation of fermentable wort from husks and residues), reducing β-glucan precipitation and decreasing mash viscosity and turbidity [79]. They also facilitate paper pulp waste bioremediation and improve β-glucan digestibility in animal feed. Also in poultry and livestock nutrition, endoglucanases improve feed digestibility by hydrolyzing β-glucans in barley-based diets, thereby mitigating the anti-nutritional effects of mixed-linkage (1,3-1,4-β-D) glucans [80]. Additionally, endoglucanases find industrial applications in textile processing (denim biofinishing and cotton softening), detergent formulations (fabric cleaning and color maintenance) and paper manufacturing (fiber modification and drainage enhancement) [77]. Endoglucanases, either alone or in combination, serve critical roles in protoplast isolation from plants and fungi [81]. Their heterologous expression further enables engineered hybrid strain development for genetic modification studies [82]. The broad industrial utility of endoglucanases necessitates their requirements with diverse pH/temperature optima, stability profiles and substrate specificities. Thermostable variants are particularly advantageous, offering enhanced activity and stability at elevated temperatures to improve process efficiency and reduce operational costs [64]. Recent biotechnological advancements have enabled *in situ* cellulose hydrolysis through heterologous expression of thermostable endoglucanases directly in plant biomass, eliminating the requirement for exogenous enzyme

supplementation during processing [83]. Transgenic plant expression systems offer distinct advantages for endoglucanase production, including scalable and cost-effective synthesis, proper post-translational folding and immediate applicability. These heterologously expressed enzymes also enable the development of hybrid strains for genetic engineering applications. Enzymatic pretreatment with endoglucanase facilitates energy-efficient production of cellulose nanofibers (CNF), yielding films with enhanced optical clarity, improved barrier properties and decreased surface wettability relative to conventional CNF materials [84].

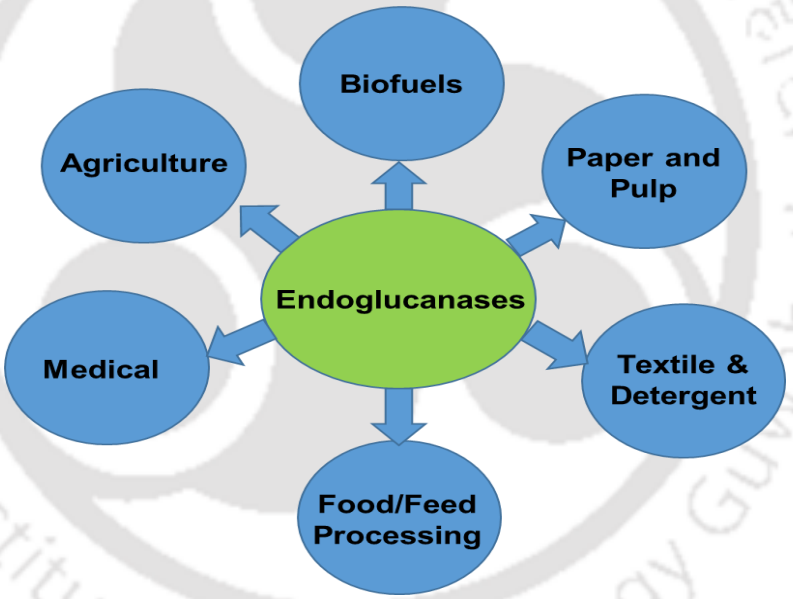


Fig. 1.10 Applications of endoglucanase in various industrial sectors.

1.8 Carbohydrate-binding modules

CBMs constitute independently folded protein modules that typically flank catalytic modules in carbohydrate-active enzymes, specializing in polysaccharide substrate targeting. Carbohydrate-binding modules (CBMs) are non-catalytic modules involved in that enhancement of the enzymatic activity by substrate targeting, catalytic

module stabilization and thermal stability improvement [85]. CBMs typically comprise ≤200 amino acid residues and occur as single or multiple domains (1-3 units) positioned at either terminus of catalytic modules. The first cellulose-binding domain was characterized in 1986 through proteolytic cleavage of cellulase, revealing distinct functional domains for catalysis and substrate binding [62]. The CAZy database (http://www.cazy.org/Carbohydrate-Binding-Modules.html) systematically categorizes CBMs into 106 phylogenetically distinct families using sequence-based classification criteria analogous to glycoside hydrolase catalytic modules.

1.8.1 Structural organization of binding sites of CBM

Carbohydrate-binding modules (CBMs) commonly exhibit β-jelly roll or β-sandwich folds and are systematically classified based on: (i) sequence homology, (ii) structural topology and (iii) functional properties. Current categorization identifies seven distinct folding families: β-sandwich (1) β-sandwich/β-jelly roll, (2) β-trefoil, (3) cysteine knot, (4) unique fold, (5) OB-fold, (6) hevein fold, (7) hevein-like/unique fold [85,86]. Carbohydrate-binding modules (CBMs) are subclassified into three functional types (Type A, Type B and Type C) according to their ligand interaction mechanisms and binding site structural features. Type A CBMs possess a flat binding interface, which mediates their affinity for crystalline polysaccharides (**Fig. 1.11**). Type B CBMs typically adopt a β-sandwich fold with binding clefts containing aromatic residues that mediate endo-type binding to soluble polysaccharides. Type C modules interact with terminal sugar residues (exo-type) as shown in **Fig. 1.11**. In addition, Type C CBMs exhibit lectin-like properties, featuring shallow binding pockets that mediate specific recognition of oligosaccharides (mono- to tri-saccharides) through combined polar and nonpolar interactions with soluble polysaccharide chains [85,86].

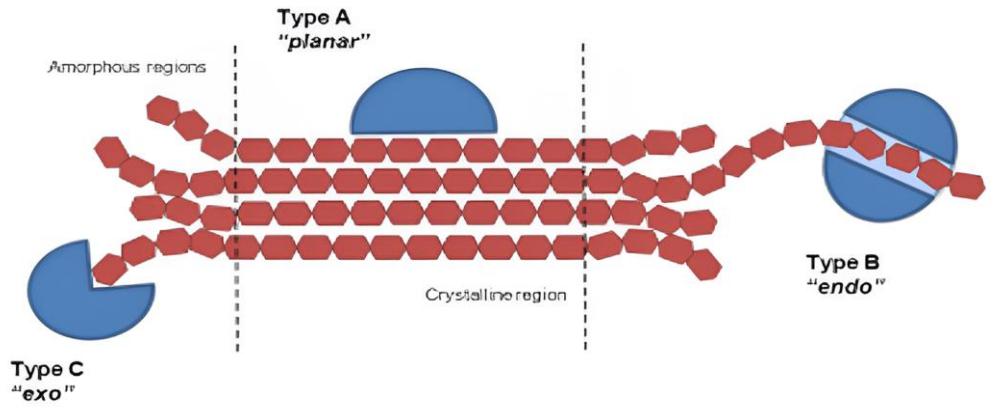


Fig. 1.11 Schematic diagam showing binding mode diversity across CBM families. Type A CBMs bind crystalline planar surfaces via hydrophobic stacking, Type B CBMs target single chains in clefts and Type C CBMs recognize terminal sugars in pockets.

1.8.2 Functions of Carbohydrate-binding modules

CBMs enhance enzymatic polysaccharide degradation by localizing enzymes onto substrate surfaces, increasing their effective concentration and promoting sustained substrate proximity [87]. CBMs fulfill three primary functional roles for their associated catalytic modules: proximity enhancement, substrate targeting and polysaccharide disruption [85]. CBMs enhance glycoside hydrolase efficiency through three mechanistic roles: (1) substrate localization by tethering enzymes to insoluble polysaccharide [88], (2) catalytic potentiation *via* substrate-binding clefts that augment enzymatic activity [89] and (3) bifunctional utility as cellulose-binding fusion tags for protein purification [90]. Experimental evidence demonstrates that CBM deletion or mutation significantly reduces or abolishes hydrolytic activity in the associated catalytic module [85]. CBMs mediate prolonged substrate proximity by anchoring catalytic modules to insoluble macromolecular structures like plant cell walls. This proximity effect is particularly evident with cellulose and xylan, maintains enzyme-

substrate association for enhanced degradation efficiency [85,91]. CBMs mediate enzyme-substrate proximity through two principal mechanisms: i) substrate targeting and ii) polysaccharide disruption, decribed in **sections 1.8.2.1 and 1.8.2.2**, respectively.

1.8.2.1 Substrate targeting of Carbohydrate-binding modules

CBMs mediate substrate targeting by specifically binding polysaccharides and positioning them near their cognate catalytic modules. Cellulose-specific CBMs exhibit divergent substrate recognition mechanisms: Type A modules target crystalline domains, while Type B modules bind amorphous regions. Type C CBMs bind terminal sugars *via* exo-type interactions and possess lectin-like shallow binding pockets for glycan recognition. Thus, distinct binding strategies for cellulose substructures by distinct CBMs are reflected [91]. The targeting effect directs hydrolytic activity to specific cellulose substructures, rather than simply mediating enzyme-substrate proximity [85].

1.8.2.2 Polysaccharide disruption by Carbohydrate-binding modules

Experimental observations demonstrate that some CBMs can physically disrupt cellulose fiber architecture, liberating microfibrillar fragments independent of hydrolytic function [92]. Studies indicate that CBMs can bind and structurally destabilize crystalline cellulose or chitin, facilitating the release of non-covalently bound fibers and exposing cleavage sites for hydrolytic enzymes. Additionally, CBMs inhibit flocculation of microcrystalline bacterial cellulose [88].

1.8.3 Multivalent CBMs in polysaccharide recognition

Glycoside hydrolases may incorporate multiple CBMs, which can either be adjacently positioned or spatially separated by intervening protein modules. These

enzymes frequently employ multiple CBMs to enhance substrate affinity and/or enable recognition of distinct polysaccharide regions or structural motifs [93]. Polysaccharides exhibit structural complexity through branching, diverse linkages and heterogeneous sugar residues. Multi-CBM systems enable recognition of these varied features, enhancing enzymatic degradation [94]. Sequential CBM action may occur, where initial binding exposes new substrate regions for secondary recognition [95]. Certain CBMs potentially crosslink adjacent polysaccharide chains, promoting inter-enzyme cooperation [88].

1.8.4 Carbohydrate-binding modules of Family 3

CBM3 represents a family of protein modules predominantly found in Gram-positive bacteria, characterized by a conserved domain of approximately, 150 amino acids. This family is further classified into three distinct subgroups: CBM3a, CBM3b and CBM3c. CBM3a (scaffoldin) and CBM3b (free enzymes) share similar primary structures and bind to cellulose. Their binding is entropically driven, involving hydrophobic interactions that release constrained water from the ligand [96,97]. CBM3b members lack crystalline cellulose binding capacity but function as helper modules, directing individual cellulose chains into adjacent catalytic modules for hydrolysis [98]. CBM3s bind chitin and xyloglucan with 500-fold lower affinity than crystalline cellulose, exhibiting enthalpically-driven soluble xyloglucan interactions that are entropically unfavourable [91].

The CBM3 crystal structure reveals a compact, prismatic β-sandwich domain composed of nine β-strands arranged in two antiparallel sheets, adopting a jelly roll topology [96,97]. The β-sheet exhibits a planar hydrophobic surface featuring five linearly arranged residues: His, Trp, Tyr and an Arg-Asp ion pair. The CBM3 molecule

displays two distinct cellulose-binding surfaces on opposing sides: a planar strip and a shallow groove, both containing conserved polar and aromatic residues [96,97].

1.8.5 Applications of CBMs

Recent years have witnessed growing implementation of CBMs in diverse biotechnological applications, accompanied by exponential growth in related scientific literature and patent applications. CBMs are ideal for diverse applications due to three key properties: (i) autonomous folding for functional independence, (ii) abundant, low-cost matrices with excellent properties and (iii) tunable binding for tailored applications [88]. Some of the applications of CBMs are described in subsequent sections.

1.8.5.1 Purification of biomolecules

CBMs are particularly valuable in bioprocessing, where large-scale recovery and purification of bioactive molecules remain key challenges. Their use in biospecific affinity chromatography has positioned them as a leading tool in immobilized ligand technology. CBMs serve as efficient purification tags, enabling low-cost isolation of bioactive peptides through fusion protein expression. Also, CBMs enable efficient plant-based production of fusion proteins by exploiting the cellulose-rich cell wall as an inherent purification matrix [88].

1.8.5.2 CBMs as molecular tools for diagnostic and research applications

CBMs were first employed analytically in cellulase-mediated pulp fiber surface assays [99]. Recent studies have revitalized this approach by engineering diverse CBM-based probes for plant cell wall polysaccharide detection. CBMs conjugated to bacterial-binding proteins (e.g. monoclonal antibodies) enable pathogen detection when immobilized on cellulosic matrices (e.g. cotton gauze), serving as microbial concentrators for rapid food safety testing [88].

1.8.5.3 Protein engineering of CBMs for diverse functional applications

Research demonstrates that targeted fusion or substitution of carbohydrate-binding modules (CBMs) with heterologous proteins enhances the enzymatic hydrolysis efficiency of these proteins. Further research implemented rational CBM modification by mutation to meet predetermined reaction criteria such as pH sensitivity, binding association, scaffolding activity, etc. [88]. Strategic incorporation of CBMs to foreign protein from the two *Clostridium* cellulosomes, *C. thermocellum* and *C. cellulolyticum,* facilitates cellulosome assembly, yielding chimeric complexes of CBMs and foreign protein with enhanced cellulolytic synergy on crystalline cellulose [100]. Surface roughening of fiber by isolated CBMs demonstrates their potential for polysaccharide structure modification in cellulose to produce nanocellulose [88].

1.9 The microorganism: *Acetivibrio thermocellus*

The most effective cellulase system known till date is produced by *Acetivibrio thermocellus* a.k.a *Clostridium thermocellum*. This ubiquitous, rod-shaped, anaerobic, thermophilic and Gram-positive bacterium exhibits the highest known rate of native cellulose degradation [101]. This bacterium can be isolated from decaying biomass under thermophilic or mesophilic conditions, including biogas plants and compost heaps. It produces the cellulosome (Molecular mass, > $2x10^6$ kDa and Diameter 18 nm), an elaborate extracellular multi-enzyme complex, among the most sophisticated systems in bacteria [102] as shown in **Fig. 1.12**.

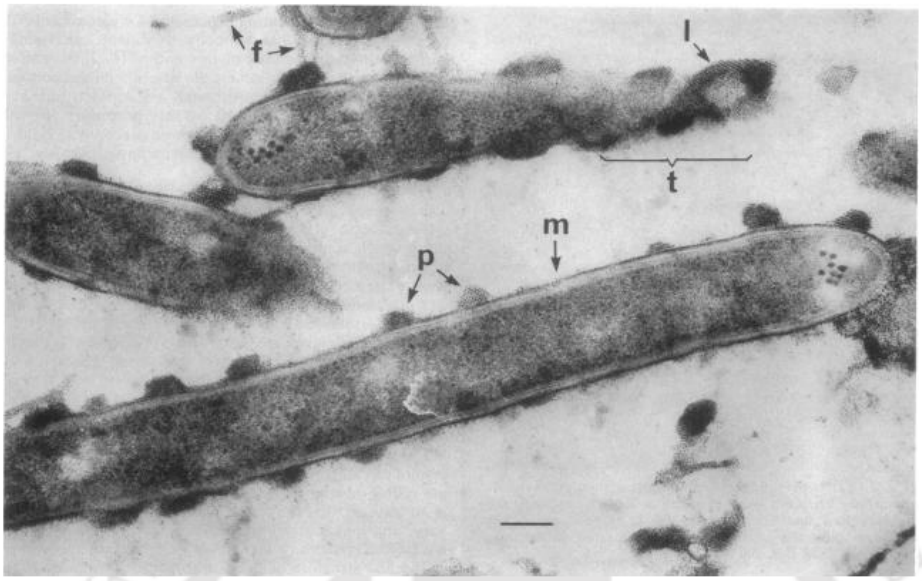


Fig. 1.12 TEM Ultrastructure of cellobiose-grown *A. thermocellus* highlighting: m: Monolayer organization; f: Intercellular bridging fibers; p: Cellulosome clusters, t: Tangent section area and I: Projections may be secondarily interconnected by basal structures at 200 nm scale [102].

Genome sequencing of by *Acetivibrio thermocellus* uncovered 74 open reading frames harboring dockerin domains, highlighting the extensive modularity of its cellulosome [103]. Cellulosomes contain a scaffoldin subunit (CipA) that lacks catalytic activity but exhibits variation in cohesin number (8–9) across different *A. thermocellus* strains. The scaffoldin protein contains nine type I cohesin domains that bind type I dockerins, linking catalytic enzyme domains *via* flexible connectors. The cellulosomal enzymes exhibit a modular architecture, with the scaffoldin CipA harboring a CBM3 domain that mediates complex binding to crystalline cellulose [102].

1.10 Research rationale and study objectives

1.10.1 Scientific relevance of the research

Fungal and bacterial endoglucanases are utilized in bioethanol production and diverse industries, from detergents to paper and pulp processing [83]. Endoglucanase-mediated hydrolysis yields eco-friendly, cost-effective products that serve as sustainable alternatives to non-renewable energy sources. Key challenges limiting industrial applications of endoglucanase include catalytic efficiency, pH and thermal stability and cost-effective production. Moreover, the mechanistic basis by which carbohydrate-binding modules (CBMs) modulate the catalytic efficiency of endoglucanase remains poorly understood. Studying efficient modular endoglucanases from thermophilic bacteria i.e. *Acetivibrio thermocellus* could provide key insights to overcome these limitations and enhance catalytic performance. Biochemical and structural findings from such research would improve enzymatic efficiency, advancing industrial feasibility. This study investigates the structure and functional analysis of the recombinant modular full-length endoglucanase *At*GH9C-CBM3A-CBM3B (GH9 family; GenBank ABN53959.1), with emphasis on elucidating the roles of its carbohydrate-binding modules (CBM3A and CBM3B) derived from *Acetivibrio thermocellus* ATCC 27405. The full-length enzyme, *At*GH9C-CBM3A-CBM3B and its truncated derivatives, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B will be cloned, expressed and purified by immobilized metal-ion affinity chromatography. The biochemical and functional properties of *At*GH9C-CBM3A-CBM3B and its truncated derivatives (*At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B) will be characterized to assess their enzymatic activity across diverse cellulosic substrates. This comparative analysis will elucidate the domain-specific contributions of CBM3A and CBM3B to

substrate binding affinity and catalytic efficiency. *In silico* structural analysis of *At*GH9C-CBM3A-CBM3B will identify key residues and elucidate the catalytic mechanism. Molecular docking of *At*GH9C-CBM3A-CBM3B will reveal active site architecture and ligand-binding interactions, while molecular dynamics simulations will assess folding stability. Solution structure and behavior will be characterized by SAXS and DLS across protein concentrations. Cellulose degradation by catalytic synergy between *At*GH9C-CBM3A-CBM3B and complementary cellulolytic enzymes: cellobiohydrolase (*At*GH5) and β-glucosidase (*At*GH1) will be evaluated. The enzymatic consortium (*At*GH9C-CBM3A-CBM3B/*At*GH5/*At*GH1) will generate nanocellulose (NC) via controlled cellulose hydrolysis, followed by morphological and physicochemical characterization. The NC will be immobilized in CMC-Na hydrogel for adsorptive removal studies of Congo red and methylene blue dyes. The scientific rationale behind investigating GH9 endoglucanase from *A. thermocellus* is given below:

1. *A. thermocellus*, a thermophilic anaerobe, produces hyperthermostable GH9 endoglucanases that retain activity under industrial biomass processing conditions (elevated temperatures, harsh solvents). This makes it a prime candidate for lignocellulosic biorefineries. The cellulosomal enzyme complex demonstrates 50-fold greater efficiency than the *Trichoderma reesei* system in crystalline substrate degradation [4].
2. The GH9-CBM fusion architecture in *A. thermocellus* is hypothesized to confer enhanced substrate affinity and processivity, contributing to synergistic module interactions due to:

- **CBMs**: Enhance binding to crystalline cellulose (e.g., Avicel, PASC) via surface-exposed aromatic residues.
- **Linker flexibility**: Optimizes spatial positioning of catalytic and binding modules for efficient hydrolysis.

3. GH9 endoglucanases act synergistically with exoglucanases and β-glucosidases in cellulosome of *A. thermocellus*, enabling near-complete cellulose depolymerization—a trait critical for efficient biofuel production.
4. GH9 enzymes show $(\alpha/\alpha)_6$ barrel fold and conserved catalytic residues (e.g., proton donor Glu, nucleophile Asp) differ from GH5/GH12 families, offering unique mechanistic perspectives on cellulose cleavage.
5. Controlled hydrolysis of cellulose by GH9 endoglucanases and exoglucanases is known to yield uniform nanocellulose, which can have diverse applications [104]. Subsequent immobilization of GH9-hydrolyzed nanocellulose to CMC-Na can form hydrogel, that will exhibit superior adsorption capacity for industrial dyes (e.g., Congo red) due to increased surface accessibility and modified fibril morphology.

1.10.2 Defined research objectives

1. Cloning, heterologous expression and subsequent purification of full-length modular endoglucanase *At*GH9C-CBM3A-CBM3B (GH9 family; GenBank ABN53959.1) and its truncated variants (*At*GH9C-CBM3A, *At*GH9C, CBM3A, and CBM3B) from *Acetivibrio thermocellus* ATCC 27405.
2. Biochemical and functional characterization of *At*GH9C-CBM3A-CBM3B and its truncation variants (*At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B) on diverse cellulose substrates with emphasis on elucidating CBM3A/CBM3B roles in substrate binding and catalytic efficiency, with their effect on the activity of catalytic *At*GH9C will be determined.
3. Determination of the structural architecture of *At*GH9C-CBM3A-CBM3B through integrated *in silico* approaches (homology modeling, molecular docking and MD simulations) and validation with solution-phase SAXS/DLS data.
4. Development of nanocellulose (NC) via an enzymatic consortium (*At*GH9C-CBM3A-CBM3B/*At*GH5) using restricted hydrolysis and evaluating its efficacy as a CMC-Na hydrogel composite for adsorptive removal of Congo red and methylene blue dyes.

1.11 References

[1] E.A. Bayer, H. Chanzy, R. Lamed, Y. Shoham, Cellulose, cellulases and cellulosomes, Curr opin struct biol 8 (1998) 548–557. https://doi.org/10.1016/S0959-440X(98)80143-7.

[2] J.C. Forbes, Plants in agriculture /, Cambridge University Press, Cambridge :, 1992.

[3] D.L., & C.M.M. Nelson, Lehninger principles of biochemistry, 7 th, 2017.

[4] C.M.G.A. Fontes, H.J. Gilbert, Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates, Annu Rev Biochem 79 (2010) 655–681. https://doi.org/10.1146/ANNUREV-BIOCHEM-091208-085603.

[5] A.R. Kermode, Plant Storage Products (Carbohydrates, Oils and Proteins), Encyclopedia of Life Sciences (2011). https://doi.org/10.1002/9780470015902.A0001325.PUB2.

[6] J.P. Flatt, Use and storage of carbohydrate and fat, American Journal of Clinical Nutrition 61 (1995). https://doi.org/10.1093/ajcn/61.4.952S.

[7] H.J. Gilbert, The Biochemistry and Structural Biology of Plant Cell Wall Deconstruction, Plant physiol 153 (2010) 444. https://doi.org/10.1104/PP.110.156646.

[8] F.A. Pettolino, C. Walsh, G.B. Fincher, A. Bacic, Determining the polysaccharide composition of plant cell walls, Nature Protocols 2012 7:9 7 (2012) 1590–1607. https://doi.org/10.1038/nprot.2012.081.

[9] P.J. Harris, B.A. Stone, Chemistry and Molecular Organization of Plant Cell Walls, Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy (2009) 61–93. https://doi.org/10.1002/9781444305418.CH4.

[10] C. Somerville, Cellulose synthesis in higher plants, Annu rev cell dev biol 22 (2006) 53–78. https://doi.org/10.1146/ANNUREV.CELLBIO.22.022206.160206.

[11] A.C. O'Sullivan, Cellulose: the structure slowly unravels, Cellulose 1997 4:3 4 (1997) 173–207. https://doi.org/10.1023/A:1018431705579.

[12] K. Keegstra, Plant cell walls, Plant Physiol 154 (2010) 483–486. https://doi.org/10.1104/PP.110.161240.

[13] C. Schadel, A. Blochl, A. Richter, G. Hoch, Short-term dynamics of nonstructural carbohydrates and hemicelluloses in young branches of temperate forest trees during bud break, Tree Physiol 29 (2009) 901–911. https://doi.org/10.1093/TREEPHYS/TPP034.

[14] D.J. Cosgrove, Enzymes and other agents that enhance cell wall extensibility, Annu rev plant physiol plant mol biol 50 (1999) 391–417. https://doi.org/10.1146/ANNUREV.ARPLANT.50.1.391.

[15] L.R.S. Moreira, E.X.F. Filho, An overview of mannan structure and mannan-degrading enzyme systems, Appl Microbiol Biotechnol 79 (2008) 165–178. https://doi.org/10.1007/S00253-008-1423-4.

[16] A.D. French, Glucose, not cellobiose, is the repeating unit of cellulose and why that is important, Cellulose 24 (2017) 4605–4609. https://doi.org/10.1007/S10570-017-1450-3/FIGURES/3.

[17] H. Seddiqi, E. Oliaei, H. Honarkar, J. Jin, L.C. Geonzon, R.G. Bacabac, J. Klein-Nulend, Cellulose and its derivatives: towards biomedical applications, Cellulose 2021 28:4 28 (2021) 1893–1931. https://doi.org/10.1007/S10570-020-03674-W.

[18] H. Seddiqi, E. Oliaei, H. Honarkar, J. Jin, L.C. Geonzon, R.G. Bacabac, J. Klein-Nulend, Cellulose and its derivatives: towards biomedical applications, Cellulose 2021 28:4 28 (2021) 1893–1931. https://doi.org/10.1007/S10570-020-03674-W.

[19] R.H. Marchessault, P.R. Sundararajan, Cellulose, (1983). https://doi.org/10.1016/B978-0-12-065602-8.50007-8.

[20] K. Okuda, S. Sekida, S. Yoshinaga, Y. Suetomo, Cellulose-synthesizing complexes in some chromophyte algae, Cellulose 2004 11:3 11 (2004) 365–376. https://doi.org/10.1023/B:CELL.0000046407.22865.03.

[21] J.E. Bidlack, W. V. Dashek, Plant cell walls, plant cells and their organelles (2017) 209–238. https://doi.org/10.1002/9781118924846.CH9.

[22] E. Macchi, A. Palma, Morphological studies on precipitated cellulose, Die Makromolekulare Chemie 123 (1969) 286–288. https://doi.org/10.1002/MACP.1969.021230127.

[23] Y. Matsuda, K. Kowsaka, K. Okajima, K. Kamide, Structural change of cellulose contained in immature cotton boll during its growth, Polym Int 27 (1992) 347–351. https://doi.org/10.1002/PI.4990270410.

[24] M. Wada, M. Ike, K. Tokuyasu, Enzymatic hydrolysis of cellulose I is greatly accelerated via its conversion to the cellulose II hydrate form, Polym Degrad Stab 95 (2010) 543–548. https://doi.org/10.1016/J.POLYMDEGRADSTAB.2009.12.014.

[25] A.D. French, The crystal structure of native ramie cellulose, Carbohydr Res 61 (1978) 67–80. https://doi.org/10.1016/S0008-6215(00)84467-5.

[26] W.E. Davis, A.J. King, A.J. Barry, F.C. Peterson, X-ray studies of reactions of cellulose in non-aqueous systems. Ii. Interaction of cellulose and primary amines, J Am Chem Soc 65 (1943) 1294–1299. https://doi.org/10.1021/JA01247A012/ASSET/JA01247A012.FP.PNG_V03.

[27] A.J. Brown, XLIII.—On an acetic ferment which forms cellulose, Journal of the Chemical Society, Transactions 49 (1886) 432–439. https://doi.org/10.1039/CT8864900432.

[28] S. Gorgieva, J. Trček, Bacterial Cellulose: Production, Modification and Perspectives in Biomedical Applications, Nanomaterials 2019, Vol. 9, Page 1352 9 (2019) 1352. https://doi.org/10.3390/NANO9101352.

[29] S. Chen, Y. Zou, Z. Yan, W. Shen, S. Shi, X. Zhang, H. Wang, Carboxymethylated-bacterial cellulose for copper and lead ion removal, J Hazard Mater 161 (2009) 1355–1359. https://doi.org/10.1016/J.JHAZMAT.2008.04.098.

[30] B. Machado, S.M. Costa, I. Costa, R. Fangueiro, D.P. Ferreira, The potential of algae as a source of cellulose and its derivatives for biomedical applications, Cellulose 31 (2024) 3353–3376. https://doi.org/10.1007/S10570-024-05816-W/METRICS.

[31] H. Rabemanolontsoa, S. Saka, Comparative study on chemical composition of various biomass species, RSC Adv 3 (2013) 3946–3956. https://doi.org/10.1039/C3RA22958K.

[32] J. Chen, J. Xu, K. Wang, X. Cao, R. Sun, Cellulose acetate fibers prepared from different raw materials with rapid synthesis method, Carbohydrate Polymer 137 (2016) 685–692. https://doi.org/10.1016/J.CARBPOL.2015.11.034.

[33] K. Hua, I. Rocha, P. Zhang, S. Gustafsson, Y. Ning, M. Strømme, A. Mihranyan, N. Ferraz, Transition from bioinert to bioactive material by tailoring the biological cell response to carboxylated nanocellulose, Biomacromolecules 17 (2016) 1224–1233.

https://doi.org/10.1021/ACS.BIOMAC.6B00053/ASSET/IMAGES/LARGE/BM-2016-00053T_0010.JPEG.

[34] D. Trache, M.H. Hussin, M.K.M. Haafiz, V.K. Thakur, Recent progress in cellulose nanocrystals: sources and production, Nanoscale 9 (2017) 1763–1786. https://doi.org/10.1039/C6NR09494E.

[35] O.A. Battista, P.A. Smith, Microcrystalline cellulose, Ind Eng Chem 54 (1962) 20–29. https://doi.org/10.1021/IE50633A003/ASSET/IE50633A003.FP.PNG_V03.

[36] V. Schuh, K. Allard, K. Herrmann, M. Gibis, R. Kohlus, J. Weiss, Impact of carboxymethyl cellulose (CMC) and microcrystalline cellulose (MCC) on functional characteristics of emulsified sausages, Meat Sci 93 (2013) 240–247. https://doi.org/10.1016/J.MEATSCI.2012.08.025.

[37] A. Pei, Q. Zhou, L.A. Berglund, Functionalized cellulose nanocrystals as biobased nucleation agents in poly(l-lactide) (PLLA) – Crystallization and mechanical property effects, Composcitech Scientific Technology 70 (2010) 815–821. https://doi.org/10.1016/J.COMPSCITECH.2010.01.018.

[38] N.L.V. Carreño, A.M. Barbosa, B.S. Noremberg, M.M.S. Salas, S.C.M. Fernandes, J. Labidi, Advances in Nanostructured cellulose-based biomaterials, Springer Briefs In Applied Sciences and Technology (2017) 1–32. https://doi.org/10.1007/978-3-319-58158-3_1.

[39] J.H. Kim, B.S. Shim, H.S. Kim, Y.J. Lee, S.K. Min, D. Jang, Z. Abas, J. Kim, Review of nanocellulose for sustainable future materials, International Journal of Precision Engineering and Manufacturing - Green Technology 2 (2015) 197–213. https://doi.org/10.1007/S40684-015-0024-9/METRICS.

[40] R.A. Ilyas, S.M. Sapuan, M.R. Ishak, Isolation and characterization of nanocrystalline cellulose from sugar palm fibres (*Arenga Pinnata*), Carbohydr Polym 181 (2018) 1038–1051. https://doi.org/10.1016/J.CARBPOL.2017.11.045.

[41] M. Jalili Tabaii, G. Emtiazi, Transparent nontoxic antibacterial wound dressing based on silver nano particle/bacterial cellulose nano composite synthesized in the presence of tripolyphosphate, J Drug Deliv Sci Technol 44 (2018) 244–253. https://doi.org/10.1016/J.JDDST.2017.12.019.

[42] R.J. Moon, A. Martini, J. Nairn, J. Simonsen, J. Youngblood, Cellulose nanomaterials review: structure, properties and nanocomposites, Chem Soc Rev 40 (2011) 3941–3994. https://doi.org/10.1039/C0CS00108B.

[43] H.V. Scheller, P. Ulvskov, Hemicelluloses, Annu Rev Plant Biol 61 (2010) 263–289. https://doi.org/10.1146/ANNUREV-ARPLANT-042809-112315/CITE/REFWORKS.

[44] C. Schadel, A. Blochl, A. Richter, G. Hoch, Short-term dynamics of nonstructural carbohydrates and hemicelluloses in young branches of temperate forest trees during bud break, Tree Physiol 29 (2009) 901–911. https://doi.org/10.1093/TREEPHYS/TPP034.

[45] A.G. Cunha, A. Gandini, Turning polysaccharides into hydrophobic materials: A critical review. Part 1. Cellulose, Cellulose 17 (2010) 875–889. https://doi.org/10.1007/S10570-010-9434-6/TABLES/4.

[46] S. Persson, K.H. Caffall, G. Freshour, M.T. Hilley, S. Bauer, P. Poindexter, M.G. Hahn, D. Mohnen, C. Somerville, The arabidopsis irregular xylem8 mutant is deficient in glucuronoxylan and homogalacturonan, Which Are Essential for Secondary Cell Wall Integrity, Plant Cell 19 (2007) 237–255. https://doi.org/10.1105/TPC.106.047720.

[47] M.C.N. Picot-Allain, B. Ramasawmy, M.N. Emmambux, Extraction, characterisation, and application of pectin from tropical and sub-tropical fruits: Annual Review, Food Reviews International 38 (2022) 282–312. https://doi.org/10.1080/87559129.2020.1733008.

[48] B.L. Ridley, M.A. O'Neill, D. Mohnen, Pectins: structure, biosynthesis, and oligogalacturonide-related signaling, Phytochemistry 57 (2001) 929–967. https://doi.org/10.1016/S0031-9422(01)00113-3.

[49] B.I. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics, Nucleic Acids Research 37 (2009). https://doi.org/10.1093/NAR/GKN663,.

[50] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res 42 (2014). https://doi.org/10.1093/NAR/GKT1178,.

[51] V. Lombard, B. Henrissat, M.L. Garron, CAZac: an activity descriptor for carbohydrate-active enzymes, Nucleic Acids Research 53 (2025) D625–D633. https://doi.org/10.1093/NAR/GKAE1045.

[52] C.B. Taylor, M.F. Talib, C. McCabe, L. Bu, W.S. Adney, M.E. Himmel, M.F. Crowley, G.T. Beckham, Computational investigation of glycosylation effects

on a family 1 carbohydrate-binding module, Journal of Biological Chemistry, 287 (2011) 3147. https://doi.org/10.1074/JBC.M111.270389.

[53] B.I. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics, Nucleic Acids Research 37 (2008) D233. https://doi.org/10.1093/NAR/GKN663.

[54] B. Henrissat, A classification of glycosyl hydrolases based on amino acid sequence similarities, Biochemical Journal 280 (1991) 309–316. https://doi.org/10.1042/BJ2800309,.

[55] G. Davies, B. Henrissat, Structures and mechanisms of glycosyl hydrolases, Structure Cell Press (1995) 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9.

[56] D.E. Koshland, Stereochemistry and the mechanism of enzymatic reactions, Biological Reviews 28 (1953) 416–436. https://doi.org/10.1111/J.1469-185X.1953.TB01386.X;WGROUP:STRING:PUBLICATION.

[57] D.L. Zechel, S.G. Withers, Glycosidase mechanisms: anatomy of a finely tuned catalyst, Acc Chem Res 33 (2000) 11–18. https://doi.org/10.1021/AR970172+.

[58] E.A. Bayer, B. Henrissat, R. Lamed, The Cellulosome: A natural bacterial strategy to combat biomass recalcitrance, biomass recalcitrance: deconstructing the plant cell wall for bioenergy (2009) 407–435. https://doi.org/10.1002/9781444305418.ch13;page:string:article/chapter.

[59] R. Lamed, E.A. Bayer, The Cellulosome of *Clostridium thermocellum*, Adv Appl Microbiol 33 (1988) 1–46. https://doi.org/10.1016/S0065-2164(08)70203-X.

[60] E.A. Bayer, Y. Shoham, R. Lamed, Cellulose-decomposing bacteria and their enzyme systems, the prokaryotes (2006) 578–617. https://doi.org/10.1007/0-387-30742-7_19.

[61] D.B. Wilson, D.C. Irwin, Genetics and Properties of Cellulases, (1999) 1–21. https://doi.org/10.1007/3-540-49194-5_1.

[62] C. Divne, J. Ståhlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J.K.C. Knowles, T.T. Teeri, T.A. Jones, The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*, Science (1979) 265 (1994) 524–528. https://doi.org/10.1126/SCIENCE.8036495,.

[63] B. Henrissat, H. Driguez, C. Viet, M. Schülein, Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose, Bio/Technology 3 (1985) 722–726. https://doi.org/10.1038/NBT0885-722.

[64] S.M. Linton, Review: The structure and function of cellulase (endo-β-1,4-glucanase) and hemicellulase (β-1,3-glucanase and endo-β-1,4-mannase) enzymes in invertebrates that consume materials ranging from microbes, algae to leaf litter, Comp Biochem Physiol B Biochem Mol Biol 240 (2020). https://doi.org/10.1016/j.cbpb.2019.110354.

[65] S. Wu, S. Wu, Processivity and the mechanisms of processive endoglucanases, Applied Biochemistry Biotechnology 190 (2020) 448–463. https://doi.org/10.1007/S12010-019-03096-W,.

[66] S. Wu, S. Wu, Processivity and the Mechanisms of Processive Endoglucanases, Applied Biochemistry Biotechnology 190 (2020) 448–463. https://doi.org/10.1007/S12010-019-03096-W/FIGURES/4.

[67] T.A. Sathya, M. Khan, Diversity of glycosyl hydrolase enzymes from metagenome and their application in food industry, J Food Sci 79 (2014) R2149–R2156. https://doi.org/10.1111/1750-3841.12677;SUBPAGE:STRING:FULL.

[68] P. Tomme, R.A.J. Warren, N.R. Gilkes, Cellulose hydrolysis by bacteria and fungi, Adv Microb Physiol 37 (1995) 1–81. https://doi.org/10.1016/S0065-2911(08)60143-5.

[69] R. Cavicchioli, K. Watson, Molecular cloning, expression, and characterization of endoglucanase genes from *Fibrobacter succinogenes* AR1, Appl Environ Microbiol 57 (1991) 359–365. https://doi.org/10.1128/AEM.57.2.359-365.1991;CTYPE:STRING:JOURNAL.

[70] X.M. Xue, A.J. Anderson, N.A. Richardson, A.J. Anderson, G.P. Xue, P.B. Mather, Characterisation of cellulase activity in the digestive system of the redclaw crayfish (*Cherax quadricarinatus*), Aquaculture 180 (1999) 373–386. https://doi.org/10.1016/S0044-8486(99)00213-6.

[71] S.M. Linton, A.J. Shirley, Isozymes from the herbivorous gecarcinid land crab, *Gecarcoidea natalis* that possess both lichenase and endo-β-1,4-glucanase activity, Comp Biochem Physiol B Biochem Mol Biol 160 (2011) 44–53. https://doi.org/10.1016/J.CBPB.2011.05.007.

[72] A. Tsuji, S. Sato, A. Kondo, K. Tominaga, K. Yuasa, Purification and characterization of cellulase from North Pacific krill (*Euphausia pacifica*). Analysis of cleavage specificity of the enzyme, Comp Biochem Physiol B

Biochem Mol Biol 163 (2012) 324–333. https://doi.org/10.1016/J.CBPB.2012.08.005.

[73] J.D. McCarter, G. Stephen Withers, Mechanisms of enzymatic glycoside hydrolysis, Curr Opin Struct Biol 4 (1994) 885–892. https://doi.org/10.1016/0959-440X(94)90271-2.

[74] J. Sakon, D. Irwin, D.B. Wilson, P. Andrew Karplus, Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*, Nat Struct Biol 4 (1997) 810–818. https://doi.org/10.1038/nsb1097-810.

[75] S. Khademi, L.A. Guarino, H. Watanabe, G. Tokuda, E.F. Meyer, Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*, Acta Crystallogr D Biol Crystallogr 58 (2002) 653–659. https://doi.org/10.1107/S0907444902002366,.

[76] T. Arimori, A. Ito, M. Nakazawa, M. Ueda, T. Tamada, Crystal structure of endo-1,4-β-glucanase from *Eisenia fetida*, Urn:Issn:0909-0495 20 (2013) 884–889. https://doi.org/10.1107/S0909049513021110.

[77] M.K. Bhat, Cellulases and related enzymes in biotechnology, Biotechnol Adv 18 (2000) 355–383. https://doi.org/10.1016/S0734-9750(00)00041-0.

[78] A.J. Ragauskas, C.K. Williams, B.H. Davison, G. Britovsek, J. Cairney, C.A. Eckert, W.J. Frederick, J.P. Hallett, D.J. Leak, C.L. Liotta, J.R. Mielenz, R. Murphy, R. Templer, T. Tschaplinski, The path forward for biofuels and biomaterials, Science (1979) 311 (2006) 484–489. https://doi.org/10.1126/SCIENCE.1114736,.

[79] K.R.S. Celestino, R.B. Cunha, C.R. Felix, Characterization of a β-glucanase produced by Rhizopus microsporus var. microsporus, and its potential for application in the brewing industry, BMC Biochem 7 (2006) 1–9. https://doi.org/10.1186/1471-2091-7-23/FIGURES/6.

[80] M. Choct, Enzyme supplementation of poultry diets based on viscous cereals., Enzymes in Farm Animal Nutrition (2001) 145–160. https://doi.org/10.1079/9780851993935.0145.

[81] R.C. Kuhad, R. Gupta, A. Singh, Microbial Cellulases and Their Industrial Applications, Enzyme Res 2011 (2011) 280696. https://doi.org/10.4061/2011/280696.

[82] M. Mandels, Applications of cellulases., Biochem Soc Trans 13 (1985) 414–416. https://doi.org/10.1042/BST0130414,.

[83] F. Akram, I. ul Haq, W. Imran, H. Mukhtar, Insight perspectives of thermostable endoglucanases for bioethanol production: A review, Renew Energy 122 (2018) 225–238. https://doi.org/10.1016/J.RENENE.2018.01.095.

[84] B. Las-Casas, V. Arantes, Endoglucanase pretreatment aids in isolating tailored-cellulose nanofibrils combining energy saving and high-performance packaging, International Journal of Biological Macromolecules 242 (2023) 125057. https://doi.org/10.1016/J.IJBIOMAC.2023.125057.

[85] A.B. Boraston, D.N. Bolam, H.J. Gilbert, G.J. Davies, Carbohydrate-binding modules: Fine-tuning polysaccharide recognition, Biochemical Journal 382 (2004) 769–781. https://doi.org/10.1042/BJ20040892,.

[86] H.J. Gilbert, J.P. Knox, A.B. Boraston, Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules, Curr Opin Struct Biol 23 (2013) 669–677. https://doi.org/10.1016/j.sbi.2013.05.005.

[87] D.N. Bolam, A. Ciruela, S. McQueen-Mason, P. Simpson, M.P. Williamson, J.E. Rixon, A. Boraston, G.P. Hazlewood, H.J. Gilbert, Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity, Biochemical Journal 331 (1998) 775. https://doi.org/10.1042/BJ3310775.

[88] O. Shoseyov, Z. Shani, I. Levy, Carbohydrate Binding Modules: Biochemical properties and novel applications, Microbiology and Molecular Biology Reviews 70 (2006) 283–295. https://doi.org/10.1128/MMBR.00028-05/ASSET/91C2FA77-D7C0-4BAD-AF52-723FB083BADC/ASSETS/GRAPHIC/ZMR0020621210001.JPEG.

[89] J.L. Henshaw, D.N. Bolam, V.M.R. Pires, M. Czjzek, B. Henrissat, L.M.A. Ferreira, C.M.G.A. Fontes, H.J. Gilbert, The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities, Journal of Biological Chemistry 279 (2004) 21552–21559. https://doi.org/10.1074/jbc.M401620200.

[90] E. Shpigel, D. Elias, I.R. Cohen, O. Shoseyov, Production and Purification of a Recombinant Human hsp60 Epitope Using the Cellulose-Binding Domain inEscherichia coli, Protein Exprression and Purification 14 (1998) 185–191. https://doi.org/10.1006/PREP.1998.0929.

[91] P. Tomme, A. Boraston, B. McLean, J. Kormos, A.L. Creagh, K. Sturch, N.R. Gilkes, C.A. Haynes, R.A.J. Warren, D.G. Kilburn, Characterization and affinity applications of cellulose-binding domains, J Chromatogr B Biomed Appl 715 (1998) 283–296. https://doi.org/10.1016/S0378-4347(98)00053-X.

[92] I. Lee, B.R. Evans, J. Woodward, The mechanism of cellulase action on cotton fibers: evidence from atomic force microscopy, Ultramicroscopy 82 (2000) 213–221. https://doi.org/10.1016/S0304-3991(99)00158-8.

[93] C.J. Duan, Y.L. Feng, Q.L. Cao, M.Y. Huang, J.X. Feng, Identification of a novel family of carbohydrate-binding modules with broad ligand specificity, Sci Rep 6 (2016) 1–8. https://doi.org/10.1038/srep19392;subjmeta=1164,2522,326,45,607,631;kwrd=applied+microbiology,hydrolases.

[94] J. Liu, D. Sun, J. Zhu, C. Liu, W. Liu, Carbohydrate-binding modules targeting branched polysaccharides: overcoming side-chain recalcitrance in a non-catalytic approach, Bioresour Bioprocess 8 (2021) 1–11. https://doi.org/10.1186/S40643-021-00381-7/METRICS.

[95] M.S. Møller, S. El Bouaballati, B. Henrissat, B. Svensson, Functional diversity of three tandem C-terminal carbohydrate-binding modules of a β-mannanase, J Biol Chem 296 (2021) 100638. https://doi.org/10.1016/J.JBC.2021.100638.

[96] J. Tormo, R. Lamed, A.J. Chirino, E. Morag, E.A. Bayer, Y. Shoham, T.A. Steitz, Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose, EMBO J 15 (1996) 5739. https://doi.org/10.1002/j.1460-2075.1996.tb00960.x.

[97] L.J.W. Shimon, S. Pagès, A. Belaich, J.P. Belaich, E.A. Bayer, R. Lamed, Y. Shoham, F. Frolow, Structure of a family IIIa scaffoldin CBD from the cellulosome of Clostridium cellulolyticum at 2.2 Å resolution, Acta Crystallogr D Biol Crystallogr 56 (2000) 1560–1568. https://doi.org/10.1107/S0907444900012889,.

[98] D. Irwin, D.H. Shin, S. Zhang, B.K. Barr, J. Sakon, P.A. Karplus, D.B. Wilson, Roles of the catalytic domain and two cellulose binding domains of Thermomonospora fusca E4 in cellulose hydrolysis, J Bacteriol 180 (1998) 1709–1714. https://doi.org/10.1128/JB.180.7.1709-1714.1998,.

[99] J.-L. Yang, B. Pettersson, K.-E. Eriksson, Development of bioassays for the characterization of pulp fiber surfaces: I. Characterization of various mechanical pulp fiber surfaces by specific cellulolytic enzymes, Nord Pulp Paper Res J 3 (1988) 19–24. https://doi.org/10.3183/NPPRJ-1988-03-01-P019-024.

[100] H.P. Fierobe, A. Mechaly, C. Tardif, A. Belaich, R. Lamed, Y. Shoham, J.P. Belaich, E.A. Bayer, Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional

complexes, Journal of Biological Chemistry 276 (2001) 21257–21261. https://doi.org/10.1074/JBC.M102082200,.

[101] V. V. Zverlov, W.H. Schwarz, Bacterial cellulose hydrolysis in anaerobic environmental subsystems - Clostridium thermocellum and Clostridium stercorarium, thermophilic plant-fiber degraders, Ann N Y Acad Sci 1125 (2008) 298–307. https://doi.org/10.1196/ANNALS.1419.008;SUBPAGE:STRING:FULL.

[102] E.A. Bayer, R. Lamed, Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose, J Bacteriol 167 (1986) 828. https://doi.org/10.1128/JB.167.3.828-836.1986.

[103] V. V. Zverlov, J. Kellermann, W.H. Schwarz, Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: Identification of the major catalytic components in the extracellular complex and detection of three new enzymes, Proteomics 5 (2005) 3646–3653. https://doi.org/10.1002/PMIC.200401199;PAGE:STRING:ARTICLE/CHAPTER.

[104] R.S.A. Ribeiro, B.C. Pohlmann, V. Calado, N. Bojorge, N. Pereira, Production of nanocellulose by enzymatic hydrolysis: Trends and challenges, Eng Life Sci 19 (2019) 279. https://doi.org/10.1002/ELSC.201800158.

Chapter 2

Cloning and expression of β-1,4 endoglucanase, *At*GH9C-CBM3A-CBM3B and its truncated derivatives from *Acetivibrio thermocellus* ATCC 27405

2.1 Introduction

Lignocellulose biomass mainly comes from the plant cell residues and composed of semi-rigid and complex backbone of primarily cellulose, hemicellulose and lignin [1]. Cellulose is the most abundant homogenous polysaccharide on earth, consisting of repeating β-1,4 linked glucose units that hold substantial carbon energy sources with properties like natural, renewable, environmentally gentle, cost-efficient, non-toxic, biodegradable and biocompatible [2]. It finds use in wide range of industries, including textiles, drug delivery systems, composite materials, nanocellulose and personal care items. However, the implementation of cellulose in a bio-based economy faces bottlenecks owing to the ineffective hydrolysis of plant cell polysaccharides due to its highly recalcitrant nature [3]. The cellulose can be broken down by the conserted action of enzymes called cellulase and it is well documented to be present in in free

from or as multi-enzyme complex termed as cellulosome various bacteria and fungi [4]. These cellulases are broadly demarcated into three major groups (i) endoglucanase (EC 3.2.1.4) which randomly cleave the β-glycosidic bonds between the two glucose units of long cellulose chain, (ii) cellobiohydrolase (EC 3.2.1.91) that produces cellobiose by processive cleavage on both reducing and non-reducing ends of cellulose chain and (iii) β-glucosidase (EC 3.2.1.21) that finally converts cellobiose into two glucose units [5]. The activity of these cellulases are occasionally assisted by one or more non-catalytic Carbohydrate Binding Modules (CBMs) [6]. These cellulases are categorized into different glycoside hydrolase (GH) families based on the amino acid sequence, three-dimensional structure and biochemical experiments [7].

Glycoside Hydrolase Family 9 (GH9), as classified in the CAZy (Carbohydrate-Active enZYmes) database, contains enzymes predominantly exhibiting endo-β-1,4-glucanase activity that catalyze the hydrolysis of β-1,4-glycosidic bonds in cellulose and other β-glucans [8]. These enzymes play a critical role in cellulose degradation, facilitating the breakdown of complex plant biomass into shorter cello-oligosaccharides. The family 9 GHs exhibits varying levels of enzymatic activity like endoglucanases, endo-xylo-glucanases, mixed linked endoglucases, cellobiohydrolases, and processive endoglucanases. These GH9 enzymes are widely distributed across various taxa, including bacteria, fungi, plants, and certain invertebrates, where they participate in diverse biological processes such as plant cell wall remodeling, microbial cellulose utilization, and insect digestion of plant material [9]. The catalytic domains of GH9 are highly analogous, displaying an $(\alpha/\alpha)_6$ barrel structure and mainly employ a catalytic mechanism involving acid-base catalysis to cleave glycosidic bonds with retention of anomeric configuration [10]. The extended

classification scheme further divides the family 9 GHs into A, B, and C, which primarily concentrates on factors like sequence similarity, codon usage, intron-exon boundary distribution, and the presence or absence of a transmembrane domain (TM, class A) or secretory peptide [11]. The non-catalytic Carbohydrate Binding Modules (CBMs) are known to assist in the GHs in hydrolysis reaction.

The non-catalytic Carbohydrate Binding Modules (CBMs) display range in the ligand specificity and assist in polysaccharide recognition [12]. Current, CAZy classification module have classified CBMs in 97 different families based on the smililarity in their amino acid sequences [13]. Furthermore, CBMs have been divided into seven families: (i) β-sandwich/β-jelly roll fold, (ii) β-trefoil, (iii) cysteine knot 1, (iv) unique, (v) OB-fold, (vi) hevein fold, and (vii) unique/hevein-like fold [12–14]. The majority of the CBMs exhibit β-sandwich fold structure and appended to the glycoside hydrolases enabling their association to the substrate and hence defining their association-kinetics [12]. Family 3 CBM modules are commonly found in multi-domain cellulases from bacteria, especially in species like *Acetivibrio thermocellus*, which are mainly subgrouped into CBM3A, CBM3B and CBM3C. CBM3A and CBM3B majorly bind to the polysaccharide cellulose and crystalline surfaces, whereas CBM3C attaches to glucan termini and regulates the endo- or exo-glucanase of family 9 GH's catalytic activity [15]. The deletion of Carbohydrate-Binding Modules (CBMs) in cellulases or cellulosomal enzymes has been shown to significantly impact their catalytic efficiency and substrate-binding capabilities [16–18].

In the present study the gene encoding GH9 endoglucanase (*At*GH9C-CBM3A-CBM3B) with a locus tag of Cthe_2760 and GenBank accession no. ABN53959.1 was recognized and retrived from the NCBI database. The gene encoding the full-length β-

1,4-endoglucanase (*At*GH9C-CBM3A-CBM3B), GH9 catalytic module with one CBM (*At*GH9C-CBM3A), only GH9 catalytic module (*At*GH9C) and the two CBMs separately i.e. CBM3A and CBM3B from *Acetivibrio thermocellus* ATCC 27405 were cloned in pET28a(+) vector and expressed in BL-21 (DE3) cells. The enzymes and its truncated derivatives were used for further purification, structural, biochemical and functional analysis.

2.2 Materials and methods

2.2.1 Chemicals, reagents, kits and bacterial strains

The genomic DNA of *Acetivibrio thermocellus* ATCC 27405 was a kind gift from the Prof. Carlos MGA Fontes, Faculdade de Medicina Veterinaria, Avenida de Universidade Técnica, Lisbon, Portugal. Bacterial expression vector pET-28a(+) was used for the cloning of the genes encoding the proteins and the amplification of recombinant plasmid was done in *E. coli* DH5α strain. *E. coli* BL-21 (DE3) cells were used for the expression of the proteins. The exprression vector, pET-28a(+), *E. coli* DH5α and *E. coli* BL-21 (DE3) cells were procured from Novagen (Madison, USA). dNTPs and $MgCl_2$ were acquired from Sigma-Aldrich, India. Other necessary chemicals like RNAse solution (20 mg/mL), glacial acetic acid (99.9% pure), Trizma base (Tris free base), ethidium bromide, nuclease-free water (pH 8.0), and polyacrylamide gel electrophoresis components are also from Sigma-Aldrich, Inc, USA. Sigma-Aldrich, Inc, USA supplied the GenElute miniprep plasmid isolation kit and GenElute gel-extraction kit. Agarose gels with low electroendosmosis (EEO) were acquired from Sigma-Aldrich, India, and were used to electrophorese DNA. The DNA ladder was acquired from New England Biolabs (NEB), USA. A Mini-PROTEAN Tetra Cell device from Bio-Rad Laboratories, USA was used for SDS-PAGE. The protein molecular weight marker was supplied by HiMedia Laboratories Private Limited, India.

2.2.2 Sequence analysis and molecular architecture of endoglucanase *At*GH9C-CBM3A-CBM3B

The amino acid sequence of the endoglucanase (*At*GH9C-CBM3A-CBM3B) from *Acetivibrio thermocellus* ATCC27405 affiliated to family 9 GH with GenBank

accession number ABN53959.1 and UniPort ID A3DJ30 was retrieved from the NCBI database (http://www.ncbi.nlm.nih). The conserved boundaries of the amino acid sequence of *At*GH9C-CBM3A-CBM3B were established by submitting the sequence in the conserved domain database (http://www.ncbi.nlm.nih.gov/cdd/) after the PSI-BLAST analysis [19]. N-terminal signal peptide position was annotated with the aid of SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP). The physico-chemical properties such as molecular mass, theoretical pI, amino acid composition, molar extinction coefficient of *At*GH9C-CBM3A-CBM3B were calculated by using ProtParam server (http://web.expasy.org/protparam/).

2.2.3 Amplification of gene fragments and cloning

The DNA sequence of the gene encoding full-length endoglucanase, *At*GH9C-CBM3A-CBM3B, spanning a length of 2886 base pairs, comprising the catalytic module and two CBMs with linker sequences among them, its truncated derivative, *At*GH9C-CBM3A, the catalytic module, *At*GH9C, carbohydrate binding module 3A, CBM3A and carbohydrate binding module 3B, CBM3B were amplified from the genomic DNA of *Acetivibrio thermocellus* ATCC27405 by polymerase chain reaction (PCR) using a pair of nucleotide primers for each gene fragment as listed in the **Table 2.1**. The genomic DNA of *Acetivibrio thermocellus* ATCC27405 was procured from DSMZ, Germany. Each pair of oligonucleotide primers was designed with a restriction enzyme site *NheI* at forward primer and *XhoI* at reverse primer. Each gene fragment of interest was amplified by using the Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific, USA).

Table 2.1 Primers used for amplification of gene encoding full-length *At*GH9C-CBM3A-BM3B and its truncated derivatives.

Gene encoding proteins	Primers
***At*GH9C-CBM3A-CBM3B**	Forward 5′-CG**GCTAGC**CCACCTGCGACTTTTACACCTGC-3′
	Reverse: 5′-CCGC**CTCGAG**TTAAACAGGCTCAACTCCATATTTGAG-3′
***At*GH9C-CBM3A**	Forward: 5′-CG**GCTAGC**CCACCTGCGACTTTTACACCTGC-3′
	Reverse: 5′-CCGC**CTCGAG**TTACGGTTCTTCACCGTAAACTTTTACACC-3′
***At*GH9C**	Forward: 5′-CG**GCTAGC**CCACCTGCGACTTTTACACCTGC-3′
	Reverse: 5′-CCGC**CTCGAG**TTACGGTTCTTCACCGTAAAC-3′
CBM3A	Forward: 5′-CG**GCTAGC**GGACAGGGACATGAACCTATTCCG-3′
	Reverse: 5′-CCGC**CTCGAG**TTACGGTTCTTCACCGTAAACTTTTACACC-3′
CBM3B	Forward: 5′-CG**GCTAGC**GGTGAAGAACCGCCAAAGCTTTC-3′
	Reverse: 5′-CCGC**CTCGAG**TTAAACAGGCTCAACTCCATATTTGAG-3′

*The bold letters in both forward and reverse primer designate the restriction sites used. *NheI* at forward primer and *XhoI* at reverse primer.

The 50 µL PCR mixure having the genomic DNA as template, Phusion polymerase enzyme, primers, dNTPs, and buffer were mixed in proportion as indicated in **Table 2.2.** Using the PCR machine, Bio-Rad T100, USA, the PCR was carried out for 30 cycles under the conditions listed in **Table 2.3.** The template DNA was initially denatured at 98°C for 10s, annealing for 30 s at 62°C for *At*GH9C-CBM3A-CBM3B and *At*GH9C-CBM3A, 58°C for *At*GH9C, 56°C for CBM3A and CBM3B; extension at 72°C for *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A and *At*GH9 for 90 s and at 72°C for 30 s for CBM3A and CBM3B and the final extension of each gene fragments was carried out at 72°C for 10 min. The amplicons obtained by PCR were detected by running on 0.8% (w/v) agarose gel.

Table 2.2 PCR mixture for amplification of *At*GH9C-CBM3A-CBM3B and its derivatives from *Acetivibrio thermocellus* ATCC27405.

PCR components	Volume (µl)	Final concentration
10X reaction buffer	5.0	1X
100 mM dNTP mix	0.2	0.33 mM
Forward primer (15 µM)	1.5	0.40 µM
Reverse primer (15 µM)	1.5	0.40 µM
Sigma water, pH 8.0	40.8	--
Genomic DNA (20.0 µg/mL)	0.5	10.0 ng
Phusion Polymerase (1 U/µl)	0.5	0.008 U
Total	50.0	--

Table 2.3 PCR conditions for *At*GH9C-CBM3A-CBM3B and its truncated derivatives.

Cycle step	Temperature	Time	Cycle
Initial denaturation	98°C	3 min	1
Denaturation	98°C	30 sec	30
Annealing	50-70°C	30 sec	
Extension	72°C	90 Sec	
Final Extension	72°C	10 min	1

2.2.4 Agarose gel electrophoresis of PCR amplified products

1X Tris-acetate-EDTA (TAE) buffer was used to generate a 0.8% (w/v) agarose gel, wherein the PCR-amplified products were electrophoresed. A TAE buffer stock solution (10X) was prepared with components 400 mM Tris-acetate, 10 mM EDTA, pH 8.0, according to the protocol of Sambrook and Russell [20]. Agarose powder (400 mg) was mixed with 50 mL of 1X TAE buffer to create the agarose gel with concentrations of 0.8% (w/v). The mixture was heated in a microwave to obtain a transparent solution. After lowering the temperature of agarose solution to 50°C, 5.0 µL of ethidium bromide (5.0 mg/mL) was added. The combs were inserted to create wells after the transparent solution had been poured over the casting device [20,21]. Following a 4:1 mixture of the DNA sample and DNA loading dye (5X), the gel was electrophoresed at 60 volts until the dye had migrated over 70% of the gel length. The gel documentation system (BioRad XR, USA) was used to visualize the DNA bands under UV light.

2.2.4.1 DNA loading dye

The ingredients listed below in **Table 2.4** were combined to prepare the DNA sample loading dye and stock solution (5X). One volume of the stock solution was combined with four volumes of DNA to create 1X before loading. The DNA loading dye's pH was adjusted to 8.0.

Table 2.4 5X DNA loading buffer composition.

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

2.2.5 Extraction of DNA from agarose gel

The PCR-amplified DNA were extracted and purified from an agarose gel using a Sigma GenElute kit, in accordance with the instructions provided by the manufacturer described in **section 2.2.5.1**. All the resulting DNA and plasmids were eluted in 30 µL of elution buffer, included with the kit (Sigma-Aldrich, USA).

2.2.5.1 Protocol for extraction of DNA from agarose gel

1. The weight of a sterile 1.5 mL microcentrifuge tube was measured, and its result was noted.
2. Using a sharp, sterile scalpel, the PCR-amplified DNA was carefully removed from the gel and placed in the microcentrifuge tube that had been previously weighed. After recording the weight of tube again, the weight of the removed gel was determined by deducting the original tube weight.
3. Then, for every volume of gel (about 100 mg to 100 µL), three volumes of Gel Solubilization Solution was added.
4. The excised gel was inserted in a microcentrifuge tube and incubated at 50°C for 10 minutes, or until the gel slice was completely dissolved.
5. An equivalent of one volume of isopropanol was added to the dissolved slice solution.
6. The GenElute binding column G was placed within the 2 mL collecting tube included with the kit. After applying 500 µL of the column preparation solution

to the column membrane, it was centrifuged for 1 min at 16,000 g. The flow-through was discarded.

7. After adding around 700 µL of the PCR-amplified DNA or plasmid solution to DNA binding columns, the flow-through was removed and the mixture was centrifuged at 16,000 g for one minute at room temperature. When the volume was more than 700 µL, the flow-through was disposed of as usual and the remaining solution was centrifuged.
8. After adding 700 µL of Wash Solution to each DNA-bound spin column, the flow-through was removed and centrifugation was performed at 16,000 g for one minute at room temperature. To guarantee total removal of any remaining ethanol, a second spin at 16,000g for one minute was carried out.
9. The bounded DNA-containing column was then moved into a brand-new, sterile 1.5 mL microcentrifuge tube. At the center of the column, 30 µL DNAse-free water (Sigma-Aldrich, USA) was added. After two minutes of incubation at ambient temperature, the column was centrifuged for one minute at 16,000 g. DNA recovery was improved by two to three times by preheating the elution solution to 65°C before adding it to the membrane.
10. After being eluted from the GenElute spin columns, the PCR-amplified DNA was collected in a sterile 1.5 mL microcentrifuge tube and kept for later use at 4°C.

2.2.6 Preparation of Luria-Bertani medium

The components listed in **Table 2.5** were dissolved in 800 mL of deionized water to make the Luria-Bertani (LB) medium required for growing *E. coli* cells (DH5α and BL-21, (DE3)) possessing recombinant plasmids. The final volume was adjusted to one liter by deionized water, and the pH was adjusted to 7.2 by NaOH (0.1 mM). After being moved to a 250 mL conical flask, 100 mL of the LB medium was autoclaved for 20 min at 121°C and 15 psi pressure. The autoclaved and cooled LB medium was combined with the filter-sterilized antibiotic (Kanamycin; 50 µg/mL) ahead of inoculation.

Table 2.5 Luria-Bertani medium composition.

Component	Final concentration (%, w/v)
Tryptone	1.0
Yeast extract powder	0.5
Sodium chloride	1.0

2.2.7 Preparation of Luria-Bertani (LB)-Agar medium

The Agar-Agar type I at concentration 2% (w/v) was dissolved in the LB-broth medium as mentioned in **section 2.2.6**, to produce the LB-Agar medium. The solution, it was autoclaved for 20 minutes at 121°C and 15 psi pressure to sterilize the same. Under a laminar air flow hood, the medium was cooled down to around 50°C and filter-sterilized antibiotic (Kanamycin, 50 µg/mL) was added. Sterile petri plate was poured with 20 mL of LB-Agar medium and each was covered, which was then allowed to solidify for 30 min and kept in an inverted posture at 4°C until use.

2.2.8 *E. coli* DH5α competent cell preparation using the calcium chloride

Day 1

1. A test tube containing 5.0 mL of LB medium and 50 µL of glycerol stock of *E. coli* DH5α cells was inoculated, and the cells were grown overnight at 37ºC and 180 rpm.
2. A 0.22 µm filter was used to filter-sterilize a 0.1 M $CaCl_2$ solution in a laminar airflow setting. The solution was then refrigerated.

Day 2

3. A 250 mL conical flask containing 100 mL of LB medium was filled with 1.0 mL of the day 1 culture. After that, the mixture was incubated at 37 ºC and 180 rpm until its cell optical density at 550 nm reached range between 0.4 and 0.6.
4. The centrifuge tubes (50 mL) were autoclaved, cooled on ice, and placed in a laminar airflow hood.
5. In laminar hood, 40 mL of the culture was transferred to round-bottom centrifuge tubes and centrifuged at 4000g and 4ºC for 10 min. The 100 mL culture was entirety centrifuged by repeating this procedure.
6. After re-suspending the cell pellet in 3–4 mL of sterile, ice-chilled 0.1 M $CaCl_2$ solution, the final volume was set at 20 mL. The cell slurry in the centrifuge tubes was kept on ice for ten minutes. The tube was centrifuged at 4000g and 4ºC for 10 min.
7. The supernatant was gently discarded, the pellet was once again suspended in 3.0 mL of sterile, ice-chilled 0.1 M $CaCl_2$ solution.
8. An aliquot of 200 µL of competent cells was transferred to separate 1.5 mL microcentrifuge tubes with a final concentration of 10% (v/v) glycerol. The tubes were then kept at -80ºC for later use.

2.2.9 Cloning of gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives into pET-28a(+) vector

2.2.9.1 The restriction map of the pET-28a(+) expression vector

pET-28a(+) is an expression vector developed for recombinant protein expression in *E. coli*. Because of its strong T7 promoter system, the T7 RNA polymerase may drive large amounts of protein expression. Studier and associates were the first to create the T7 promoter system [22]. Encoded by the T7 RNA polymerase of the host cell, the T7 bacteriophage promoter controls the expression of genes inserted into pET plasmids. The pET vectors' cloned genes are transcriptionally dormant in the uninduced state. Since these cloned genes encode proteins with a His_{6x}-Tag, affinity chromatography can be used to purify them in a single step. The pET-28a(+) vector allows an expressed protein to be integrated with an N-terminal His_6-Tag/thrombin/T7-Tag and optionally with a C-terminal His_{6x}-Tag sequence (**Fig. 2.1**). The positions of the sequences encoding the His_{6x}-Tag, T7 promoter, T7 terminator, kanamycin resistance, and f1 origin are shown in the figure.

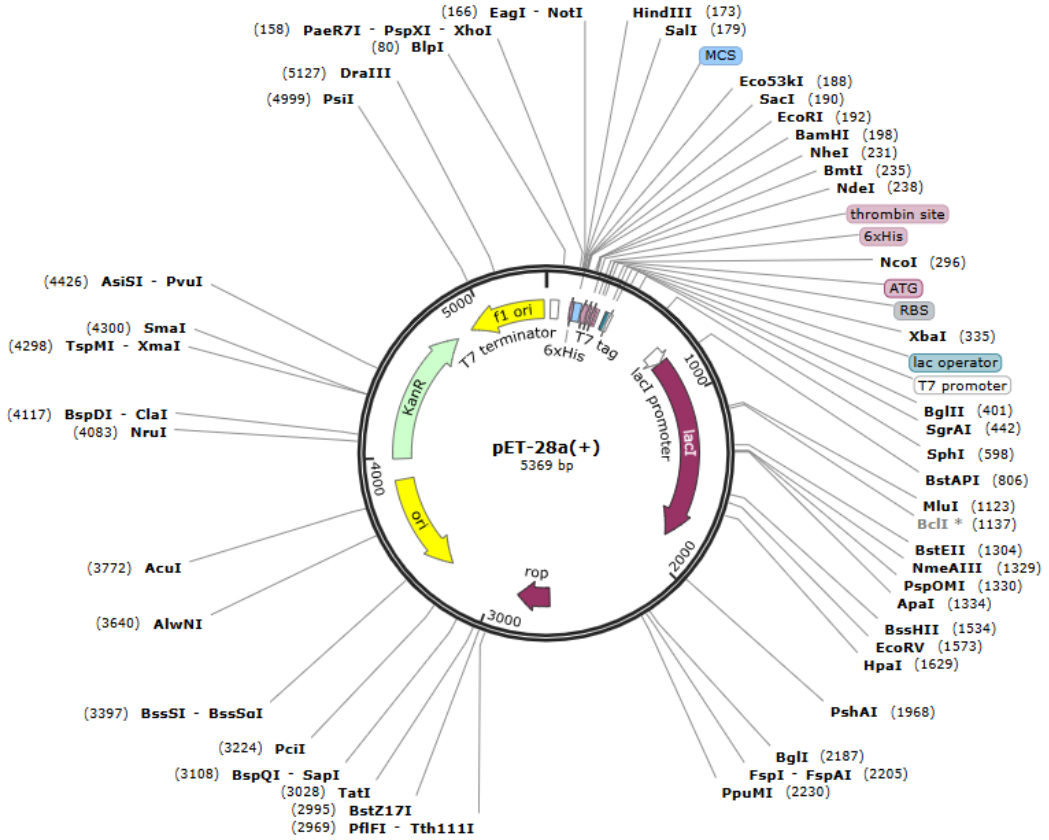


Fig. 2.1 The restriction map of the pET-28a(+) expression vector showing multiple cloning sites and restriction enzyme sites .

2.2.9.2 Restriction digestion of pET-28a(+) plasmid DNA and PCR-amplified gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives

The *Nhe*I and *Xho*I restriction enzymes were used to digest the pET-28a(+) vector (see **Table 2.6**). These same enzymes facilitated the digestion of PCR-amplified genes encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives to prepare for ligation with the restriction-digested pET-28a(+) vector (refer to **Table 2.7**). The restriction digestion reactions were conducted by incubating the digestion reaction mixture at 37°C in a water bath for 90 min. After *Nhe*I and *Xho*I digestion, the PCR-amplified gene and the digested pET-28a(+) vector were run on an agarose gel and extracted from the gel using the procedure described in **section 2.2.5.1**.

Table 2.6 Restriction enzyme digestion of pET-28a (+) plasmid DNA.

Reaction components	Volume (µL)
10x buffer	3.0
Nuclease free water	4.5
Bovine serum albumin (10 mg/mL)	0.5
Plasmid DNA (approx. 40 ng/µl)	20.0
*Nhe*I (10 U/µl)	1.0
*Xho*I (10 U/µl)	1.0
Total	30.0

Table 2.7 Restriction enzyme digestion of PCR amplified gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives.

Reaction component	Gene encoding *At*GH9C-CBM3A-CBM3B and its derivatives (µL)
10x buffer	3.0
Nuclease-free water	4.5
Bovine serum albumin (10 mg/mL)	0.5
PCR amplified product	20
*Nhe*I (10 $U.\mu l^{-1}$)	1
*Xho*I (10 $U.\mu l^{-1}$)	1
Total	30

2.2.9.3 Ligation of gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives into pET-28a(+) vector

The *Nhe*I and *Xho*I digested gene, encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives were ligated into the pET-28a(+) expression vector. The ligation reaction was developed from the *Nhe*I and *Xho*I digested gene, encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives with pET-28a(+) digested vector using the ingredients listed in **Table 2.8**. After that, this reaction was incubated at 16°C for 12 h to maximize the number of transformants. A 3:1 insert: vector molar ratio was used in the setup, and the amount of insert was determined using the procedure described below:

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

$$\frac{80\ (ng) \times 2.483\ (kb)}{5.369\ (kb)} \times \frac{3}{1} = 110\ ng\ (AtGH9C\text{-}CBM3A\text{-}CBM3B)$$

$$\frac{80\ (ng) \times 1.973\ (kb)}{5.369\ (kb)} \times \frac{3}{1} = 88\ ng\ (AtGH9C\text{-}CBM3A)$$

$$\frac{80\ (ng) \times 1.430\ (kb)}{5.369\ (kb)} \times \frac{3}{1} = 64\ ng\ (AtGH9C)$$

$$\frac{80\ (ng) \times 2.483\ (kb)}{5.369\ (kb)} \times \frac{3}{1} = 110\ ng\ (AtGH9C\text{-}CBM3A\text{-}CBM3B)$$

$$\frac{80\ (ng) \times 0.449\ (kb)}{5.369\ (kb)} \times \frac{3}{1} = 20\ ng\ (CBM3A)$$

$$\frac{80\ (ng) \times 0.452\ (kb)}{5.369\ (kb)} \times \frac{3}{1} = 20\ ng\ (CBM3B)$$

Table 2.8 Components of reaction for ligating gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives to pET-28a (+) expression vector.

Reaction component	Gene encoding *At*GH9C-CBM3A-CBM3B and its derivatives (µL)
10x Rapid Ligation Buffer	1.8
pET-28a(+) Vector	5 (~80 ng)
Restriction digested product	8 (~20-110 ng)
T4 DNA Ligase (3 Units/µl)	1
Nuclease-free water	2.2
Total	18

2.2.10 Screening of recombinant plasmid DNA for positive clones

2.2.10.1 Transformation of *E. coli* DH5α cells by ligated recombinant DNA

Following the overnight ligation, ligation reactions were used to transform in the competent cells, *E. coli* DH5α. **section 2.2.8** outlines the process of generating competent *E. coli* DH5α cells. The stepwise transformation protocol is given below:

1. 200 µL of competent cells contained in microcentrifuge tubes were taken out from the ultrafreezer and placed on ice for 5 min.
2. 10 µL of overnight ligation mixture was added to the 200 µL competent cells. The tube was gently tapped four or five times and left on ice for 30 min. Over the course of the 30-minute incubation period, the cells were gently tapped periodically. Meanwhile, the water bath is set to 42°C.
3. The cells were subjected to heat shock at 42°C for 40s.
4. After the heat shock treatment the cells were promptly kept on the ice and left on it for 5 min.
5. The transformed cells were supplemented with 800 µL of sterile LB-medium that was pre-incubated at 37°C.

6. The transformed cells were kept on a shaking incubator for 1 h at 37°C and 220 rpm for growth.
7. The cells were centrifuged for 5 min at 4000g and 25°C and 800 µL of supernatant was disposed. The cell pellet was subsequently suspended in the 200 µL of supernatant that remained.
8. 200 µL of cells were spread on an LB agar plate containing a suitable antibiotic. The LB agar plates were incubated at 37°C for 12 h.
9. The transformation efficiency was calculated using the following formula,

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

The transformation process described above was followed to insert 10 µL of the ligation mixture to 200 µL of *E. coli* DH5α competent cells. The transformed *E. coli* DH5α cells were grown on LB plates supplemented with 50 µg/mL of kanamycin at 37°C for 12 h.

2.2.10.2 Isolation of plasmid DNA from transformed colonies by miniprep kit

Colonies were picked up from the plates that had been incubated for 12 h. The colonies varying in size were randomly selected from plates in a laminar airflow hood and grown in 10 mL of LB medium supplemented with kanamycin (50 µg/mL) for 12 h. The plasmid DNA from this 10 mL culture was extracted by using a miniprep kit from Sigma-Aldrich, Inc, USA following the procedure outlined in **section 2.2.10.2.1** below.

2.2.10.2.1 Plasmid isolation protocol by miniprep kit

1. In a 1.5 mL microcentrifuge tube, the grown culture containing recombinant plasmid was centrifuged at 6500g for 5 min.
2. The above process was repeated to fully pellet down 10 mL of grown culture.
3. Each recombinant cell pellet was resuspended in 200 µL of resuspension solution and vortexed. Before usage, the resuspension solution was treated with RNAse at a final concentration of 0.3 mg/mL.
4. 200 µL of the lysis solution was added to every tube and gently inverted 5-6 times to ensure proper mixing for lysis to happen. It was then allowed to stand for 2-5 min.

5. After lysis 350 µL of neutralization solution was added and the mixture was thoroughly mixed by inverting the tubes 4-6 times. The resulting mixture was then spun at 14,000g for 10 min.
6. 500 µL of column preparation solution was added to the binding column, and it was centrifuged at 14,000g for 1 min in order to prepare and activate the DNA binding column. The flow through that had gathered in the collection tube was discarded.
7. The clear lysate was added to the activated DNA binding column and centrifuged at 14000g for 1 min. The flow through in the collecting tube was discarded.
8. The plasmid DNA that had adhered to the column was washed with 700 µL of wash solution and spun for 1 min at 14,000g. In order to fully remove the wash solution, the flow-through was disposed of and the column was spun for an additional minute at 14,000g.
9. The DNA binding column was placed on a new, sterile microcentrifuge tube, and 30 µL of DNAse-free water was added. After letting the microcentrifuge tube remain at room temperature (25°C) for 10 min, the plasmid DNA was extracted using centrifugation at 14,000 g for 1 min.
10. Plasmid DNA eluted in a clean microcentrifuge tube was kept at -20°C until further use.

2.2.10.3 Verification of positive clones using plasmid DNA restriction digestion

The isolated plasmid DNA from the transformed colonies containing the gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives was subjected to restriction digestion for verification of positive clones. Each plasmid DNA was then placed in fresh, sterile micro-centrifuge tube for evaluation using restriction enzyme digestion. For confirmation of positive clones, the recombinant plasmid DNA of each of the derivative was digested by the restriction enzymes, *Nhe*I and *Xho*I in a 30 µL reaction mixture, which was set up as shown in **Table 2.6.**

2.2.11 Overexpression of the gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives

2.2.11.1 *E. coli* BL-21 (DE3) competent cells preparation

The calcium chloride method of transformation was used to prepare the competent *E. coli* BL-21 (DE3) cells as described in **section 2.2.8**. 10% (v/v) glycerol (the final concentration) was added to competent cells and 200 µL aliquots were prepared in sterile microcentrifuge tubes and kept at -80°C for later use.

2.2.11.2 Transformation of *E. coli* BL-21 (DE3) competent cells by recombinant plasmid encoding genes for *At*GH9C-CBM3A-CBM3B and its derivatives

The transformation procedure outlined in **section 2.2.10.1** was followed. The *E. coli* BL-21 (DE3) cells (200 µL) were transformed for protein expression using 2 µL of the recombinant plasmid of the positive pET-28a(+) clone obtained in **section 2.2.10.3**. After transformation of *E. coli* BL-21 (DE3) cells, the recombinant plasmid carrying the gene expressing *At*GH9C-CBM3A-CBM3B and its truncated derivatives was plated on LB agar plates, supplemented with 50 µg/mL of kanamycin, and cultured for whole night at 37°C for 12 h. The grown colonies were used for checking protein expression.

2.2.11.3 Overexpression of recombinant *At*GH9C-CBM3A-CBM3B and its truncated derivatives

The host *E. coli* BL21(DE3) cells harboring the recombinant plasmid encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives were cultivated in 100 mL of LB medium supplemented with kanamycin (50 µg/mL) at 37°C and 180 rpm in a shaking incubator (ORBITEK, Scigenics Pvt. Ltd., India). Isopropyl-β-D-thiogalactopyranoside (IPTG) was added at varied concentrations (0.25 mM, 0.50 mM and 1.0 mM) at 24°C when the cell growth reached the mid-exponential phase and

absorbance at 600 nm was 0.6 (A_{600} = 0.6). The culture was then incubated at 24°C for 20 hours at 180 rpm after IPTG addition.

2.2.11.4 Investigation of optimum IPTG concentration for overexpression of *At*GH9C-CBM3A-CBM3B and its truncated derivatives

In 10 mL of LB medium supplemented with kanamycin (50 µg/mL), *E. coli* BL-21 (DE3) cells carrying the recombinant plasmid harboring the gene expressing *At*GH9C-CBM3A-CBM3B or its truncated derivatives were cultured at 37°C and 180 rpm in a shaking incubator. The culture temperature was brought down to 24°C once it had reached the mid-exponential phase (A_{600} = 0.6). Isopropyl-β-D-thiogalactopyranoside (IPTG) was added at a final concentration of 0.25 to 1.0 mM to induce recombinant protein expression. Cells were incubated for 16 hours at 180 rpm at 24°C after IPTG induction. For every culture, 200 µL of broth was taken out and centrifuged for 5 min at 6500g. The supernatant was disposed off and the cell pellet was dissolved in 200 µL of distilled water, centrifuged at 6500g for 5 min. It was then resuspended in 40 µL of distilled water. Equal volumes (10 µL) of uninduced and all IPTG-induced cell samples were loaded onto SDS-PAGE gel for quantitative analysis. The expression level of recombinant *At*GH9C-CBM3A-CBM3B and its truncated derivatives were compared by visual analysis.

2.2.12 Analysis of recombinant protein overexpression by SDS-PAGE

The overexpression of the recombinant protein *At*GH9C-CBM3A-CBM3B and its truncated derivatives was examined using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) technique by using 12% (w/v) gel.

2.2.12.1 Preparation of SDS-PAGE gel

The components involved in making SDS-PAGE gel include 30 % (w/v) acrylamide, a resolving gel (pH 8.8), a stacking gel (pH 6.8), 10% (w/v) sodium dodecyl sulfate (SDS), 10% (w/v) ammonium persulfate (APS), N,N,N′,N′ - Tetramethylethylenediamine (TEMED), glycerol, glycine, β-marcapto ethanol, a 5X sample loading buffer (pH 6.8), and an electrophoretic running or tank buffer (pH 8.3-8.5). The detailed description on the SDS-PAGE gel and buffer components are given in **sections 2.2.12.2 to 2.2.12.5**.

2.2.12.2 Preparation of acrylamide solution

Initially, 0.8 g of bis-acrylamide was dissolved in 50 mL of ultra-pure deionized water (MilliQ) in an amber-color bottle on a magnetic stirrer (IKA, C-MAG HS7, Germany) to prepare the acrylamide solution (30% w/v). When the solution turned transparent, 29.2 g of acrylamide was added to the solution. Finally, a measuring cylinder was used to adjust the final volume to 100 mL by adding deionized water. After that, the acrylamide solution was filtered through Whatman No. 1 filter paper, and kept in an amber-color bottle at 4°C.

2.2.12.3 Polymerization of SDS-PAGE gel

The composition of reagents listed in **Tables 2.9 and 2.10** was used to make the resolving gel and stacking gels in accordance with the procedures described by Sambrook et al. [23]. As indicated in **Table 2.9**, each component was added one at a time to a 50 mL conical flask to create the resolving gel whereas maintaining the final acrylamide concentration at 12.0% (w/v). Similarly the stacking gel (4%, w/v) was developed by mixing all the components listed in **Table 2.10**.

Table 2.9 SDS-PAGE component composition for resolving gel preparation.

Component	Gel (12%, w/v) 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
TEMED	0.01
APS (10%,w/v)	0.1
Total	10

Table 2.10 SDS-PAGE component composition for stacking gel preparation.

Components	Gel (4%, w/v) volume (mL)
Acrylamide solution (30%, w/v)	0.7
Deionised water	2.8
SDS (10%, w/v)	0.5
0.5 M Tris-HCl (pH 6.8)	1.0
TEMED	0.005
APS (10%, w/v)	0.05
Total	5.055

2.2.12.4 Preparation of SDS-PAGE running buffer

The running buffer, 1X Tris-Glycine (pH 8.3), was used to run the gel at a steady current of 5 mA per lane. The 5X stock solution of running buffer was made as indicated in **Table 2.11**, from which 1X Tris-Glycine was made. In a 1 L beaker, 15.14 g of Tris-HCl and 94 g of glycine were added and dissolved in 800 mL of deionized water to produce 1 L of 5X Tris Glycine buffer. The buffer pH was brought to 8.3 after 50 mL of 10% (w/v) SDS was added. The final volume was made to 1.0 L and kept at 4°C.

Table 2.11 Composition of 5X Tris-Glycine, running buffer.

Component	Final concentration (5X buffer)	Final concentration (1X buffer)
Tris-HCl	0.125 M	0.025 M
Glycine	1.25 M	0.25 M
SDS	0.5 % (w/v)	0.125% (w/v)

2.2.12.5 Preparation of sample loading buffer

In order to prepare the sample loading buffer (5X), the components were dissolved sequentially at the concentrations listed in **Table 2.12**, and the pH of the buffer was set to 6.8. When loading onto an SDS-PAGE gel, the sample buffer's final concentration was consistently kept at 1X.

Table 2.12 Composition of 5X sample loading buffer.

Component	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)

2.2.12.6 Preparation of staining and destaining solutions

The staining solution for the SDS-PAGE gel was made by dissolving 250 mg (0.25% w/v) of Coomassie Brilliant Blue (CBB R-250) in 50 mL of deionized water on a magnetic stirrer. A final ratio of 5:4:1 (deionized water:methanol:acetic acid) was then achieved by adding 40 mL of methanol and 10 mL of glacial acetic acid. In an amber-colored bottle, the solution was stirred with a magnetic stirrer for 12 h. The solution was filtered using Whatman Filter No. 1 and kept at 4°C. The destaining mixture (100 mL) was made by mixing deionized water, methanol, and glacial acetic acid in a 5:4:1 ratio. Gels were destained by shaking in submerged state in the destaining solution, which was changed every 30 min until protein bands were visible.

2.3 Results and Discussion

2.3.1 Molecular architecture of *At*GH9C-CBM3A-CBM3B and its truncated derivatives

The quantitative proteomic analysis of the cellulocytic bacterium *Acetivibrio thermocellus* revealed a total of 41 cellulosomal proteins, out of which 20 belong to the family 9 glycoside hydrolase [24]. *At*GH9C-CBM3A-CBM3B is listed in the UniPort with ID A3DJ30 and have GenBank accession number ABN53959.1. Sequence analysis of gene locus Cthe_2760 by BLASTP tool revealed the presence of an initial signal peptide (1–39 aa), at its N-terminal (**Fig. 2.2a**). It is followed by the catalytic module (*At*GH9C) of 476 amino acid (58–534 aa) in length, which is flanked by two linkers at 40–57 aa and 535–564 aa. Downstream to the catalytic domain of the enzyme there are two different CBMs, CBM3A (565–715 aa) and CBM3B (736–885 aa) with a linker sequence (716–735 aa) in between them. After CBM3B there is a short stretch of linker peptide (886–893 aa) (**Fig. 2.2a**). At the C-terminus of the sequence, a domain resembling Doc I (894–961 aa) is present. The gene encoding catalytic module (*At*GH9C), with two CBMs i.e. *At*GH9C-CBM3A-CBM3B (**Fig. 2.2b**) comprising 827 aa (58–885 aa), its truncated derivative, *At*GH9C-CBM3A (**Fig. 2.2c**) of 657 aa in length (58–715 aa), the catalytic module, *At*GH9C (**Fig. 2.2d**) of 476 aa in length (58–534 aa), carbohydrate binding module 3A, CBM3A (**Fig. 2.2e**) of 147 aa in length (565–715 aa) and carbohydrate binding module 3B, CBM3B (**Fig. 2.2f**) of 149 aa in length (885–736 aa) were cloned separately in pET-28a(+) expression vector. A stretch of 21 amino acid long pET-28a(+) vector sequence at N-terminal (1–21 aa) containing the His_6-tag and 8 amino residues at C-terminal was included in each clone that contributed to the extra molecular mass of the each protein.

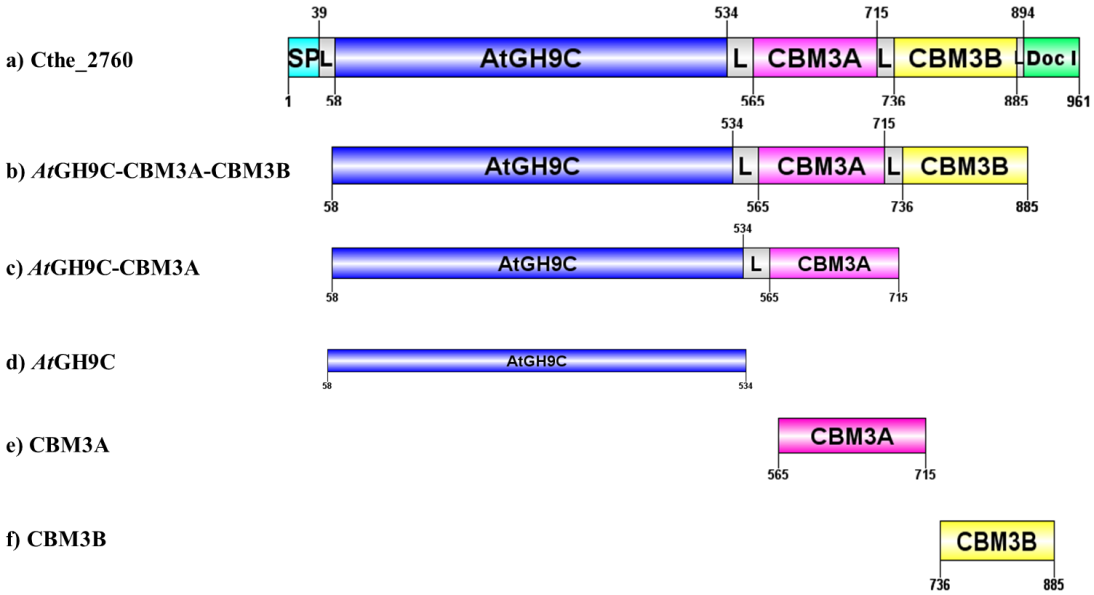


Fig. 2.2 Schematic representation of the molecular architecture of the gene locus a) Cthe_2760 from *Acetivibrio thermocellus* showing its truncated derivatives b) *At*GH9C-CBM3A-CBM3B, c) *At*GH9C-CBM3A, d) *At*GH9C with the carbohydrate binding modules, e) CBM3A and f) CBM3B.

2.3.2 PCR amplification of genes encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives

The genes encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives were amplified by PCR from *Acetivibrio thermocellus* ATCC 27405 genomic DNA following the procedures outlined in **section 2.2.3**. It was subsequently detected on a 0.8% (w/v) agarose gel and illustrated in **Fig. 2.3** below. As summarized in **section 2.2.5**, the PCR products were extracted from the gel using a gel extraction kit and kept at -20°C. The PCR products of size 2.5 kb, 2.0 kb, 1.4 kb, 0.5 kb and 0.5 kb were obtained for *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B respectively.

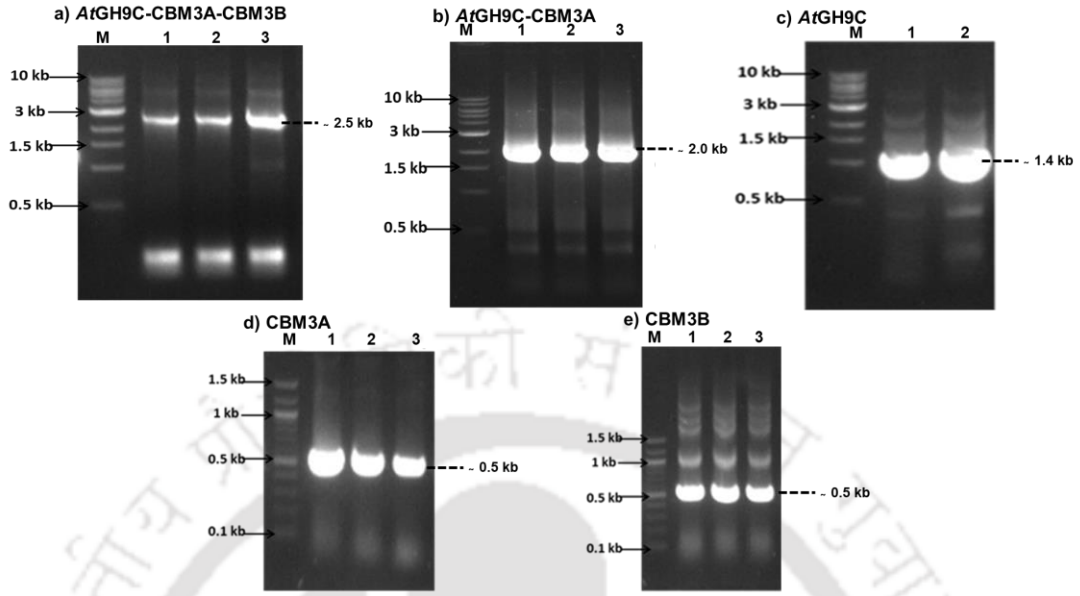


Fig. 2.3 Analysis of PCR amplified gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives on 0.8% (w/v) Agarose gel. The gel images of a) *At*GH9C-CBM3A-CBM3B; b) *At*GH9C-CBM3A; c) *At*GH9C; d) CBM3A and e) CBM3B is shown where, Lane M- DNA Ladder (NEB, USA) and Lane 1- 56 °C, Lane 2- 58°C and Lane 3- 56°C are respective annealing temperature.

2.3.3 Digestion of pET28a(+) vector DNA and PCR insert DNA by restriction enzymes

The process described in **section 2.2.9.2** was followed to enzymatically digest the vector pET28a(+) and PCR insert DNA. The double-digested DNA insert and vector DNA were run on 0.8% (w/v) agarose gel and purified using the gel extraction kit as indicated in **section 2.2.5.1**.

2.3.4 Ligation of genes encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives into pET-28a (+) vector and its transformation

The linearized fragments of the pET-28a(+) vector were used to ligate the restriction enzyme-digested gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives in accordance with the procedure outlined in **section 2.2.10.3.** Under stable conditions, the ligated product was transformed into *E. coli* DH5

competent cells, which were then cultured for the entire night at 37°C on LB agar plates. The transformation efficiency was calculated using the formula mentioned in **section 2.2.10.1** and shown in **Table 2.13**.

Table 2.13 Transformation efficiency of the *E. coli* (DH5α) cells.

Clone	Insert DNA (ng)	No. of colonies	Transformation efficiency (cfu/µg)
Vector (pET28a(+))	112.3	255	2.8×10^4
*At*GH9C-CBM3A-CBM3B	135.2	275	1.05×10^5
*At*GH9C-CBM3A	115.2	201	2.5×10^4
*At*GH9C	80.6	130	1.3×10^3
CBM3A	145.6	302	2.6×10^5
CBM3B	167.72	330	2.9×10^5

2.3.4.1 Isolation of recombinant plasmid DNA

After cloning into pET-28a(+), the plasmid DNA from growing colonies was extracted using a Plasmid miniprep kit following the procedure outlined in **section 2.2.10.2.** The separated plasmids were observed on a 0.8% (w/v) agarose gel following the electrophoresis. Further by carrying out the restriction digestion on the extracted plasmid DNA, the positive clones were confirmed and identified.

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

*Nhe*I and *Xho*I restriction enzymes were used to digest the extracted plasmid in order to verify the positive clones. Following restriction digestion, the reaction products were run on 0.8% (w/v) agarose gels (**Fig. 2.4**). The DNA fragments of ~2.5 kb, ~2 kb, ~0.5 kb, ~0.5 kb and ~1 kb were detected on 0.8% agarose gel for *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B, respectively (Lane 1, **Fig. 2.4 a-e**). The linearized pET-28a(+) vector was also detected at around 5.3 kb after the restriction digestion. All cloned constructs were validated by

sequencing using both forward and reverse primers, and the sequences showed 100% identity with the corresponding *Acetivibrio thermocellus At*GH9C-CBM3A-CBM3B reference gene sequence (GenBank accession no. ABN53959.1).

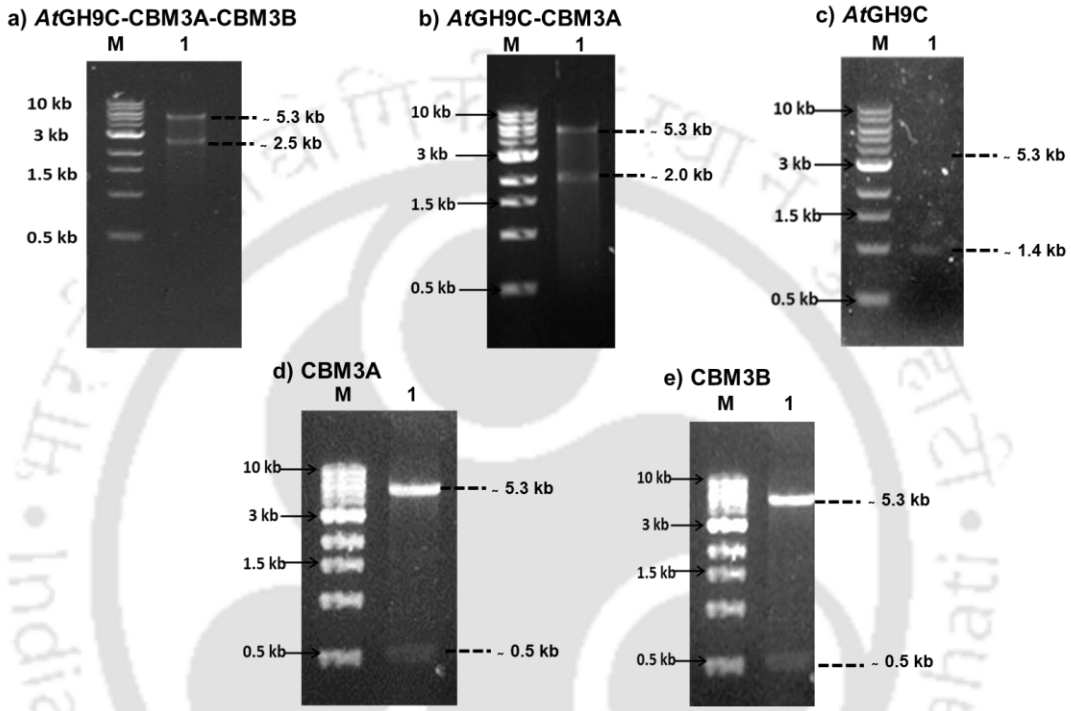


Fig. 2.4 The agarose gel (0.8 %, w/v) images displaying *Nhe*I-*Xho*I digested recombinant plasmid containing gene encoding a) *At*GH9C-CBM3A-CBM3B; b) *At*GH9C-CBM3A; c) *At*GH9C; d) CBM3A and e) CBM3B where, Lane M- DNA Ladder (NEB, USA) and Lane 1- digested vector and insert.

2.3.5 Protein overexpression analysis

The final concentration of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at the mid-exponential stage was used to induce the production of the protein *At*GH9C-CBM3A-CBM3B and its truncated derivatives from the transformed *E. coli* BL-21 (DE3) cells, using the procedure outlined in **section 2.2.11.3.** SDS-PAGE was used to analyze the expression of the protein *At*GH9C-CBM3A-CBM3B and its

truncated derivatives using a 12% (w/v) gel, as shown in **Fig. 2.5**. The molecular mass of the overexpressed *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B were around 100 kDa, 73 kDa, 54 kDa, 16 kDa and 18 kDa respectively (**Fig. 2.5 a-e**).

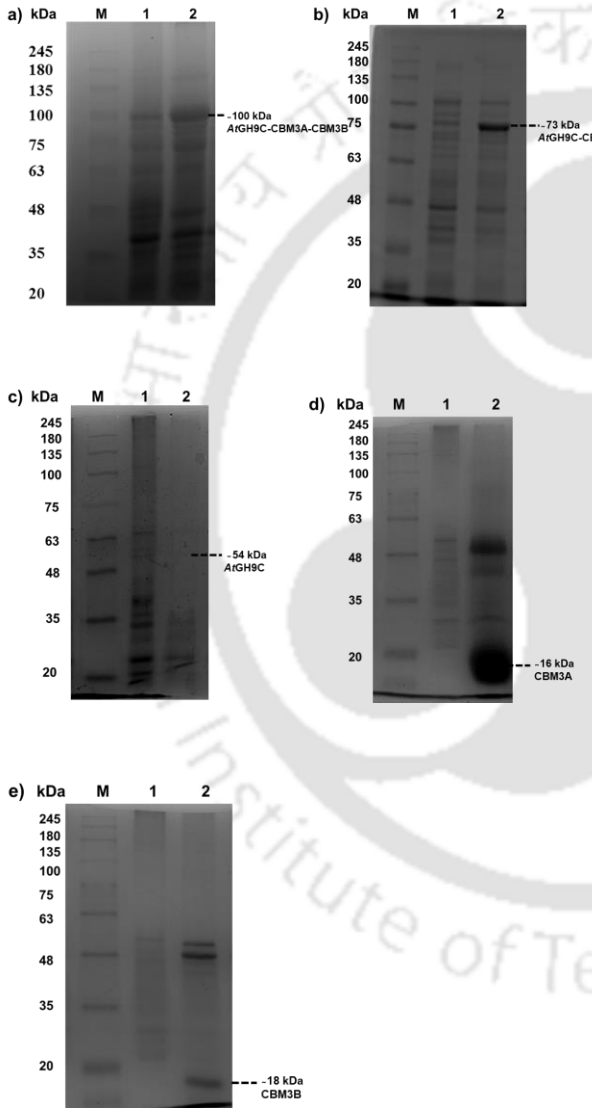


Fig. 2.5 Expression analysis of a) *At*GH9C-CBM3A-CBM3B; b) *At*GH9C-CBM3A; c) *At*GH9C; d) CBM3A and e) CBM3B on 12%, w/v gel SDS-PAGE. Lanes M: Protein molecular-mass marker, Lane 1: Un-induced *E. coli* BL-21 (DE3) cells pellet, Lane 2: Induced *E. coli* BL-21 (DE3) cells pellet.

2.3.6 Optimization of IPTG concentration for protein overexpression

The transformed *E. coli* BL-21 (DE3) cells were induced by using different IPTG concentration ranging from 0.25 to 1.0 mM IPTG to determine its optimum IPTG concentration for overexpression of the recombinant proteins for *At*GH9C-CBM3A-CBM3B and its truncated derivatives. 1.0 mM IPTG was found to be the ideal concentration for maximum protein overexpression of *At*GH9C-CBM3A-CBM3B and its truncated derivatives (**Fig. 2.6 a-e**). Thus , the cells for all proteins were induced with 1 mM IPTG for all further studies.

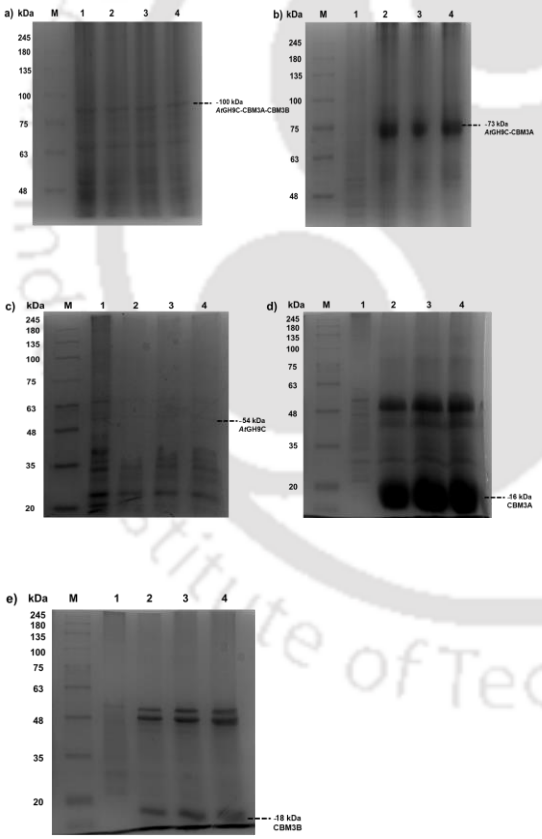


Fig. 2.6 Investigation of optimum IPTG concentration analysis for a) *At*GH9C-CBM3A-CBM3B; b) *At*GH9C-CBM3A; c) *At*GH9C; d) CBM3A and e) CBM3B using SDS-PAGE (12%, w/v gel) analysis. Lanes M: Protein molecular-mass marker (HiMedia Labs Pvt. Ltd., India), 1: Un-induced cells pellet, 2: Induced cells pellet (0.25 mM), 3: Induced cells pellet (0.5 mM) and 4: Induced cells pellet (1.0 mM IPTG).

2.4 Conclusion

The gene encoding *At*GH9C-CBM3A-CBM3B (ABN53959.1) and its truncated derivatives (*At*GH9C-CBM3A, *At*GH9C, CBM3A, CBM3B) belonging to the glycoside hydrolase family 9 (GH9) from *Acetivibrio thermocellus* ATCC 27405 were cloned in pET28a(+) and subsequently expressed in *E. coli* BL21 cells. The molecular architecture analysis of *At*GH9C-CBM3A-CBM3B revealed modular structure a signal peptide (1–39 aa), at its N-terminal followed by catalytic module from GH9 family (*At*GH9C) of 476 amino acids. Downstream there are two different CBMs, CBM3A (565–715 aa) and CBM3B (736–885 aa) with the sequence, a domain resembling Doc I (894–961 aa) at the C-terminus. PCR amplification of *At*GH9C-CBM3A-CBM3B and its truncated derivatives, *At*GH9C-CBM3A, *At*GH9C, CBM3Aand CBM3B showed 2483 bp, 1973 bp, 1430 bp, 449 bp and 452 bp, respectively. The PCR amplified products with suitable restriction sites were ligated to pET28a(+) expression vector for each clone. The recombinant pET-28a(+) expression vector carrying the gene encoding *At*GH9C-CBM3A-CBM3B and its truncated derivatives were used for transformation of *E. coli* DH5α competent cells. The recombinant plasmids were isolated and the positive clones were confirmed by restriction enzyme digestion using *Nhe*I and *Xho*I restriction enzymes. After the restriction digestion, the recombinant plasmid displayed pET-28a(+) vector DNA, of approximately 5.3 kb size, and insert DNA fragments of 2.5 kb, 2.0 kb, 1.4 kb, 0.5 kb and 0.5 kb for *At*GH9C-CBM3A-CBM3B and its truncated derivatives, *At*GH9C-CBM3A, *At*GH9C, CBM3Aand CBM3B respectively. All the recombinant proteins were overexpressed by *E. coli* BL-21 (DE3) cells using an optimum concentration of 1.0 mM IPTG and quanlitatively analyzed by SDS-PAGE gel.

2.5 References

[1] F.H. Isikgor, C.R. Becer, Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers, Polym Chem 6 (2015) 4497–4559. https://doi.org/10.1039/C5PY00263J.

[2] D. Roy, M. Semsarilar, J.T. Guthrie, S. Perrier, Cellulose modification by polymer grafting: a review, Chem Soc Rev 38 (2009) 2046–2064. https://doi.org/10.1039/B808639G.

[3] M.E. Himmel, S.Y. Ding, D.K. Johnson, W.S. Adney, M.R. Nimlos, J.W. Brady, T.D. Foust, Biomass recalcitrance: Engineering plants and enzymes for biofuels production, Science (1979) 315 (2007) 804–807. https://doi.org/10.1126/SCIENCE.1137016/ASSET/53180CC7-C2AC-411F-A286-B238D5DC153A/ASSETS/GRAPHIC/315_804_F3.JPEG.

[4] H.G. Lawford, J.D. Rousseau, Cellulosic Fuel Ethanol, Biotechnology for Fuels and Chemicals (2003) 457–469. https://doi.org/10.1007/978-1-4612-0057-4_38.

[5] E.A. Bayer, R. Lamed, M.E. Himmel, The potential of cellulases and cellulosomes for cellulosic waste management, Curr Opin Biotechnol 18 (2007) 237–245. https://doi.org/10.1016/J.COPBIO.2007.04.004.

[6] Y. Shoham, R. Lamed, E.A. Bayer, The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides, Trends Microbiol 7 (1999) 275–281. https://doi.org/10.1016/S0966-842X(99)01533-4.

[7] D.B. Wilson, D.C. Irwin, Genetics and Properties of Cellulases, (1999) 1–21. https://doi.org/10.1007/3-540-49194-5_1.

[8] R.M. Yennamalli, A.J. Rader, A.J. Kenny, J.D. Wolt, T.Z. Sen, Endoglucanases: Insights into thermostability for biofuel applications, Biotechnol Biofuels 6 (2013) 1–9. https://doi.org/10.1186/1754-6834-6-136/TABLES/1.

[9] S. Kundu, R. Sharma, Origin, evolution, and divergence of plant class C GH9 endoglucanases, BMC Evol Biol 18 (2018) 79. https://doi.org/10.1186/S12862-018-1185-2.

[10] N.R. Gilkes, B. Henrissat, D.G. Kilburn, R.C. Miller, R.A.J. Warren, Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families, Microbiol Rev 55 (1991) 303–315. https://doi.org/10.1128/MR.55.2.303-315.1991.

[11] E. Libertini, Y. Li, S.J. McQueen-Mason, Phylogenetic analysis of the plant endo-β-1,4-glucanase gene family, J Mol Evol 58 (2004) 506–515. https://doi.org/10.1007/S00239-003-2571-X/METRICS.

[12] A.B. Boraston, D.N. Bolam, H.J. Gilbert, G.J. Davies, Carbohydrate-binding modules: fine-tuning polysaccharide recognition, Biochem J 382 (2004) 769–781. https://doi.org/10.1042/BJ20040892.

[13] E. Drula, M.L. Garron, S. Dogan, V. Lombard, B. Henrissat, N. Terrapon, The carbohydrate-active enzyme database: functions and literature, Nucleic Acids Res 50 (2022) D571–D577. https://doi.org/10.1093/NAR/GKAB1045.

[14] H.J. Gilbert, J.P. Knox, A.B. Boraston, Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules, Curr Opin Struct Biol 23 (2013) 669–677. https://doi.org/10.1016/J.SBI.2013.05.005.

[15] S. Petkun, I.R. Grinberg, R. Lamed, S. Jindou, T. Burstein, O. Yaniv, Y. Shoham, L.J.W. Shimon, E.A. Bayer, F. Frolow, Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: Structural and functional significance of the intermodular linker, PeerJ 2015 (2015) e1126. https://doi.org/10.7717/PEERJ.1126/FIG-7.

[16] K. Kumar, S. Singal, A. Goyal, Role of carbohydrate binding module (CBM3c) of GH9 β-1,4 endoglucanase (Cel9W) from *Hungateiclostridium thermocellum* ATCC 27405 in catalysis, Carbohydr Res 484 (2019). https://doi.org/10.1016/J.CARRES.2019.107782.

[17] M.H. Foley, G. Déjean, G.R. Hemsworth, G.J. Davies, H. Brumer, N.M. Koropatkin, A cell-surface GH9 endo-glucanase coordinates with surface glycan binding proteins to mediate xyloglucan uptake in the gut symbiont *Bacteroides ovatus*, J Mol Biol 431 (2019) 981. https://doi.org/10.1016/J.JMB.2019.01.008.

[18] O. Article, Applied Microbiology and Biotechnology Highly active and processive endoglucanase from GH9 family of *Bacillus licheniformis* and the role of CBM in its processivity Protein Engineering Laboratory , Department of Biological Sciences , Indian Institute of Ce, (n.d.).

[19] G.M. Boratyn, C. Camacho, P.S. Cooper, G. Coulouris, A. Fong, N. Ma, T.L. Madden, W.T. Matten, S.D. McGinnis, Y. Merezhuk, Y. Raytselis, E.W. Sayers, T. Tao, J. Ye, I. Zaretskaya, BLAST: a more efficient report with usability improvements, Nucleic Acids Res 41 (2013) W29–W33. https://doi.org/10.1093/NAR/GKT282.

[20] J. and R.D.W. Sambrook, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Woodbury, New York., 2001.

[21] J., F.E.F. and M.T. Sambrook, Molecular Cloning: A Laboratory Manual, Vol. 1. Plainview, Cold Spring Harbor Laboratory Press, 2nd ed., Cold Spring Harbor Laboratory Press, Woodbury, New York., 1989.

[22] F.W. Studier, B.A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, J Mol Biol 189 (1986) 113–130. https://doi.org/10.1016/0022-2836(86)90385-2.

[23] J. and R.D.W. Sambrook, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Woodbury, New York., 2001.

[24] N.D. Gold, V.J.J. Martin, Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis, J Bacteriol 189 (2007) 6787. https://doi.org/10.1128/JB.00882-07.

Chapter 3

Role of carbohydrate binding modules, CBM3A and CBM3B in stability and catalysis by a β-1,4 endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus* ATCC 27405

3.1 Introduction

The plant cell wall is composed of well-structured and complex network of polysaccharides like cellulose, hemicellulose, pectin and their derived carbohydrate polymers [1]. These polysaccharides possess copious source of carbon and energy that cues to the advantage of utilization of the plant polysaccharides as a major renewable energy source [2]. Cellulose is the most abundant, straight chain and unbranched polymer of repeating units of D-glucopyranose joined by β-1,4 glycosidic linkage. Cellulose is present in the primary cell wall of plants, algae and in some bacterial biofilms. This cellulose can be hydrolyzed to simple glucopyranose and successively converted to bioethanol by the action of yeast [3]. Albeit abundant, cellulose is recalcitrant to degradation owing to its insoluble nature and hydrogen-bonded

crystalline structure. Cellulose can be converted into glucose by the concerted action of a variety of enzymes that exist in many bacteria and fungi in free form or in a large extracellular multi-enzyme complex called cellulosome [4]. These cellulosomes are modular in nature and in common arrangement comprises catalytic module(s), accountable for hydrolysis reaction with one or more non-catalytic Carbohydrate Binding Modules (CBMs) [5].

Cellulases cleave the β-1,4-glycosidic linkage of cellulosic chain with the involvement of a water molecule. They are categorized into different glycoside hydrolase (GH) families based on their amino acid sequence [6]. GHs represent one of the five enzyme classes present in the Carbohydrate-Active enZymes (CAZy) database with rationalized information with respect to their substrate specificities, catalytic mechanisms and three-dimensional structures. Depolymerisation of cellulose to glucose is done by the concerted action of three enzymes *viz*. endo-β-1,4-glucanase, exo-β-1,4-glucanase/cellobiohydrolase and β-1,4-glucosidase. Endoglucanase (EC 3.2.1.4) initiates the hydrolysis by randomly cleaving internal β-1,4 glycosidic linkages of linear chain and increase the chain ends. Then cellobiohydrolase (EC 3.2.1.91) acts on the chain end and cleave glycosidic bond processively two glucose residues from the reducing or the non-reducing end, to form cellobiose, the smallest structural repeating unit of cellulose. At last β-glucosidase (EC 3.2.1.21) breaks down the cellobiose into two glucose units [3]. Endoglucanase being the initial player in the cellular hydrolysis dictates the hydrolysis process and found in various glycoside hydrolase families.

GH 9 is the second-largest family containing endoglucanase. The catalytic domains among them are very analogous, showing an $(\alpha/\alpha)_6$-barrel fold having

equivalent catalytic machinery. However, the enzymatic activity among the family 9 GHs differs and the enzymes are primarily distinct as endoglucanases, endo-xylo-glucanases, mixed-linked endoglucases, cellobiohydrolases and processive endoglucanases. The family 9 GHs are further classified as A, B, and C following the extant classification scheme that principally focuses on the conditions like sequence similarity, codon usage, distribution of intron-exon boundaries, and presence/absence of a transmembrane domain, (TM, class A) or secretory peptide [7]. The role of endoglucanases in hydrolysis can be assisted by carbohydrate binding modules (CBMs).

CBMs, are non-catalytic proteins appended to the catalytic modules, preferentially binds the substrate and increase the enzyme-substrate proximity, facilitating their association thereby dictating their association-dissociation kinetics [8]. CBMs have also been shown to dictate substrate specificity, mechanism of enzyme action, thermo-stability and assist the catalytic module in bacterial cell wall attachment [9]. Family 3 CBMs are approximately, 150 amino acid residue long, showing primarily cellulose-binding function and are sub-grouped into CBM3A, CBM3B, CBM3C and CBM3D. CBM3A and CBM3B bind the crystalline surface of polysaccharide cellulose and chitin, while CBM3C fixes to the termini of glucans and modulates the catalytic activity of glucanase (endo- or exo-) of family 9 GH [10]. The detachment of family 3 CBM from the catalytic module of the thermophilic *Thermomonospora fusca* E4 leads to reduced enzymatic activity with incomplete digestion of filter paper [11]. Deletion and mutation analysis of the CBM3c-CBM3b tandem dyad in *cel9B* gene from *Acetivibrio cellulolyticus* indicated that the whole dyad is needed for the active hydrolysis of the cellulosic substrate [12]. The effect of location of CBM3a on the

activity of endoglucanase CelA of *Clostridium thermocellum* before and after the dockerin module points the differential activity modulation of CelA by CBM3a [13]. The deletion of CBM3c from its associated catalytic module, GH9 of *Clostridium thermocellum* Cel9I [14], *Hungateiclostridium thermocellum* Cel9W [15] displayed significantly reduced catalytic activity of the enzyme showing the involvement of CBMs in effective catalysis.

The central impediments to the technological advancement of endoglucanase in the industrial utilization are its catalytic efficiency, pH stability, thermo-tolerance and low-cost maintenance. Moreover, the milieu of understand concerning the role of CBM3A and CBM3B on the activity of catalytic module of an endoglucanase is yet not explored. Investigation on efficient endoglucanases from thermophilic bacteria may be the essential roadmap to overcome such barriers and achieve effective catalysis. In present study, a multimodular enzyme, *At*GH9C-CBM3A-CBM3B, of *Acetivibrio thermocellus* from class C endoglucanase of GH9 family, containing a GH9 catalytic module (*At*GH9C) at the N-terminal followed by two different CBMs, CBM3A and CBM3B joined by a linker sequence in between them at its C-terminus were studied and the role of CBMs on the catalysis was elucidated. The gene encoding the full-length β-1,4-endoglucanase (*At*GH9C-CBM3A-CBM3B), GH9 catalytic module with one CBM (*At*GH9C-CBM3A) only GH9 catalytic module (*At*GH9C) and two CBMs i.e., CBM3A and CBM3B from *Acetivibrio thermocellus* ATCC 27405 were cloned and expressed in *E. coli* BL21 cells, purified and biochemically characterized. The role of CBM3A and CBM3B in the catalysis was explored by studying their physical interactions with the catalytic module *At*GH9C in combination with different cellulosic substrates.

3.2 Materials and methods

3.2.1 Bacterial strain and plasmids

Bacterial expression vector pET-28a(+) was used for the cloning of the genes encoding the proteins and amplification of plasmid was done in *E.coli* DH5α strain. *E.coli* BL21 (DE3) cells were used for the expression of the proteins. pET-28a(+) vector, *E. coli* DH5α and *E. coli* BL21 (DE3) cells were procured from Novagen (Madison, USA).

3.2.2 Substrate and chemicals

Avicel, β-D-Glucan from Barley, Lichenan, Laminarin (from *Laminaria digitata*), Sodium salt of Carboxy methylcellulose (CMC-Na) and Hydroxy ethyl cellulose (Medium viscosity, FLUKA) were purchased from Sigma-Aldrich Chemical Co. (USA). Carob galactomannan, Curdlan (from *Alcaligenes faecalis*) was obtained from Megazyme Ltd., Ireland. Beechwood xylan was procured from SRL India Pvt. Ltd. Metal ion salts: NaCl, KCl, $CaCl_2$, $MgCl_2$, $MnCl_2$, $CoCl_2$, $NiSO_4$, $FeSO_4$, $ZnCl_2$, $CuSO_4$, Ethylene diamine tetraacetic acid (EDTA) disodium salt dihydrate, Ethylene glycol tetraacetic acid (EGTA) and the dialysis tube (pore size, 14 kDa) were from Hi-Media Laboratories Pvt. Ltd., India. All the chemicals used are of analytical grade and also acquired from Hi-Media Laboratories Pvt. Ltd., India. Isopropyl-β-D thiogalactopyranoside (IPTG), kanamycin, purchased from Sigma-Aldrich Co. LLC., USA. The thin layer chromatography plate (TLC Silica gel 60 F_{254}, 20 × 20 cm) was obtained from Merck KGaA, Darmstadt, Germany.

3.2.3 Expression and purification of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B

The expression and purification of *At*GH9C-CBM3A-CBM3B and its truncated derivatives were carried out using the protocol given in the literature [15]. The *E. coli* BL-21 cells expressing the recombinant plasmids for *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B were separately grown in 500 mL Luria broth medium supplemented with kanamycin (50 µg/mL) at 37°C with shaking at 180 rpm, till the absorbance at 600 nm (A_{600}) reached 0.6. The cells were then induced for protein expression by Isopropyl 1-thio-ß-D-galactopyranoside (IPTG) at final concentration of 1 mM and the cultures were further grown at 24°C, shaking at 180 rpm for 18 h. The recombinant protein, containing a His_6-tag at the N-terminus, was purified using a single-step process based on immobilized metal-ion affinity chromatography (IMAC), as described in s**ection 3.2.3.1.**

3.2.3.1 Purification protocol of recombinant proteins by IMAC

1. The induced bacterial cells containing the recombinant proteins, *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A or CBM3B was harvested by centrifugation at 6,000***g*** at 4°C for 12 min. The cells were resuspended in 10 mL, sodium phosphate buffer, pH 7.5, sonicated on ice for 15 min (10s on and 15s off pulse; with 33% amplitude) and further centrifuged at 12,000***g*** for 1h at 4°C to obtain the cell-free extract as supernatant.
2. The Hi-Trap IMAC column (5 mL) was pre-washed with 5 volumes of Whatman paper filtered and pump degassed water. Then, 5.0 mL of 0.1 M $NiSO_4$ solution was passed to charge the column, and 10 mL of water was used to wash away the unbound Ni^{2+} ions.

3. Then, the column was equilibrated with equilibration buffer (by passing 25 mL) (**Table 3.1**).
4. The cell-free extract obtained after sonication was passed through filter membrane (pore size, 0.45 µm) using syringe filter and subsequently loaded on to the column at a flow rate of 0.5 mL/min.
5. The column was the washed extensively with 100 mL of equilibration buffer to remove the unbound proteins.
6. After washing the protein of interest was eluted using elution buffer, and 1 mL fractions (a total in 5 mL) were collected.
7. After protein elution the column was washed by passing 25 mL of degassed water and removed the Ni^{2+} ions by using 10 mL of cleaning buffer as listed in **Table 3.1** followed by washing with 10 mL of water. Then 25 mL 1N NaOH ws passed thorugh the column and and incubated for two hours at 4°C. After that the NaOH from the column was removed with 50 mL of water followed by passing 10 mL of 20% (v/v) ethanol and stored at 4°C.

The eluted fractions of the purified protein were dialyzed in 4.0 L of 50 mM sodium phosphate, pH 7.5 with 4 changes (1.0 L buffer each time). The molecular mass and purity of the recombinant protein were confirmed using 12% (w/v) SDS-PAGE.

Table 3.1 Buffers used for purification of proteins by affinity chromatography.

Protein	Buffer	Composition
*At*GH9C-CBM3A-CBM3B or *At*GH9C-CBM3A	Equilibration buffer	50 mM sodium phosphate, pH 7.5 500 mM NaCl, 60 mM Imidazole
	Elution buffer	50 mM sodium phosphate, pH 7.5 300 mM NaCl, 300 mM Imidazole
*At*GH9C or CBM3A or CBM3B	Equilibration buffer	50 mM sodium phosphate, pH 7.5 500 mM NaCl, 40 mM Imidazole
	Elution buffer	50 mM sodium phosphate, pH 7.5 300 mM NaCl, 300 mM Imidazole
All	Column cleaning	50 mM sodium phosphate, pH 7.5 300 mM NaCl, 50 mM EDTA

3.2.4 Determination of the purified recombinant protein concentration using Bradford and UV methods

3.2.4.1 Protein estimation by Bradford method

The concentration of the purified protein was estimated by Bradford method [16] using Bovine serum albumin (BSA) as standard. Initially, the standard curve was generated with the help of serially diluting BSA solution stock of 1.0 mg/mL in the concentration range of 10-100 µg/mL. The sample protein, 100 µL was mixed with 1.0 mL of Bradford reagent and kept for 30 min in dark at 25°C. After incubation the color changed from reddish-brown to blue and its absorbance was measured at 595 nm, where only buffer without the protein served as blank. The method developed for the formulation of Bradford's reagent is given in the next section.

3.2.4.2 Preparation of the Bradford reagent

Bradford's reagent quantifies the concentration of the protein in solution based on the spectral signatures of Coomassie Brilliant Blue G-250 [16]. Initially in an amber colour bottle, 100 mg of 0.01% (w/v) Coomassie Brilliant Blue G-250 was weighed and dissolved in 50 mL 95% ethanol. 100 mL of 85% (w/v) phosphoric acid was added to the earlier solution and mixed on magnetic stirrer until the dye completely dissolved.

The dye was diluted with 1 liter of deionized water, filtered using Whatman, No. 1 paper and stored at 4 °C. Commercial Bradford reagent (Sigma-Aldrich Co. LLC, USA) was also used for protein concentration determination. The following formula was used to estimate the concentration of recombinant protein,

$$\text{Protein concentration (mg/mL)} = \frac{\Delta A_{595} \times C \times V}{v}$$

Where,

ΔA_{595} = change in absorbance of the sample
V = volume of the protein-buffer mixture (mL)
C = 1 OD equivalent of BSA from the standard plot (mg/mL)
v = volume of the protein used for assay (mL)

3.2.4.3 Protein estimation by UV spectroscopy method

The concentration of the purified protein was also ascertained from its corrospoding absorbance at 280 nm (A_{280}) by the UV-Vis spectroscopy method [17]. The protein was appropriately diluted and the absorbance, A_{280} was measured using a UV-Vis spectrophotometer (Gene Quant, GE healthcare, USA) with a 1 cm path length. The following formula was used for the calculation of protein concentration,

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

The molar extinction coefficient of the different proteins used in the calculation of their concentration is given in **Table 3.2.**

Table 3.2 Molar extinction co-efficient of different recombinant proteins.

Recombinant protein	**Molar extinction co-efficient ($M^{-1}cm^{-1}$)**
*At*GH9C-CBM3A-CBM3B	186880
*At*GH9C-CBM3A	167845
*At*GH9C	127380
CBM3A	35870
CBM3B	19035

3.2.5 Assay of enzyme activity for recombinant proteins

The biochemical assays of the different recombinant proteins comprising the catalytic modules *viz.*, *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A and *At*GH9C were carried out with different cellulosic substrates at its 1.0% (w/v) final concentration with 10 µg/mL as the final enzyme concentrations. For enzymatic assay, the reaction mixture contatining of 10 µL of recombinant enzymes (10 µg/mL) and 90 µL of 1.1% (w/v) substrate in sodium phosphate buffer (50 mM, pH 7.5), was incubated at 55°C for 1 min. The substrate at 1.0% (w/v) final concentration in 100 µL of sodium phosphate buffer (50 mM, pH 7.5) without the enzyme served as the negative control. After incubation, the concentration of reducing sugars released was estimated by method developed by Nelson [18] and Somogyi [19] with D-glucose as standard. Nelson-Somogyi-D (NS-C) was prepared by mixing 25 parts of the copper reagent A (NS-A) as given in **section 3.2.5.1** with 1 part of copper reagent B (NS-B), as given in **section 3.2.5.1**. 100 µL of NS-C reagent was added to the 100 µL of the reaction mixture and heated in boiling water bath for 20 min to deactivate the enzyme and for the reaction between NS-C and reducing sugars. The reaction mixture was cooled to 25°C and 100 µL of arsenomolybdate colour reagent (NS-D) reagent was added. Further, 700 µL of sodium phosphate (50 mM, pH 7.5) buffer was added to the mixture to bring its volume to 1.0 mL and the absorbance at 500 nm (A_{500}) was recorded using a UV-Vis spectrophotometer (Multiskan SkyHigh, Thermo Fisher Scientific, Waltham, MA, USA).

3.2.5.1 Preparation of NS-reagents for reducing sugar estimation

Reagent A

Chemicals	Amount (g)
Sodium carbonate anhydrous	6.25
Sodium potassium tatarate	6.25
Sodium bicarbonate	5.0
Sodium sulphate anhydrous	50.0

After dissolving the above-mentioned ingredients in 100 mL of deionized water it was mixed properly using a magnetic stirrer and the final volume was adjusted to 250 mL. Whatman No. 1 filter paper was used to filter the resulting solution and kept at 30°C.

Reagent B

The Reagent B was prepared by taking 15 g of copper sulfate ($CuSO_4$) was dissolved in 50 mL of deionized water and one or two drops of concentrated sulfuric acid were added. The solution was filtered through Whatman No. 1 filter paper and the final volume was adjusted to 100 mL. The solution was kept at 30°C for further use.

Reagent C

Reagents A and B were mixed in a 25:1 ratio to prepare Reagent C. The reagent C was always freeashly prepared for all assays.

Reagent D

A two-step procedure was used for making Reagent D in dark condition. A beaker, 100 mL was used to dissolve 2.5 g of ammonium molybdate in 45 mL of deionized water. Concentrated sulfuric acid, 2.1 mL was added to it. Separately, 0.3 g of sodium arsenate was dissolved in 2.5 mL of deionized water in another beaker. The total volume was made uo to 50 mL after mixing the sodium arsenate solution with the

ammonium molybdate solution. The resulting solution was filtered through Whatman No. 1 filter paper kept at 37°C for 24 h. Following the incubation, the solution was stored at 30°C for further use.

3.2.5.2 Generation of the standard plot of D-glucose

D-glucose standard plot was generated by varying its concentration between 10 and 100 µg/mL with a gradient of 10 µg/mL. The NS-reagent C, 100 µL of was added separately to all the 1.5 mL microcentrifuge tube containing 50 mM sodium phosphate buffer, pH 7.5 with varying concentrations of D-glucose in a constant volume of 100 µL. It was then kept boiling water bath for 20 min. After heating, the reaction mixture was cooled to 25°C. Finally, 100 µL of NS-reagent D was added and the final volume was adjusted to 1.0 mL by adding 700 µL of sodium phosphate (50 mM, pH 7.0) buffer. The absorbance at 500 nm (A_{500}) of the mixture was recorded on an UV-Visible spectrophotometer (Thermo, Multiskan) taking the same buffer as blank. One A_{500} equivalent of D-xylose (µg/mL) was determined from the standard plot of A_{500} versus D-xylose concentration (µg/mL). One A_{500} equivalent of D-xylose (µg/mL) was converted to mg/mL in order to calculate the enzyme activity.

3.2.5.3 Calculation of enzyme activity of recombinant proteins

The enzyme activity was determined in units per millilitre (U/mL) whilst, the specific activity of the enzyme is expressed in units per milligram (U/mg). The amount of enzyme required to produce one µmole of glucose per minute was established to be one unit (U) of enzyme (1 µmol/min). Subsequently, the enzyme activity (U/mL or µmol/min/mL) is defined as the quantity of enzyme per milliliter required to produce one µmole of glucose every minute. Whereas, the specific enzyme activity (U/mg or µmol/min/mg) is defined as the amount of enzyme per milligram needed to produce

one µmole of glucose per minute under the ideal reaction conditions. The enzyme activity of the purified enzymes were determined by using the equation as described below,

$$\text{Enzyme activity (U/mL)} = \frac{\Delta A_{500} \times C \times V}{150 \times t \times v} = (\mu\text{mole/min/mL}) = (\text{U/mL})$$

$$\text{Specific activity (U/mL)} = \frac{\text{Enzyme activity (U/mL)}}{\text{Protein concentration (mg/mL)}} = (\mu\text{mol/min/mg}) = (\text{U/mg})$$

Where,

ΔA_{500} = change in absorbance of the sample at 500 nm
C = 1 A_{500} equivalent D-glucose concentration from the standard plot
V = volume of the reaction mixture (mL)
t = time of reaction (min)
150 = molecular weight of D-glucose
v = volume of the enzyme taken in assay (mL) for reducing sugar estimation.

3.2.6 Substrate specificity of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A and *At*GH9C

The hydrolysing activity of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A and *At*GH9C was examined against different carbohydrate polymers (Carboxy methylcellulose (CMC), Lichenan, β-D-glucan from Barley, Hydroxy ethyl cellulose (Medium viscosity), Avicel, Beechwood xylan, Laminarin, Carob galactomannan and Curdlan with β-1,3: β-1,4 or β-1,6 linkages. The reaction mixture (100 µL) comprised 90 µL of 1.1 % (w/v) of substrate in 50 mM sodium phosphate buffer, pH 7.5 and 10 µL of purified *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A or *At*GH9C at final concentration of 10 µg/mL. Each of the reaction was incubated at 55°C for 1 min and the reducing sugar produced was quantified as described earlier [20,21]. For Avicel, the reaction was carried out at 55°C for 1 h with shaking at 200 rpm. Referring D-Glucose as standard, the specific activity was calculated. One unit of specific activity of the enzymes was defined as the µmole of glucose produced from the substrate per minute

per mg of the respective enzyme (µmol/min/mg) under the optimum reaction conditions. Statistical significance was estimated through one-way ANOVA with the aid of MINITAB software at 95 % confidence interval.

3.2.7 Biochemical characterization of *At*GH9C-CBM3A-CBM3B

3.2.7.1 Optimum pH, temperature and stability

The influence of pH on the enzymatic activity of *At*GH9C-CBM3A-CBM3B was studied by subjecting the purified enzyme to different buffer and pH conditions: citrate-phosphate (pH 3.0–7.0), sodium phosphate (pH 6.0–8.0) and Tris-HCl (pH 7.0–9.0) at 50 mM buffer concentration. For determination of optimum pH, the reaction mixture (100 µL) consisted of 90 µL of 1.1 % (w/v) of CMC-Na as substrate and 10 µL of purified *At*GH9C-CBM3A-CBM3B (10 µg/mL) dissolved in respective pH buffers, was incubated at 55°C for 1 min. The reducing sugar released was quantified as described earlier in **section 3.2.5** and the enzyme activity was determined. For optimum temperature, the activity of *At*GH9C-CBM3A-CBM3B was determined with in a temperature range, 25°C -100°C by incubating its 100 µL reaction mixture containing 10 µg/mL enzyme and 1% (w/v) final CMC concentration in 50 mM sodium phosphate pH 7.5 for 1 min. The concentration of released reducing sugar and the enzyme activity was determined as mentioned earlier. The stability of the enzyme at different pH conditions was determined by incubating 100 µL enzyme (10 µg/mL) in various buffers of 50 mM concentration (citrate-phosphate, pH 3.0 - 7.0; sodium phosphate, pH 6.0 - 8.0 and Tris-HCl, pH 7.0 - 9.0) at 55°C for 90 min. An aliquot of 10 µL of the incubated enzyme (10 µg/mL) was taken for 100 µL reaction mixture containing 90 µL of 1.1 % (w/v) of CMC-Na dissolved in 50 mM sodium phosphate pH 7.5. The specific activity was determined from the released reducing sugar by incubating the reaction mixture at

optimum condition of 55°C for 1 min. The released reducing sugar was estimated and the specific activity was determined as described in **section 3.2.5**. The thermostability of *At*GH9C-CBM3A-CBM3B was analysed by incubating 100 µL (10 µg/mL) enzyme in 50 mM sodium phosphate buffer (pH 7.5) at different temperatures between 25°C and 100°C for 90 min. 10 µL of enzyme was taken for assay and the specific activity was calculated under optimized conditions of pH 7.5 and 55°C.

3.2.7.2 Determination of kinetic parameters

The kinetic parameters of *At*GH9C-CBM3A-CBM3B were determined under optimum condition using different substrate concentrations of CMC, β-D-glucan and lichenan in 50 mM sodium phosphate buffer (pH 7.5) at optimum temperature of 55°C and incubating for 1 min. The equivalent concentration of substrate was treated as blank and the enzymatic activity was calculated accordingly. Michaelis-Menten equation and its double reciprocal plot (Lineweaver-Burk plots) were created by using GraphPad Prism v8.0.1 to determine the kinetic parameters like V_{max}, K_m, turnover (K_{cat}) and catalytic efficiency (K_{cat}/K_m).

3.2.7.3 Effect of metal-ions and additives on enzyme activity

The effect of metal-ions and additives on the activity of *At*GH9C-CBM3A-CBM3B was explored by incubating the enzyme in presence of various metal ions (Na^+ (NaCl), K^+ (KCl), Ca^{2+} ($CaCl_2$), Mg^{2+} ($MgCl_2$), Mn^{2+} ($MnCl_2$), Co^{2+} ($CoCl_2$) and Ni^{2+} ($NiSO_4$), Fe^{2+} ($FeSO_4$), Zn^{2+} ($ZnCl_2$) and Cu^{2+} ($CuSO_4$) and chelating agents (EDTA and EGTA), SDS at 1 and 10 mM and Urea at 1M and 5M concentrations in 100 µL reaction mixture containing 1% (w/v) of CMC-Na in 50 mM sodium phosphate buffer pH 7.5. The reaction was carried out at 55°C for 1 min. The released reducing sugar was estimated as described earlier. The enzyme activity of *At*GH9C-CBM3A-CBM3B,

without any additive, was considered as 100%, to calculate the relative activity of the enzyme with additive.

3.2.8 Catalytic mechanism of *At*GH9C-CBM3A-CBM3B

Thin layer chromatography (TLC Silica gel 60 $F_{254,}$ size 10 × 10 cm, Merck) analysis of the hydrolysed products of the CMC-Na by *At*GH9C-CBM3A-CBM3B was performed to decipher its catalytic mode of action of *At*GH9C-CBM3A-CBM3B. One hundred microliter of reaction mixture containing 10 µL of *At*GH9C-CBM3A-CBM3B (100 µg/mL) and 90 µL of 1.1% (w/v) CMC dissolved in 50 mM sodium phosphate buffer (pH 7.5) was incubated at 55°C for various time intervals from 1 min to 24 h (1, 2, 5, 10, 20, 30 min and 1, 3, 6, 12, 24 h). One microliter from each concentrated enzyme hydrolysate after ethanol precipitation [22] and 1 µL (1.0 mg/mL) of mixture of standard (glucose, cellobiose, cellotriose and cellotetraose) were loaded onto the TLC plate. The mobile phase used was a mixture of n-butanol/acetic acid/water in the ratio of 2:1:1. The chromatogram was visualized under visualization solution (sulphuric acid: methanol 5:95 (v/v) and α-naphthol 0.5%, w/v) as reported earlier [15]. An aliquot of 10 µL from each hydrolysate at different incubation time was subjected to reducing sugar analysis by the method described in **section 3.2.5**.

3.2.9 Thermal denaturation analysis of *At*GH9C-CBM3A-CBM3B

The unfolding transition midpoint (T_m) of *At*GH9C-CBM3A-CBM3B was analysed by incubating 75 µg/mL enzyme in 50 mM sodium phosphate buffer (pH 7.5) at different temperatures, ranging from 25°C to 100°C on a CD spectrophotometer equipped with peltier temperature controller (J-1500, Jasco Corporation, Tokoyo, Japan) by using the temperature-wavelength scan program. The secondary structure was determined in the far UV range (190-260 nm) at scanning rate of 50 nm/min and 1

nm bandwidth with an average of 5 scans. The plot of the molar ellipticity (Δε) at 222 nm vs temperature was generated to find the T_m [23]. The analyses were separately carried out by adding 10 mM $MgCl_2$, $CaCl_2$ or 1 mM EDTA to elucidate their effects on the unfolding transition midpoint (T_m) of enzyme, *At*GH9C-CBM3A-CBM3B.

3.2.10 Interaction of carbohydrate binding modules, CBM3A and CBM3B with catalytic module *At*GH9C

The modulation in activity of the catalytic module *At*GH9C in association with the carbohydrate binding modules, CBM3A and CBM3B was investigated to decipher their role in catalysis. The activity of catalytic module, *At*GH9C (7 µM) was analyzed in presence and in absence of an equimolar concentration of CBM3A (7 µM) or CBM3B (7 µM) against different substrates (Carboxy methylcellulose (CMC), Lichenan, β-D-glucan from Barley and Avicel) to investigate the role of CBM3A and CBM3B in the enzyme catalysis. Fixed concentration of *At*GH9C (7 µM) with different concentrations of CBM3A (7 µM–35 µM), CBM3B (7 µM–35 µM), CBM3A+CBM3B (7 µM–35 µM) was incubated for 1h at 25°C. The enzyme volume of 10 µL was taken from preincubated enzyme mixture and incubated in 100 µl reaction mixture containing 1% (w/v) final concentration of CMC-Na in sodium phosphate buffer (pH 7.5) at 55°C for 1 min to investigate the influence of different molar concentrations of CBM3A and CBM3B on *At*GH9C activity. Statistical significance was estimated through one-way ANOVA with the aid of MINITAB software at 95 % confidence interval.

3.2.11 Effect of CBM3A and CBM3B on the thermostability of catalytic module *At*GH9C

The thermostability of *At*GH9C was analysed in presence and absence of CBM3A, CBM3B or CBM3A+CBM3B. This was performed by incubating 100 µL purified 7 µM *At*GH9C only or in a mixture of 7 µM *At*GH9C and 7 µM of each of

CBM3A, CBM3B or CBM3A+CBM3B at different temperatures ranging from 30°C to 100°C for 90 min. 10 µL from each enzyme mixture was taken for assay and the specific activity was determined against 1% (w/v) carboxymethyl cellulose in 50 mM sodium-phosphate buffer (pH 7.5) at optimum conditions of pH 7.5 and 55°C for 1 min. The amount of reducing sugar generated was determined from the procedure mentioned in **section 3.2.5** and the specific activity was calculated.

3.3. Results and Discussion

3.3.1 Expression and purification of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B

All the recombinant proteins, *At*GH9C-CBM3A-CBM3B and its truncated derivatives, *At*GH9C-CBM3A, *At*GH9C and CBMs i.e., CBM3A, CBM3B for this study were expressed using *E. coli* BL21 (DE3) cells and method of immobilized metal ion affinity chromatography (IMAC) was applied to purify the proteins to homogeneity. The purified proteins were further analysed by 12% (w/v) SDS-PAGE electrophoresis for molecular mass and purity (**Fig. 3.1a**). The purified recombinant proteins *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B on 12% (w/v) SDS-gel depicted an agreement with the calculated theoretical molecular mass from their amino acid sequence of 97.2 kDa, 72.3 kDa, 54.4 kDa, 16.4 kDa and 17.9 kDa respectively. The *At*GH9C-CBM3A-CBM3B was analyzed under native PAGE (7.5%, w/v) conditions (**Fig. 3.1b**), which revealed a single homogeneous band, confirming the absence of degradation and the structural stability of the purified proteins. The purified proteins were further utilized for biochemical characterization. The enzyme activity and protein yields of all the catalytic modules (*At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C) after of purification are displayed in **Table 3.3**. The purification fold of full-length *At*GH9C-CBM3A-CBM3B, truncated derivative *At*GH9C-CBM3A and the catalytic module *At*GH9C were 19.8, 15.4 and 2.3, respectively. The lower purification fold for *At*GH9C observed could be due to its lower or ineffective catalysis of the substrate CMC-Na.

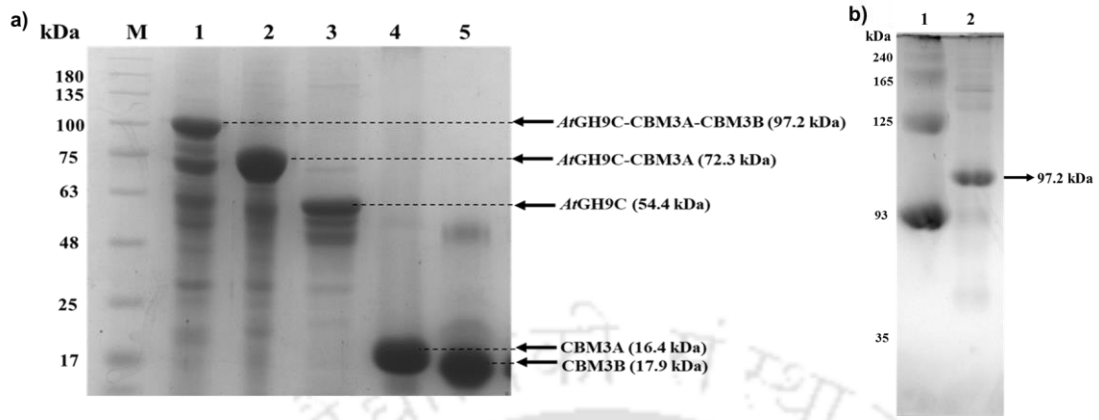


Fig. 3.1 Purified recombinant proteins on a) SDS-PAGE gel (12%, w/v): Lane M− Molecular mass marker (11–180 kDa, Hi-media, Pvt. Ltd. India), lane 1- *At*GH9C-CBM3A-CBM3B, lane 2- *At*GH9C-CBM3A, lane 3- *At*GH9C, lane 4- CBM3A and lane 5- CBM3B and b) Native gel (7.5%, w/v): Lane 1- BSA and lane 2- *At*GH9C-CBM3A-CBM3B.

Table 3.3 Purification table for the recombinant proteins.

Purification step	Volume (mL)	Enzyme Activity*			Protein		Specific activity (U/mg)	Purification fold
		U/mL	Total Activity (U)	Activity yield (%)	mg/mL	Total protein (mg)		
***At*GH9C-CBM3A-CBM3B**								
Cell extract	15	74.6	1119.2	0	27.2	408	2.7	-
IMAC	6	87.2	523.2	46.7	1.6	9.6	54.5	19.8
***At*GH9C-CBM3A**								
Cell extract	15	62.2	931.7	0	23.8	357	2.6	-
IMAC	5	56.2	281	30.2	1.4	7	40.2	15.4
***At*GH9C**								
Cell extract	12	52.6	631.2	0	21.5	258	2.4	-
IMAC	4	4.5	18	2.8	0.8	3.2	5.6	2.3

*Enzyme activity was measured using CMC-Na (1 %, w/v) as substrate dissolved in 50 mM sodium phosphate buffer (pH 7.5) under optimized conditions.

3.3.2 Substrate specificity of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C

The hydrolysing ability of the recombinant enzymes *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C in terms of specific activity, against different cellulosic polysaccharides is listed in **Table 3.4**. The full-length endoglucanase

*At*GH9C-CBM3A-CBM3B displayed maximum hydrolysing ability against β-1,4 linked anionic cellulose derivative CMC-Na (58.8 ± 2.6 U/mg), followed by β-1,3; β-1,4 linked lichenan (44.5 ± 2.0 U/mg), β-D-glucan (36.2 ± 1.7 U/mg) and hydroxy ethyl cellulose (17.9 ± 1.2 U/mg). It depicted marginal enzymatic activity against crystalline substrate avicel (0.80 ± 0.2 U/mg). The truncated derivative without one terminal CBM, *At*GH9C-CBM3A exhibited reduction in enzymatic activity as compared with the full-length module but showed similar trend in substrate hydrolysis showing highest activity with CMC-Na (53.4 ± 2.4 U/mg), trailed by lichenan (40.2 ± 2.8 U/mg), β-D-glucan (35.2 ± 1.6 U/mg), hydroxy ethyl cellulose (11.8 ± 0.4 U/mg) and avicel (0.24 ± 0.1 U/mg). Both, the full-length and truncated modules were unable to act on substrate laminarin, beechwood xylan, carob galactomannan and curdlan which are either β-1,3; β-1,6 linked β-D-glucose or not having the β-D-glucose as its simple sugar unit.

Table 3.4 Substrate specificity of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A and *At*GH9C against different substrates.

Substrate (1%, w/v)	**Specific Activity* (U/mg)**		
	***At*GH9C-CBM3A-CBM3B**	***At*GH9C-CBM3A**	***At*GH9C**
Carboxy methylcellulose (CMC)	58.8 ± 2.6	53.4 ± 2.4	2.3 ± 0.6
Lichenan	44.5 ± 2.0	40.2 ± 2.8	1.9 ± 0.3
β-Glucan	36.2 ± 1.7	35.2 ± 1.6	0.8 ± 0.3
Hydroxy ethyl cellulose	17.9 ± 1.2	11.8 ± 0.4	--
Avicel	0.8 ± 0.2	0.24 ± 0.11	--
Beechwood xylan	--	--	--
Laminarin	--	--	--
Carob galactomannan	--	--	--
Curdlan	--	--	--

*All the experiments were performed in triplicate (n=3) and mean ± SD for each experiment is depicted here. Significant differences ($p \leq 0.05$) were noted between *At*GH9C-CBM3A-CBM3B/*At*GH9C-CBM3A and *At*GH9C through statistical analysis using one-way ANOVA. Non-significant differences ($p > 0.05$) were observed through statistical analysis using one-way ANOVA between *At*GH9C-CBM3A-CBM3B and *At*GH9C-CBM3A.

The results indicated that the full-length, *At*GH9C-CBM3A-CBM3B and the truncated derivative *At*GH9C-CBM3A, both are able to cleave either β-1,4 cellulosic substrate (CMC-Na) or mixed β-1,3; β-1,4 cellulosic substrates (lichenan and β-D-glucan) but not β-1,3 linkage containing glucan, curdlan confirming their β-1,4-glucanase activity. None of the enzyme was able to breakdown the insoluble substrate (avicel) as efficiently the soluble one. Similar results were reported for from endoglucanase EglA from *Bacillus pumilus* belonging to family 9 GH with CBM3, where it was able to hydrolyse cellulosic CMC-Na *in-vitro* but not avicel and xanthan gum [24]. CBP105, an endoglucanase from *Cellulomonas flavigena* of GH9 family with family 3 CBM also indicated highest activity on CMC-Na and lowest on avicel [25]. In contrast, some family 9 endoglucanases from *Clostridium thermocellum* and *Caldicellulosiruptor bescii* exhibited highest hydrolysing capacity against crystalline cellulose [26,27]. The enzyme activity of only catalytic module (*At*GH9C) was significantly low in comparison to the full-length enzyme (*At*GH9C-CBM3A-CBM3B) and exhibited only 2.3 ± 0.6 U/mg, 1.9 ± 0.3 U/mg and 0.8 ± 0.3 U/mg against CMC-Na, lichenan and β-D-glucan respectively. *At*GH9C was unable to hydrolyse hydroxy ethyl cellulose and avicel. The results indicated the obligatory requirement of carbohydrate binding modules (CBM3A or CBM3B) for enzyme hydrolysis of different substrate. Similar results from the GH9 modular endoglucanase from *Clostridium thermocellum* Cel9W [15], *Clostridium cellulolyticum* Cel9G [28] and *Acetivibrio cellulolyticus* Cel9B [12] were reported.

3.3.3 Biochemical properties of *At*GH9C-CBM3A-CBM3B

3.3.3.1 Analysis of optimum pH, temperature and stability

The maximum relative activity of full-length enzyme endoglucanase, *At*GH9C-CBM3A-CBM3B was noted at pH 7.5 (**Fig. 3.2a**) in both 50 mM of Tris-HCl and sodium phosphate buffer and at temperature 55°C (**Fig. 3.2b**). The optimal pH and temperature of the enzyme *At*GH9C-CBM3A-CBM3B was comparable with the pH 7.0 and temperature 60°C of endoglucanases *Bl*Cel9, from mesophilic bacteria *Bacillus licheniformis* [29], pH 7.5 and temperature 60°C of endoglucanase CPB105 from *Cellulomonas flavigena* [25] and a lower pH 5.5 and temperature 60°C of endoglucanase cel9K from *Paenibacillus sp.* X4 [30]. The hydrolytic efficiency of the enzyme towards cellulosic substrate remained stable in wide range of pH (6.0-9.0) in 50 mM of both citrate phosphate and sodium phosphate buffer (**Fig. 3.2c**). The enzyme retained 80% of the activity at pH 6.0 and displayed 75% of activity at pH 9.0, after 90 min of incubation. The enzyme, *At*GH9C-CBM3A-CBM3B retained activity over 85% in the temperature range of 4°C to 60°C after 90 min of incubation (**Fig. 3.2d**). Above 55°C, the activity of the enzyme dropped significantly and tend to be inactive at higher temperatures between 80°C and 100°C. These results demonstrated the thermophilic nature of the enzyme and its stability at alkaline pH. The characteristic of being stable over inclusive range of temperature and pH makes the enzyme *At*GH9C-CBM3A-CBM3B a suitable candidate for industrial applications, which require enzymes functioning at elevated temperatures, having longer half-life with reduced risk of microbial contamination [31].

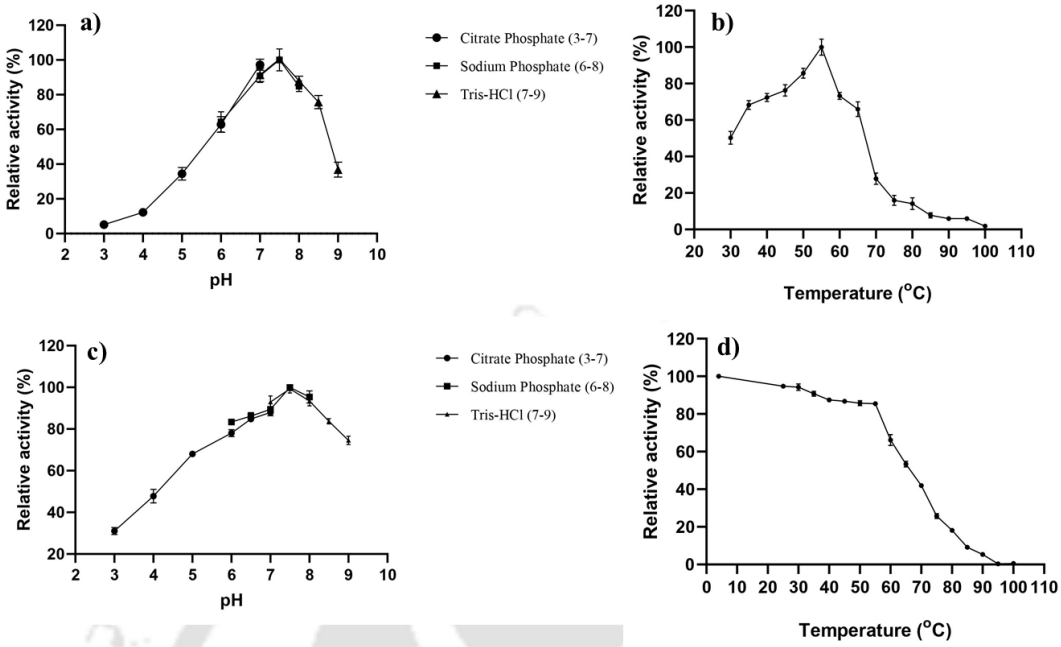


Fig. 3.2 pH and temperature optimum and stability profile of *At*GH9C-CBM3A-CBM3B using 1.0% (w/v) Carboxymethyl cellulose and 0.01 mg/ml of enzyme *At*GH9C-CBM3A-CBM3B. (a) Optimum pH for *At*GH9C-CBM3A-CBM3B activity, (b) Optimum temperature for *At*GH9C-CBM3A-CBM3B activity, (c) pH stability and (d) thermostability of *At*GH9C-CBM3A-CBM3B. All experiments were carried out in triplicates. The plots were generated using Graphpad prism 8.0.1 software and error bars represent the standard deviation (±SD) of each datum point.

3.3.3.2 Analysis of kinetic parameters of *At*GH9C-CBM3A-CBM3B

The kinetic parameters of endoglucanase, *At*GH9C-CBM3A-CBM3B against various cellulosic polysaccharides (CMC-Na, lichenan and β-D-glucan) was assessed and are summarized in **Table 3.5**. The enzyme displayed maximum velocity, V_{max} of 61 ± 4 U/mg against CMC-Na (**Fig. 3.3a**) followed by V_{max} of 50 ± 4 U/mg and 39 ± 3 U/mg against lichenan (**Fig. 3.3b**) and β-D-glucan (**Fig. 3.3c**). Various other kinetic parameters like K_m, catalytic efficiency (K_{cat}/K_m) and turnover (K_{cat}) against different substrate was deciphered after analysing the data using Michaelis-Menten plot and Lineweaver-Burk plots. The K_m of 1.5 ± 0.03 mg/mL was obtained for CMC-Na. The values of K_m against CMC-Na obtained from the earlier characterized endoglucanases

CenC, CelT and Cel9W from *Acetivibrio thermocellus* were 7.14 mg/mL, 16.7 mg/mL and 2.7 mg/mL, respectively [15,32,33]. The lower K_m of *At*GH9C-CBM3A-CBM3B towards CMC-Na, demonstrated its strong substrate affinity. The turnover (K_{cat}) and catalytic efficiency (K_{cat}/K_m) of *At*GH9C-CBM3A-CBM3B against CMC-Na was analysed to be 98.4 ± 0.9 s^{-1} and 65.5 ± 0.5 $mL.mg^{-1}s^{-1}$ respectively, highest among all substrates used for the study. The earlier reports on endoglucanases from GH9 family displayed higher affinity towards the natural substrates (lichenan and β-D-glucan) than the CMC-Na [15,32].

Table 3.5 Kinetic parameters of *At*GH9C-CBM3A-CBM3B against different substrates.

Substrate	V_{max} (U/mg)	K_m(mg/mL)	K_{cat} (s^{-1})	K_{cat}/K_m (mL/mg/s)
CMC	61 ± 4	1.5 ± 0.3	98.4 ± 0.9	65.5 ± 0.5
Lichenan	50 ± 4	1.7 ± 0.4	80.3 ± 0.6	47.2 ± 0.6
β-D-glucan	39 ± 3	2.7 ± 0.4	63.6 ± 0.5	23.3 ± 0.6

The data shown here are the mean of three experimental replicates (n=3) ± SD.

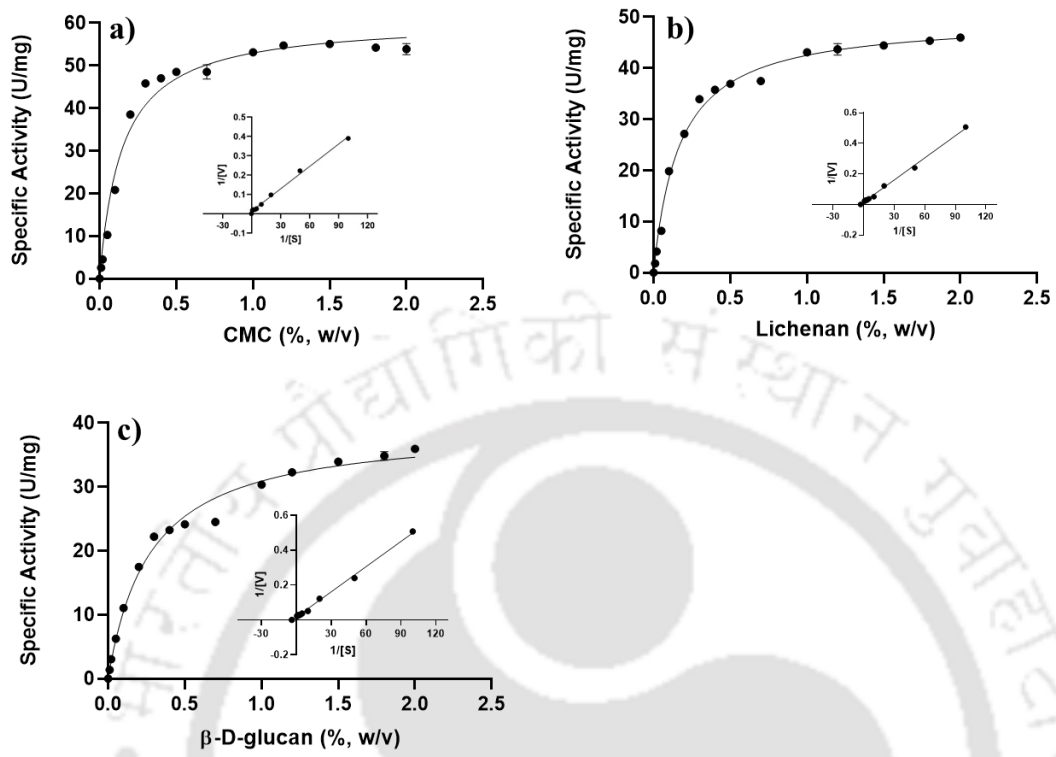


Fig. 3.3 Michaelis-Menten and Lineweaver-Burk plots (inset) for determining the kinetic parameters of *At*GH9C-CBM3A-CBM3B against (a) Carboxy methylcellulose, (b) Lichenan and (c) β-glucan. The data shown here is the mean of three experimental replicates with error bars showing ±SD.

3.3.3.3 Analysis of metal-ion and additive effects on the enzyme activity

The mono and divalent metal-ions and additives at tested 1 mM and 10 mM concentration had significant influence on the enzymatic activity of *At*GH9C-CBM3A-CBM3B on assaying against CMC-Na (**Table 3.6**). At 1 mM and 10 mM, Mg^{2+} had elevated the activity by 19% and 29%, respectively as compared with the control. The crystal structure analysis of endoglucanase *Bo*GH9 from Bacteroidetes phylum pointed the presence of Mg^{2+} ion on the surface of the enzyme coordinated with two aspartate residues and three water molecules help in catalysis [34]. However, in case of divalent ions Ca^{2+} and Mn^{2+}, the enhancement in enzyme activity was not observed, while

comparing with the control. This result corroborated with the findings of the endoglucanase, *Bl*cel9 from *Bacillus licheniformis* which showed no stimulatory effect of calcium ion on the enzyme activity [29]. Nonetheless, in endoglucanase Cel9K from *Paenibacillus sp.* X4 [30] and CenC from *C.thermocellum* [32], Ca^{2+} ion increased the activity by 20% and 25%, respectively. The presence of divalent cations Zn^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+} or Fe^{2+} adversely affected the CMCase activity of *At*GH9C-CBM3A-CBM3B. An abrupt decrease in catalytic activity of the enzyme over 70% was observed in the presence of 1 mM chelating agents EDTA or EGTA. Evidences in activity reduction of the enzyme in presence of metal-ion chelator and surge in activity in presence of Mg^{2+} indicates that the native enzyme, *At*GH9C-CBM3A-CBM3B contains an intrinsically bound Mg^{2+} ion, which contributes to the catalysis by the enzyme. Interestingly, urea at 1M and 5M concentration exhibited the similar inhibitory effect on the enzyme activity by 50%, representing the invariance of concentration of urea after 1M.

Table 3.6 Effect of metal-ions/reagents on the enzymatic activity of *At*GH9C-CBM3A-CBM3B. Mean ± SD of the triplicate (n=3) result is shown here.

Metal ion/reagent	Relative activity (%) at 1.0 mM	Relative activity (%) at 10.0 mM
Control	100 ± 3.0	100 ± 4.0
Mg^{2+}	119.7 ± 1.2	129.4 ± 3.0
Ca^{2+}	97.4 ± 1.8	113.4 ± 1.6
Mn^{2+}	88.4 ± 2.2	86.9 ± 3.8
Zn^{2+}	60.6 ± 0.8	60.8 ± 1.8
Ni^{2+}	52.1 ± 1.8	50.1 ± 3.0
Co^{2+}	44.4 ± 1.2	42.7 ± 2.3
Fe^{2+}	38.2 ± 2.8	38.5 ± 2.6
Cu^{2+}	23.2 ± 1.0	17.5 ± 0.6
Na^{+}	60.2 ± 0.4	67.4 ± 1.4
EDTA	26.4 ± 1.2	8.5 ± 0.6
EGTA	22.2 ± 0.8	14.8 ± 1.1
SDS	39.7 ± 1.7	33.8 ± 1.8
Urea	53.7 ± 2.6 (1M)	51.3± 2.6 (5M)

3.3.4 Thermal denaturation analysis of *At*GH9C-CBM3A-CBM3B

Protein unfolding transition curve analysis of *At*GH9C-CBM3A-CBM3B was performed to investigate the unfolding transition midpoint (T_m) of protein using the data curated from the Far UV-Circular Dichroism at 222 nm [23]. T_m of 65°C was shown by the native *At*GH9C-CBM3A-CBM3B (**Fig. 3.4a**). The T_m of *At*GH9C-CBM3A-CBM3B was elevated by 5°C to 70°C with the addition of 10 mM Mg^{2+} ions (**Fig. 3.4b**), but it remained unaffected by 10 mM Ca^{2+} (**Fig. 3.4c**). However, in the presence of chelating agent (EDTA) at 1 mM, the transition midpoint dropped noticeably to 60°C showing a shift of 5°C as compared with the native enzyme (**Fig. 3.4d**). The results indicate that Mg^{2+} enhances the catalytic activity and thermal stability of *At*GH9C-CBM3A-CBM3B, suggesting a possible stabilizing interaction. However, the intrinsic presence or binding of Mg^{2+} ion(s) within the native enzyme remains to be confirmed through ICP-MS or related metal-analysis techniques. On addition of EDTA the removal of bound metal-ion of native *At*GH9C-CBM3A-CBM3B takes place, introducing the conformational instability in the protein. This observation is in concurrence with the previous results where incubation of *At*GH9C-CBM3A-CBM3B with 10 mM Mg^{2+} ions increased the activity by 29% and the presence of Ca^{2+} ions doesn't interfere with the activity, but 1 mM EDTA significantly drops the activity of the enzyme. The results of thermal denaturation analysis and enzyme activity revealed a role for naturally occurring Mg^{2+} ion(s) in the structural and conformational integrity and activity of the native endoglucanase *At*GH9C-CBM3A-CBM3B.

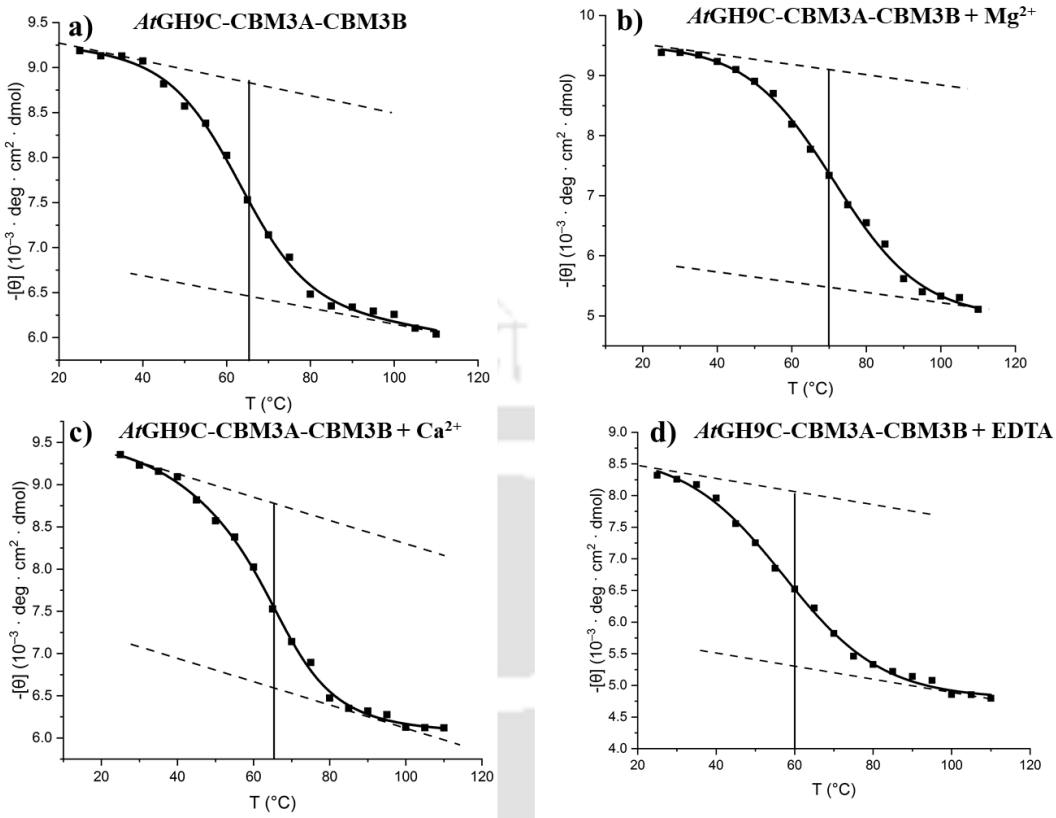


Fig. 3.4 Unfolding transition midpoint (T_m) curve of (a) *At*GH9C-CBM3A-CBM3B (75 µg/mL), (b) *At*GH9C-CBM3A-CBM3B + Mg^{2+} (10 mM), (c) *At*GH9C-CBM3A-CBM3B + Ca^{2+}(10 mM), and (d) *At*GH9C-CBM3A-CBM3B + EDTA (1 mM). The unfolding transition midpoint (T_m) were generated by determining the secondary structure of protein *At*GH9C-CBM3A-CBM3B at 222 nm and the mid-point of the sigmoidal curve gives the T_m. The data represent the average of 5 scan with error bars depicting ± SD.

3.3.5 Catalytic mechanism of *At*GH9C-CBM3A-CBM3B

A time dependent (1 min to 24 h) chromatogram of hydrolysed products of *At*GH9C-CBM3A-CBM3B generated against CMC-Na on a TLC plate, revealed the presence of cellotriose, cellotetraose and higher cello-oligosaccharides after 2 min of reaction (**Fig. 3.5a**). Cellotetraose and higher cello-oligosaccharides (DP>4) were prevalent in the hydrolysed CMC chromatogram at all time intervals, demonstrating the endo-1,4-glucanase activity of *At*GH9C-CBM3A-CBM3B. The estimation of the

reducing sugar produced by the hydrolysis of CMC-Na by *At*GH9C-CBM3A-CBM3B at different time interval (1 min to 24 h) showed that the reaction reaches saturation at 12 h at 355 µg/mL reducing sugar concentration (**Fig. 3.5b**).

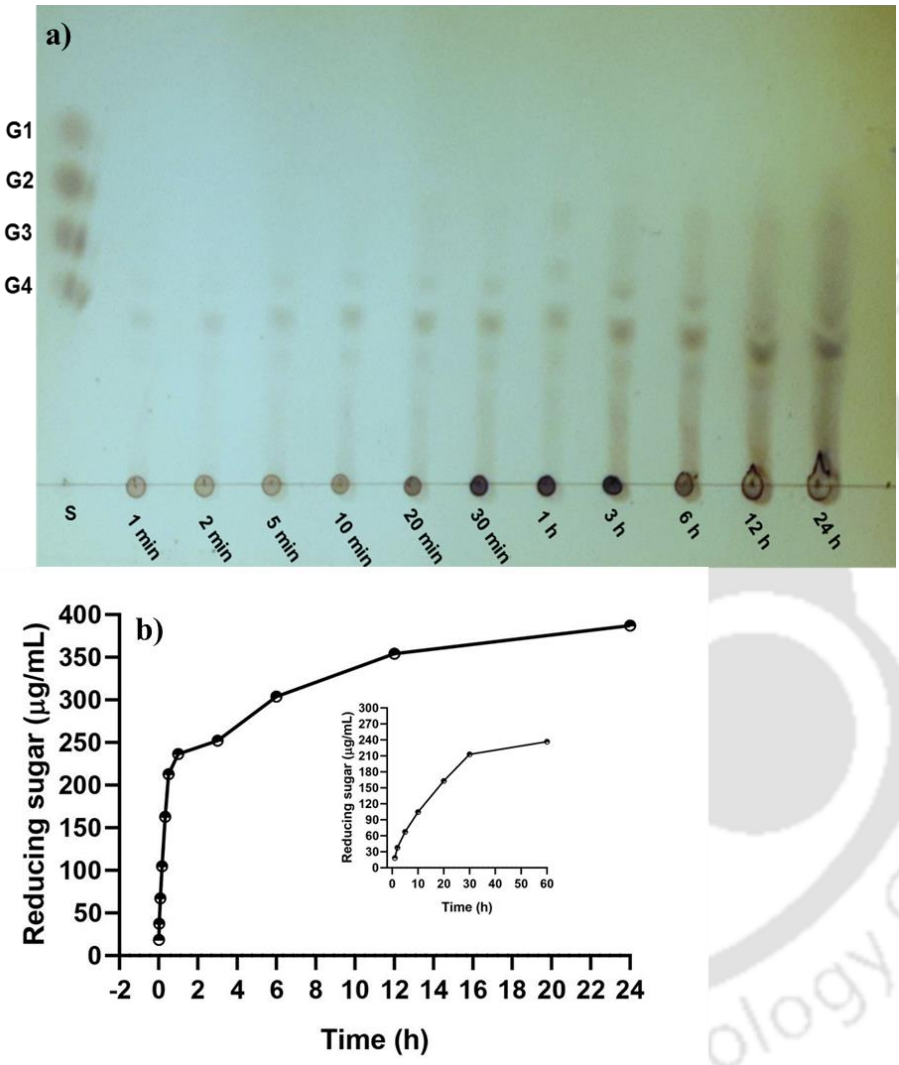


Fig. 3.5 TLC analysis of CMC hydrolyzed products by *At*GH9C-CBM3A-CBM3B at different time intervals (1 min- 24 h). (a) TLC chromatogram of time dependent hydrolysis of CMC-Na by *At*GH9C-CBM3A-CBM3B. The lane S represent standard (G1: Glucose, G2: Cellobiose, G3: Cellotriose and G4: Cellotetraose), (b) Plot of reducing sugar produced with time. Inset contains the reducing sugar released up to 60 min. The assay was performed with the 100 µg/mL final enzyme concentration of *At*GH9C-CBM3A-CBM3B at 55°C and pH 7.5 (50 mM sodium-phosphate) for 1 min to 24 h. The data represented here are the mean of three experimental replicates with error bars showing ±SD.

3.3.6 Interaction of carbohydrate binding modules, CBM3A and CBM3B with catalytic module *At*GH9C

The delineation in variation of the activity of catalytic module *At*GH9C in physical association with the carbohydrate binding modules CBM3A or CBM3B was carried out at different and equimolar concentration. Different molar ratio of individual *At*GH9C and CBM3A (*At*GH9C:CBM3A), *At*GH9C and CBM3B (*At*GH9C:CBM3B) and *At*GH9C and individual combination of CBM3A and CBM3B i.e. CBM3A+CBM3B (*At*GH9C:CBM3A+CBM3B) were used. Earlier in **section 3.3.2** it was reported that the catalytic module (*At*GH9C) alone is insufficient in effective hydrolysis of the substrate with maximum activity of 2.3 ± 0.6 U/mg against CMC-Na. The catalytic efficiency of the module *At*GH9C thus uplifted by physical association through means of mixing the catalytic module (*At*GH9C) and cognate carbohydrate binding modules (CBM3A and CBM3B) in solution. The equimolar mixture of *At*GH9C and CBM3A (*At*GH9C:CBM3A) displayed maximum activity of 27.8 ± 1.2 U/mg with CMC at 55°C, whereas, the full length *At*GH9C-CBM3A-CBM3B gave maximum activity of 58.8 ± 2.6 U/mg at 55°C (**Table 3.7**). The association of CBM3A with catalytic module (*At*GH9C) in equimolar concentration also partially recovered the activity of *At*GH9C by 47%, 57% and 31% against CMC-Na, lichenan and β-D-glucan, respectively (**Table 3.7**). Further, with increase in the molar ratio of *At*GH9C:CBM3A from 1:1 to 1:2 the enzymatic activity increased from 27.8 U/mg to 36.5 U/mg at 55°C against CMC-Na, shifting the percent recovery from 47% to 62% (**Fig. 3.6a**) with no catalytic activity observed against avicel. The mixture's enzyme activity gradually grew past the equimolar ratio of *At*GH9C:CBM3A and reached saturation at a molar ratio of 1:3. These results indicated the need of physical

association of carbohydrate binding module CBM3A with *At*GH9C and the importance of linker between them for effective catalysis.

Table 3.7 Revival of *At*GH9C activity in presence of CBMs.

Substrate (1%, w/v)	Specific Activity* (U/mg)				
	***At*GH9C-CBM3A-CBM3B**	***At*GH9C**	***At*GH9C:CBM3A (1:1 Molar ratio)**	***At*GH9C:CBM3B (1:1 Molar ratio)**	***At*GH9C:CBM3A+CBM3B (1:1 Molar ratio)**
Carboxy methyl cellulose (CMC)	58.8 ± 2.6	2.3 ± 0.6	27.8 ± 1.2	7.5 ± 0.7	29.6 ± 0.2
Lichenan	44.5 ± 2.0	1.9 ± 0.3	25.8 ± 0.8	5.8 ± 0.5	27.8 ± 0.3
β-glucan	36.2 ± 1.7	0.8 ± 0.3	11.4 ± 0.6	5.5 ± 0.3	17.6 ± 0.4
Avicel	0.8 ± 0.2	--	--	--	0.14 ± 0.1
Hydroxy ethyl cellulose	17.9 ± 1.2	--	--	--	--
Beechwood xylan	--	--	--	--	--
Laminarin	--	--	--	--	--
Carob galactomannan	--	--	--	--	--
Curdlan	--	--	--	--	--

*The data represent mean of the triplicates (n=3) ± SD for each experiment. Significant differences ($p \leq 0.05$) were noted between *At*GH9C:CBM3A /*At*GH9C:CBM3A + CBM3B and *At*GH9C:CBM3B through statistical analysis using one-way ANOVA. Non-significant differences ($p > 0.05$) were observed through statistical analysis using one-way ANOVA between *At*GH9C:CBM3A and *At*GH9C:CBM3A + CBM3B

A study on the fusion of CBM3A with the family 9 endoglucanase *Bp*GH9 from *Bacillus pumilus* increased its activity on soluble cellulose corroborates our findings [35]. On contrary to the above results, when the association of CBM3B and catalytic module (*At*GH9C) at equimolar concentration was tested, the catalytic activity revival against different substrates was not that prominent. Equimolar concentration *At*GH9C and CBM3B (*At*GH9C:CBM3B) recovered the activity of *At*GH9C against CMC-Na, lichenan and β-D-glucan by 13% (7.5 ± 0.7 U/mg), 13% (5.8 ± 0.5 U/mg) and 15% (5.5 ± 0.3 U/mg), respectively and no catalytic activity observed against avicel (**Table 3.7**). Perhaps the reason for reduced revival of activity resides, in the physical location of the CBMs with corresponding catalytic module, the number of aromatic amino acid residues that are present in CBMs and on the nature of the substrate used for study [8].

However, increase in the molar ratio of *At*GH9C:CBM3B from 1:1 to 1:3 the enzymatic activity increased from 7.2 U/mg to 13.7 U/mg at 55°C against CMC-Na and that hits the saturation after 1:3 molar ratio (**Fig. 3.6a**). This indicated the need of higher molar concentration of a farther positioned CBM3B for better revival of the catalytic activity of enzyme. The mixture containing CBM3A and CBM3B (CBM3A+CBM3B) in equimolar ratio and catalytic module *At*GH9C i.e. *At*GH9C: (CBM3A+CBM3B) showed higher percentage of revival of the catalytic activity as compared with the individual CBM3A or CBM3B. Nevertheless, the activity revival was only marginally higher than the mixture *At*GH9C: CBM3A against CMC-Na, lichenan and β-D-glucan, respectively (**Table 3.7**). It was noteworthy here that the mixture *At*GH9C:CBM3A+CBM3B at equimolar concentration was able to hydrolyse crystalline avicel and revive its activity to 0.14 (U/mg) by 18% as compared with 0.80 (U/mg) for the native endoglucanase *At*GH9C-CBM3A-CBM3B. This finding may be settled with the information that CBM3B have higher propensity to bind the crystalline cellulose compared to the amorphous one because of π- π interaction between the large proportion of aromatic amino acids in CBM3B and planer crystalline surface [36]. Interestingly, by increasing the molar ratio of *At*GH9C:CBM3A+CBM3B beyond 1:3, the enzymatic activity reaches close to the native *At*GH9C-CBM3A-CBM3B enzyme against particular soluble substrates (CMC-Na, lichenan and β-D-glucan) with an exception hydroxy ethyl cellulose was used as substrate. This finding further

strengthens the contention of the cross-talk between the CBMs for enhancing the activity of the catalytic module *At*GH9C.

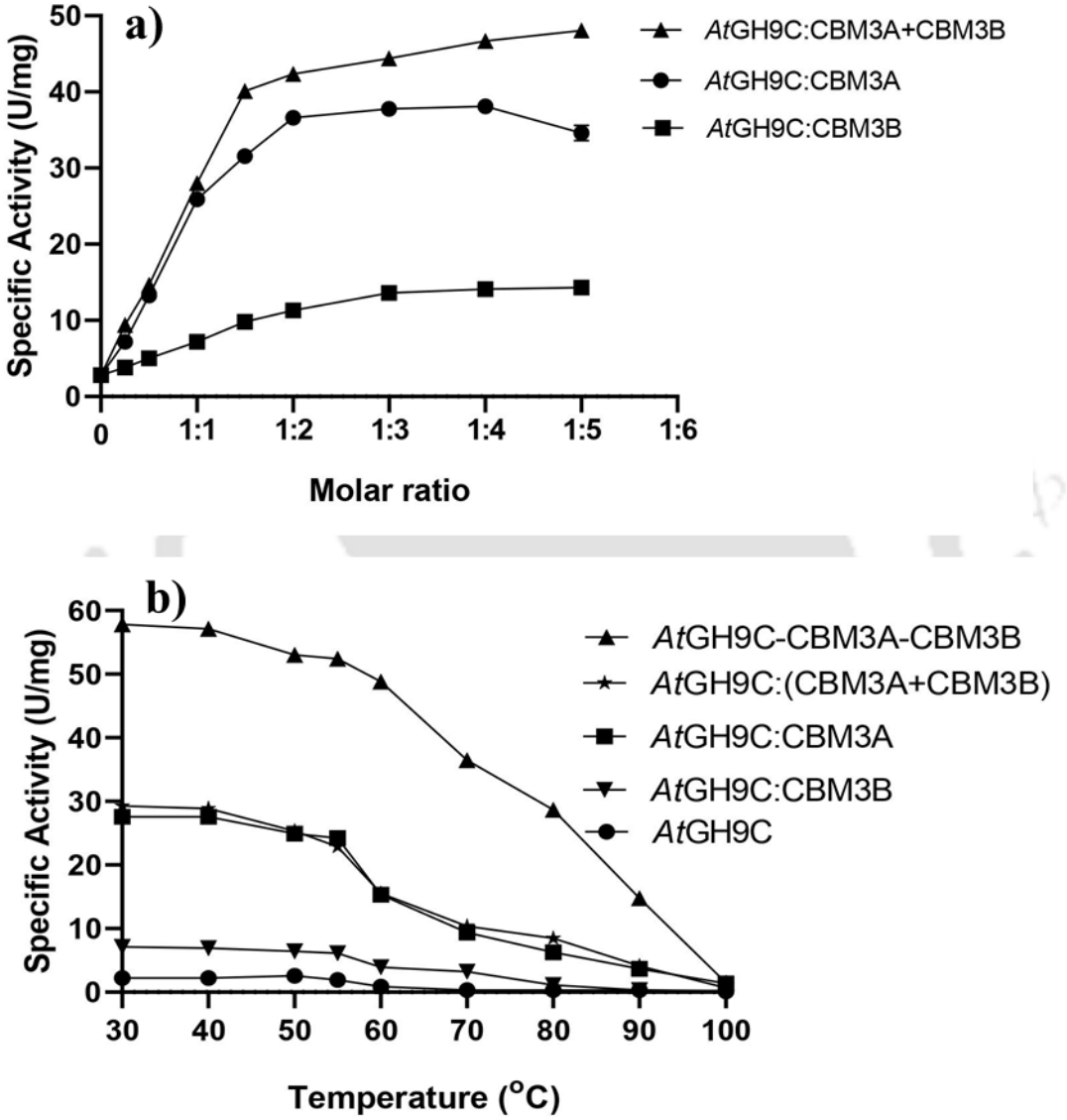


Fig. 3.6 Interaction study of catalytic module, *At*GH9C and CBMs (CBM3A, CBM3B or CBM3A+CBM3B). (a) The effect of different molar concentrations of CBM3A, CBM3B and CBM3A+CBM3B on the activity of *At*GH9C against CMC-Na. The enzyme activity was determined at optimal conditions. (b) Activity of *At*GH9C with the equimolar concentrations of CBM3A, CBM3B or CBM3A+CBM3B against CMC-Na at different temperature ranging from 30°C to 100°C. The data represented here is the mean of three experimental replicates with error bars showing ±SD.

3.3.7 Thermostability of carbohydrate binding modules, CBM3A and CBM3B with catalytic module *At*GH9C

Thermostability of catalytic module *At*GH9C with equimolar concentration of carbohydrate binding modules CBM3A, CBM3B or CBM3A+CBM3B was investigated by incubating the enzyme mixture at different temperature for 90 min followed by reducing sugar analysis at optimum temperature 55°C and pH 7.5. Native endoglucanase, *At*GH9C-CBM3A-CBM3B showed 55 U/mg specific activity at 50°C (**Fig. 3.6b**). The thermostability of the catalytic module *At*GH9C at all temperatures was minimal or negligible as compared with *At*GH9C-CBM3A-CBM3B (**Fig. 3.6b**) as it showed only 2.4 U/mg activity up to 50°C. The equimolar mixture of *At*GH9C and CBM3A (*At*GH9C:CBM3A) showed significant increase in the thermostability of *At*GH9C by displaying activity up to 24.9 U/mg at 50°C. On contrary, the equimolar mixture of *At*GH9C and farther located CBM3B could revive the thermostability only to 6.4 U/mg at 50°C. Moreover, the equimolar mixture of *At*GH9C and CBM3A+CBM3B, (*At*GH9C: CBM3A+CBM3B) displayed activity only 28.8 U/mg at 50°C showing similar thermostability as by *At*GH9C:CBM3A mixture. All the results indicated that the associated neighbouring CBMs impart thermostability to the catalytic module.

3.4 Conclusion

A new member of GH9 β-1,4-glucanase from *Acetivibrio thermocellus* *At*GH9C-CBM3A-CBM3B and its truncated forms *At*GH9C-CBM3A, *At*GH9C accompanying its carbohydrate binding modules CBM3A, CBM3B were cloned, expressed and characterized. The full-length *At*GH9C-CBM3A-CBM3B displayed maximum activity at pH 7.5 and temperature 55°C with capability to hydrolyse both β-1,4 cellulosic substrate (CMC-Na) and mixed β-1,3; β-1,4 cellulosic substrates (lichenan and β-D-glucan). Endoglucanase *At*GH9C-CBM3A-CBM3B is stable in wide temperature range 4°C to 60°C and pH range 6.0-9.0 indicating its possible utilization in biomass saccharification for bioethanol production. Zn^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+} and Fe^{2+} ions and EDTA and anionic detergent adversely affected the activity of the endoglucanase but Mg^{2+} ions elevated the enzyme activity by 29%. Thermal denaturation study of *At*GH9C-CBM3A-CBM3B revealed the unfolding transition midpoint (T_m) to be 65°C. Mg^{2+} ions increased its T_m by 5°C indicating its inherent presence playing role in catalysis and imparting thermal stability to the enzyme. The TLC analysis of *At*GH9C-CBM3A-CBM3B of treated CMC hydrolysate revealed the presence of oligosaccharides greater than DP3 demonstrating the endo acting nature of the enzyme. The native *At*GH9C-CBM3A-CBM3B showed maximum activity 58.8 U/mg against CMC-Na, however the activity of only catalytic module *At*GH9C was only 2.3 U/mg. The activity of the *At*GH9C was recovered in presence of equimolar concentrations of CBM3A (47%), CBM3B (13%) and CBM3A+CBM3B (50%) against CMC-Na but no revival was observed against hydroxy ethyl cellulose. The molar ratio of *At*GH9C:CBM3A+CBM3B at 1:3 displayed the enzyme activity close to the native *At*GH9C-CBM3A-CBM3B enzyme against soluble substrates confirming the cross-

talk between the CBMs. The thermostability analysis of equimolar mixtures of *At*GH9C and CBMs indicated that the associated CBMs impart thermostability to the catalytic module. This study established endoglucanase *At*GH9C-CBM3A-CBM3B as an efficient enzyme against cellulosic substrates and essential requirement of physical association between catalytic module *At*GH9C and CBMs (CBM3A, CBM3B) for its functional activity. Therefore, the cellulolytic activity of endoglucanases can be improved by physical association or by fusing with the suitable CBMs.

3.5 References

[1] N. Karaki, A. Aljawish, C. Humeau, L. Muniglia, J. Jasniewski, Enzymatic modification of polysaccharides: Mechanisms, properties, and potential applications: A review., Enzyme Microb Technol 90 (2016) 1–18. https://doi.org/10.1016/j.enzmictec.2016.04.004.

[2] F.H. Isikgor, C.R. Becer, Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers, Polym Chem 6 (2015) 4497–4559. https://doi.org/10.1039/C5PY00263J.

[3] L.R. Lynd, P.J. Weimer, W.H. van Zyl, I.S. Pretorius, Microbial Cellulose Utilization: Fundamentals and Biotechnology, Microbiology and Molecular Biology Reviews 66 (2002) 506–577. https://doi.org/10.1128/MMBR.66.3.506-577.2002/ASSET/38C13E20-0334-4477-8327-F000A9C877DC/ASSETS/GRAPHIC/MR032001411T.JPEG.

[4] R.H. Doi, A. Kosugi, Cellulosomes: plant-cell-wall-degrading enzyme complexes, Nature Reviews Microbiology 2004 2:7 2 (2004) 541–551. https://doi.org/10.1038/nrmicro925.

[5] Y. Shoham, R. Lamed, E.A. Bayer, The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides, Trends Microbiol 7 (1999) 275–281. https://doi.org/10.1016/S0966-842X(99)01533-4.

[6] D.B. Wilson, D.C. Irwin, Genetics and Properties of Cellulases, (1999) 1–21. https://doi.org/10.1007/3-540-49194-5_1.

[7] E. Libertini, Y. Li, S.J. McQueen-Mason, Phylogenetic analysis of the plant endo-beta-1,4-glucanase gene family., J Mol Evol 58 (2004) 506–15. https://doi.org/10.1007/s00239-003-2571-x.

[8] A.B. Boraston, D.N. Bolam, H.J. Gilbert, G.J. Davies, Carbohydrate-binding modules: fine-tuning polysaccharide recognition, Biochemical Journal 382 (2004) 769. https://doi.org/10.1042/BJ20040892.

[9] P. Tomme, A. Boraston, B. McLean, J. Kormos, A.L. Creagh, K. Sturch, N.R. Gilkes, C.A. Haynes, R.A.J. Warren, D.G. Kilburn, Characterization and affinity applications of cellulose-binding domains, J Chromatogr B Biomed Sci Appl 715 (1998) 283–296. https://doi.org/10.1016/S0378-4347(98)00053-X.

[10] S. Petkun, I.R. Grinberg, R. Lamed, S. Jindou, T. Burstein, O. Yaniv, Y. Shoham, L.J.W. Shimon, E.A. Bayer, F. Frolow, Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: Structural and functional significance of the intermodular linker, PeerJ 2015 (2015) e1126. https://doi.org/10.7717/peerj.1126.

[11] D. Irwin, D.H. Shin, S. Zhang, B.K. Barr, J. Sakon, P.A. Karplus, D.B. Wilson, Roles of the catalytic domain and two cellulose binding domains of

Thermomonospora fusca E4 in cellulose hydrolysis, J Bacteriol 180 (1998) 1709–1714. https://doi.org/10.1128/JB.180.7.1709-1714.1998.

[12] S. Jindou, Q. Xu, R. Kenig, M. Shulman, Y. Shoham, E.A. Bayer, R. Lamed, Novel architecture of family-9 glycoside hydrolases identified in cellulosomal enzymes of Acetivibrio cellulolyticus and Clostridium thermocellum, FEMS Microbiol Lett 254 (2006) 308–316. https://doi.org/10.1111/j.1574-6968.2005.00040.x.

[13] M. Sajjad, M.I.M. Khan, R. Zafar, S. Ahmad, U.H.K. Niazi, M.W. Akhtar, Influence of positioning of carbohydrate binding module on the activity of endoglucanase CelA of Clostridium thermocellum, J Biotechnol 161 (2012) 206–212. https://doi.org/10.1016/J.JBIOTEC.2012.05.023.

[14] T. Burstein, M. Shulman, S. Jindou, S. Petkun, F. Frolow, Y. Shoham, E.A. Bayer, R. Lamed, Physical association of the catalytic and helper modules of a family-9 glycoside hydrolase is essential for activity, FEBS Lett 583 (2009) 879–884. https://doi.org/10.1016/J.FEBSLET.2009.02.013.

[15] K. Kumar, S. Singal, A. Goyal, Role of carbohydrate binding module (CBM3c) of GH9 β-1,4 endoglucanase (Cel9W) from Hungateiclostridium thermocellum ATCC 27405 in catalysis, Carbohydr Res 484 (2019). https://doi.org/10.1016/J.CARRES.2019.107782.

[16] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal Biochem 72 (1976) 248–254. https://doi.org/10.1016/0003-2697(76)90527-3.

[17] J.E. Noble, Quantification of protein concentration using UV absorbance and Coomassie dyes, Methods Enzymol 536 (2014) 17–26. https://doi.org/10.1016/B978-0-12-420070-8.00002-7.

[18] J. Hospital, METHOD NELSON for Medical, (1944) 375–381.

[19] M. Somogyi, DETERMINATION OF BLOOD SUGAR, Journal of Biological Chemistry 160 (1945) 69–73. https://doi.org/10.1016/S0021-9258(18)43098-0.

[20] N. Nelson, A PHOTOMETRIC ADAPTATION OF THE SOMOGYI METHOD FOR THE DETERMINATION OF GLUCOSE, Journal of Biological Chemistry 153 (1944) 375–380. https://doi.org/10.1016/S0021-9258(18)71980-7.

[21] M. Somogyi, DETERMINATION OF BLOOD SUGAR, Journal of Biological Chemistry 160 (1945) 69–73. https://doi.org/10.1016/S0021-9258(18)43098-0.

[22] A. Karnchanatat, A. Petsom, P. Sangvanich, J. Piapukiew, A.J.S. Whalley, C.D. Reynolds, G.M. Gadd, P. Sihanonth, A novel thermostable endoglucanase from the wood-decaying fungus Daldinia eschscholzii (Ehrenb.:Fr.) Rehm, Enzyme Microb Technol 42 (2008) 404–413. https://doi.org/10.1016/J.ENZMICTEC.2007.11.009.

[23] N.J. Greenfield, Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, Nature Protocols 2006 1:6 1 (2006) 2527–2535. https://doi.org/10.1038/nprot.2006.204.

[24] A.O.S. Lima, M.C. Quecine, M.H.P. Fungaro, F.D. Andreote, W. Maccheroni, W.L. Araújo, M.C. Silva-Filho, A.A. Pizzirani-Kleiner, J.L. Azevedo, Molecular characterization of a β-1,4-endoglucanase from an endophytic Bacillus pumilus strain, Appl Microbiol Biotechnol 68 (2005) 57–65. https://doi.org/10.1007/S00253-004-1740-1.

[25] T. Mejia-Castillo, M.E. Hidalgo-Lara, L.G. Brieba, J. Ortega-Lopez, Purification, characterization and modular organization of a cellulose-binding protein, CBP105, a processive β-1,4-endoglucanase from Cellulomonas flavigena, Biotechnol Lett 30 (2008) 681–687. https://doi.org/10.1007/s10529-007-9589-x.

[26] Z. Yi, X. Su, V. Revindran, R.I. Mackie, I. Cann, Molecular and Biochemical Analyses of CbCel9A/Cel48A, a Highly Secreted Multi-Modular Cellulase by Caldicellulosiruptor bescii during Growth on Crystalline Cellulose, PLoS One 8 (2013) e84172. https://doi.org/10.1371/JOURNAL.PONE.0084172.

[27] I. ul Haq, F. Akram, M.A. Khan, Z. Hussain, A. Nawaz, K. Iqbal, A.J. Shah, CenC, a multidomain thermostable GH9 processive endoglucanase from Clostridium thermocellum: cloning, characterization and saccharification studies, World J Microbiol Biotechnol 31 (2015) 1699–1710. https://doi.org/10.1007/S11274-015-1920-4/TABLES/5.

[28] L. Gal, C. Gaudin, A. Belaich, S. Pages, C. Tardif, J.P. Belaich, CelG from Clostridium cellulolyticum: a multidomain endoglucanase acting efficiently on crystalline cellulose, J Bacteriol 179 (1997) 6595–6601. https://doi.org/10.1128/JB.179.21.6595-6601.1997.

[29] E.A. de Araújo, M. de Oliveira Neto, I. Polikarpov, Biochemical characterization and low-resolution SAXS structure of two-domain endoglucanase BlCel9 from Bacillus licheniformis, Appl Microbiol Biotechnol 103 (2019) 1275–1287. https://doi.org/10.1007/s00253-018-9508-1.

[30] J.P. Lee, Y.A. Kim, S.K. Kim, H. Kim, Characterization of a multimodular endo-β-1,4-glucanase (Cel9K) from paenibacillus sp. x4 with a potential additive for saccharification, J Microbiol Biotechnol 28 (2018) 588–596. https://doi.org/10.4014/JMB.1712.12047.

[31] B.L. Zamost, H.K. Nielsen, R.L. Starnes, Thermostable enzymes for industrial applications, J Ind Microbiol 8 (1991) 71–81. https://doi.org/10.1007/BF01578757.

[32] I. ul Haq, F. Akram, M.A. Khan, Z. Hussain, A. Nawaz, K. Iqbal, A.J. Shah, CenC, a multidomain thermostable GH9 processive endoglucanase from Clostridium thermocellum: cloning, characterization and saccharification

studies, World J Microbiol Biotechnol 31 (2015) 1699–1710. https://doi.org/10.1007/s11274-015-1920-4.

[33] J. Kurokawa, E. Hemjinda, T. Arai, T. Kimura, K. Sakka, K. Ohmiya, Clostridium thermocellum cellulase CelT, a family 9 endoglucanase without an Ig-like domain or family 3c carbohydrate-binding module, Appl Microbiol Biotechnol 59 (2002) 455–461. https://doi.org/10.1007/s00253-002-1048-y.

[34] M.H. Foley, G. Déjean, G.R. Hemsworth, G.J. Davies, H. Brumer, N.M. Koropatkin, A cell-surface GH9 endo-glucanase coordinates with surface glycan binding proteins to mediate xyloglucan uptake in the gut symbiont Bacteroides ovatus, J Mol Biol 431 (2019) 981. https://doi.org/10.1016/J.JMB.2019.01.008.

[35] L.A. Ogonda, A. Saumonneau, M. Dion, E.K. Muge, B.M. Wamalwa, F.J. Mulaa, C. Tellier, Characterization and engineering of two new GH9 and GH48 cellulases from a Bacillus pumilus isolated from Lake Bogoria, Biotechnol Lett 43 (2021) 691–700. https://doi.org/10.1007/s10529-020-03056-z.

[36] P. Tomme, A. Boraston, B. McLean, J. Kormos, A.L. Creagh, K. Sturch, N.R. Gilkes, C.A. Haynes, R.A.J. Warren, D.G. Kilburn, Characterization and affinity applications of cellulose-binding domains, J Chromatogr B Biomed Sci Appl 715 (1998) 283–296. https://doi.org/10.1016/S0378-4347(98)00053-X.

Chapter 4

Structure elucidation of endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus* ATCC 27405 and its substrate binding analysis

4.1 Introduction

21st century experienced rapid depletion of fossil fuels and other non-renewable resources. This attracts energy generation from lignocellulosic biomass in the form of potentially low-carbon biofuels, as the primary substitute for non-renewable fuels [1]. The cohort of plants lignocellulosic biomass composed of organic biopolymers i.e. cellulose, hemicellulose and lignin, represent a reserve of significant carbon source, that signals its utilization benefit as a possible renewable energy source by its bioconversion to biofuel [2]. Cellulose is a straight chain, unbranched microfibrillar structure composed of β-1,4 glycosidic bond linked adjacent D-glucopyranose residues, and is the major shareholder in plant biomass composition [3]. Despite cellulose being considered the inexhaustible carbon source, the packing of cellulose chains into insoluble bundles and the intricacies of the hydrogen-bonded inter-glucan make it recalcitrance for hydrolysis in bioconversion. The cellulose can be hydrolyzed or

depolymerized utilizing enzymes produced by bacteria or fungi, which can either be in free form or as part of a multi-enzyme complex called a cellulosome [4]. Owing to their industrial importance, notably to the current biofuels sector, cellulose deconstructing enzymes, particularly cellulases, have received extensive attention in recent years [5].

Cellulases cleave the β-1,4-glycosidic bond of a cellulose chain with the aid of a water molecule. Endoglucanase/endocellulase (EC 3.2.1.4), cellobiohydrolase/exocellulase (EC 3.2.1.91), and β-glucosidase (EC 3.2.1.21), together known as cellulases, work in tandem to completely hydrolyze cellulosic chains by converting the cellulose polymer into cello-oligosaccharides, cellobiose and glucose respectively [6]. The endocellulase active-site comprises an open binding cleft that can fit arbitrarily at any point along the cellulose main chain and yields oligosaccharides as a byproduct. Whereas, a tunnel-shaped structure at the exocellulase active-site connects to one end of the cellulose chain and cleaves cellobiose either sequentially or processively [7]. Cellulases are grouped into different Glycoside Hydrolase (GH) families in the Carbohydrate Enzyme (CAZy) database (http://www.cazy.org/Glycoside-Hydrolases.html) on the ground of their amino acid arrangements [8]. These GH family enzymes have modular architecture comprising catalytic modules that are responsible for the hydrolysis reaction, associated with one or more non-catalytic accessory Carbohydrate Binding Modules (CBMs) [9]. The primary enzyme in cellulose hydrolysis, endoglucanase is present in 13 different glycoside hydrolase families. Carbohydrate binding modules (CBMs), are accessory proteins, appended to the catalytic modules that primarily aid in substrate turnover by binding its surface. CBMs are shown to mandate substrate specificity, enzyme reaction mechanism, thermo-stability and assist in bacterial cell wall attachment [10].

The family GH9 is the second-largest for endoglucanase and some endoglucanases of GH families 9 and 5 can perform both endo- and exo-cellulase activities. The catalytic module of the GH9 family enzymes displays a conserved $(\alpha/\alpha)_6$-barrel structure having different cellulolytic and hemi-cellulolytic enzymatic activities. The present classification module further classifies the family 9 GHs into class A, B, and C and proposes the general acid-base mechanism of catalysis with aspartic-acid (D) and/or glutamic-acids (E) as alternating proton donors/acceptors [11]. Furthermore, the extant arrangement of GH9 enzymes suggests the catalysis of crystalline cellulose occurs in entirely enclosed tunnels (processive), or open solvent-accessible clefts (non-processive), or both of them. Broadly, the cellulose binding site of enzymes is designated between subsites +1 to -1 and the subsites -7 to +2 are known to make contact with the glucan chain through the hydrogen bonding [12]. However, the lack of formation of an active enzyme-substrate (ES) complex between the catalytic cleft of GH9 enzymes and ligands with degree of polymerization, DP ≥ 6 from polymeric crystalline cellulose, makes it difficult to interpret the enzymatic activity of GH9 enzymes with higher oligosaccharides [13]. Therefore, a glance at the inter-dependence and quasi-allostericity of GH9 modules and carbohydrate-binding modules (CBMs) may provide the clue to the above issue.

In general, the GH9 families are associated with family 3 carbohydrate binding module which elevates the enzymatic performance of GH9 enzymes. Family 3 CBMs are around 150 amino acid residue long, and are sub-grouped into CBM3A, CBM3B, CBM3C and CBM3D based on the amino acid arrangements [14]. CBM3A and CBM3B strongly bind to the surface of cellulose and chitin, while CBM3C is attached to the ends of crystalline cellulose and weakly contributes to the catalytic activity of

glucanase (endo- or exo-) of GH9 family [15]. CBM3D most likely originated from CBM3A, but it is attached to a glycoside hydrolase of a different family [16]. Family 3 CBMs usually adopt β-sandwich fold and the number of CBMs, their position and conserved amino acids are detrimental to the range of substrate accommodation to the GH9 enzymes. Association of the module-X2 and CBM3 (X2-CBM3) with endoglucanase displays 4.6-fold increase in the activity of towards insoluble substrate [17]. The impact of CBM3a location on *Clostridium thermocellum* endoglucanase CelA activity, before and after the dockerin module indicates CBM3a differently modifies CelA activity [18]. The removal of CBM3c from its cognate catalytic module, endoglucanase GH9 of *Clostridium thermocellum* Cel9I [19] and Cel9W of *Hungateiclostridium thermocellum* [20] displayed a marked decrease in the enzyme's catalytic activity, indicating the important role of CBMs in the efficient enzyme catalysis. The crystallization and preliminary diffraction studies indicate the requirement of conserved amino acid residues in CBMs, for effective substrate recognition and binding [21].

Most of the GH9 family cellulases are endocellulase, but, a small number are reported to have exocellulase and bifunctional cellulase/xylanases activities [22]. The exocellulases from the GH9 do not contain the tunnel forming loop, but contain a barrier at the active-site. The processive endoglucanse from the *Thermobifida fusca* producing cellotetraose as major product also represents a barrier at -4 subsite of the active-site [23]. Moreover, comparative profiling of 70 cellulases from *Clostridium thermocellum* (a.k.a *Acetivibrio thermocellus*) demonstrated four distinct hydrolysis patterns namely (i) exo-mode, (ii) endo-mode with non-specific hydrolysis, (iii) processive endoglucanases with cellotetraose as main product (pEG4) and (iv) processive

endoglucanases with cellobiose as main product (pEG2) [24]. The structural basis and mechanism behind variations in the activity among the members of family 9 GH is not well understood. Therefore, it is important to explore further in depth on the structure and functional relationship of the family 9 GH with the possible effects of CBMs on their activity.

In Chapter 3 it was shown that the physical association of catalytic module, *At*GH9C of recombinant β-1,4 endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus* ATCC27405 requires its associated CBMs for effective hydrolysis of cellulosic substrates [25]. In the present study the structural and functional basis of such behavior, displayed by *At*GH9C-CBM3A-CBM3B was explored through *in silico* approach in order to understand its organization, stability and the folding pattern. The prediction on the binding interaction and the processivity of *At*GH9C-CBM3A-CBM3B were validated by molecular docking against various cellulosic ligands and the processivity analysis against phosphoric acid swollen cellulose (PASC), respectively. Endoglucanase, *At*GH9C-CBM3A-CBM3B in both ligand-bound and unbound form was subjected to MD simulation analysis in order to elucidate the conformational dynamics of the its modules, the role of loops and the key residues involved in catalysis or binding. The secondary structure prediction analysis of *At*GH9C-CBM3A-CBM3B was validated by circular dichroism (CD) and its solution structure was analyzed by SAXS and DLS studies. The results unfolded the structural and functional insights into the catalytic mechanism, stability and functional behavior of endoglucanase, *At*GH9C-CBM3A-CBM3B against various cello-oligosaccharides and the behavior of associated CBMs in aiding the stability.

4.2 Materials and methods

4.2.1 Substrate and chemicals

Sodium salt of carboxymethyl cellulose (CMC-Na) was purchased from Sigma-Aldrich Chemicals Inc., (USA). Sodium phosphate monobasic and sodium phosphate dibasic used for the preparation of sodium phosphate buffer was procured from Hi-Media Laboratories Pvt. Ltd., India.

4.2.2 Amino acid sequence and evolutionary analyses

The amino acid sequence details of *At*GH9C-CBM3A-CBM3B were fetched from the genome of *Acetivibrio thermocellus* ATCC 27405 existing in the NCBI database (https://www.ncbi. nlm.nih.gov/protein/WP_009984467). Full-length *At*GH9C-CBM3A-CBM3B is 961 amino acids long modular enzyme having the GenBank ID, ABN53959.1 with UniPort id: A3DJ30. The Conserved Domain Database of NCBI (http://www.ncbi.nlm.nih.gov/cdd/) was used to explore and establish the boundaries of the Conserved Domains (CDs) in *At*GH9C-CBM3A-CBM3B amino acid sequence. The annotation of the N-terminal signal peptide is done with the assistance of SignalP 4.1 server (http://www.cbs.dtu. dk/services/SignalP). Due to query cover below 20% for full-length *At*GH9C-CBM3A-CBM3B (857 amino acids), the PDB homologs of *At*GH9C-CBM3A-CBM3B were determined from the amino acid sequence equivalent to only catalytic module i.e. *At*GH9C (498 amino acids) after its NCBI-BLAST analysis against the PDB database [26]. The amino acid sequence *At*GH9C was subjected to multiple sequence alignment (MSA) with the aforementioned PDB homologs using the EMBL-EBI Clustal Omega server (http://www.ebi.ac.uk/Tools/msa/clustalo/) [27] and analyzed using the ESPript 3.0 server (http://espript.ibcp.fr/ESPript /ESPript/) [28]. The evolutionary relationship

between endoglucanase *At*GH9C-CBM3A-CBM3B and the glycoside hydrolases of the GH9 family was ascertained through phylogenetic analysis employing the Neighbour-Joining method and MEGA11 software [29].

4.2.3 Secondary structure analysis of *At*GH9C-CBM3A-CBM3B

The secondary structure element of *At*GH9C-CBM3A-CBM3B was evaluated by submitting its amino acid sequence in different online servers PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/), GOR4 (https://npsa-pbil.ibcp.fr/NPSA/npsa_gor4.html), APSSP (http://crdd.osdd.net/raghava/apssp/) and SOPMA (https://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). Furthermore, NetSurfP 2.0 (http://www.cbs.dtu.dk/services/NetSurfP/) predicted the relative surface accessibility of the amino acid residues. The proportion of secondary structural elements in *At*GH9C-CBM3A-CBM3B was ascertained using the Circular Dichroism (CD) technique. The recombinant protein *At*GH9C-CBM3A-CBM3B (75 µg/mL) was dissolved in 50 mM sodium phosphate buffer (pH 7.5) and the secondary structure was determined at 37°C in the far UV range (190-250 nm) using Circular Dichroism spectroscopy equipped with a Peltier temperature controller. The CD spectrum was measured with a scanning rate of 50 nm/min, 1 nm bandwidth, and an average of ten scans. The CD result was depicted as a difference in molar ellipticity (Δε, decilitre/mol/cm) in the specified range of wavelength [30]. Structural percentages of α-helix and β-sheet percentages in *At*GH9C-CBM3A-CBM3B were estimated by uploading the molar ellipticity difference value (Δε) to the K2D3 server (http://k2d3.ogic.ca/).

4.2.4 Homology modeling of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A and *At*GH9C

The full length endoglucanase, *At*GH9C-CBM3A-CBM3B comprising 857 amino acids containing the pET-28a(+) vector sequence at N-terminal (1–21 aa) containing the His_6-tag and 8 amino acid residues at the C-terminal as reported earlier [25] was subjected to homology modeling. The amino acid sequence of *At*GH9C-CBM3A-CBM3B was supplied to various online servers namely AlphaFold[31], I-TASSER[32], RaptorX [33], Robetta [34], Phyre2 [35] and Swiss-Model [36] for homology modeling and corresponding 3D model structures were developed in the form of PDB files from all the servers mentioned above. The homology modeling of catalytic module *At*GH9C and the derivative *At*GH9C-CBM3A were carried out using the cloned sequence as reported [25] earlier by utilizing aforementioned servers in order to assess their binding characteristics with cellotetraose.

4.2.5 Refinement, energy minimization and quality assessment

The 3D modeled PDB structure obtained from different servers for *At*GH9C-CBM3A-CBM3B was refined using online energy minimization server, YASARA (http://www.yasara.org/minimizations erver.php) [37]. The quality assurance of energy minimized PDB structures of *At*GH9C-CBM3A-CBM3B generated from different servers were validated using structural and verification online tool: SAVES v6.0 server (https://saves.mbi.ucla.edu/). The SAVES v6.0 (https://saves.mbi.ucla.edu/) is a comprehensive toolkit, that envisages different types of stereochemical parameters of the protein structure using five tools (ERRAT, VERIFY3D, PROVE, PROCHECK-Ramachandran, and WHAT CHECK) and evaluate the fitness of model. The ProSA server was utilized to compute the statistical Z-score deviation of the modeled structures, relying on previously deposited high-resolution PDB and NMR structures

[38]. The 2D topology profile of *At*GH9C-CBM3A-CBM3B was generated using the PDBSum database (http ://www.ebi.ac.uk/thorn-ton-srv/databases/pdbsum/Generate.html). The stereochemical evaluation of the modeled catalytic module *At*GH9C and the derivative *At*GH9C-CBM3A was performed on the server, SAVES v6.0.

4.2.6 Molecular dynamics simulation of *At*GH9C-CBM3A-CBM3B

Molecular dynamics simulation was used to evaluate the behavior of *At*GH9C-CBM3A-CBM3B at fine temporal resolution. The Robetta modeled *At*GH9C-CBM3A-CBM3B was subjected to molecular dynamic simulation using GROMACS v2020.4 software [39] and the PARAM-ISHAN High Performance Computing (HPC) supercomputer facility available at Indian Institute of Technology Guwahati (IITG), India to compute the interacting protein forces. CHARMM27 all-atom force field was applied to compute the accuracy of atomistic force field of *At*GH9C-CBM3A-CBM3B modeled structure. The modeled structure was positioned inside a dodecahedron box of dimension 12.4 nm x 9.2 nm x 12.7 nm and volume 1445.3 nm^3 and solvated using TIP3P water model containing water with roughly 650400 electrons and 150961 solvent molecules. The protein was electrostatically neutralized with the help of 55 Na^+ ions. Using the steepest descent method using 50,000 iteration steps with a cut-off at 1000 kJ/min/nm of force the whole system was electrostatically minimized to sustain the equilibrium of the aqueous environment. Following the energy minimization, the entire system was then equilibrated with two-phase equilibration process. The canonical ensemble NVT phase was run first for 500 ps at 300K with time constant of 2 fs. Along with it, isothermal-isobaric NPT was carried out secondly for 500 ps at 1 bar pressure with an iteration time of 2 fs [40]. The equilibrated structure with 208870 atoms in

system was then finally deployed for MD simulation for 200 ns with time step of 10 fs. Lastly, the simulated protein *At*GH9C-CBM3A-CBM3B was analyzed for various factors determining its stability and structural properties. The root mean square deviation (RMSD) presenting the conformational change in the protein backbone was assessed by least square fitting method using *gmx_mpi rms* command. The root mean square fluctuation (RMSF) of Cα-atom was determined using *gmx_mpi rmsf* command line. The solvent accessible surface area (SASA) and radius of gyration (R_g) were determined by employing the *gmx_mpi sasa* and *gmx_mpi* gyrate commands in system, respectively.

4.2.7 Protein ligand interaction and active-site analysis of *At*GH9C-CBM3A-CBM3B

The molecular docking study was carried out to establish the interaction of different cellooligosaccharies of varied lengths with the active-site pocket of modeled *At*GH9C-CBM3A-CBM3B structure by the Autodock 4.2.1 [41] compiled with MGLTools (http://mgltools. scripps.edu/). The cellooligosaccharide ligands (cellobiose to cellopentaose) were retrieved in their 3D SDF canonical format from the PubChem database (http://pubchem.ncbi.nlm.nih.gov) and cellohexaose was constructed using the smiles from the PubChem database in the UCSF Chimera software [42]. All the ligands were then converted from SDF to PDB format using Openbabel [43]. The molecular docking was executed in Autodock 4.2.1 after defining the grid box parameter using 50 cycles of Genetic Algorithm (GA). The docked protein-ligand complexes showing highest negative binding free energy were considered as best dock complexes at the active-site. The best dock complexes were further assessed for 2D polar-nonpolar interactions between protein-ligand docked complexes using Discovery Studio v 5.4 (https://discover.3ds.com/discovery-studio-visualizer-

download). The molecular docking analysis of catalytic module, *At*GH9C with cellotetraose and the derivative *At*GH9C-CBM3A with cellotetraose was performed using Autodock 4.2.1 to assess the binding properties.

4.2.8 Processivity index analysis of *At*GH9C-CBM3A-CBM3B

The catalytic signature of endoglucanase on amorphous cellulosic substrate, phosphoric acid swollen cellulose (PASC) predicts its mode of action in hydrolyzing the substrate via processive (continuous breakage of higher oligosaccharides to smaller one) or non-processive (random cleavage of oligosaccharides) way [44]. PASC was prepared by using cellulose as described by Zhang et al. 2018 [45] The reaction mixture (100 µL) consisting of 10 µL of purified 100 µg/mL of *At*GH9C-CBM3A-CBM3B and 1% (w/v) PASC (10 mg/mL) in 50 mM sodium phosphate buffer, pH 7.5 was incubated at 37°C for a varied time period 1, 2, 5, 10, 15, 30 minutes and 1, 4, 12, and 24 hours. After completion of the reaction, the reaction mixture was centrifuged at 6500g and the soluble supernatant [S] fraction containing cleaved smaller DP oligosaccharides and the insoluble pellet [I] containing higher DP oligosaccharides was separated. The soluble fraction and insoluble fraction (after dissolving in 100 µL 50 mM sodium phosphate buffer, pH 7.5) were analyzed for reducing sugar present by Nelson [43] and Somogyi [46,47] sugar estimation method. The ratio of the estimated reducing sugar obtained in soluble fraction [S] and insoluble fraction [I] gave the processivity index, calculated from the formula [44] given below:

$$\text{Processivity index (PI)} \frac{\text{Reducing sugar in soluble fraction [S]}}{\text{Reducing sugar in insolube fraction [I]}} =$$

4.2.9 Molecular dynamics simulation of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex

The docked complex of *At*GH9C-CBM3A-CBM3B with cellotetraose showed the highest binding energy, and was investigated for observed changes in conformational stability of protein-ligand interaction through MD simulation using GROMACS v2020.4. The CHARMM27 force field was used to determine the protein forces. The ligand topology and parameters for cellotetraose were built using the SwissParam server (http://swissparam.ch/) to make it suitable with GROMACS. The *At*GH9C-CBM3A-CBM3B-Cellotetraose complex was placed in the dodecahedron box with dimensions 10 nm x 7.5 nm x 10.4 nm along with 65246 water molecules corresponding to the volume of 2121 nm^3. The system was immersed in TIP3 water and then balanced out with 57 Na^+ ions to neutralize the electrostatic charge. The entire system was first equilibrated for 500 ps in an NVT ensemble, where the number of particles, volume, and temperature remained stable. This was followed by an additional second equilibration for 500 ps in an NPT ensemble, where the number of particles, pressure, and temperature remained constant. A 2 fs iteration time was used in both equilibrations. The equilibrated system containing 295843 atoms in the system was finally subjected to an MD simulation with a 200 ns runtime and an integration time of 10 fs. The temporal analysis of the complex structure assesses its stability in aqueous system. The *gmx_mpi rms* command was used to estimate the RMSD of the ligand-bound *At*GH9C-CBM3A-CBM3B after simulation, and the *gmx_mpi rmsf* command was administered to estimate the residue fluctuations in the complex. Other conformational stability parameters like radius of gyration (R_g) and solvent accessible surface area (SASA) was determined using *gmx_mpi gyrate* and *gmx_mpi sasa*

command respectively. The final MD simulated complex structure obtained was visualized in PyMOL V2.5.5 software.

4.2.10 Binding free energy analysis of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex

The Gibbs free energy ($\Delta G_{binding}$) of the *At*GH9C-CBM3A-CBM3B-Cellotetraose complex was computed by the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method using the MD simulated 200 ns trajectory of the complex. The g_mmpbsa tools [48] was used at its default parameters to determine the the total binding free energy ($\Delta G_{binding}$) represented as summation of three energy components i.e. 1. potential energy (ΔE_{MM}) that comprises the electrostatic (ΔE_{elec}) and van der Waals interactions (ΔE_{vdw}) in vacuum, 2. polar solvation energy ($\Delta G_{polar)}$ depend on Assisted Poison Boltzmann Solver (APBS) program and 3. apolar energy (ΔG_{apolar}) by using Solvent Accessible Surface Area (SASA) model.

4.2.11 Solution structure analysis of *At*GH9C-CBM3A-CBM3B by SAXS

The Small-Angle X-ray Scattering (SAXSpace, Anton Paar GmbH, Graz Austria) profiles were used to evaluate the conformational behavior of *At*GH9C-CBM3A-CBM3B in solution at two different concentrations, 2 mg/mL and 5 mg/mL. A CMOS Mytheme detector (Dectris, Baden, Switzerland) was used to measure the scattering patterns for the protein and corresponding buffer (50 mM sodium phosphate, pH-7.5) at an incident wavelength of 1.54 Å. The protein *At*GH9C-CBM3A-CBM3B and matched buffer (50 mM sodium phosphate, pH 7.5) were centrifuged at 14,000***g*** for 20 min at 4°C to prevent interference from aggregation and precipitation. The sample-to-detector distance was 317 mm, covering the scattering vector range ($S=4*\pi*Sin\theta/\lambda$) of 0.11-7.6 nm-1 [49]. A Peltier system maintained a stable temperature of 10°C during data processing. Each sample was subjected to two

consecutive 30-minute frames in order to assess radiation damage and beam stability. The final scattering pattern of *At*GH9C-CBM3A-CBM3B was ascertained by subtracting the scattering profile of the matched buffer from the recorded individual protein sample profile. The scattering patterns of *At*GH9C-CBM3A-CBM3B were evaluated for scattering distributions and interparticle interactions using the Primus package of ATSAS 3.2.1 suite at each concentration [50]. Based on the treated and normalized scattering profile, scattering profile at 5 mg/mL concentration is used for further processing.

The method of Guinier approximation and indirect Fourier transformation was administered simultaneously to evaluate the radius of gyration using PrimusQt tool [51]. Protein folding and global compactness of *At*GH9C-CBM3A-CBM3B were measured from Kratky plot analysis. The maximum dimension (D_{max}) and distance distribution plot, P(R) plot were obtained analytically using GNOM software [52]. The plausible molecular mass of *At*GH9C-CBM3A-CBM3B was estimated using the SAXSMoW software [53]. Finally, twenty independent *ab initio* models of *At*GH9C-CBM3A-CBM3B in solution were created employing ATSAS DAMMIF tool. The models are subjected to further averaging, clustering and filtering using DAMAVER and DAMMIN and final *ab initio* structure of *At*GH9C-CBM3A-CBM3B was created [54]. The final *ab initio* structure and homology modeled *At*GH9C-CBM3A-CBM3B were superimposed using PyMOL v2.5.5 and UCSF Chimera software. The homology model of *At*GH9C-CBM3A-CBM3B and the model generated by DAMMIF were fitted using the CRYSOL program [55], and their accuracy was compared against experimental SAXS data.

4.2.12 Dynamic light scattering analysis of *At*GH9C-CBM3A-CBM3B

The apparent hydrodynamic diameter of *At*GH9C-CBM3A-CBM3B (D_h) in aqueous environment was analyzed using dynamic light scattering (DLS) with a particle analyzer (Lite-sizer 500, Anton Paar, Graz, Austria). Purified *At*GH9C-CBM3A-CBM3B protein at three concentrations (1.0, 2.0, and 5.0 mg/mL) was centrifuged at 13,000g at 4 °C for 10 min and the supernatant was filtered by a syringe filter using a 0.25 μm membrane to get rid of any protein aggregates. The three concentrations of the enzyme were subjected to DLS and zeta potential (ξ) analysis as reported earlier [56]. The apparent particle size and zeta potential (ξ) of *At*GH9C-CBM3A-CBM3B were derived from the analysis of average 50 runs using Kalliope software.

4.3. Results and Discussion

4.3.1 Sequence and phylogenetic analysis of *At*GH9C-CBM3A-CBM3B

Detailed proteomic analysis of the genome of thermophilic bacterium *Acetivibrio thermocellus* detected 41 cellulosomal proteins including 36 type I dockerin-associated proteins with 16 new subunits [57]. *At*GH9C-CBM3A-CBM3B is a novel modular endoglucanase form *Acetivibrio thermocellus*, representing one catalytic module (*At*GH9C) and two adjunct Carbohydrate Binding Modules (CBMs) *viz.*, CBM3A and CBM3B [25]. It is enlisted in the UniPort comprehensive database having ID A3DJ30 and in CAZy database with GenBank Accession Number, ABN53959.1. The amino acid sequence of *At*GH9C-CBM3A-CBM3B revealed the presence of a signal peptide of 39 aa long (1-39 aa) at N-terminal as reported earlier [25]. Subsequent to the signal peptide, amino acid of 476 in length (58-534 aa) is annotated as catalytic module (*At*GH9C) having a linker sequence (40-57 aa) between them. Following the catalytic module, the enzyme is structured with two different CBMs i.e CBM3A (565-715 aa) and CBM3B (736-885 aa) having a short stretch of linker peptide (716-735 aa) present between the two CBMs. The C-terminal of *At*GH9C-CBM3A-CBM3B is constituted with Doc I resembling module (894-961 aa) [25]. The gene encoding full-length enzyme i.e. *At*GH9C-CBM3A-CBM3B and its truncated derivative modules namely, *At*GH9C-CBM3A, *At*GH9C, CBM3A and CBM3B were separately cloned, expressed and purified as described in previous report [25]. The recombinant endoglucanase *At*GH9C-CBM3A-CBM3B, consists of 857 amino acids having 828 amino acids from natural protein, rest 29 amino acids are constituted from the vector sequence: 21 amino acids at its N-terminal including His_6-tag and 8 amino acids at its C-terminal [25].

Owing to low query coverage, below 20% resulting from the amino acid BLAST of *At*GH9C-CBM3A-CBM3B against the PBD database, PSI-BLAST analysis with the sequence of catalytic module i.e. *At*GH9C was performed. The PSI-BLAST analysis of recombinant amino acids sequence of only catalytic modules containing 6X-His tag at N-terminal i.e. *At*GH9C against the PDB database exhibited the highest score of 317 with 96% of query cover and 38.38 % sequence identity against cellulase Cel9M from *Ruminiclostridium cellulolyticum* (PDB ID: 1IA6) (**Table 4.1**).

Table 4.1 Sequence investigation of homologs *At*GH9C-CBM3A-CBM3B using PSI-BLAST analysis against PDB database.

Organism	Enzyme	PDB ID	Max. Score	Query Cover (%)	Identity (%)	E-Value	PDB Resolution (Å)
Ruminiclostridium cellulolyticum	Cellulase Cel9M	1IA6	317	96	38.38	6e-103	1.8
Caldicellulosiruptor bescii	1,4-beta-glucanase	4DOD	235	100	35.55	2e-71	1.7
Thermobifida fusca	Endo/Exocellulase E4	1JS4	239	95	34.68	2e-71	2.0
Perinereis brevicirris	Endoglucanase	4ZG8	202	95	31.91	3e-59	1.4
Eisenia fetida	Endoglucanase	8IHW	209	95	31.69	1e-61	1.7
Nasutitermes takasagoensis	Endo-β-1,4-glucanase	1KS8	199	95	30.55	9e-58	1.4

The BLAST analysis further revealed sequence identity in the range of 35-24% with other endoglucanases from different bacterial sources *viz.*, 1,4-beta glucanase from *Caldicellulosiruptor bescii* (PDB-ID: 4DOD), endocellulase E4 form *Thermobifida fusca* (PDB-ID: 1JS4) and chain A, endoglucanase from *Perinereis brevivirris* (PDB-ID: 4ZG8). The homologs amino acid sequence retrieved from BLAST analysis and the *At*GH9C sequences were aligned with Clustal Omega and viewed with EsPript. Furthermore, the conserved and the semi-conserved residues of *At*GH9 were envisioned from the multiple sequence alignment (MSA) are shown in white letters under dark red background and red letter with white background, respectively (**Fig. 4.1**).

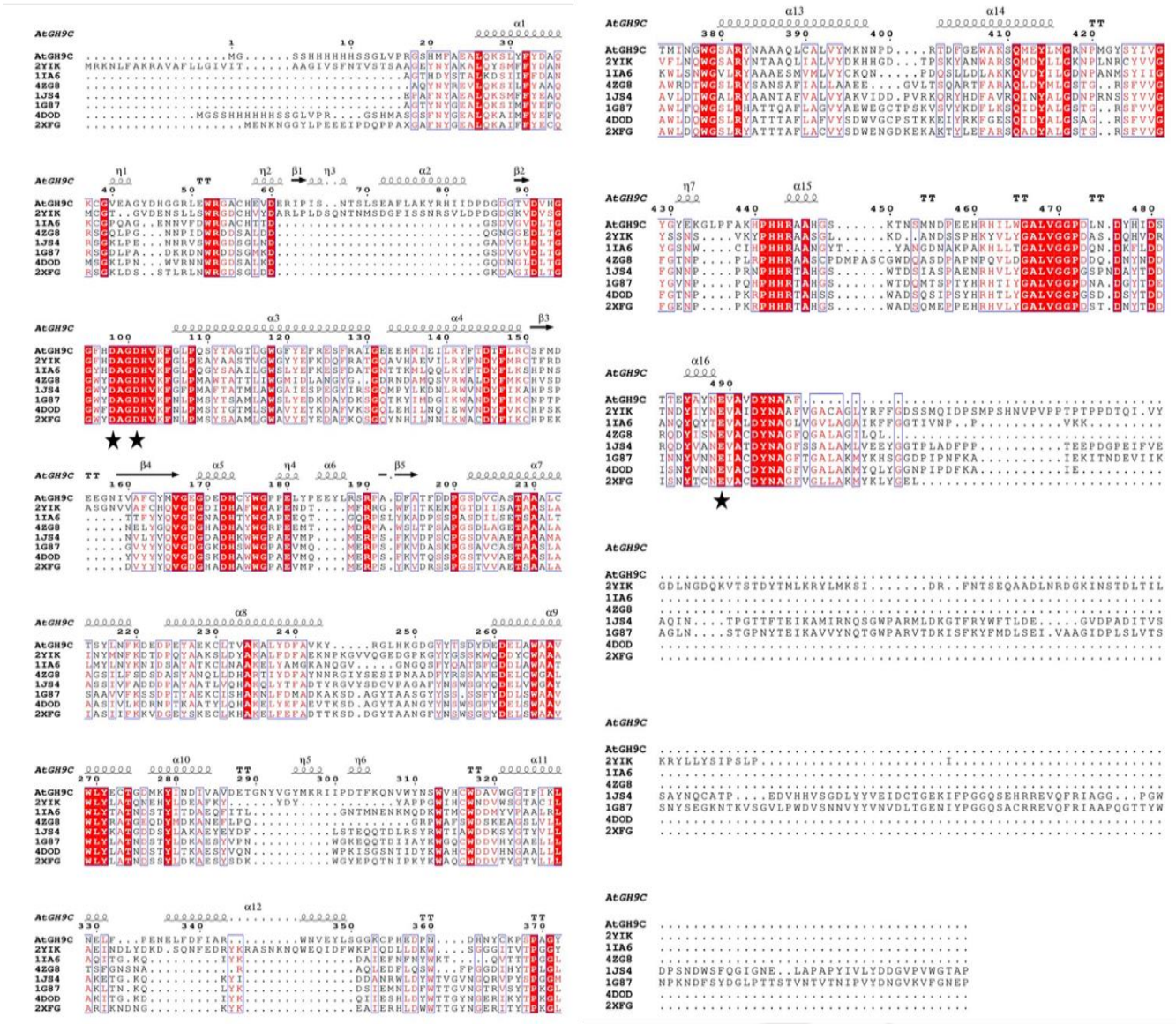


Fig. 4.1 Multiple sequence analysis of catalytic module i.e. *At*GH9C of endoglucanse *At*GH9C-CBM3A-CBM3B with its PDB-homologs and its visualization using EsPript 3.0. Amino acid single-letter codes that are embedded in red patches indicate conserved residues while single-letter codes that are colored red indicate semi-conserved residues. The residues forming the catalytic triad (Asp98, Asp101 and Glu489) are indicated with black asterisk.

The MSA result demonstrated three conserved amino acid residues Asp98, Asp101 and Glu489 as the principal amino acids involved in catalysis and it is conserved throughout other homologs and GH9 family. It is also perceivable from the MSA analysis that the residues Leu26, Phe32, Gly39, Trp53, Arg54, Asp60, Gly94,

His102, Val 103, Thr198, Pro200, Asp261, Trp265, Lys328, Pro369, Ala491 are conserved residues (**Fig. 4.1**). MSA visualization through EsPript showed the presence of total of 23 helices (16α + 7η), 5 β-strands and 10 turns (TT) in the *At*GH9C. The phylogenetic tree analysis of *At*GH9C-CBM3A-CBM3A against non-redundant protein sequences established its evolutionary relationships with different GH9 endoglucanases from bacterial and fungal sources (**Fig. 4.2**). The phylogenetic analysis depicted that Glycoside hydrolase from *Clostridium sp.* (Locus ID: NLM57935) is closely evolutionary related to *At*GH9C-CBM3A-CBM3B followed by a family 9 endoglucanase from *Acetivibrio mesophilus.* (Locus ID: WP128706291).

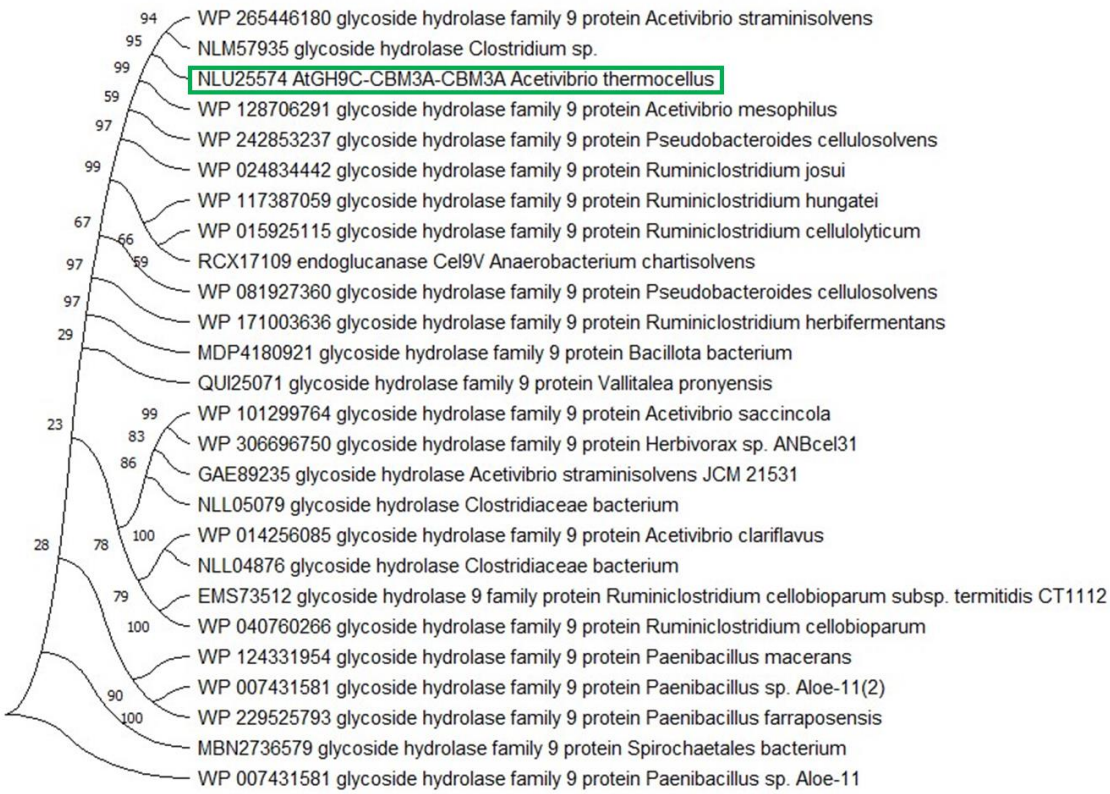


Fig. 4.2 Phylogenetic analysis using the neighbour-joining method using Mega11 to explain the evolutionary link between the *At*GH9C-CBM3A-CBM3B (marked in green-box) sequence and its homologs.

4.3.2 Secondary structural elements analysis of *At*GH9C-CBM3A-CBM3B

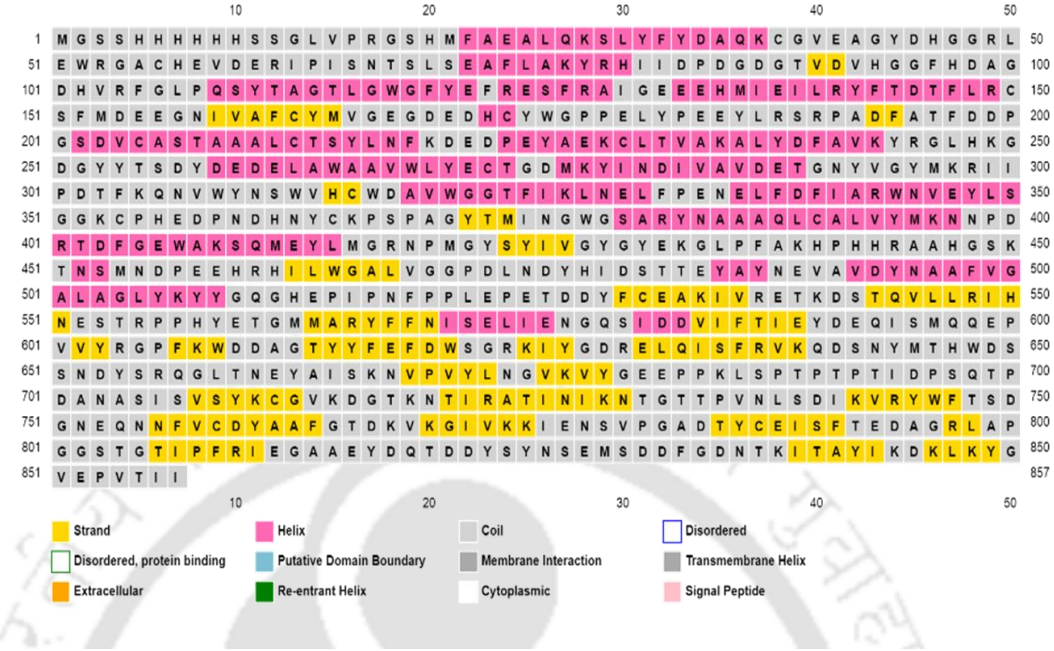
The secondary structure elemental composition of *At*GH9C-CBM3A-CBM3B was predicted from its amino acid sequence with the aid of online servers PSIPRED, GOR4 and SOPMA. The assessment from the servers indicated that *At*GH9C-CBM3A-CBM3B is marked by the presence of average 25.5% of α-helix, 20.7% of β-strands and 53.8% of random coils (**Table 4.2**).

Table 4.2 Secondary structure elements estimation of *At*GH9C-CBM3A-CBM3B.

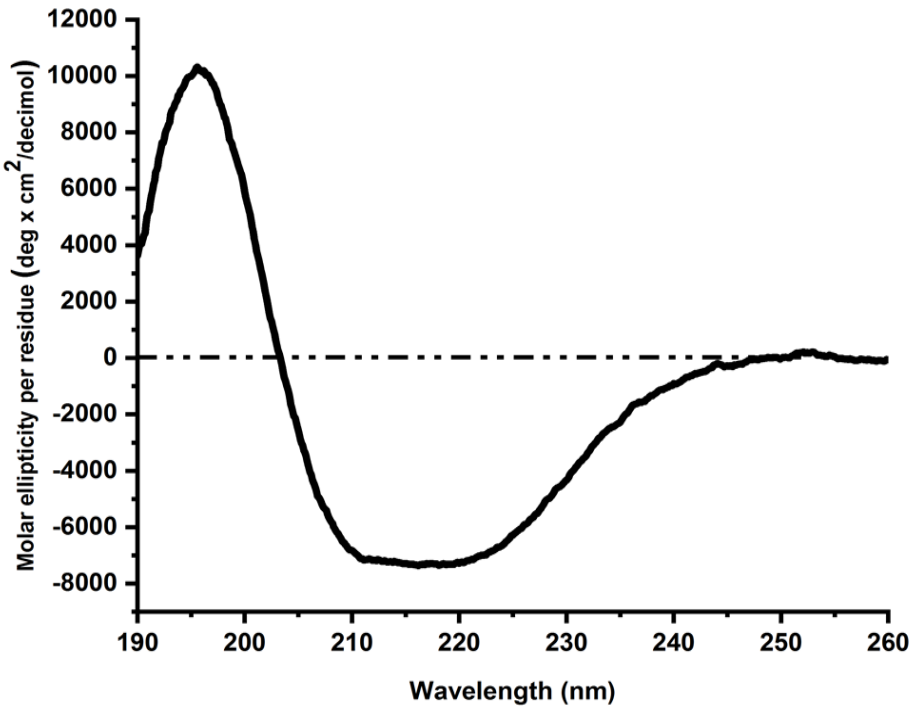
Server	α-helix (%)	β-Strand (%)	Coils (%)
PSIPRED	26.1	17.5	56.4
GOR4	24.5	23.4	52.1
SOPMA	26.3	21.4	52.3
APSSP	26.9	16.9	56.2
K2D3 (from CD)	25.2	18.4	56.4

PSIPRED server displayed presence of 224 amino acid residues in α-helix region, 150 residues in β-sheet and 483 amino acid of *At*GH9C-CBM3A-CBM3B in random coils (**Fig. 4.3a**). The CD data generated with purified *At*GH9C-CBM3A-CBM3B in the Far-UV region were assessed by using K2D3 server (http://k2d3.ogic.ca/). The result displayed nearly 25.2% absorption in α-helix, 18.4% in β-sheet area and rest 56.4% of random coils (**Fig. 4.3b**). The predicted secondary structure elements of *At*GH9C-CBM3A-CBM3B from online servers PSIPRED, GOR4 and SOPMA were consistent with the CD results. The state-of-the-art prediction of surface accessibility of the residues assessed from NetSurfP-2.0 server (**Fig. 4.3c**) indicated that mostly, the α-helix and random coils are present at the surface, whereas the β-strands are buried inside the core of the protein and constitutes the Carbohydrate Binding Modules (CBMs).

a)



b)



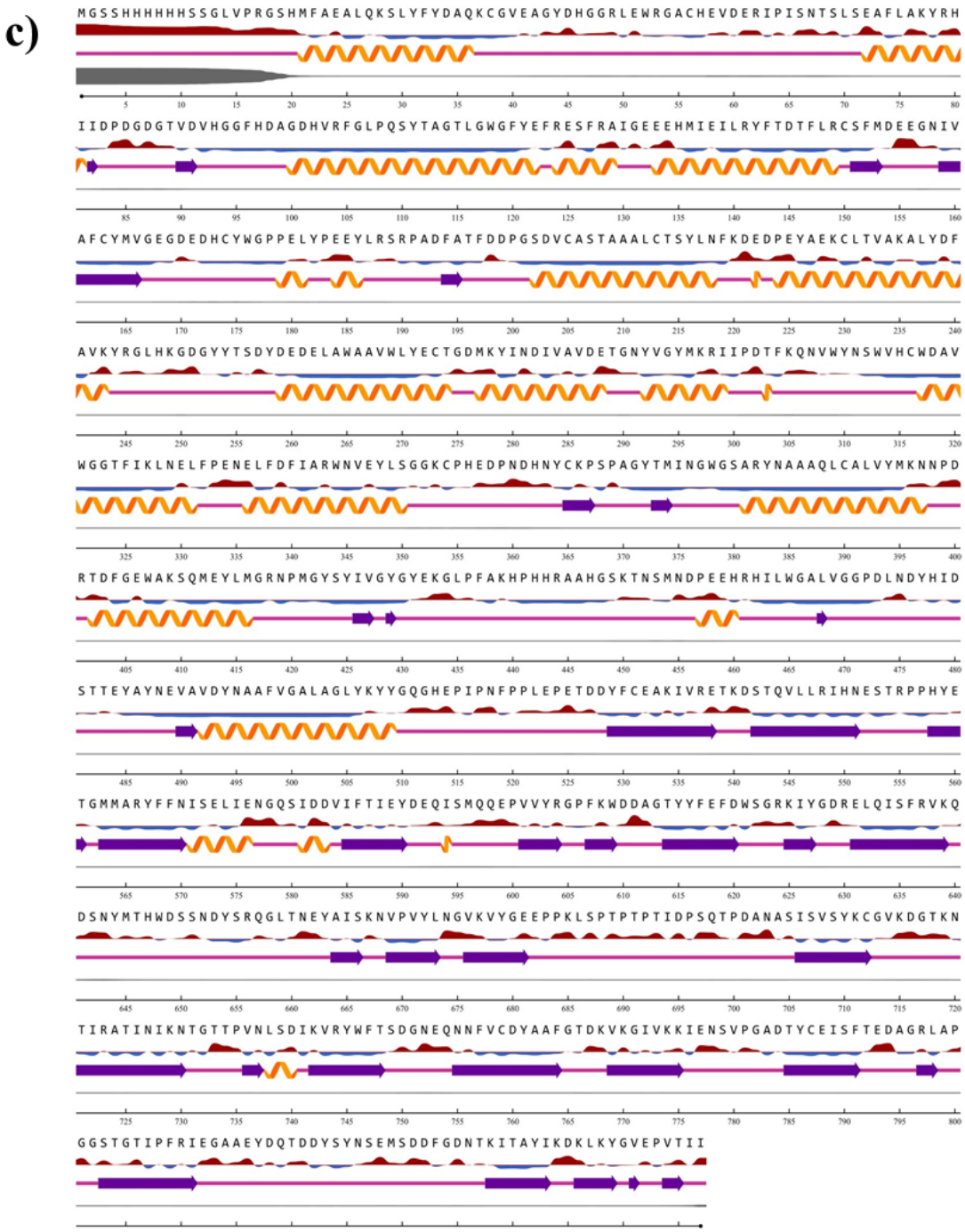


Fig. 4.3 Secondary structure estimation of *At*GH9C-CBM3A-CBM3B by (a) PSIPRED tools, (b) Far-UV CD spectrum and (c) NetSurfP-2.0 server. In PSIPRED, the alpha helix is shown by a pink line, the beta strands by a yellow line, and the random coils by a black line. After procurement of CD data, the proportion of α-helices, β-sheets, and random coils present in *At*GH9C-CBM3A-CBM3B three-dimensional structure was calculated by K2D3 software NetSurfP-2.0 represents the buried and exposed amino acid residues in red and blue color, respectively.

4.3.3 Homology modeling of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A, *At*GH9C

Putative domains and structural models of *At*GH9C-CBM3A-CBM3B were created by using various automated protein structure tools available online, mentioned in **section 4.2.4**. Robetta server predicted the best fitted 3D structure (**Table 4.3**) of *At*GH9C-CBM3A-CBM3B using *de novo* prediction method [34]. Robetta server also proved to provide better quality model than AlphaFold attributing membrane compatible topologies [57]. The model was energy minimized by the YASARA tool to define the significant topologies in space.

Table 4.3 Stereochemical quality assessment of models of *At*GH9C-CBM3A-CBM3B using SAVES v6.0.

Server	**Verify-3D (%)**	**ERRAT (%)**	**PROCHECK Ramachandran Plot (% and Residue Number)**				
			Most Favoured (%)	*Additional Allowed (%)*	*Generously Allowed (%)*	*Total Allowed*	*Disallowed (%)*
I-TASSER	89.30 PASS	72.08	74.2	17.4	7.3	98.9	1.1
Phyre2	79.20 Fail	66.76	75.4	17.7	5.6	98.7	1.3
SWISS Model	85.16 PASS	83.36	85.2	13.2	0.5	98.9	1.1
AlphaFold	90.74 PASS	91.21	88.2	11.5	0.3	100	0
Robetta	**91.74 PASS**	**93.68**	**89.7**	**10**	**0.3**	**100**	**0**
RaptorX	86.74 PASS	87.25	85.5	12.7	0.8	99	1.0

The modeled structure revealed the presence of N-terminal catalytic domain, composed of $(\alpha/\alpha)_6$- barrel topology and its two appended carbohydrate binding modules i.e. CBM3A and CBM3B, present at the C-terminal displaying an antiparallel β-sandwich fold. The CBMs are connected by a linker peptide. PyMOL 2.0 software

was employed to generate and analyze the cartoon view (**Fig. 4.4a**) and surface view (**Fig. 4.4b**) of the modeled structure of *At*GH9C-CBM3A-CBM3B. The α-helix, β-sheets and loops were numbered by following the 2D topology file obtained from the PDBSum server (**Fig. 4.4c**). 2-D topology analysis of *At*GH9C-CBM3A-CBM3B form PBDSum server revealed the presence of total 27 α-helices and 26 β-sheet contributing to the surface and core topological arrangement of 3D modeled structure (**Fig. 4.4d**). The surface view of homology modeled *At*GH9C-CBM3A-CBM3B showed a shallow substrate binding cleft and the active-site of catalytic module (*At*GH9C) was guarded with the arrangement of four loops, namely loop 1 (244-259 aa), loop 2 (291-320 aa), loop 3 (342-379 aa) and loop 4 (448-484 aa) (**Fig. 4.4e**). Loop 3 is anticipated to block the end of the active-site. The formation of the $(\alpha/\alpha)_6$-barrel structure at N-terminal, comprises the 12 longest α-helices out of the 27 α-strands. This $(\alpha/\alpha)_6$-barrel topology forms the core of the catalytic module (*At*GH9C), giving a funnel-like shape. Similar structural arrangement of the catalytic module is evident in crystal structure of Cel9M form *Ruminiclostridium cellulolyticum* [59]. The C-terminal CBM3A and CBM3B modules are composed of evenly distributed 26 β-sheet over two modules, contributing to the antiparallel β-sandwich structure. Other family 3, CBM proteins also displayed a similar type of structural arrangement [60].

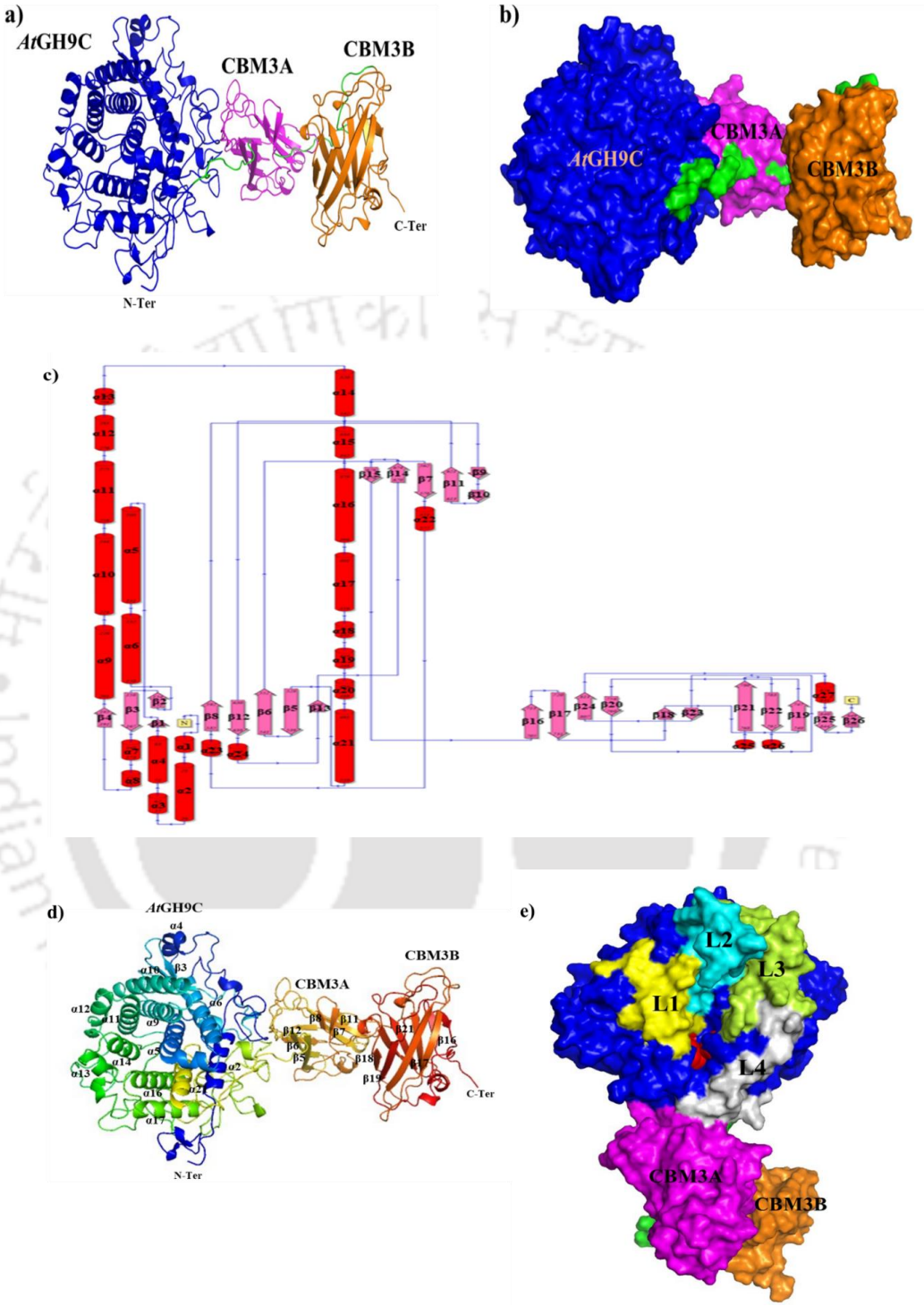


Fig. 4.4 Homology modelled 3-D structure of *At*GH9C-CBM3A-CBM3B. (a) Cartoon view and (b) Surface view of modelled structure, the catalytic domain (*At*GH9C) of modelled *At*GH9C-CBM3A-CBM3B is shown in blue, while CBM3A in magenta and CBM3B in yellow with linkers in green color. (c) Topology diagram of modelled *At*GH9C-CBM3A-CBM3B generated using PDBSum showing total total 27 α-helices and 26 β-sheet, (d) Cartoon view of the model highlighting α-helices and β-sheet localization, (e) Surface view of the overall *At*GH9C-CBM3A-CBM3B showing the positions of loops present in the catalytic module.

Structural homology modeling of the catalytic module, *At*GH9C (**Fig. 4.5a**) and the derivate *At*GH9C-CBM3A (**Fig. 4.5b**) revealed $(\alpha/\alpha)_6$-barrel topology in their catalytic core and antiparallel β-sandwich arrangement in CBM3A.

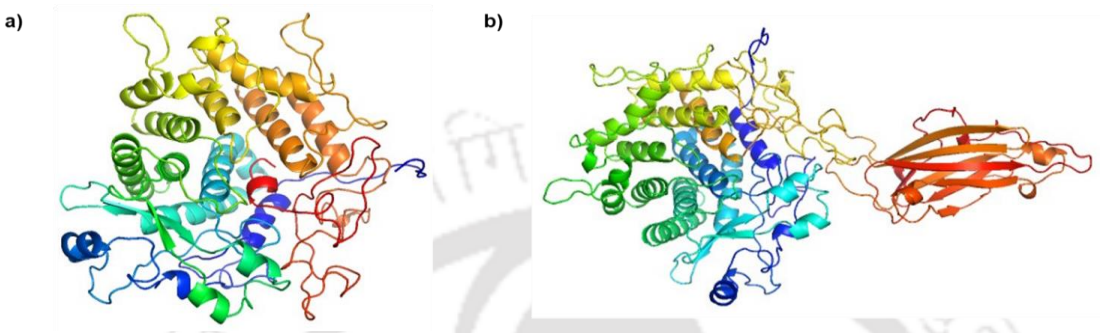


Fig. 4.5 Cartoon view of homology modeled a) *At*GH9C and b) *At*GH9C-CB3A from Robetta server.

The modeled *At*GH9C and *At*GH9C-CBM3A structures were validated using SAVES v6.0. The stereochemical quality assessment of *At*GH9C and *At*GH9C-CBM3A is depicted in **Tables 4.4a** and **Table 4.4b** respectively. Following which, the Robetta modeled structures of both were considered for further analysis.

Table 4.4a Quality assessment of *At*GH9C model by SAVES v6.0.

Server	**Verify-3D (%)**	**ERRAT (%)**	**PROCHECK Ramachandran Plot (% and Residue Number)**				
			Most Favoured (%)	*Additional Allowed (%)*	*Generously Allowed (%)*	*Total Allowed*	*Disallowed (%)*
I-TASSER	82.45 PASS	70.12	78.5	15.8	4.3	98.6	1.4
Phyre2	81.40 PASS	71.34	80.5	12.3	5.9	98.7	1.3
SWISS Model	91.30 PASS	84.65	88.1	10.4	0.6	99.1	0.9
Robetta	**96.55 PASS**	**85.78**	**89.4**	**10.0**	**0.6**	**100.0**	**0.0**
AlphaFold	95.66 PASS	85.80	88.8	9.5	0.7	100.0	0.0
Raptor X	89.24 PASS	81.44	86.4	11.5	1.0	98.9	0.9

Table 4.4b Quality assessment of *At*GH9C-CBM3A model by SAVES v6.0.

Server	Verify-3D (%)	ERRAT (%)	PROCHECK Ramachandran Plot (% and Residue Number)				
			Most Favoured (%)	*Additional Allowed (%)*	*Generously Allowed (%)*	*Total Allowed*	*Disallowed (%)*
I-TASSER	84.30 PASS	67.08	70.3	22.4	5.3	98.0	2.0
Phyre2	81.66 PASS	63.56	82.4	11.7	3.9	98.0	2.0
SWISS Model	90.24 PASS	85.45	87.2	11.3	0.5	99.0	1.0
Robetta	**95.91 PASS**	**82.75**	**88.4**	**10.7**	**0.9**	**100.0**	**0.0**
AlphaFold	95.65 PASS	83.22	89.7	9.5	0.8	100.0	0.0
Raptor X	90.74 PASS	80.22	85.5	12	0.8	98.5	1.7

4.3.4 Structure quality assessment of *At*GH9C-CBM3A-CBM3B

The energy minimized 3D modeled structure of *At*GH9C-CBM3A-CBM3B was validated through Ramachandran plot (**Fig. 4.6a**) for the backbone phi (φ) and psi (ψ) dihedral angles of the amino acid residues through PROCHECK server. The plot depicted 89.7% residues in most favored region, 10.0% residues in additional allowed region and 0.3% residues in generously allowed region. As a result, all residues of *At*GH9C-CBM3A-CBM3B were found in the permitted region, while none were found in the disallowed region. Based on the Ramachandran plot, it was determined that the *At*GH9C-CBM3A-CBM3B structure has a proper backbone having allowable phi (φ) and psi (ψ) angles. Furthermore, a Z-score of -10.86, from the ProSA analysis also validated the accuracy and compatibility of *At*GH9C-CBM3A-CBM3B modeled structure (**Fig. 4.6b**). ProSA uses the nuclear magnetic resonance (NMR) and X-ray crystallography databases to compare the modeled structure parameters with those of

the solved protein structures. A high negative Z-value indicates a better protein model with little energy separation between the native fold and the ensemble of misfolded protein. The assessment of local model quality from ProSA demonstrates only few amino acids have positive knowledge-based values, indicating the overall correctness of the model. The model topology correctness of *At*GH9C-CBM3A-CBM3B was further established from the Verify-3D tool, displaying 91.74% percent of the total amino acids in *At*GH9C-CBM3A-CBM3B above the average 3D-1D score (≥ 0.2) (**Fig. 4.6c**). The fitness credibility of the model is thus verified from the Verify-3D analysis. Overall quality assessment by ERRAT server indicated the threshold for *At*GH9C-CBM3A-CBM3B to be 93.68%, further confirming the excellent quality structure (**Fig. 4.6d**). The modeled *At*GH9C-CBM3A-CBM3B structure was considered satisfactory by several means of evaluations, therefore it was endorsed for further analysis.

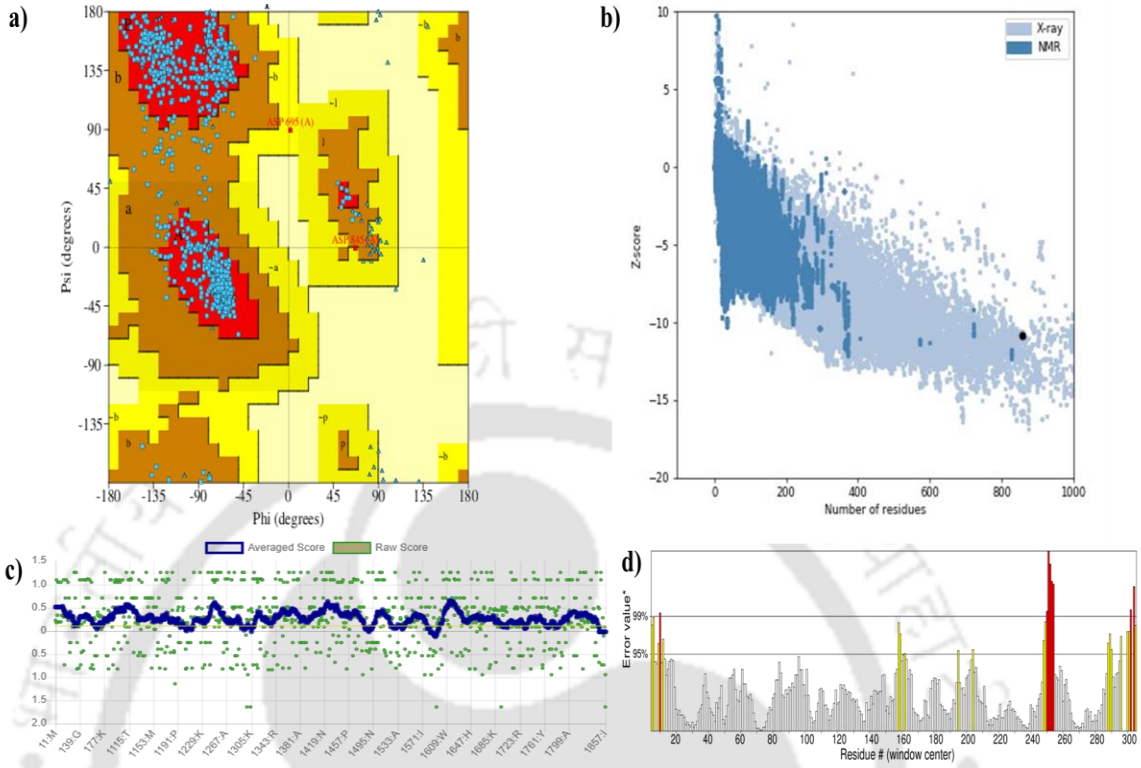


Fig. 4.6 Structural quality assessment of 3-D model of *At*GH9C-CBM3A-CBM3B on different stereochemical parameters. showing (a) Ramachandran plot of *At*GH9C-CBM3A-CBM3B generated by PROCHECK server, (b) ProSA plot displays Z-score of -10.86, indicating positive confidence of the model, (c) *At*GH9C-CBM3A-CBM3B, 3D-1D average score examined using VERIFY3D, (d) ERRAT plot displaying 89% of amino acid residues of *At*GH9C-CBM3A-CBM3B below threshold error values.

4.3.5 Active-site analysis and mechanism of hydrolysis

The acid and base, active-site residues of *At*GH9C-CBM3A-CBM3B were enumerated by superposing the homology model of *At*GH9C-CBM3A-CBM3B with its evolutionary-related PBD homologs 1IA6 (Cel9M) from bacterium *Ruminiclostridium cellulolyticum* [59] and 1,4-beta glucanase (CelA) from *Caldicellulosiruptor bescii* (PDB-ID: 4DOD) [61]. Albeit 1IA6 and 4DOD represent the closest PDB homologs of *At*GH9C-CBM3A-CBM3B, their superimposition with validated 3D modeled of *At*GH9C-CBM3A-CBM3B in PyMol 2.0 resulted in RMSD

value of 4.3 Å and 4.5 Å, respectively (**Fig. 4.7a**). The RMSD around 4.5 Å represents the distant homology modeling case for modeled *At*GH9C-CBM3A-CBM3B, conferring the uniqueness of the model structure [62]. The superimposition results further displayed three conserved amino acid residues namely, Asp98, Asp101 and Glu489 forming the catalytic triad involved in catalysis. Two amino acid residues namely Asp98 and Asp101 of *At*GH9C-CBM3A-CBM3B correspond to Asp56, Asp59 of Cel9M and Asp77, Asp80 of CelA are considered to act as a putative catalytic base, while Glu489 of *At*GH9C-CBM3A-CBM3B shows equivalence to Glu410 of Cel9M and Glu443 of CelA that performs the function of catalytic acid and present as DAGD motifs [59,61]. The superimposition of homology model depicts catalytic cleft of *At*GH9C-CBM3A-CBM3B is narrow in structure with enough space to accommodate the 4 to 6 glucose residues at subsites -4 to +2 with the help of subsite binding amino acid residues His173, Trp176, Trp265, Trp269, Trp313, Trp378, IIe463 and Trp465 [63]. Catalysis of glycosidic bonds of glucans is supposed to occur via the binding of water molecule between the catalytic base Asp98 and Asp101, favoring the initiation of attack resulting in the formation of oxocarbonium-like transition state and release of the intermediate from the other side of the active cleft via catalytic acid Glu489 [64]. The catalytic mechanism (retaining or inverting) of the enzyme was elucidated by computing the average distance between the two carboxyl groups of the catalytic residues using PyMol 2.0. Experimental interpretation of glycosidases suggested that, the average separation between catalytic residues varies in range 9.0-9.5 Å in inverting mechanism to accommodate a water molecule, whereas in retaining the range lies between 4.8-5.3 Å [65]. The *in-silico* analsyis of *At*GH9C-CBM3A-CBM3B estimated the distance of 8.5 Å between the two carboxyl groups, Asp98/Asp101 (catalytic base)

and Glu489 (catalytic acid), indicating an inverting-type of reaction mechanism (**Fig. 4.7b**).

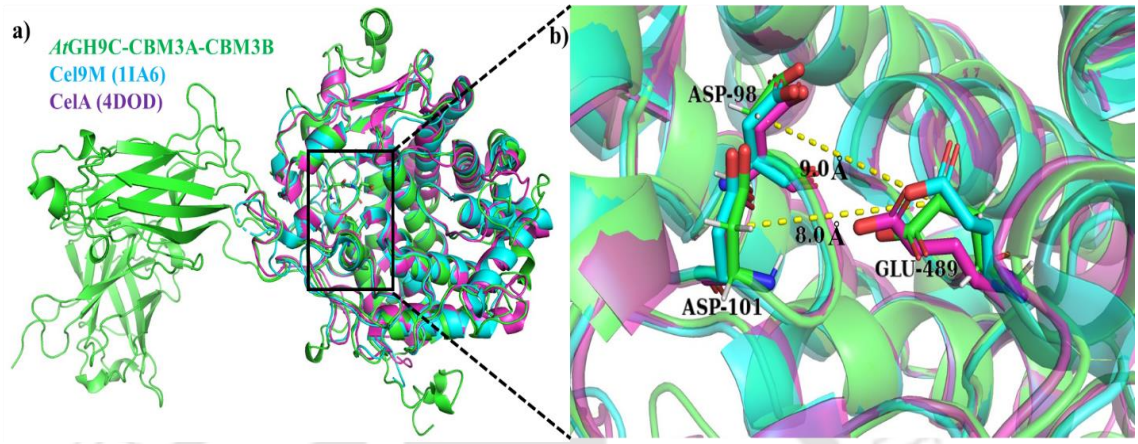


Fig. 4.7 ***At*GH9C-CBM3A-CBM3B (Green color) superposed with its (a) PDB homologs, 1IA6 (Sky color) from *Ruminiclostridium cellulolyticum* (Cel9M) and 4DOD (Magenta) from *Caldicellulosiruptor bescii* (CelA), (b) The active-site residues of *At*GH9C-CBM3A-CBM3B (green) are superimposed on top of the active-site of Cel9M (Sky color) and CelA (Magenta). The average measured distance of 8.5 Å between catalytic base (Asp98 and Asp101) and catalytic acid (Glu489) supporting the inverting-type catalytic mechanism.**

4.3.6 Molecular dynamics simulation of modeled *At*GH9C-CBM3A-CBM3B structure

The spatial and temporal properties at the nanoscale level of the 3D modeled protein structure of *At*GH9C-CBM3A-CBM3B was assessed for stability, global compactness, interatomic interaction and conformational behavior by subjecting the same for MD simulation for 200 ns. The root mean square deviation (RMSD) of simulated *At*GH9C-CBM3A-CBM3B from its primary structure at 0 ns was analyzed in due course in the form of a time-dependent function over 200 ns. The RMSD measures the changes in the conformation of the protein on the application of external processing stress *viz.* temperature, pressure and electric field. The RMSD analysis projected the structural topology of the modeled *At*GH9C-CBM3A-CBM3B structure showing the fluctuation in the range, 0.3 to 2.1 nm until 140 ns, and it was stabilized at

an average RMSD value of 1.8 nm after 140 ns up to 200 ns (**Fig. 4.8a**). As protein tumbling motions are inferred in range of nanoseconds [66], a stabilized RMSD value confirms the structural stability of modeled *At*GH9C-CBM3A-CBM3B pointing the initial and final coordinates of atom remaining same at 1.8 ns after 140 ns. The Radius of gyration (R_g) fluctuated in the range, 3.4 nm to 3.6 nm up to 30 ns, and after that it abruptly increased to 4.6 nm at 40 ns, However, it remained remarkably stable from 150 ns to 200 ns at an average R_g of 4.2 nm, indicating its global compactness and less change in moment of inertia of secondary structure conformation (**Fig. 4.8b**). The fluctuations of backbone C_α-atoms of the amino acid residues in different regions are further assessed with the aid of Root Mean Square Fluctuations (RMSF) analysis to determine atomic displacement in simulated structure in reference of original topologies. The RMSF analysis revealed, that the residues forming the buried structural elements viz., α-helices and β-sheet are prone to less fluctuations. The accessible and flexible residues mainly found in loops (loop 1 (244-259 aa), loop 2 (291-320 aa), loop 3 (342-379 aa) and loop 4 (448-484 aa)) were abruptly fluctuating possibly accounting for its interaction to exposed aqueous environment with RMSF values more than 1.0 nm (**Fig. 4.8c,** blue dotted circles). It is worth mentioning that the residues that form the core of CBM3B, displaying β-sheet topologies were highly fluctuating with moderate fluctuations in CBM3A amino acid residues. Perhaps, this behavior can be better explained by the redundancy of CBM3B in revival of catalytic activity [25]. However, the relative position and orientation of all the catalytic residues (Asp98, Asp 101 and Glu489) are recorded to be relatively stable with RMSF value of 0.15 over simulation time of 200 ns, signifying the buried and core forming nature (**Fig. 4.8c,** green circles). Therefore, the catalytic residues remained stable while only the loops of

*At*GH9C-CBM3A-CBM3B were observed to be variable in RMSF. Additionally, the SASA of the simulated protein over 200 ns remained constant in the range 400 nm^2 to 440 nm^2, giving an average value, 425 nm^2 (**Fig. 4.8d**). This observation demonstrated that the hydrophobic interactions between the amino acid clusters remained intact with no variation in solvent accessibility to the catalytic residues, ensuring the fitness of the structure. Furthermore, the main chain of *At*GH9C-CBM3A-CBM3B contained an average of 590 intramolecular H-bonds during 200 ns simulation (**Fig. 4.8e**) which significantly provided the structural stability to the protein. The compactness of *At*GH9C-CBM3A-CBM3B was further confirmed by superposing the MD simulated structure with the original modeled structure with RMSD of 1.3 nm, which revealed that the location of catalytic residues remained intact (**Fig. 4.8f**). Molecular simulation of the model revealed overall stability and global compactness throughout the time-span of 200 ns.

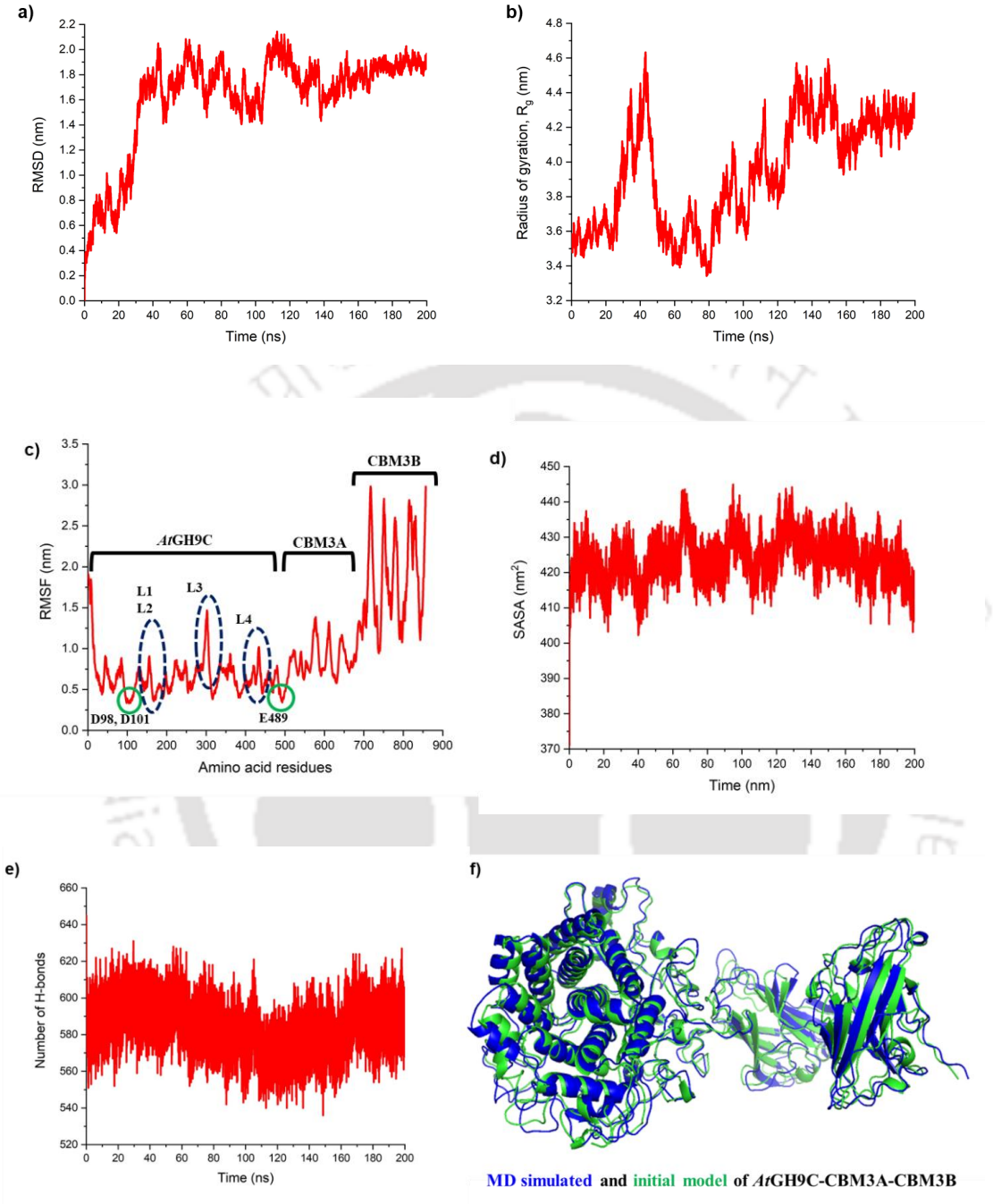


Fig. 4.8 Molecular dynamics simulation analysis of Robetta modeled *At*GH9C-CBM3A-CBM3B for 200 ns showing a) RMSD, b) Radius of gyration (R_g), c) RMSF, d) SASA and e) Intramolecular hydrogen bonds and f) Superposed final 200 ns simulated model (blue) with initial input model (green) of *At*GH9C-CBM3A-CBM3B.

4.3.7 Molecular docking analysis of *At*GH9C-CBM3A-CBM3B and its substrate binding

The docking interaction of various cello-oligosaccharides with the amino acid residues present in the catalytic cleft of *At*GH9C-CBM3A-CBM3B was performed with the aid of Autodock 4.2.1 tool to facilitate the understanding of its substrate binding. Cellulosic ligands with different degree of polymerization mainly, cellobiose, cellotriose, cellotetraose, cellopentaose and cellohexaose were used to visualize docking resolution with the receptor protein *At*GH9C-CBM3A-CBM3B. Specific docking analysis result (ΔG, kcal/mol) of various cellulosic ligands with the *At*GH9C-CBM3A-CBM3B in AutoDock 4.2.1 displaying active-site residues and its sub-site interacting residues of *At*GH9C-CBM3A-CBM3B are listed in **Table 4.5**

Table 4.5 Molecular Docking analysis of *At*GH9C-CBM3A-CBM3B, *At*GH9C-CBM3A and *At*GH9C with cello-oligosaccharides.

Ligands	Binding Energy, ΔG (kcal/mol)	Residues making polar interactions	Residues making hydrophobic interactions
***At*GH9C-CBM3A-CBM3B**			
Cellobiose	-2.45	His173, His442, Glu489	Asp98, Asp101, Trp176, Tyr253, Tyr254, Trp378, Tyr485
Cellotriose	-3.40	Asp101, Tyr253, His442, Tyr485	Ala99, His173, Trp176, Tyr254, Trp378, Glu489
Cellotetraose	-5.05	His173, Tyr253, Ser453, Met454, Asn455, Thr483	Trp176, Tyr254, Trp378, His442, Arg444, Asn452, Tyr485
Cellopentaose	-3.45	Gly377, Trp378, Glu484	Ser312, Trp313, Tyr348, Cys365, Lys366, Met374, Ile375, Asn376, Asp473, Ala486, Tyr487
Cellohexaose	+1.10	Trp378, Thr483	Trp313, Ile375, Asn376, Gly377, Asn455, Ser481, Glu484, Tyr485, Ala486
***At*GH9C-CBM3A**			
Cellotetraose	-4.98	His173, Tyr253, Met454, Asn455, Thr483	Trp176, Tyr254, Asn376, Trp378, His442, Arg444, Asn452, Tyr485
***At*GH9C**			
Cellotetraose	-4.65	His173, Ser453, Met454, Asn455, Thr483	Trp176, Tyr254, Trp378, His442, Arg444, Asn452, Tyr485

The endoglucanase, *At*GH9C-CBM3A-CBM3B showed maximum binding affinity with cellotetraose with concomitant release of free energy (ΔG) of -5.05 kcal/mol. The lower binding affinity with the catalytic module *At*GH9C (ΔG of -4.65 kcal/mol) and the derivative *At*GH9C-CBM3A (ΔG of -4.98 kcal/mol) on docking against cellotetraose was observed (**Table 4.5**). These results indicated the contributory roles of CBMs in the ligand binding. Moreover, the near values of binding affinity of *At*GH9C-CBM3A-CBM3B and *At*GH9C-CBM3A suggested the limited role of CBM3B in substrate binding. An important observation was noted while calculating the docking score of *At*GH9C-CBM3A-CBM3B with its associated ligands. A surge in the exothermic release of free energy (ΔG) was noted with the increment of the degree of polymerization (DP) of the cellulosic ligand from cellobiose (DP2) to cellotetraose (DP4). The binding of cellooliogosaccharides, of DP ≥ 5 to catalytic cleft of *At*GH9C-CBM3A-CBM3B tends to be more endothermic, possessing positive ΔG (**Table 4.5**).

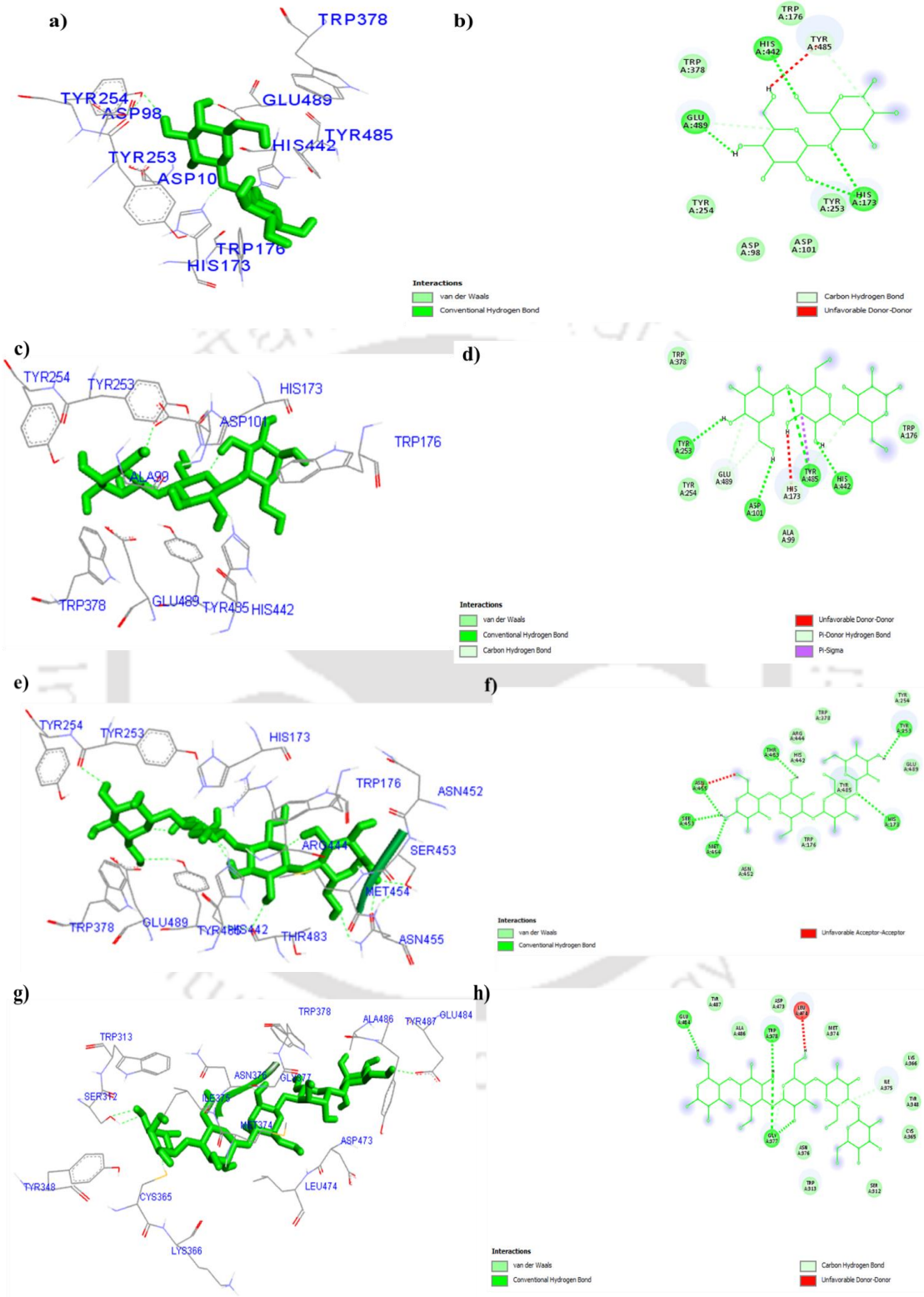
The binding free energy observed for ligands with DP ≤ 4 *viz.*, cellobiose, cellotriose and cellotetroase were 2.45, -3.40, -5.05 kcal/mol (**Fig. 4.9 a-f**), whereas with DP ≥ 5 ligands i.e., cellopentaose and cellohexaose, the values of ΔG were -3.45 and +1.10 (**Fig. 4.9 g-j**). This possibly can be explained by the observation that the catalytic cleft of the modeled *At*GH9C-CBM3A-CBM3B has enough space to accommodate four glucose residues occupying subsite -3 to +1 and get hydrolyzed. This interpretation is also supported by the fact the primary product of crystalline cellulose digestion is cellotetraose by family 9 enzyme Cel9M. [67]. Moreover, cellotetraose was detected as the major product in the TLC analysis of *At*GH9C-CBM3A-CBM3B as mentioned in our earlier report [25]. A comparative analysis of cellulases from *Clostridium thermocellum* revealed two types of processive

endoglucanases: one releases cellotetraose as an intermediate product (pEG4) and other one produces cellobiose as the major product (pEG2) [24]. Processivity analysis of *At*GH9C-CBM3A-CBM3B revealed, a mild processive behavior (**Table 4.6**) on PASC. The processivity index (PI) rises minimally along the time, supporting its classification in the pEG4 category of cellulases. However, with an increase of ligand DP > 5, the compact catalytic groove of the active-site may be unable to accommodate it due to the barrier formed by the loops surrounding the catalytic cleft. The loop 3 (342-379 aa) is anticipated to form a barrier referred to as tower as reported earlier by Guerriero et al., 2018 [68] at the non-reducing end of the catalytic site of *At*GH9C-CBM3A-CBM3B resulting in the cleavage of cellotetraose from cellulosic chain. Binding energy for cellopentaose could be described by its attachment to catalytic cleft of *At*GH9C-CBM3A-CBM3B and instant hydrolysis in cellobiose and cellotriose, leading to docking energy comparable to cellotriose.

Table 4.6 Processivity Index (PI) analysis of *At*GH9C-CBM3A-CBM3B with PASC.

Reaction Time	Reducing sugar in soluble fraction (µg/mL) [S]	Reducing sugar in insoluble fraction (µg/mL) [I]	Processivity Index (PI) [S/I]
Control	0.0	14.8	0.0
1 min	18.9	33.2	0.57
2 min	22.7	38.9	0.58
5 min	32.2	47.9	0.66
10 min	37.8	52.9	0.71
15 min	41.9	57.3	0.73
30 min	48.2	64.4	0.74
1 h	56.2	71.7	0.78
4 h	68.7	86.6	0.79
12 h	83.7	93.4	0.89
24 h	100.7	112.6	0.89

Three amino acid residues of aromatic nature mainly the Trp313, Trp378 and Tyr485 were believed to stack the sugar moieties at the +1, -1 and -3 subsite of catalytic cleft of *At*GH9C-CBM3A-CBM3B (**Fig. 4.9k**). This is in perfect agreement with the previously published crystal structure of cellulase Cel9M from *Clostridium celluloyticum* [69]. Employing Discovery Studio v 5.4 software, the docked complexes were further investigated to determine the amino acid residues involved in hydrophobic, hydrogen bond and Van der Waals interactions with the corresponding cellooligosaccharide by generating a 2-D plot. The interpretation of the 2-D plot revealed contribution of Asp101, His173, Tyr253 His442, Tyr458 Glu489 and Arg319 in the H-bond formation or polar interactions. The residues Asp98, Ala99 Asp101, His173 Trp176, Tyr253, Tyr254, Trp378, Tyr485, Glu489, Ser312, Trp313, Tyr348, Cys365, Lys366, Met374, Ile375, Asn376, Asp473, Ala486, Tyr487 were involved in the hydrophobic interaction with ligands. Needless to mention here, that one among the three amino acid residues involved in catalysis is also involved in polar interaction and assisted in catalysis. The outcome of the docking analysis revealed that *At*GH9C-CBM3A-CBM3B has a higher affinity towards cellotetraose and the same represents its major product after cleavage of cellulosic substrate.


a)
TRP378
TYR254
ASP98
GLU489
TYR485
HIS442
TYR253
ASP101
TRP176
HIS173
b)
TRP A:176
HIS A:442
TYR A:485
TRP A:378
GLU A:489
TYR A:254
TYR A:253
HIS A:173
ASP A:98
ASP A:101
Interactions
van der Waals
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Unfavorable Donor-Donor
c)
TYR254
TYR253
HIS173
ASP101
TRP176
ALA99
TRP378
GLU489
TYR485
HIS442
d)
TRP A:378
TYR A:253
TYR A:254
GLU A:489
TYR A:485
HIS A:173
HIS A:442
ASP A:101
ALA A:99
TRP A:176
Interactions
van der Waals
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Unfavorable Donor-Donor
Pi-Donor Hydrogen Bond
Pi-Sigma
e)
TYR254
TYR253
HIS173
TRP176
ASN452
ARG444
SER453
MET454
TRP378
GLU489
TYR485
HIS442
THR483
ASN455
f)
Interactions
van der Waals
Conventional Hydrogen Bond
Unfavorable Acceptor-Acceptor
g)
TRP378
ALA486
TYR487
GLU484
TRP313
SER312
ASN371
GLY372
MET374
ASP473
TYR348
CYS365
LEU474
LYS366
h)
Interactions
van der Waals
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Unfavorable Donor-Donor

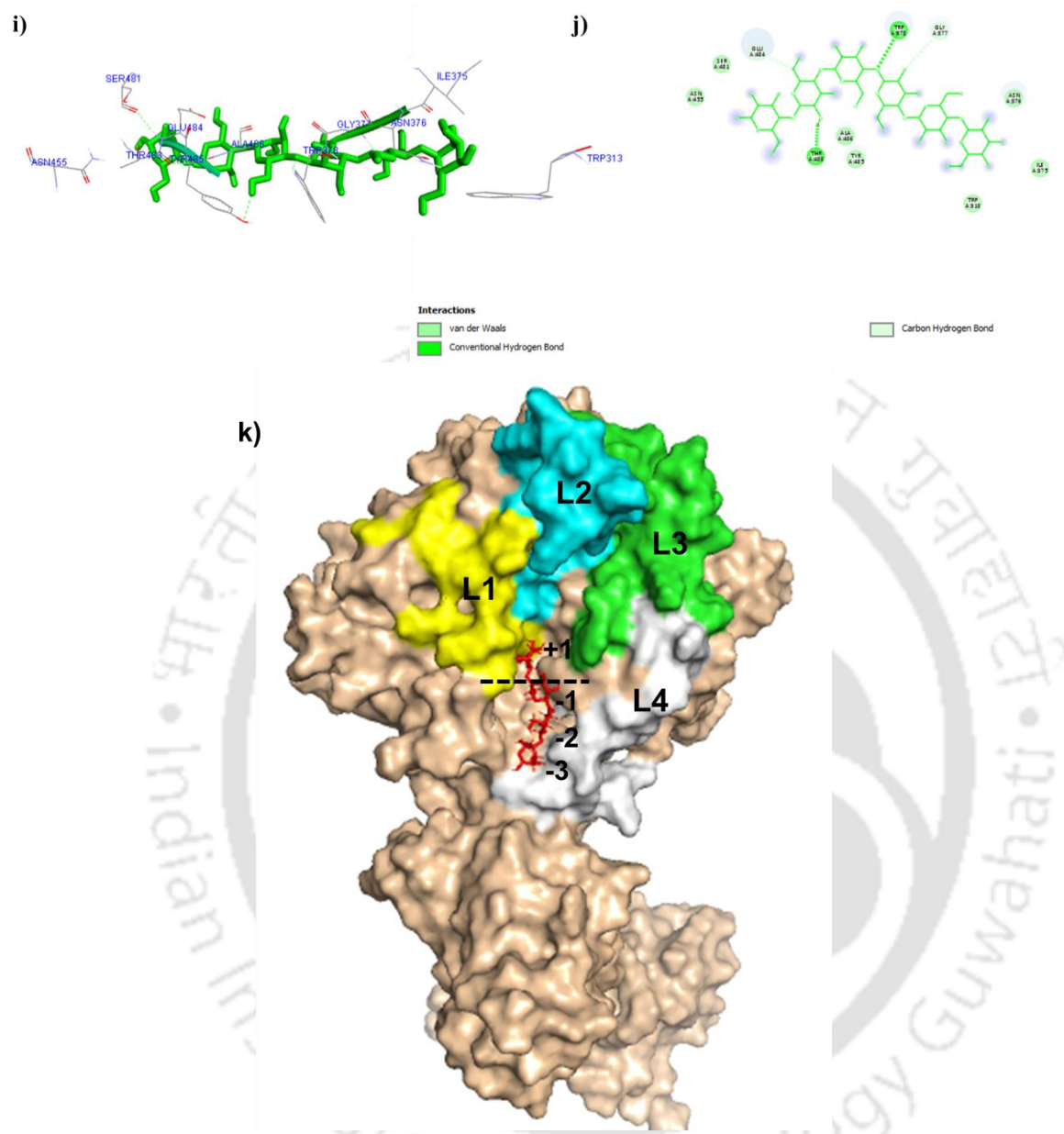


Fig. 4.9 Molecular docking analysis of *At*GH9C-CBM3A-CBM3B with various oligosaccharides ligands. (a, c, e, g, i) Interaction interface of catalytic residues of *At*GH9C-CBM3A-CBM3B with ligand cellobiose, cellotiose, cellotetraose, cellopentaose and cellohexaose, respectively. (b, d, f, h, j) 2D-representation of interaction between active-site residues of *At*GH9C-CBM3A-CBM3B with ligand cellobiose, cellotiose, cellotetraose, cellopentaose and cellohexaose respectively, showing conventional hydrogen bonds and Van der waals interactions. k) sub-site numbering for *At*GH9C-CBM3A-CBM3B-Cellotetraose complex.

4.3.8 Molecular dynamics simulation of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex

Energy-minimized modeled *At*GH9C-CBM3A-CBM3B forming the complex with cellotetraose (*At*GH9C-CBM3A-CBM3B-Cellotetraose) was subjected to MD simulation for 200 ns in order to understand the conformational dynamics, the compactness and stability. The simulated *At*GH9C-CBM3A-CBM3B-Cellotetraose complex was assessed by different structural parameters and dynamics datasets and compared with only *At*GH9C-CBM3A-CBM3B simulated structure to examine the possible impacts of *At*GH9C-CBM3A-CBM3B on binding to the substrate. The initial assessment of the variation in the structural composition of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex was done through the RMSD plot, depicting the system fluctuating in the initial 20 ns and then remaining stable and constant at an average of 1.5 nm up to 200 ns. The depicted average RMSD value of *At*GH9C-CBM3A-CBM3B-Cellotetraose was 0.3 nm lesser than the only *At*GH9C-CBM3A-CBM3B, simulated in the same time frame of 200 ns (**Fig. 4.10a**). The result pointed out that, probably the association of cellotetraose in the catalytic cleft of *At*GH9C-CBM3A-CBM3B stabilizes the protein locally and assists in attaining the incisive equilibrium state. Analysis of the structural compactness and stability of the docked complex was inferred from the Radius of gyration (R_g) plot, which showed the complex fluctuated for an initial 20 ns and after it remained stable at an average R_g of 3.3 nm throughout 200 ns (**Fig. 4.10b**). Unlike the only protein, where the R_g fluctuated, a stable R_g of the docked complex specifies the apparent binding of ligand plausibly decreased the changes in the rotational inertia of the CBMs, which were believed to bind with the substrate, contributing to the compactness and tightness of the system. The Root Mean Square Fluctuation (RMSF) provides information on both flexible and rigid regions of protein

structures. The analysis of C_α conformation of each amino acid residue in the docked complex was envisioned from RMSF analysis revealed, that the catalytically active residues (Asp98, Asp101 and Glu489) remained more stable in the docked complex (RMSF ≤ 0.1) as compared with the only protein (RMSF ≤ 0.2) (**Fig. 4.10c,** green circles). Moreover, the fluctuation of residues in the docked complex pertaining to the α-helices and β-strands and binding residues of overall protein dropped significantly as compared with the only protein demonstrating the increased residue stiffness when bonded to the ligand. Additionally, comprehensive decrease in fluctuation of the loop region was recorded in the RMSF plot of the docked complex compared to the individual *At*GH9C-CBM3A-CBM3B, indicating decreased flexibility of the loops (**Fig. 4.10c** blue dotted circles). The loops L1-L4 that clusters around the catalytic pocket located at -3 subsite position, showed a decrease in RMSF value by 0.5 nm, signifying a possible role of loop acting as gate for entry of the substrate. After conducive attachment of the ligand to the active-site, probably the loop become rigid, until the release of the substrate. There was reduced fluctuation in the RMSF value nearly by 1 nm in amino acid residues in CBM3A of the docked complex as compared with the RMSF value of CBM3A of only *At*GH9C-CBM3A-CBM3B. Moreover, there was significant decrease in fluctuation in RMSF value of CBM3B by 2.0 nm with respect the only *At*GH9C-CBM3A-CBM3B. This decrease in RMSF values of CBM3A and CBM3B upon binding the cellotetraose suggested the stability and compactness of the docked complex. The lower decrease in RMSF (1 nm) of CBM3A than CBM3B (2 nm) may be due to the close proximity of CBM3A to catalytic module, *At*GH9C again suggesting more stability and compactness of CBM3A upon ligand binding whereas CBM3B displaying the extended flexibility. The SASA of the docked complex

decreased by approximately, 20 nm^2 from the SASA of only the *At*GH9C-CBM3A-CBM3B simulated structure (**Fig. 4.10d**) and stayed close to an average 405 nm^2 for the duration of the trajectory up to 200 ns. This minor drop in SASA was expected because the presence of the ligand cellotetraose, somewhat restricted the solvent accessibility to the catalytic cleft. A mean of 610 intramolecular hydrogen bonds were detected in the docked complex, approximately 30 more than only simulated *At*GH9C-CBM3A-CBM3B suggesting effective ligand binding at *At*GH9C-CBM3A-CBM3B catalytic cleft (**Fig. 4.10e**). The comparative analysis of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex and only *At*GH9C-CBM3A-CBM3B simulated structures showed the stable and compact structure formation of the docked complex.

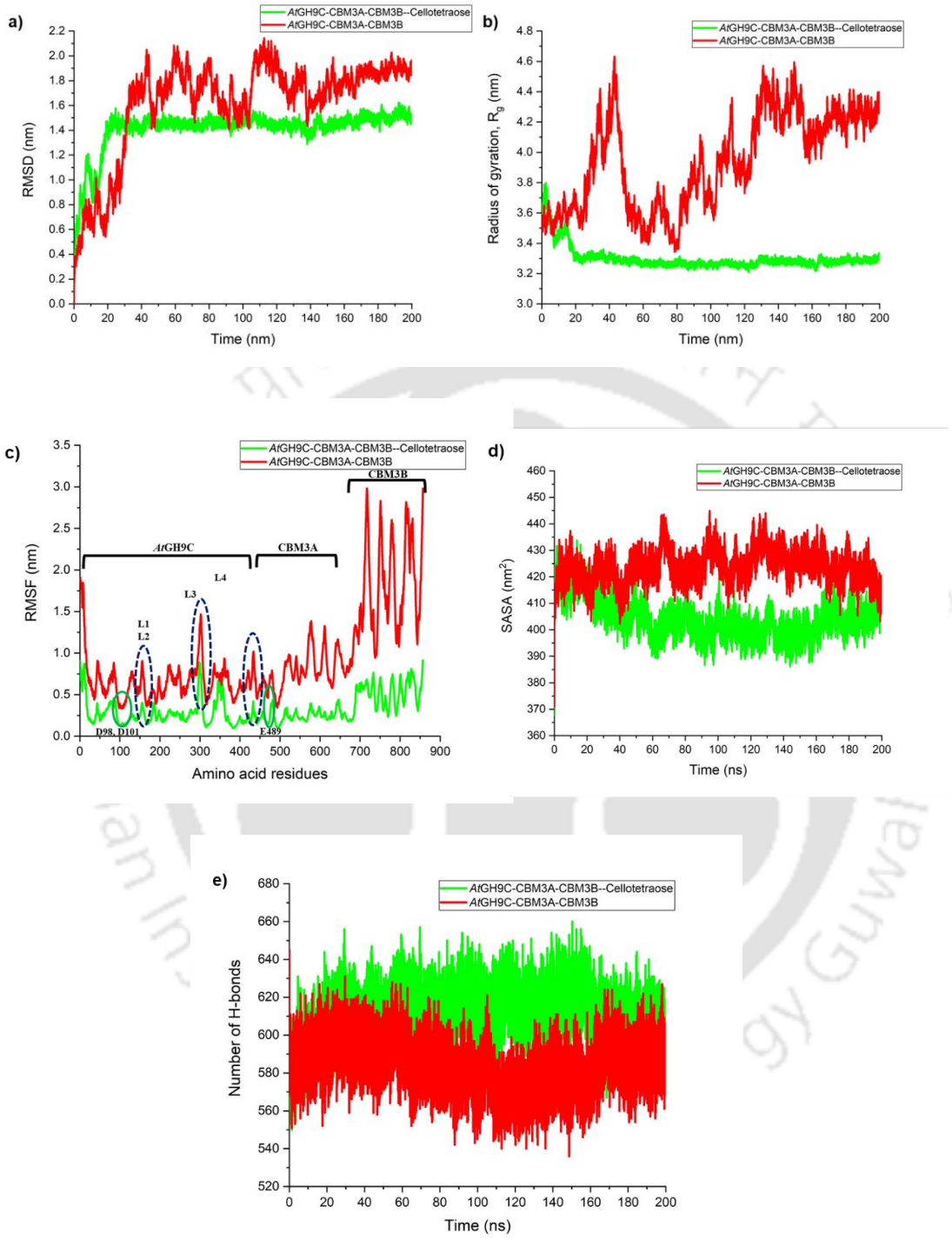


Fig. 4.10 Comparative analysis of Molecular dynamics simulation of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex with only Robetta modeled *At*GH9C-CBM3A-CBM3B structure for 200 ns displaying a) RMSD, b) Radius of gyration (R_g), c) RMSF, d) SASA and e) Intramolecular hydrogen bonds.

4.3.9 Binding free energy calculations of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex

The Gibbs free binding energy ($\Delta G_{binding}$) of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex was determined by using the MM-PBSA method using the g_mmbpsa tool. Among the three components of the binding energy (ΔE_{MM}, ΔG_{polar} and ΔG_{apolar}), negative binding energy is contributed by molecular mechanics energy (ΔE_{MM}) and apolar solvation energy (ΔG_{polar}) with values 199 ± 6.8 kJ/mol and -97 ± 1.2 kJ/mol respectively (**Table 4.7**).

Table 4.7 Binding free energy analysis of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex.

Complex	ΔE_{elec} (kJ/mol)	ΔE_{vwd} (kJ/mol)	ΔG_{polar} (kJ/mol)	ΔG_{apolar} (kJ/mol)	$\Delta G_{binding}$ (kJ/mol)
*At*GH9C-CBM3A-CBM3B-Cellotetraose	-132 ± 4.5	-67 ± 2.3	232 ± 5.6	-97 ± 1.2	-65 ± 1.8

ΔE_{MM} comprises electrostatic interactions (ΔE_{elec}) and van der Waals interactions (ΔE_{vdw}) computed to be -132 ± 4.5 kJ/mol and -67 ± 2.3 kJ/mol, respectively. These high negative values show that the electrostatic interactions and solvation effects between the receptor (*At*GH9C-CBM3A-CBM3B) and the ligand (Cellotetraose) are stable. However, a positive binding energy was observed with polar solvation energy (ΔG_{polar}) with value, 232 ± 5.6 kJ/mol indicating repulsive effect of receptor and ligand, probably for helping the exit of ligand from the active-site. The total negative Gibbs free binding energy, -65 ± 1.8 kJ/mol represents the stable formation of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex corroborating the docking results of **section 4.3.7**.

4.3.10 Solution structure of *At*GH9C-CBM3A-CBM3B by SAXS

The molecular arrangements and the structural characterization of *At*GH9C-CBM3A-CBM3B in solution form (50 mM sodium phosphate buffer, pH 7.5) were deduced by employing the SAXS technique at 2 mg/mL and 5 mg/mL concentrations. The SAXS data for both concentrations were processed for the scattering profile (**Fig. 4.11a**). The investigation of the processed data displaying the concentration-dependent scattering profile at both concentrations showed similar pattern and monodispersed nature in solution with no inter-particle interaction or aggregation. Considering scattering homogeneity, the SAXS data recorded at 5 mg/mL was considered for further downstream processing and analysis. An overview of the structural parameters of *At*GH9C-CBM3A-CBM3B at 5 mg/mL as determined by SAXS data processing and evaluation is provided in **Table 4.8.** The fit of scattering profile: ln(I) vs. q^2 at low q region depicted linearity (**Fig. 4.11b**), signifying Guinier scaling thus confirms the essentially monodisperse nature of *At*GH9C-CBM3A-CBM3B in solution at 5 mg/mL. The radius of gyration for the globular shape (R_g) and the rod shape (R_c) were estimated by using the Guinier plot approximation to be roughly 4.50 ± 0.33 nm and 1.05 ± 0.93 nm, respectively. The R_g of the globular shape roughly corroborated with the MD simulation result of only *At*GH9C-CBM3A-CBM3B, where the maximum R_g obtained was 4.6 nm. Employing, ($\sqrt{12}[(R_g)^2-(R_c)^2]$) [70], the persistence length of *At*GH9C-CBM3A-CBM3B was calculated to be 15.2 nm, pointing to the limited flexibility of the protein. The indirect Fourier transformation of the smooth regularized scattering profile of *At*GH9C-CBM3A-CBM3B at 5 mg/mL provides the electron pair distribution in form of P(R) plot (**Fig. 4.11c**) displaying an asymmetric profile asserting oligomeric nature of the protein. Moreover, a shoulder appeared in form of a plateau on the P(R)

plot implying the multi-modular nature of the protein with a split between modules. Based on the P(R) plot analysis the maximum inter-particle dimension (D_{max}) and radius of gyration (R_g) was calculated to be 16.2 and 5.10 nm respectively. Approximately, three-fold higher value of D_{max} over R_g revealed that *At*GH9C-CBM3A-CBM3B has an elongated structure as opposed to a globular structure.

Table 4.8 SAXS data collection and parameter analysis of *At*GH9C-CBM3A-CBM3B.

Data-collection parameters	***At*GH9C-CBM3A-CBM3B**
Instrument	SAXSpace Anton-Paar
Wavelength (Å)	1.54 Å
S range (nm^{-1})	0.12-7
Exposure time (min)	2 x 30
Temperature (°C)	10
Protein Concentration (mg/mL)	5
Structural parameters	
Q range (nm^{-1}) used for Rg analysis	0.04-12
I(0) au from Guinier	51024.5 ± 1623.5
R_g nm from Guinier	4.50 ± 0.33
Rc nm from Guinier	1.05 ± 0.93
R_g nm from P(r)	5.10
D_{max} (nm)	16.2
Porod volume estimate (nm^3)	186.01
Persistent length (nm)	15.2
Molecular mass determination	
Theoretical molecular mass (kDa)	97.2
Molecular mass from Qp (kDa)	94.3
Modeling parameters	
Normalized spatial discrepancy (NSD)	1.80 ± 0.14
Resolution	5.8 ± 0.3
χ^2	1.13
Normalized special discrepancy (NSD)	2.10 ± 0.76
Software employed	
Data processing	Primus
P(r) function calculation	GNOM
Ab initio modeling	DAMMIF
Validation and averaging	DAMAVER
3-D graphical representation	PyMOL

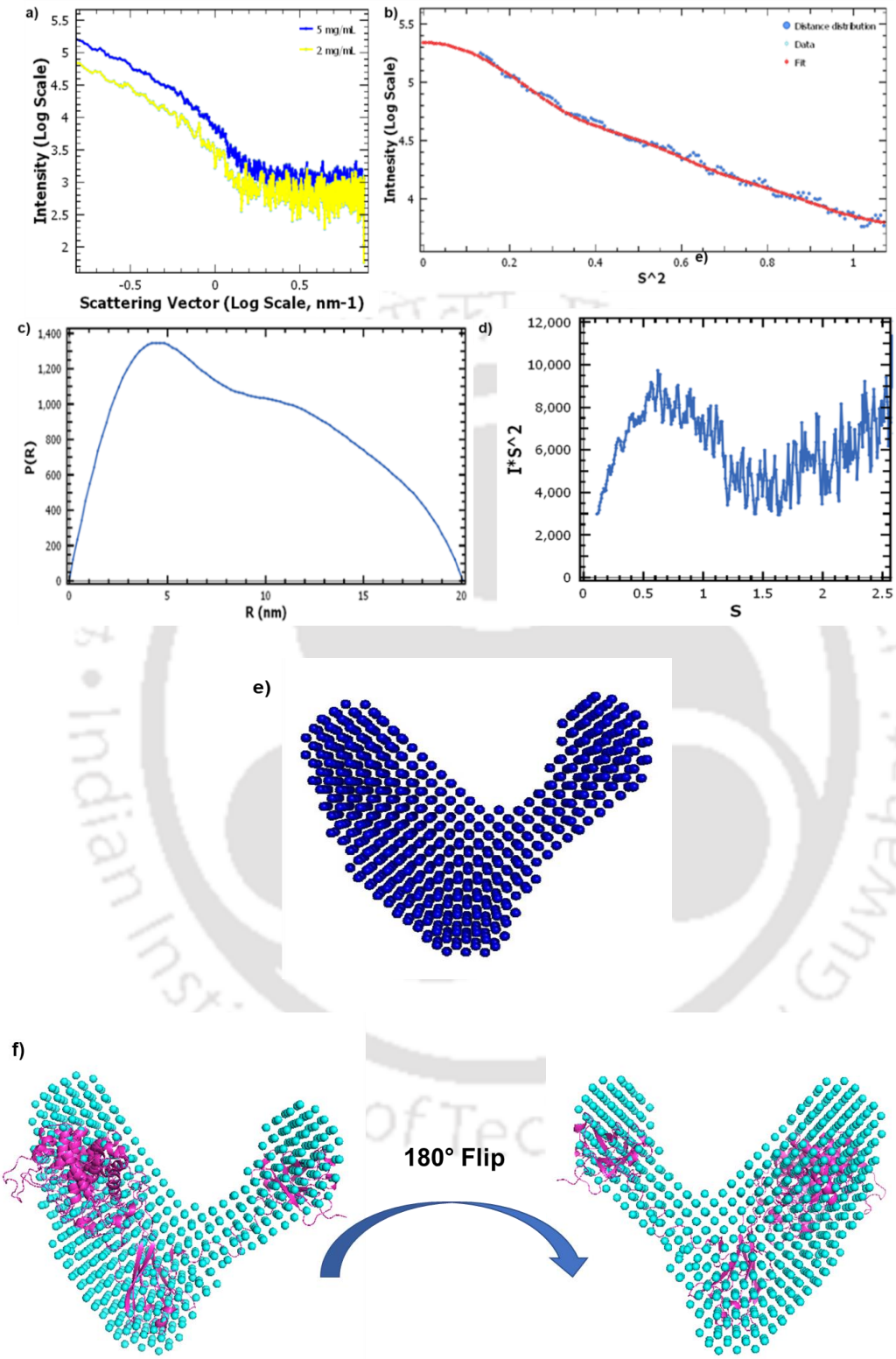


Fig. 4.11 **Solution-based structure analysis of *At*GH9C-CBM3A-CBM3B using SAXS. (a) Scattering profile of *At*GH9C-CBM3A-CBM3B at 2 mg/mL and 5 mg/mL, (b) Guinier plot, (c) P(R) curve as a function of R, (d) Kratky plot and (e) Solution based model of *At*GH9C-CBM3A-CBM3B, (f) Superposition of *Ab initio* model with homology modeled *At*GH9C-CBM3A-CBM3B.**

The qualitative assessment of the degree of flexibility and global rigidity of *At*GH9C-CBM3A-CBM3B by Kratky plot analysis displayed a bell-shaped curve at low Q region, with a rise at S of range 1.5-2.5 in x-axis, ascertaining its compact folding and degree of flexibility, respectively (**Fig. 4.11d**). A molecular mass of 94.3 kDa was recorded using the SAXS scattering profile in SAXSMoW server, which was comparable to the theoretical and experimental molecular mass of 97.2 kDa that was previously reported [25]. A similar molecular mass represents *At*GH9C-CBM3A-CBM3B to exist as a monomer as native form in the solution. Rapid *ab initio* modeling approach was used to construct the molecular conformation of the solution derived *At*GH9C-CBM3A-CBM3B using ATSAS DAMMIF software. A total of 20 independent models were averaged and refined by DAMAVER and DAMMIN with a normalized spatial discrepancy (NSD) value of $<1.80 \pm 0.14$ and resolution of 5.8 ± 0.3 nm was taken into consideration. The modeled structure resembles a fist-and-elbow-like shape represented as two dissimilar independent units (**Fig. 4.11e**). The superposition of the modeled *At*GH9C-CBM3A-CBM3B with *ab initio* derived model obtained from the SAXS data demonstrated excellent fit with a minor deviation in the C-terminal part. The fitting depicted two initial modules from the N-terminal sequence i.e. *At*GH9C and CBM3A existed together in bigger independent units and CBM3B remained in the smaller one (**Fig. 4.11f**). Perhaps it corroborated with the redundancy of CBM3B in assisting the catalysis of *At*GH9C by close association [25].

4.3.11 DLS and Zeta analysis of *At*GH9C-CBM3A-CBM3B by SAXS

DLS analysis of *At*GH9C-CBM3A-CBM3B for determination of particle diameter at 3 different concentrations *viz.* 1, 2 and 5 mg/mL showed a single symmetric peak at all concentrations with, a polydispersity index of 14.53% signifying the monodispersity of the protein (**Fig. 4.12a**). The apparent hydrodynamic diameter (D_h) of *At*GH9C-CBM3A-CBM3B obtained at 1, 2 and 5 mg/mL were 9.2, 10.5 and 11.6 nm respectively, suggesting their hydrodynamic radius, (R_h) to be 4.6 nm, 5.3 nm and 5.8 nm at their respective concentrations. The gradual increase in the radius with an increase in protein concentration could probably be defined by oligomerization at higher concentrations. The hydrodynamic radius, R_h (5.8 nm) found at 5 mg/mL form the DLS analysis was in agreement with the *R*g, obtained by the SAXS analysis (5.1 nm). The reflection of surface charge of *At*GH9C-CBM3A-CBM3B at its different concentration (1 mg/mL, 2 mg/mL and 5 mg/mL) was noted by the zeta potential (ξ) analysis. A remarkably higher negative zeta potential of *At*GH9C-CBM3A-CBM3B with values -18.6 ± 2.1 mV, - 21.6 ± 1.3 mV and – 24.2 ± 1.7 mV for 1 mg/mL, 2 mg/mL and 5 mg/mL in the 50 mM sodium phosphate buffer at pH 7.5 is possibly accountable for its better stability in aqueous form (**Fig. 4.12b**). Since zeta potential is the charge density on a protein surface, a charged protein indicates its solubility and, consequently, its resistance to aggregation [71], therefore no aggregation behavior was observed in the SAXS data of *At*GH9C-CBM3A-CBM3B at 5 mg/mL. Therefore, a high zeta potential of *At*GH9C-CBM3A-CBM3B can be utilized for its long-term storage, lyophilization and modulation of its activity with cation and anions at industrial scale.

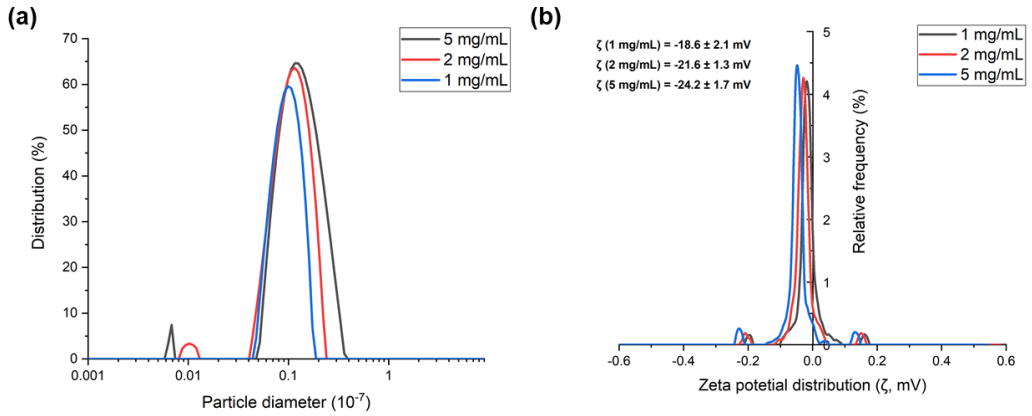


Fig. 4.12 DLS analysis of *At*GH9C-CBM3A-CBM3B at 5 mg/mL, 2 mg/mL and 1 mg/mL concentrations for determination of (a) Hydrodynamic diameter (D_h) and (b) Surface charge potential (Zeta potential, ζ).

4.4 Conclusion

Cellulases from the GH9 family vary in their enzymatic activity and exhibit endo, exo and processive endocellulase activities. Therefore, studying the structure and functional relationship of GH9 enzymes, particularly endoglucanases is important in the context of enzymatic hydrolysis in biofuel sector. The homology-modeled three-dimensional structure of a GH9 family endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus*, showed $(\alpha/\alpha)_6$-barrel structure for catalytic module (*At*GH9C) and its associated carbohydrate binding modules i.e. CBM3A and CBM3B resembled antiparallel β-sandwich fold. The MSA analysis and structure superposition with its homologs suggested a catalytic triad comprising Asp98, Asp101 and Glu489 for substrate catalysis. The carboxyl group of catalytic bases (Asp98 and Asp101) and catalytic acid (Glu489) were separated by 8.5 Å, suggesting inverting-type mechanism of hydrolysis. The analysis of secondary structure composition of *At*GH9C-CBM3A-CBM3B by CD revealed 25.2% α-helix, 18.4% β-sheet and 56.4% random coils which also corroborated with the predicted data obtained from various servers. Molecular dynamic simulation of modeled *At*GH9C-CBM3A-CBM3B up to 200 ns at 300 K showed an average RMSD value of 1.8 nm after 140 ns ensuring the structure stability compactness of the structure. The stable R_g value of 4.2 nm from 140 ns to 200 ns indicated its global compactness. A high fluctuation in RMSF was observed in the area corresponding to CBM3B indicating its possible redundancy in substrate binding. The molecular docking of *At*GH9C-CBM3A-CBM3B with different cello-oligosaccharides showed strong binding affinity with cellotetraose suggesting that its open groove of active-site can accommodate ligands up to DP4. The MD simulation analysis of *At*GH9C-CBM3A-CBM3B-Cellotetraose complex recorded a lowered RMSD value

(1.5 nm) as compared with the only protein, *At*GH9C-CBM3A-CBM3B (1.8 nm) displaying its compactness in bound form. The processivity analysis on PASC classified *At*GH9C-CBM3A-CBM3B as pEG4 type of processive endoglucanase with loop 3 (342-379 aa) possibly blocking the non-reducing cellulosic end resulting in cellotetraose release, as determined by molecular docking analysis. The SAXS analysis of *At*GH9C-CBM3A-CBM3B at 5 mg/mL showed its fully folded, monodispersed and elongated form. The *ab initio* derived DAM model of the enzyme *At*GH9C-CBM3A-CBM3B displaying two dissimilar independent units having a fist-and-elbow-like shape in solution form superposed well with its homology modeled 3-dimensional structure. The surface charge of -24 mV on *At*GH9C-CBM3A-CBM3B at 5 mg/mL indicated its higher solubility and monodispersity. The overall results on structure determination of *At*GH9C-CBM3A-CBM3B showed the involvement of loop region in determining the cleavage pattern of cellotetraose-producing endoglucanase displaying the possible redundancy of the appended CBM3B module in substrate binding. This *in silico* study cannot elucidate the catalytic nature of *At*GH9C-CBM3A-CBM3B against different substrates. However, a thorough understanding of *At*GH9C-CBM3A-CBM3B structure will pave the way for rational protein engineering by mutating the residues in the loop region or the catalytic residues possibly improving its performance. Furthermore, modifying the aromatic residues of CBM3B or fusing catalytic module, *At*GH9C with suitable CBMs may increase the cellulolytic activity of endoglucanase, *At*GH9C-CBM3A-CBM3B. Thus, the improved endoglucanases will be beneficial in advancing biofuel production and also help in reduction of carbon footprint.

4.5 References

[1] M. V. Rodionova, A.M. Bozieva, S.K. Zharmukhamedov, Y.K. Leong, J. Chi-Wei Lan, A. Veziroglu, T.N. Veziroglu, T. Tomo, J.S. Chang, S.I. Allakhverdiev, A comprehensive review on lignocellulosic biomass biorefinery for sustainable biofuel production, Int J Hydrogen Energy 47 (2022) 1481–1498. https://doi.org/10.1016/J.IJHYDENE.2021.10.122.

[2] B.A. Simmons, D. Loque, H.W. Blanch, Next-generation biomass feedstocks for biofuel production., Genome Biol 9 (2008) 242. https://doi.org/10.1186/GB-2008-9-12-242/FIGURES/3.

[3] D. Klemm, B. Heublein, H.P. Fink, A. Bohn, Cellulose: Fascinating Biopolymer and Sustainable Raw Material, Angewandte Chemie International Edition 44 (2005) 3358–3393. https://doi.org/10.1002/ANIE.200460587.

[4] E.A. Bayer, L.J.W. Shimon, Y. Shoham, R. Lamed, Cellulosomes—Structure and Ultrastructure, J Struct Biol 124 (1998) 221–234. https://doi.org/10.1006/JSBI.1998.4065.

[5] S.P.S. Chundawat, G.T. Beckham, M.E. Himmel, B.E. Dale, Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals, Annu Rev Chem Biomol Eng 2 (2011) 121–145. https://doi.org/10.1146/annurev-chembioeng-061010-114205.

[6] L.R. Lynd, P.J. Weimer, W.H. van Zyl, I.S. Pretorius, Microbial Cellulose Utilization: Fundamentals and Biotechnology, Microbiology and Molecular Biology Reviews 66 (2002) 506–577. https://doi.org/10.1128/MMBR.66.3.506-577.2002/ASSET/6F6A28CF-2C07-40BC-B7A7-BAF0F705A33E/ASSETS/GRAPHIC/MR032001411T.JPEG.

[7] G. Guerriero, K. Sergeant, S. Legay, J.F. Hausman, H.M. Cauchie, I. Ahmad, K.S. Siddiqui, Novel Insights from Comparative In Silico Analysis of Green Microalgal Cellulases, International Journal of Molecular Sciences 2018, Vol. 19, Page 1782 19 (2018) 1782. https://doi.org/10.3390/IJMS19061782.

[8] D.B. Wilson, D.C. Irwin, Genetics and Properties of Cellulases, (1999) 1–21. https://doi.org/10.1007/3-540-49194-5_1.

[9] Y. Shoham, R. Lamed, E.A. Bayer, The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides, Trends Microbiol 7 (1999) 275–281. https://doi.org/10.1016/S0966-842X(99)01533-4.

[10] A.B. Boraston, D.N. Bolam, H.J. Gilbert, G.J. Davies, Carbohydrate-binding modules: fine-tuning polysaccharide recognition, Biochem J 382 (2004) 769–781. https://doi.org/10.1042/BJ20040892.

[11] E. Libertini, Y. Li, S.J. McQueen-Mason, Phylogenetic analysis of the plant endo-beta-1,4-glucanase gene family., J Mol Evol 58 (2004) 506–15. https://doi.org/10.1007/s00239-003-2571-x.

[12] S. Kundu, Insights into the mechanism(s) of digestion of crystalline cellulose by plant class C GH9 endoglucanases, J Mol Model 25 (2019). https://doi.org/10.1007/s00894-019-4133-1.

[13] D. Mandelman, A. Belaich, J.P. Belaich, N. Aghajari, H. Driguez, R. Haser, X-Ray Crystal Structure of the Multidomain Endoglucanase Cel9G from *Clostridium cellulolyticum* Complexed with Natural and Synthetic Cello-Oligosaccharides, J Bacteriol 185 (2003) 4127. https://doi.org/10.1128/JB.185.14.4127-4135.2003.

[14] O. Yaniv, G. Fichman, I. Borovok, Y. Shoham, E.A. Bayer, R. Lamed, L.J.W. Shimon, F. Frolow, Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti-σI factors, Volume 70, Issue 2, Pages 522 - 534 70 (2014) 522–534. https://doi.org/10.1107/S139900471302926X.

[15] S. Petkun, I.R. Grinberg, R. Lamed, S. Jindou, T. Burstein, O. Yaniv, Y. Shoham, L.J.W. Shimon, E.A. Bayer, F. Frolow, Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: Structural and functional significance of the intermodular linker, PeerJ 2015 (2015) e1126. https://doi.org/10.7717/peerj.1126.

[16] S. Cai, X. Zheng, X. Dong, CBM3d, a novel subfamily of family 3 carbohydrate-binding modules identified in Cel48A exoglucanase of *Cellulosilyticum ruminicola*, Volume 193, Issue 19, Pages 5199 - 5206 193 (2011) 5199–5206. https://doi.org/10.1128/JB.05227-11.

[17] N. Pasari, N. Adlakha, M. Gupta, Z. Bashir, G.H. Rajacharya, G. Verma, M. Munde, R. Bhatnagar, S.S. Yazdani, Impact of Module-X2 and Carbohydrate Binding Module-3 on the catalytic activity of associated glycoside hydrolases towards plant biomass, Sci Rep 7 (2017). https://doi.org/10.1038/S41598-017-03927-Y.

[18] M. Sajjad, M.I.M. Khan, R. Zafar, S. Ahmad, U.H.K. Niazi, M.W. Akhtar, Influence of positioning of carbohydrate binding module on the activity of endoglucanase CelA of *Clostridium thermocellum*, J Biotechnol 161 (2012) 206–212. https://doi.org/10.1016/J.JBIOTEC.2012.05.023.

[19] T. Burstein, M. Shulman, S. Jindou, S. Petkun, F. Frolow, Y. Shoham, E.A. Bayer, R. Lamed, Physical association of the catalytic and helper modules of a family-9 glycoside hydrolase is essential for activity, FEBS Lett 583 (2009) 879–884. https://doi.org/10.1016/J.FEBSLET.2009.02.013.

[20] K. Kumar, S. Singal, A. Goyal, Role of carbohydrate binding module (CBM3c) of GH9 β-1,4 endoglucanase (Cel9W) from *Hungateiclostridium thermocellum* ATCC 27405 in catalysis, Carbohydr Res 484 (2019). https://doi.org/10.1016/J.CARRES.2019.107782.

[21] S. Jindou, S. Petkun, L. Shimon, E.A. Bayer, R. Lamed, F. Frolow, Crystallization and preliminary diffraction studies of CBM3b of cellobiohydrolase 9A from *Clostridium thermocellum*, Acta Crystallogr Sect F Struct Biol Cryst Commun 63 (2007) 1044. https://doi.org/10.1107/S1744309107054644.

[22] S. Kundu, R. Sharma, Origin, evolution, and divergence of plant class C GH9 endoglucanases, BMC Evolutionary Biology 2018 18:1 18 (2018) 1–19. https://doi.org/10.1186/S12862-018-1185-2.

[23] W. Zhou, D.C. Irwin, J. Escovar-Kousen, D.B. Wilson, Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes, Volume 43, Issue 30, Pages 9655 - 9663 43 (2004) 9655–9663. https://doi.org/10.1021/bi049394n.

[24] B. Leis, C. Held, F. Bergkemper, K. Dennemarck, R. Steinbauer, A. Reiter, M. Mechelke, M. Moerch, S. Graubner, W. Liebl, W.H. Schwarz, V. V. Zverlov, Comparative characterization of all cellulosomal cellulases from *Clostridium thermocellum* reveals high diversity in endoglucanase product formation essential for complex activity, Biotechnol Biofuels 10 (2017) 1–16. https://doi.org/10.1186/S13068-017-0928-4/FIGURES/5.

[25] A. Mandal, A. Thakur, A. Goyal, Role of carbohydrate binding modules, CBM3A and CBM3B in stability and catalysis by a β-1,4 endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus* ATCC 27405, Int J Biol Macromol 242 (2023). https://doi.org/10.1016/j.ijbiomac.2023.125164.

[26] G.M. Boratyn, C. Camacho, P.S. Cooper, G. Coulouris, A. Fong, N. Ma, T.L. Madden, W.T. Matten, S.D. McGinnis, Y. Merezhuk, Y. Raytselis, E.W. Sayers, T. Tao, J. Ye, I. Zaretskaya, BLAST: a more efficient report with usability improvements, Nucleic Acids Res 41 (2013) W29–W33. https://doi.org/10.1093/NAR/GKT282.

[27] F. Sievers, D.G. Higgins, Clustal Omega for making accurate alignments of many protein sequences, Protein Science 27 (2018) 135–145. https://doi.org/10.1002/PRO.3290.

[28] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res 42 (2014) W320–W324. https://doi.org/10.1093/NAR/GKU316.

[29] K. Tamura, G. Stecher, S. Kumar, MEGA11: Molecular Evolutionary Genetics Analysis Version 11, Mol Biol Evol 38 (2021) 3022–3027. https://doi.org/10.1093/MOLBEV/MSAB120.

[30] S.M. Kelly, T.J. Jess, N.C. Price, How to study proteins by circular dichroism, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1751 (2005) 119–139. https://doi.org/10.1016/J.BBAPAP.2005.06.005.

[31] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 2021 596:7873 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[32] J. Yang, Y. Zhang, Protein Structure and Function Prediction Using I-TASSER, Curr Protoc Bioinformatics 52 (2015) 5.8.1-5.8.15. https://doi.org/10.1002/0471250953.BI0508S52.

[33] M. Källberg, G. Margaryan, S. Wang, J. Ma, J. Xu, Raptorx server: A resource for template-based protein structure modeling, Methods in Molecular Biology 1137 (2014) 17–27. https://doi.org/10.1007/978-1-4939-0366-5_2/FIGURES/4.

[34] D.E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server, Nucleic Acids Res 32 (2004) W526. https://doi.org/10.1093/NAR/GKH468.

[35] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nature Protocols 2015 10:6 10 (2015) 845–858. https://doi.org/10.1038/nprot.2015.053.

[36] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res 46 (2018) W296–W303. https://doi.org/10.1093/NAR/GKY427.

[37] H. Land, M.S. Humble, YASARA: A tool to obtain structural guidance in biocatalytic investigations, Methods in Molecular Biology 1685 (2018) 43–67. https://doi.org/10.1007/978-1-4939-7366-8_4/TABLES/5.

[38] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res 35 (2007) W407–W410. https://doi.org/10.1093/NAR/GKM290.

[39] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, J Chem Theory Comput 4 (2008) 435–447. https://doi.org/10.1021/CT700301Q/ASSET/IMAGES/LARGE/CT700301QF00006.JPEG.

[40] J. Lemkul, From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0], Living J Comput Mol Sci 1 (2019). https://doi.org/10.33011/livecoms.1.1.5068.

[41] G.M. Morris, H. Ruth, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J Comput Chem 30 (2009) 2785–2791. https://doi.org/10.1002/JCC.21256.

[42] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis, J Comput Chem 25 (2004) 1605–1612. https://doi.org/10.1002/JCC.20084.

[43] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: An Open chemical toolbox, J Cheminform 3 (2011) 1–14. https://doi.org/10.1186/1758-2946-3-33/TABLES/2.

[44] W. Wang, T. Archbold, J.S. Lam, M.S. Kimber, M.Z. Fan, A processive endoglucanase with multi-substrate specificity is characterized from porcine gut microbiota, Sci Rep 9 (2019). https://doi.org/10.1038/S41598-019-50050-1.

[45] X. Zhang, T. Qu, N.S. Mosier, L. Han, W. Xiao, Cellulose modification by recyclable swelling solvents, Biotechnol Biofuels 11 (2018). https://doi.org/10.1186/s13068-018-1191-z.

[46] N. Nelson, A PHOTOMETRIC ADAPTATION OF THE SOMOGYI METHOD FOR THE DETERMINATION OF GLUCOSE, Journal of Biological Chemistry 153 (1944) 375–380. https://doi.org/10.1016/S0021-9258(18)71980-7.

[47] M. Somogyi, DETERMINATION OF BLOOD SUGAR, Journal of Biological Chemistry 160 (1945) 69–73. https://doi.org/10.1016/S0021-9258(18)43098-0.

[48] R. Kumari, R. Kumar, A. Lynn, G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations, J Chem Inf Model 54 (2014) 1951–1962. https://doi.org/10.1021/CI500020M/SUPPL_FILE/CI500020M_SI_001.PDF.

[49] A.G. Kikhney, D.I. Svergun, A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins, FEBS Lett 589 (2015) 2570–2577. https://doi.org/10.1016/J.FEBSLET.2015.08.027.

[50] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H.D.T. Mertens, A.G. Kikhney, N.R. Hajizadeh, J.M. Franklin, C.M. Jeffries, D.I. Svergun, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, Urn:Issn:1600-5767 50 (2017) 1212–1225. https://doi.org/10.1107/S1600576717007786.

[51] A. Guinier, G. Fournet, C.B. Walker, G.H. Vineyard, Small-Angle Scattering of X-Rays, Phys Today 9 (1956) 38–39. https://doi.org/10.1063/1.3060069.

[52] D.I. Svergun, IUCr, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, Urn:Issn:0021-8898 25 (1992) 495–503. https://doi.org/10.1107/S0021889892001663.

[53] H. Fischer, M. De Oliveira Neto, H.B. Napolitano, I. Polikarpov, A.F. Craievich, Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale, J Appl Crystallogr 43 (2010) 101–109. https://doi.org/10.1107/S0021889809043076/CE5058SUP1.PDF.

[54] V. V. Volkov, D.I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering, J Appl Crystallogr 36 (2003) 860–864. https://doi.org/10.1107/S0021889803000268.

[55] D. Svergun, C. Barberato, M.H. Koch, CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, Urn:Issn:0021-8898 28 (1995) 768–773. https://doi.org/10.1107/S0021889895007047.

[56] S. Singh, J. Ahmed, P.V. Gavande, C.M.G.A. Fontes, A. Goyal, Structural and functional insights into the glycoside hydrolase family 30 xylanase of the rumen *bacterium Ruminococcus flavefaciens*, J Mol Struct 1272 (2023) 134155. https://doi.org/10.1016/J.MOLSTRUC.2022.134155.

[57] N.D. Gold, V.J.J. Martin, Global View of the *Clostridium thermocellum* Cellulosome Revealed by Quantitative Proteomic Analysis, J Bacteriol 189 (2007) 6787. https://doi.org/10.1128/JB.00882-07.

[58] F. Azzaz, N. Yahi, H. Chahinian, J. Fantini, The Epigenetic Dimension of Protein Structure Is an Intrinsic Weakness of the AlphaFold Program, Biomolecules 12 (2022). https://doi.org/10.3390/BIOM12101527.

[59] G. Parsiegla, A. Belaïch, J.P. Belaïch, R. Haser, Crystal structure of the cellulase Ce19M enlightens structure/function relationships of the variable catalytic modules in glycoside hydrolases, Biochemistry 41 (2002) 11134–11142. https://doi.org/10.1021/BI025816M/ASSET/IMAGES/LARGE/BI025816MF00005.JPEG.

[60] O. Yaniv, G. Fichman, I. Borovok, Y. Shoham, E.A. Bayer, R. Lamed, L.J.W. Shimon, F. Frolow, Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti-σI factors, Volume 70, Issue 2, Pages 522 - 534 70 (2014) 522–534. https://doi.org/10.1107/S139900471302926X.

[61] R. Brunecky, M. Alahuhta, Q. Xu, B.S. Donohoe, M.F. Crowley, I.A. Kataeva, S.J. Yang, M.G. Resch, M.W.W. Adams, V. V. Lunin, M.E. Himmel, Y.J. Bomble, Revealing nature's cellulase diversity: The digestion mechanism of *Caldicellulosiruptor bescii* CelA, Science (1979) 342 (2013) 1513–1516. https://doi.org/10.1126/SCIENCE.1244273/SUPPL_FILE/BRUNECKY.SM.V2.PDF.

[62] I. Kufareva, R. Abagyan, Methods of protein structure comparison, Methods Mol Biol 857 (2012) 231. https://doi.org/10.1007/978-1-61779-588-6_10.

[63] J. Sakon, D. Irwin, D.B. Wilson, P. Andrew Karplus, Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*, Volume 4, Issue 10, Pages 810 - 818 4 (1997) 810–818. https://doi.org/10.1038/nsb1097-810.

[64] C.S. Rye, S.G. Withers, Glycosidase mechanisms, Curr Opin Chem Biol 4 (2000) 573–580. https://doi.org/10.1016/S1367-5931(00)00135-6.

[65] Q. Wang, S.G. Withers, R.W. Graham, D. Trimbur, R.A.J. Warren, Changing Enzymic Reaction Mechanisms by Mutagenesis: Conversion of a Retaining Glucosidase to an

Inverting Enzyme, J Am Chem Soc 116 (1994) 11594–11595. https://doi.org/10.1021/JA00104A060/ASSET/JA00104A060.FP.PNG_V03.

[66] D. V. Buonomano, The biology of time across different scales, Nature Chemical Biology 2007 3:10 3 (2007) 594–597. https://doi.org/10.1038/nchembio1007-594.

[67] A. Belaich, G. Parsiegla, L. Gal, C. Villard, R. Haser, J.P. Belaich, Cel9M, a New Family 9 Cellulase of the *Clostridium cellulolyticum* Cellulosome, J Bacteriol 184 (2002) 1378. https://doi.org/10.1128/JB.184.5.1378-1384.2002.

[68] G. Guerriero, K. Sergeant, S. Legay, J.F. Hausman, H.M. Cauchie, I. Ahmad, K.S. Siddiqui, Novel Insights from Comparative In Silico Analysis of Green Microalgal Cellulases, International Journal of Molecular Sciences 2018, Vol. 19, Page 1782 19 (2018) 1782. https://doi.org/10.3390/IJMS19061782.

[69] G. Parsiegla, A. Belaïch, J.P. Belaïch, R. Haser, Crystal structure of the cellulase Ce19M enlightens structure/function relationships of the variable catalytic modules in glycoside hydrolases, Biochemistry 41 (2002) 11134–11142. https://doi.org/10.1021/BI025816M/ASSET/IMAGES/LARGE/BI025816MF00005.JPEG.

[70] B.A. Khorramian, S.S. Stivala, Small-angle x-ray scattering of high- and low-affinity heparin, Arch Biochem Biophys 247 (1986) 384–392. https://doi.org/10.1016/0003-9861(86)90597-7.

[71] S. Salgin, U. Salgin, S. Bahadir, Zeta Potentials and Isoelectric Points of Biomolecules: The Effects of Ion Types and Ionic Strengths, Int J Electrochem Sci 7 (2012) 12404–12414. https://doi.org/10.1016/S1452-3981(23)16554-0.

Chapter 5

Enzymatic synthesis of nanocellulose from biomass using recombinant cellulases: Application in dye adsorption via nanocellulose-based hydrogel

5.1 Introduction

The increasing environmental burden of industrial dye pollution necessitates the development of sustainable and efficient adsorbent materials. The dye industry generates approximately 8×10^5 tons of synthetic dyes annually, most of which are organic compounds [1]. These dyes pose significant ecological and health risks upon environmental release due to their toxic and carcinogenic nature [2]. The textile and printing industries extensively utilize synthetic azo dye Congo red (CR) and thiazine Methylene blue (MB), which exhibit high environmental pollution owing to their complex aromatic structures and inherent toxicity. These xenobiotic compounds resist microbial degradation, enabling environmental accumulation. Their ecological impacts include aquatic phototoxicity, genotoxic effects and human health risks through dermal irritation and sensitization[3,4]. Numerous natural and synthetic adsorbents have been developed for the removal of aqueous dyes in recent decade. Common remediation

approaches include adsorption, membrane filtration, coagulation-flocculation, ozonation and biological treatment [5]. However, large-scale implementation of these adsorbents remains constrained by economic and technological limitations. These challenges include high costs, limited reusability and unresolved disposal issues, which continue to hinder industrial adoption. Thus, natural material-based adsorbents with optimized surface chemistry, selective binding sites and regeneration capacity, represent a promising example for advanced wastewater treatment. Nanocellulose, derived from renewable biomass, has emerged as a promising candidate for toxic dyes removal due to its high surface area, biodegradability and tuneable surface chemistry [6]. Nanocelluloses hold exceptional properties including high mechanical strength, extensive surface area, and abundant surface hydroxyl groups. These characteristics enable versatile surface modifications while maintaining inherent hydrophilicity, making them ideal for advanced applications [6]. Additionally, nanocellulose-based hydrogels offer an enhanced remediation strategy for eliminating toxic synthetic dyes from environmental systems [7,8].

Nanocellulose is primarily classified into three structural variants: cellulose nanofibers (CNF), cellulose nanocrystals (CNC) and bacterial cellulose (BC), each exhibiting distinct morphological characteristics, dimensional parameters and synthesis protocols [9]. The morphological and dimensional characteristics of CNC and CNF are governed by both the lignocellulosic source material, (including wood, bast fibers, seed fibers and graminaceous plants such as bamboo and bagasse) and the extraction methodology employed [6,10]. Cellulose nanofibers (CNF) or nanofibrillated cellulose, typically exhibit diameters of 5-60 nm and micrometer-scale lengths, featuring flexible fibrillar structures with alternating crystalline-amorphous domains. These

nanomaterials, produced through various methodologies, display dimension-dependent properties [6,8]. For CNF production, cellulose is first extracted from natural biomass *via* pre-treatment to remove lignin and hemicellulose, followed by bleaching [11]. sugarcane trash (SCT) represents a cost-effective and industrially viable renewable feedstock for the synthesis of high-value materials, including nanocellulose [12]. Sugarcane fields yield 7–12 tons of SCT per hectare annually in India [13]. Due to the substantial production volume and prohibitive collection/transportation costs, most trash is burnt on field, resulting in significant nutrient and organic matter losses that contribute to the environmental pollution [14]. One tonne of sugarcane, generates approximately, 140 kg of bagasse and 250 kg of sugarcane trash (SCT) with up to 50% of SCT being sustainably extractable for high-value material synthesis without affecting soil fertility [15]. Composition analysis of sugarcane trash (SCT) revealed 40% cellulose, 33% hemicellulose and 17% lignin content, demonstrating its suitability as a lignocellulosic substrate [16]. Therefore, utilizing sugarcane trash (SCT) for nano-polymer production represents a carbon-neutral approach to valorising residual lignocellulosic waste. Following cellulose isolation *via* biomass pre-treatment, mechanical defibrillation is typically achieved using homogenizers, microfluidizers or grinders. Although these conventional methods are well-established, they present significant limitations, particularly their excessive energy requirements [17]. Moreover, mechanical processing typically yields CNF aqueous dispersions with low solid content (<5%), leading to elevated transportation expenses and storage difficulties [18]. The isolated cellulose can undergo various chemical treatments, including cationization, carboxymethylation and TEMPO-mediated oxidation to diminish energy requirements and enhance mechanical processing efficiency [19,20]. The use of these chemicals in

cellulose nanofiber (CNF) production, despite their higher efficiency, pose significant environmental concerns. The combination of enzymatic and mechanical treatments for cellulose fibrillation adheres to principles of green chemistry and circular economy, offers high substrate specificity, thus accelerating its widespread implementation in response to increasing sustainability requirements [21,22].

Numerous studies have explored the application of enzyme endoglucanases, either independently or in synergistic combination with cellobiohydrolase, to optimize cellulose fibrillation for enhanced CNF production [23–25]. Recent research indicates that auxiliary enzymes, particularly xylanases and lytic polysaccharide monooxygenases (LPMOs), improve cellulose fibrillation *via* complementary mechanisms and their synergistic interaction demonstrates enhanced catalytic efficiency [26]. While xylanases improve cellulose accessibility by cleaving surface xylan, LPMOs oxidatively modify the polysaccharide matrix. Multiple investigations have documented the application of these enzymes either individually or in combinations for the production of cellulose nanofibrils (CNF) [23]. Limited research has specifically addressed the utilization of sugarcane bagasse as a lignocellulosic substrate [27–29]. Their findings demonstrated that sugarcane bagasse (SCB)-derived CNF required reduced enzyme dosages. Moreover, integrating enzymatic pre-treatment with mechanical fibrillation enabled the production of tailored CNFs and offered an eco-friendly approach to lower energy consumption in downstream mechanical processing. However, to date no studies have investigated the feasibility of using sugarcane trash (SCT) for CNF synthesis *via* enzymatic hydrolysis.

In present study, the synergistic potential of two recombinant enzymes, full-length endoglucanase, *At*GH9C-CBM3A-CBM3B and cellobiohydrolase, *At*CBH5A

from *Acetivibrio thermocellus,* through restricted hydrolysis-mediated enhancement of cellulose fibrillation in sugarcane trash (SCT) for nanocellulose (NC) production was evaluated. The cellulose-hydrolysing capability of *At*GH9C-CBM3A-CBM3B has been esatlblished earlier as described in **chapter 3, section 3.3.3** [30]. Concurrently, cellobiohydrolase, *At*CBH5A has found applications in saccharification studies and enzyme research [31,32]. It was hypothesized that the controlled enzymatic hydrolysis of sugarcane trash (SCT) cellulose by using full-length endoglucanase, *At*GH9C-CBM3A-CBM3B and cellobiohydrolase, *At*CBH5A would enhance cellulose fibrillation through two key mechanisms: (1) selective cleavage of amorphous cellulose domains and (2) partial disentanglement of cellulose microfibril bundles. The fibre morphology, crystallinity, functional group content of enzymatically-produced nanocellulose (NCs) was characterized and compared with TEMPO-oxidized NCs. Further, the nanocellulose–carboxymethyl cellulose (NC–CMC) hydrogel was developed as an efficient adsorbent for the removal of anionic Congo red and cationic Methylene blue dyes from textile effluent. The adsorption of CV and MB dyes onto NCC-CH was optimized through response surface methodology (RSM), examining critical parameters including initial dye concentration, contact time, pH and adsorbent dosage. RSM was employed to overcome limitations of conventional optimization methods, which require extensive experimental runs and fail to elucidate parameter interactions. Adsorption mechanisms were further validated through comprehensive kinetic and isotherm modeling.

5.2 Materials and methods

5.2.1 Reagents and lab chemicals

All reagents, chemicals and materials employed in this study were of analytical grade, procured from commercial suppliers. The following analytical-grade chemicals and reagents were procured from HiMedia Pvt. Ltd., India: isopropyl-β-D-thiogalactopyranoside (IPTG), Trizma base, NaCl, NaOH, KOH, $CaCl_2$, HCl, HNO_3, H_3BO_3, Bradford reagent, isopropanol, Sodium hydroxide pellets, sodium azide,and protein ladder. Substrate sodium salt of carboxymethyl cellulose (CMC-Na) was procured from Sigma Aldrich Co. LLC, USA. Ethanol, acetic acid, and Thin Layer Chromatography (TLC) plates were obtained from Merck Chemical Co. (Darmstadt, Germany). Whatman No. 1 and TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl) were purchased from GE healthcare, USA and SRL Pvt. Ltd., India, respectively. SCT samples were collected from an agricultural field near Guwahati, Assam, India.

5.2.2 The enzymes and their activity

The full-length endoglucanase (*At*GH9C-CBM3A-CBM3B) (**Chapter 2, section 2.3.4.2**) and cellobiohydrolase (*At*CBH5A), a kind gift from Professor Carlos M.G.A. Fontes, NZYTech Ltd., Lisbon, Portugal, belong to *Acetivibrio thermocellus*. These enzymes were heterologously expressed in Escherichia coli BL21(DE3) cells using the pET28a(+) expression vector. *At*GH9C-CBM3A-CBM3B was purified using the procedure mentioned earlier [30]. The enzyme-expressing *E. coli* BL21(DE3) strains for *At*GH9C-CBM3A-CBM3B and *At*CBH5A, were cultivated in 500 mL Luria-Bertani (LB) medium supplemented with 50 µg/mL kanamycin at 37°C with orbital shaking of 180 rpm. Enzyme production for *At*GH9C-CBM3A-CBM3B and *At*CBH5A was induced with 1.0 mM IPTG upon reaching mid-exponential phase growth (OD_{600}

= 0.6). Post-induction, recombinant cells were cultured at 24°C for 16 h (180 rpm), harvested by centrifugation (6,500g, 4°C, 10 min), and lysed via sonication (33% amplitude, 10 s on/off cycles, 15 min). Following sonication, cellular debris was pelleted by centrifugation (12,000g, 4°C, 1h). The supernatant was filtered (0.45 µm membrane) using a syringe filter and subsequently purified using a Ni^{2+}-chelating column (HiTrap, GE Healthcare, USA). The resuspension buffer, equilibration buffer and wash buffer used for the purification of *At*GH9C-CBM3A-CBM3B and *At*CBH5A was 50 mM sodium-phosphate buffer, pH 7.5, 500 mM NaCl, 60 mM imidazole and 50 mM sodium-phosphate buffer, pH 7.0, 500 mM NaCl, 40 mM imidazole, respectively. Protein elution was performed using 50 mM sodium phosphate buffer (pH 8.5) containing 300 mM NaCl and 300 mM imidazole. The eluted proteins were dialyzed (cut-off size 12-14 kDa) against 50 mM Tris-HCl (pH 8.5) and analyzed for homogeneity by SDS-PAGE using 12% (w/v) gel as described in **Chapter 2, section 2.2.12**.

The hydrolytic activity of *At*GH9C-CBM3A-CBM3B and *At*CBH5A was assayed using carboxymethylcellulose (CMC) as substrate. Reaction mixtures (100 µL total volume) contained 90 µL of 1.1% (w/v) CMC and 10 µL of purified enzyme, *At*GH9C-CBM3A-CBM3B at 10 µg/mL in 50 mM sodium phosphate, pH 7.5 or *At*CBH5A at 100 µg/mL in 50 mM sodium phosphate, pH 7.0. The reactions mixture for *At*GH9C-CBM3A-CBM3B was inclubated at 55°C for 1 min and for *At*CBH5A at 65°C for 5 min and the reducing sugar released was quantified as previously described [33,34] in **Chapter 3, section 3.2.5**.

5.2.3 Enzyme-mediated nanofibrillation

The initial stage involved cellulose extraction from SCT using an optimized alkaline-based separation protocol as earlier reported from our laboratory [32]. In short, SCB underwent chlorite-acetic acid delignification (60°C, 3 h), followed by filtration (Whatman No. 1) and neutralization. The dried delignified SCT (80°C, 24 h) was treated with $NaOH/H_3BO_3$, and cellulosic residues were recovered and washed to neutrality. The SCB-cellulose was further used in enzyme-mediated nanofibrillation.

SCB-derived cellulose was subjected to enzymatic hydrolysis using *At*GH9C-CBM3A-CBM3B and *At*CBH5A (1.5 mg enzyme/g substrate) under controlled conditions. Enzymatic nanofibrillation reaction was performed in 50 mM sodium phosphate buffer (pH 7.5, 3 mL reaction mixture volume) containing 1% (w/v) SCB-cellulose, 5% (v/v) sodium azide, and enzymes (*At*GH9C-CBM3A-CBM3B and *At*CBH5A, 1.5 mg/mL each). Hydrolysis proceeded at 50°C (pH 7.5, 180 rpm) for 6 h under restrictive conditions. Reactions were terminated by heating to 95°C (15 min) after 24 h. Post-hydrolysis, samples were centrifuged (6,500 × g), and the hydrolysate and the fibres were separated. The fibres were washed sequentially with 0.1 M HCl and water until neutral pH. The resulting material was sonicated (30% amplitude, 20 min), dried (35°C, 48 h), labelled as EN-NC (Enzymatic nanocellulose) and stored at 25°C for subsequent use. The hydrolysate was analysed using TLC technique, where the mobile phase consisted of n-butanol/acetic acid/water (2:1:1 v/v). Chromatograms were visualized using a detection reagent (5% H_2SO_4 in methanol with 0.5% α-naphthol) as previously described in **Chapter 3, section 3.2.8**.

5.2.4 Oxidative nanofibrillation using TEMPO catalyst

A suspension of SCB-cellulose (2 g in 150 mL distilled water) was added with 0.025 g of TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl) and 0.25 g NaBr on a magnetic stirring at 25°C. The reaction was initiated by slowly adding $NaClO_2$ solution (19.4 mmol/g, 0.13 L, pH 10), while maintaining pH at 10.5 with 0.5 M NaOH. Reaction completion was confirmed by pH stabilization after 2 h [35]. The product was filtered, washed and resuspended in 5 mL deionized water followed by sonication (30% amplitude, 20 min) under identical conditions to enzymatically treated samples to yield TEMPO Oxidised-NC dispersion. Following TEMPO-mediated oxidation, the samples were dried (35°C, 48 h), labelled as TO-NC and then subjected to characterization.

5.2.5 Characterization of nanocelluloses

5.2.5.1 FESEM analysis

The morphological characteristics of both nanocelluloses (NCs), EN-NC and TO-NC were examined by scanning electron microscopy (SEM). For analysis, the dried EN-NC and TO-NC samples were sputter-coated with a 10-nm gold layer using a Hitachi E-1010 ion sputter. Surface morphology characterization was conducted by field-emission scanning electron microscopy (FESEM; Zeiss Sigma, Germany) at an accelerating voltage of 5.0 kV and micrographs acquired across a magnification range of 15–50 kX.

5.2.5.2 XRD-analysis

The crystallinity indices (*CrI*) of EN-NC and TO-NC were evaluated using X-ray diffraction (XRD). The samples were analyzed with an X-ray diffractometer (XRD, SmartLab, Rigaku Technologies, Japan) across a 2θ range of 5°–40° at a step size of

0.05°. The *CrI* was calculated according to the established method [36] by following the equation

$$CrI\ (\%) = \frac{(I\_crystalline - I\_amorphous)}{I\_crystalline} \times 100$$

Where, I_crystalline = peak intensity at 2θ = 22° (crystalline cellulose) I_amorphous = minimum intensity at 2θ = 18° (amorphous background).

5.2.5.3 FTIR analysis

The functional groups present in the EN-NC and TO-NC were characterized by Fourier-transform infrared (FT-IR) spectroscopy. For analysis, samples of NCs from enzymatic hydrolysis and TEMPO-oxidation were homogenized with potassium bromide (KBr) at a 1:100 (w/w) ratio using an agate mortar and pestle. The blended mixture was then compressed into pellets under 15-ton hydraulic press (Hydraulic pellet press, KP, Kimaya engineers, India). FT-IR spectra were acquired using a spectrophotometer (PerkinElmer Spectrum Two, USA) across the wavenumber range, 4000–400 cm^{-1}.

5.2.6 Preparation of EN-NC-CMC-Na hydrogel

The synthesized nanocellulose (NC) *via* enzymatic route (EN-NC) was combined with carboxymethyl cellulose-sodium salt to produce hydrogel (EN-NC-CMC-Na). TEMPO-oxidized nanocellulose (TO-NC) has excess $–COO^-$ groups, leading to high electrostatic repulsion, resulting in poor self-assembly in gel formation with reduced gel strength. Thus, only EN-NC was used for hydrogel preparation. Enzyme-synthesized nanocellulose, EN-NC (0.5 g) was dispersed in 50 mL of 2% (v/v) acetic acid solution in a 200 mL beaker and stirred on a magnetic stirrer at 500 rpm for 1 h at 25°C to promote swelling and partial dispersion of nanocellulose (NC). This

process yielded a uniform, paste-like NC suspension. Separately, 0.5% (w/v) carboxymethyl cellulose sodium salt (CMC-Na) solution was made by dissolving 1.0 g in 50 mL of distilled water in a 100 mL beaker on a magnetic stirring at 500 rpm at 60°C until a homogeneous solution was obtained. After complete dissolution, the CMC-Na solution was cooled to 25°C. The NC-acetic acid suspension was added dropwise to the CMC-Na solution using a 50 mL syringe pump with gentle stirring (20 rpm) on the magnetic stirrer to ensure uniform incorporation and to avoid premature gelation. The acetic acid containing, acidic NC suspension lowered the pH of the CMC solution, facilitating intermolecular interactions and inducing the hydrogel formation. The resulting EN-NC–CMC-Na mixture was incubated at 35°C until a constant weight was achieved, indicating stabilization of the hydrogel network. The final hydrogel, designated as EN-NC–CMC-Na hydrogel, was stored at 25°C for further characterization and applications.

5.2.7 Characterization of EN-NC-CMC-Na hydrogel

5.2.7.1 Swelling studies

The swelling capacity of the EN-NC-CMC-Na hydrogel, was investigated by immersing a predetermined mass of gel in 100 mL deionized water in a stationary incubator at a temperature (25±1°C) for 12 h. At 2-hour intervals, samples were removed, surface-dried with filter paper and weighed to determine the water uptake. The extent or percentage of swelling was quantified using the following equation:

$$\text{Swelling (\%)} = \frac{(W_s - W_0)}{W_0} \times 100$$

where, W_0 = initial dry weight of hydrogel (g) and
W_s = swollen weight at measurement time (g).

5.2.7.2 Determination of point of zero charge (pH_{zpc}) of EN-NC-CMC-Na hydrogel

The point of zero charge (pH_{pzc}) of the adsorbent hydrogel was determined by using the pH drift method. The point of zero charge (pH_{pzc}) is the pH at which the net surface charge of a material (e.g., an adsorbent, nanoparticle, or mineral) is neutral. Below pH_{pzc}, the surface is positively charged favoring anion adsorption, while above pH_{pzc}, it becomes negatively charged favoring cation adsorption. The experiment employed ten of 10 mL samples of 0.01 M NaCl solution, with pH systematically adjusted from 1 to 10 by 0.1 M NaOH or H_2SO_4 titration. The pH point of zero charge (pH_{pzc}) was determined by introducing a precise mass of gel into each flask, followed by 48-hour equilibration under constant agitation of 200 rpm at 25°C in an orbital shaker incubator. The final pH values (pH_{final}) from initial pH values ($pH_{initial}$) values were noted. The difference of the pH_{final} from $pH_{initial}$ was plotted against their corresponding initial pH values $pH_{initial}$, with the pHpzc identified as the intersection point of the resulting curve with the line $pH_{initial} = pH_{final}$ [37].

5.2.7.3 Brunauer-Emmett-Teller (BET) analysis of EN-NC-CM-Na hydrogel

The specific surface area and pore volume of the CMC, EN-NC and EN-NC-CMC-Na hydrogels were evaluated using Brunauer-Emmett-Teller (BET) analysis. Prior to measurements, the dried CMC, EN-NC and EN-NC-CMC-Na hydrogel samples were degassed at 150°C for 3 h. Nitrogen (N_2) adsorption-desorption isotherms were subsequently recorded at −196 °C using a Quantachrome instrument (Autosorb-iQ MP, USA).

5.2.8 Dye removal studies using EN-NC-CMC-Na hydrogel

5.2.8.1 Process variables and the response

Previous studies have indicated that optimizing process parameters for dye adsorption by maintaining one factor constant while varying others is both time-consuming and labour-intensive. Response surface methodology (RSM) has been employed for parameter optimization, enhancing dye removal efficiency to circumvent these limitations [38]. RSM has proven to be an effective strategy, minimizing both experimental duration and associated costs. Consequently, this investigation utilized response surface methodology (RSM) to determine the optimal adsorption conditions for MB and CV elimination from aqueous medium. The optimization study incorporated four independent variable, *viz.* initial dye concentration (mg/L), adsorbent (EN-NC-CMC-Na Hydrogel) dosage (g), contact time (h) and pH, which were selected based on findings from previous report [37]. The temperature and rotation speed, rpm were kept constant at 25°C and 200 rpm to study the effect of the primary variables. The interactions among process parameters and their influence on the response (dye removal efficiency) were examined by using Central Composite Design (CCD-RSM). The percentage removal of the dye from aqueous medium was the primary response parameter for process optimization [39]. CCD-RSM proposed a randomized experimental matrix comprising 29 runs, each executed in triplicate to minimize variability. To derive optimal process conditions, the response data were statistically evaluated using DESIGN EXPERT software (Stat-Ease Inc., USA, version 13).

5.2.8.2 Batch adsorption experiment

A stock solution of each dye (Congo Red and Methylene Blue) with a concentration of 1000 ppm was prepared by dissolving 0.1 g of the respective dye in 100 mL of distilled water. Working solutions with target concentrations (15, 30, 45, and 60 mg/L) were subsequently prepared through serial dilution of the stock solution. For the experimental trials, 10 mL aliquot of each concentration were transferred to 25 mL black cap glass bottle. Batch adsorption experiments were performed under controlled temperature conditions at 25.0 ± 0.5 °C. Post-adsorption, the dye concentrations were analyzed spectrophotometrically (Thermo Scientific™ Genesys 180 UV-Vis) with predetermined λ_{max} values. The characteristic absorption maxima (λ_{max}) for MB and CR were determined to be 664 nm and 498 nm, respectively, verified through standard calibration procedures [3,4]. The removal efficiency (%) and adsorption capacity (mg/g) of the EN-NC-CMC-Na hydrogel were calculated using the following equations:

$$\textbf{Dye removal efficiency (\%)} = \frac{(C_0 - C_t)}{C_0} \times 100$$

$$\text{Adsorption capacity at time t (mg/g), } \mathbf{q_t} = \frac{(C_0 - C_t) * V}{M}$$

$$\text{Adsorption capacity at equilibrium (mg/g), } \mathbf{q_e} = \frac{(C_0 - Ce) * V}{M}$$

Where, C_o represents the initial dye concentration (mg/L), while C_t and C_e denote the concentrations at time t and equilibrium (mg/L), respectively. V signifies the solution volume (L) and M corresponds to the mass of dry adsorbent beads (g). The terms q_t and q_e indicate the adsorption capacities (mg/g) at time t and equilibrium, respectively.

5.2.9 Adsorption isotherm of EN-NC-CMC-Na hydrogel

Adsorption isotherm studies were performed using batch experiments, with initial dye concentrations of CR and MB were systematically varied between 25 and 60 mg/L. The experimental conditions were standardized with a fixed contact time of 180 min and an adsorbent mass of 0.1 g. For each tested concentration of CR or MB dye, reactions were carried out in 10 mL batches at 25°C. Multiple mathematical models (Langmuir, Freundlich, Temkin and Sips) were employed to assess the effect of dye removal efficiency (%) of the EN-NC-CMC-Na hydrogel.

5.2.10 Adsorption kinetics of EN-NC-CMC-Na hydrogel

Adsorption experiments were conducted in batch mode, maintaining a fixed EN-NC-CMC-Na hydrogel adsorbent mass of 0.1 g and an initial dye concentration of 60 mg/L of CR or MB. The contact time systematically varied from 20 to 180 min and the dye removal efficiency (%) was assessed. All dye (CR/MB) experiments used 10 mL solutions at 25°C for each concentration tested. The temporal variation in dye uptake by the adsorbent was analyzed by plotting adsorption data against time and fitting it with various kinetic models (pseudo-first-order, pseudo-second-order, Elovich, and intra-particle diffusion) to elucidate the adsorption mechanism.

5.2.11 Desorption and regeneration studies

The regeneration capacity and reusability of adsorbent materials represent critical parameters for their practical application in wastewater treatment. These characteristics were assessed through successive adsorption-desorption cycles employing batch experimental methodology. EN-NC-CMC-Na hydrogel (0.1 g) was mixed in 10 mL of dye solution (60 mg/L) until an equilibrium was attained. Following adsorption equilibrium, the dye-saturated hydrogel was subjected to separate desorption

analysis utilizing 10 mL of multiple elution media: double-distilled water, 50% ethanol, 1 M NaCl, 1 M NaOH, 1 M KOH, 1 M $CaCl_2$, 0.1 M HCl and 0.1 M HNO_3 for 1 h. After, the hydrogel adsorbent was isolated through filtration and reactivated by heating at 35°C for 1 h for complete removal of solvent before reuse. The desorption efficiency was quantitatively determined by spectrophotometric analysis of the eluents by measuring the absorbance at 498 nm for CR and 665 nm for MB at 25°C. The dye concentrations were calculated from the 1 O.D equivalence to standard calibration curves of each dye.

5.3. Results and Discussion

5.3.1 Purification and activity analysis of *At*GH9C-CBM3A-CBM3B and *At*CBH5A

The purified enzymes, *At*GH9C-CBM3A-CBM3B and *At*CBH5A, exhibited a single band on SDS-PAGE, corresponding to their predicted molecular masses of 97.2 kDa and 62 kDa, respectively (**Fig. 5.1**), consistent with their theoretical molecular masses. The yields of the purified *At*GH9C-CBM3A-CBM3B and *At*CBH5A from 500 mL LB cultures were determined to be 10.8 ± 0.2 mg and 12.5 ± 0.5 mg, respectively. The maximum specific activity of 58.8 ± 0.8 U/mg for *At*GH9C-CBM3A-CBM3B was achieved under optimum conditions: enzyme concentration of 10 mg/mL, reaction time of 1 min and a temperature of 55°C, using 1.0% (w/v) CMC-Na as the substrate (**Table 5.1**). In contrast, *At*CBH5A exhibited its highest activity of 96.2 ± 1.2 U/mg under optimum conditions of enzyme concentration, 100 mg/mL, reaction duration, 5 min and temperature, 65°C, with the same substrate, 1.0% (w/v) CMC-Na (**Table 5.1**).

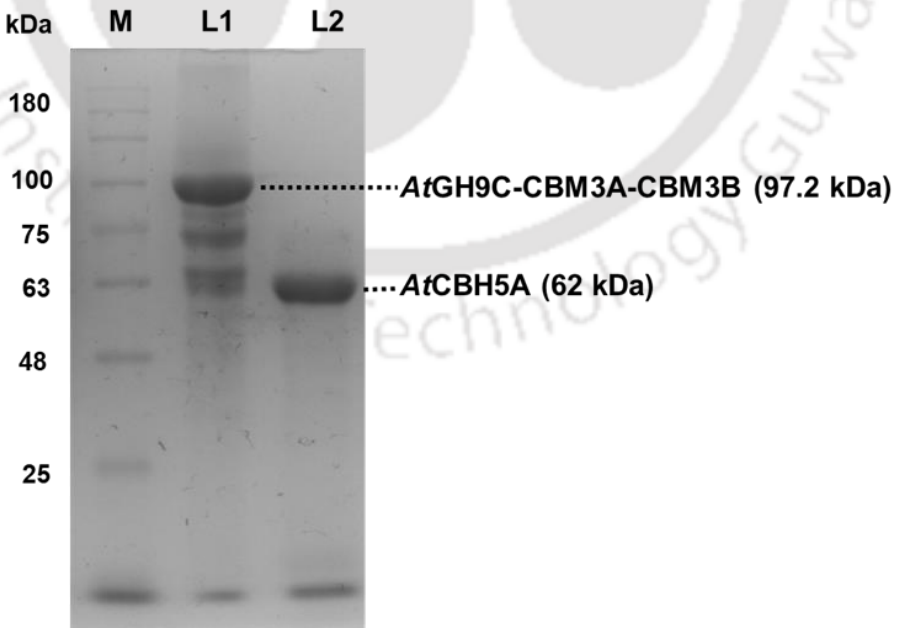


Fig. 5.1 Purified recombinant proteins on SDS-PAGE gel (12%, w/v). Lane M− Molecular mass marker (11–180 kDa, Hi-media, Pvt. Ltd. India), lane 1- *At*GH9C-CBM3A-CBM3B and lane 2- *At*CBH5A.

Table 5.1 Optimum conditions and Specific activity of recombinant enzymes.

Recombinant enzyme	Optimum Conditions			Specific activity (U/mg)
	Concentration (µg/mL)	Time (min)	Temperature (°C)	
Endoglucanase, *At*GH9C-CBM3A-CBM3B	10	1	55	58.8 ± 0.8
Cellobiohydrolase, *At*CBH5A	100	5	65	96.2 ± 1.2

5.3.2 Synthesis of nanocellulose by enzymatic treatment and TEMPO-oxidation of SCT

Cellulose was extracted from sugarcane trash (SCT) by using a previously optimized alkali-based protocol, involving delignification (chlorite-acetic acid, 60°C, 3 h), neutralization and NaOH/H_3BO_3 treatment [32]. The purified SCB-cellulose was enzymatically nanofibrillated using *At*GH9C-CBM3A-CBM3B and *At*CBH5A (1.5 mg enzyme/g substrate) in 50 mM phosphate buffer (pH 7.5) at 50°C for 6 h. Post-hydrolysis, fibers were washed, sonicated (30% amplitude, 20 min), and dried to yield EN-NC (Enzymatic Nanocellulose) (**Fig. 5.2a**). In contrast, for TEMPO-mediated oxidative nanofibrillation, SCB-cellulose (2g) was oxidized using TEMPO/NaBr/$NaClO_2$ (pH 10.5, 2 h), followed by washing, sonication (30% amplitude, 20 min), and drying to obtain TO-NC (TEMPO-Oxidized Nanocellulose) (**Fig. 5.2b**). The resulting dispersions were characterized for further analysis.

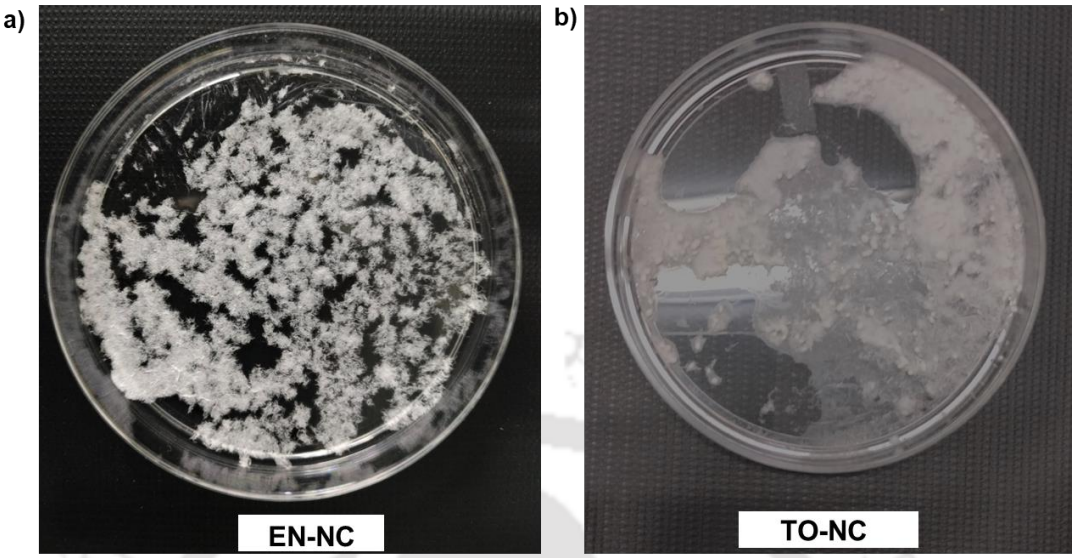


Fig. 5.2 Digital images of dried nanocellulose (NCs) derived from sugarcane trash (SCT) cellulose (a) Enzymatically processed NC (EN-NC) and (b) TEMPO-oxidized NC (TO-NC).

Thin-layer chromatography (TLC) analysis of the enzymatic nanocellulose hydrolysate demonstrated the predominant presence of cellotetraose and higher cello-oligosaccharides (DP > 4) after the 6 h reaction period (**Fig. 5.3**). The absence of glucose residues or shorter than DP4 cello-oligosaccharides in the chromatogram confirmed the restricted hydrolysis of SCT-cellulose, indicating selective fiber fragmentation without significant saccharification.

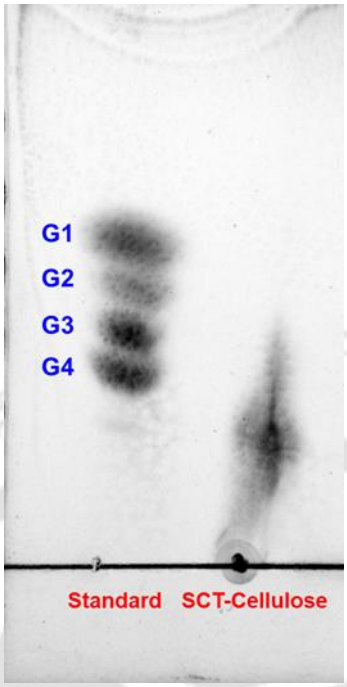


Fig. 5.3 TLC analysis of enzymatic nanocellulose hydrolysate. Lane assignment: G1 (glucose), G2 (cellobiose), G3 (cellotriose), and G4 (cellotetraose) as standards. The hydrolysate shows predominant oligosaccharides with degree of polymerization ≥4 (DP≥4), confirming restricted hydrolysis of cellulose without significant glucose release.

5.3.3 Characterization of EN-NC and TO-NC by FESEM, FTIR and XRD

5.3.3.1 FESEM analysis

FESEM analysis of raw SCT-cellulose, EN-NC and TO-NC revealed distinct structural transformations during nanocellulose production from sugarcane trash (SCT). NC diameters were analyzed using ImageJ software. The raw SCT cellulose exhibited coarse, micron-scale architecture, ~10 µm (**Fig. 5.4a**). Enzymatic treatment produced EN-NC with uniform, ultra-thin fibrils, ≤20 nm diameter (**Fig. 5.4b**), demonstrating the substrate-specific action of cellulases. TEMPO-mediated oxidation yielded TO-NC with slightly broader fibrils, 20-100 nm (**Fig. 5.4c**), characteristic of the chemical oxidation process. The dimensional differences between EN-NC and TO-NC highlight the contrasting mechanisms: enzymatic precision versus oxidative

random cleavage. Notably, EN-NC's narrower size distribution reflects the selective hydrolysis of amorphous regions, while TO-NC's varied sizes result from non-specific carboxylation. Both nanocelluloses showed complete morphological transformation from the raw material, confirming effective fibrillation. These results demonstrated method-dependent control over nanocellulose morphology.

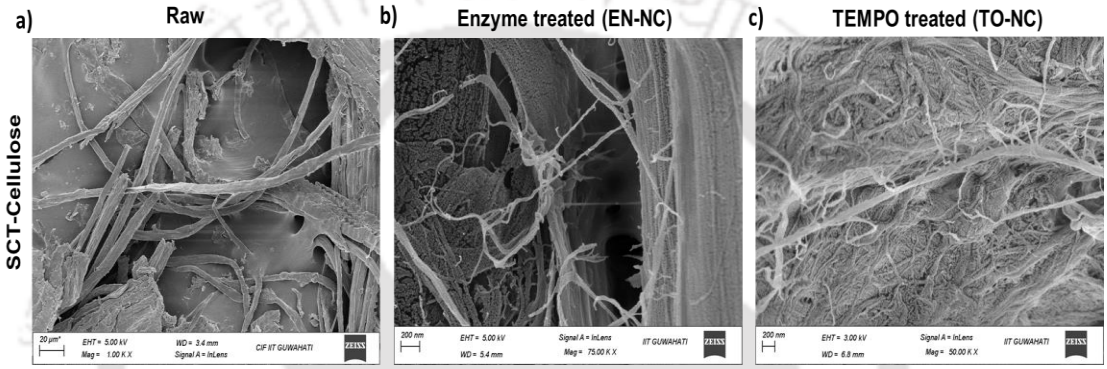


Fig. 5.4 Field Emission Scanning Electron Microscopy (FESEM) analysis of raw, enzyme-treated and TEMPO-oxidized cellulose from sugarcane trash (SCT). The micrographs reveal the morphological changes in cellulose fibers by different treatments (a) Raw SCT cellulose displaying intact, bundled fibrils with a rough surface; (b) Enzyme-treated cellulose (EN-NC) showing partial fibrillation and surface erosion due to enzymatic action and (C) TEMPO-oxidized cellulose (TO-NC) exhibiting excessive disintegration.

5.3.3.2 FTIR analysis

The FTIR profiles of raw SCT-cellulose, enzyme-treated nanocellulose (EN-NC) and TEMPO-oxidized nanocellulose (TO-NC) are presented in **Fig. 5.5** for structural comparison. A prominent absorption band at approximately, 3400 cm^{-1} corresponds to O–H stretching vibrations, while the band at 2800 cm^{-1} arises from aliphatic C–H stretching. Peaks observed at 1480 cm^{-1} and 1250 cm^{-1} are assigned to $–CH_2$ scissoring vibrations, whereas absorptions in the range of 1165–1145 cm^{-1} signify C–O–C asymmetric stretching. Additionally, peaks between 1120–1000 cm^{-1} are

attributed to C–O bond stretching. The presence of bands near 910 cm^{-1} and 890 cm^{-1} indicates β-D-glucosyl groups. Notably, the spectra of EN-NC and TO-NC closely resemble that of raw SCT-cellulose, confirming that the enzymatic hydrolysis and TEMPO-mediated oxidation processes did not alter the fundamental molecular structure of cellulose.

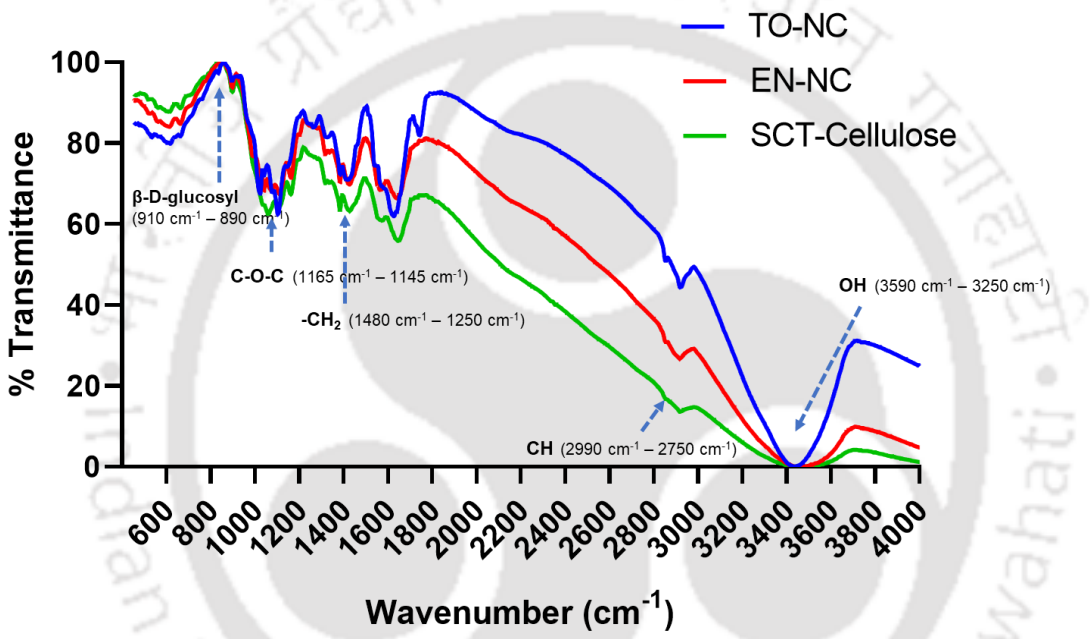


Fig. 5.5 FTIR spectroscopic analysis of SCT-Cellulose, EN-NC and TO-NC. Key vibrational bands are annotated, demonstrating structural integrity after the treatments.

5.3.3.3 XRD analysis

The crystallinity indices (*CrI*s) of all samples were quantitatively compared (**Fig. 5.6**). The analysis revealed no significant differences between enzymatically treated nanocellulose (EN-NC; 46.8%) and TEMPO-oxidized nanocellulose (TO-NC; 47.5%), with both values remaining higher than untreated SCT-Cellulose (41.2%**)** (**Table 5.2**). These observations suggested that the mechanical defibrillation process, rather than the pretreatment methodology (enzymatic vs. oxidative), primarily

determined the final nanocellulose crystallinity. The marginal *CrI* variations may reflect treatment-induced removal of amorphous matrix components (hemicellulose/lignin), potentially enabling cellulose microfibril reorganization into larger crystallite domains.

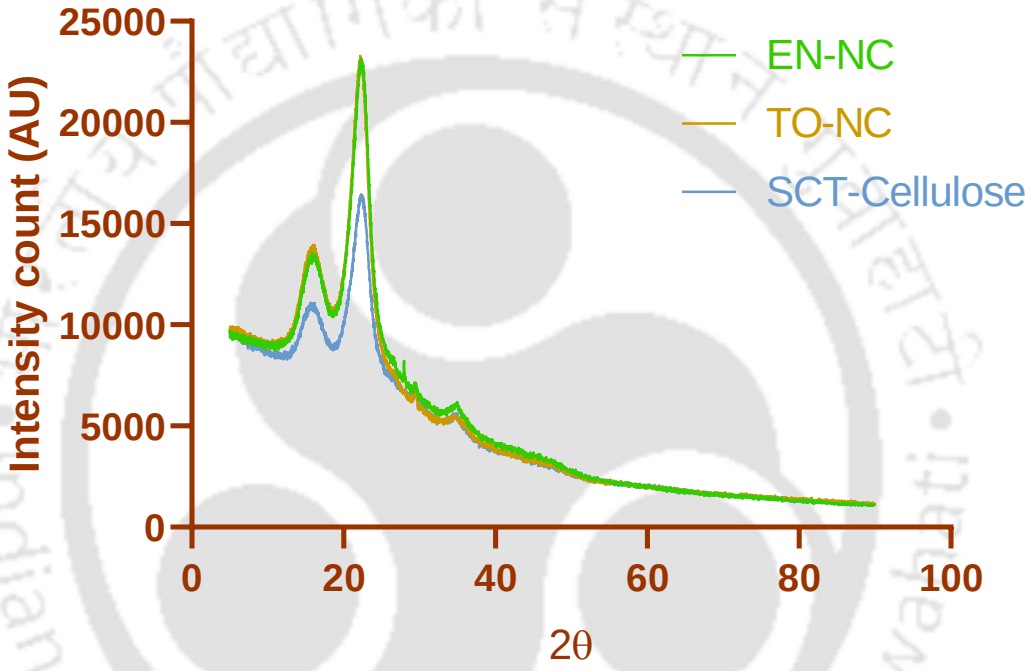


Fig. 5.6 XRD-analysis of SCT-Cellulose, EN-NC and TO-NC from 2θ range of 5°–90°. The similar peak positions and intensities among all samples indicate preservation of the native cellulose crystalline structure following both enzymatic and chemical treatments.

Table 5.2 *CrI* analysis of SCT-Cellulose, EN-NC and TO-NC.

Sample	*CrI*
SCT-Cellulose	41.2
EN-NC	46.8
TO-NC	47.5

5.3.4 Synthesis of EN-NC-CMC-Na hydrogel

Carboxymethylcellulose (CMC), an abundant and economical biopolymer, serves as a sustainable precursor for synthesizing superabsorbent hydrogels. However, CMC-derived hydrogels often exhibit inferior mechanical strength, necessitating reinforcement with fillers. Nanocellulose (NC) emerges as an effective reinforcing agent, enhancing physicochemical properties such as structural integrity, crosslinking efficiency and water retention. While CMC-Na alone cannot form stable hydrogels, NC incorporation significantly improves the mechanical stability and functional performance, making these hydrogels particularly suitable for applications in wastewater treatment (e.g., dye adsorption) and controlled drug delivery. The hydrogel was formed through pH-driven addition of enzyme-treated nanocellulose (EN-NC) and carboxymethyl cellulose (CMC) as described in **section 5.2.6**.

EN-NC dispersion in acetic acid swelled fibrils while acidifying the system, when mixed with CMC solution, protonation of carboxylate groups ($–COO^- \rightarrow –COOH$) enabled electrostatic crosslinking with NC hydroxyl groups. Simultaneously, entangled NC fibrils reinforced the matrix through hydrogen bonding and physical interpenetration. Controlled incubation at 35°C optimized the network density by enhancing polymer mobility while evaporating excess water. The resulting hydrogel (**Fig. 5.7**) structure derived stability from synergistic interactions: electrostatic forces, hydrogen bonds, hydrophobic associations and physical entrapment. Gradual acidification by dropwise addition and mild thermal treatment ensured homogeneous, reproducible network formation while preserving NC nanofibrillar architecture, yielding a mechanically robust yet porous hydrogel with hierarchical organization. This

approach balanced chemical crosslinking with physical reinforcement for functional material design.

Fig. 5.7 Digital image of the EN-NC-CMC-Na hydrogel obtained via pH-driven addition of EN-NC to CMC-Na solution.

5.3.5 Characterization of EN-NC-CMC-Na hydrogel

5.3.5.1 Swelling ratio analysis

The swelling behavior of the hydrogel is governed by both its porous architecture and surface chemical properties. The hydrogel exhibited distinct time-dependent swelling kinetics, characterized by an initial gradual increase (20% at 2 h to 30% at 4 h), followed by rapid water uptake (70% at 6 h) as shown in **Fig. 5.8**. A transient plateau (~72% at 8 h) suggested partial saturation, while prolonged incubation to 12 h revealed a secondary swelling phase (82%), indicative of matrix relaxation. This biphasic profile reflects the hydrogel's hierarchical porosity, where the initial fast absorption through interconnected macropores occurs, followed by slower diffusion into the nanofibrillar network. The osmotic pressure difference dropped, weakening the force that was fighting against the natural shrinking of the hydrogel.

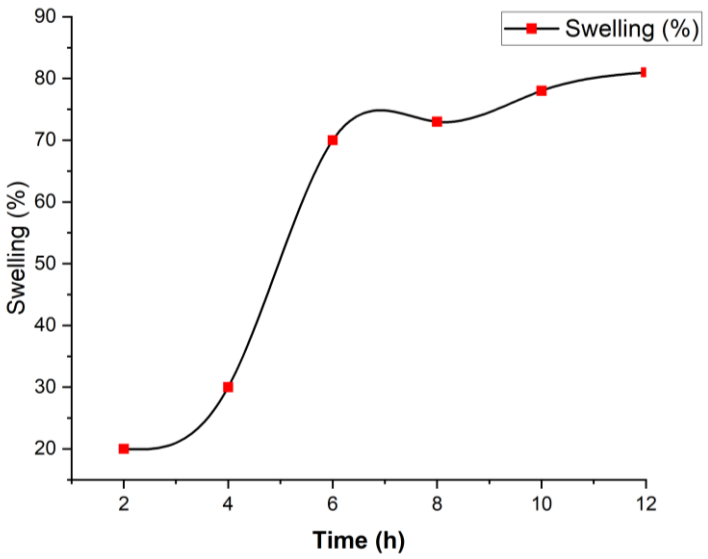


Fig. 5.8 Time-dependent swelling kinetics of the EN-NC-CMC-Na hydrogel. The hydrogel exhibits biphasic swelling behaviour, demonstrating the interplay between macropore filling and nanoscale polymer hydration.

5.3.5.2 Point of zero charge (pH_{pzc}) analysis

The point of zero charge (pH_{zpc}) of the EN-NC-CMC-Na hydrogel was determined to be 6.6 (**Fig. 5.9**), representing the pH at which the net surface charge neutralizes. Beyond this critical pH value, the adsorbent surface charge becomes pH-dependent: (1) at pH > pH_{zpc}, deprotonation generates a negatively charged surface favorable for cationic dye adsorption through electrostatic attraction, while (2) at pH < pHzpc, proton adsorption creates a net positive surface charge. This charge reversal behavior demonstrates the pH-responsive adsorption properties of the hydrogel, with the pH_{zpc} serving as the thermodynamic transition point between dominant surface charge states.

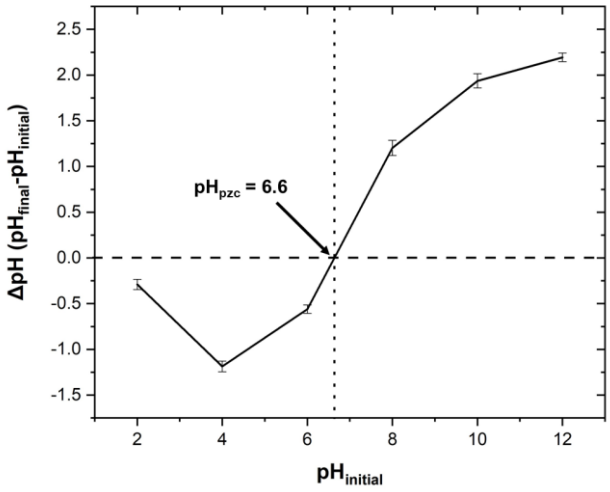


Fig. 5.9 pH$_{pzc}$ plot of the prepared NC-hydrogel showed pH$_{pzc}$ = 6.6.

5.3.5.3 BET analysis

Nitrogen physisorption measurements by BET analysis revealed significant improvements in pore size in the EN-NC-CMC-Na hydrogel as compared with the individual components. The EN-NC-CMC-Na hydrogel exhibited over 3-fold increase in BET surface area (4.81 m²/g) versus pure CMC-Na (1.57 m²/g), while NC (EN-NC) alone showed intermediate porosity (2.16 m²/g). The hydrogel, EN-NC-CMC-Na with (4.81m^2/g) showed 30% increase in BET surface area than both, EN-NC and CMC-Na taken together (1.57 + 2.16 = 3.73 m^2/g). Pore volume followed a similar trend, with the EN-NC-CMC-Na achieving 0.17 cm³/g, 240% enhancement over CMC (0.05 cm³/g) and 70% greater than NC (0.10 cm³/g). A 13% enhancement in pore volume was observed for the EN-NC-CMC-Na hydrogel as compared to the sum of individual pore volumes of EN-NC and CMC-Na. These results demonstrated the nanocellulose's role in creating a more open porous network within the CMC-Na matrix. The data suggest successful integration of NC fibrils, which disrupt CMC-Na's dense packing to generate

additional surface area and void spaces, potentially enhancing performance in adsorption-driven applications.

Table 5.3 BET analysis of CMC-Na, NC and EN-NC-CMC-Na hydrogel.

Parameter	CMC-Na	NC	CMC-Na+NC	EN-NC-CMC-Na	% Increase
BET Surface area (m^2/g)	1.57	2.16	3.73	4.81	206
Total pore volume (cm^3/g)	0.05	0.10	0.15	0.17	13.3

5.3.6 RSM-CCD optimization of adsorption parameters for enhanced dye removal by EN-NC-CMC-Na hydrogel

5.3.6.1 Run parameters and responses

The adsorption process parameters for Congo Red and Methylene Blue were optimized using Response Surface Methodology (RSM) with a Central Composite Design (CCD). Employing RSM for experimental design enhances accuracy by refining process parameters across multiple levels, thereby reducing experimental error [39]. The individual and interactive influences of these parameters on the response variable were analyzed using three-dimensional surface plots generated by CCD-RSM design.

The process parameters were defined in coded terms as follows: **A**: Initial dye concentration (15–60 mg/L); **B**: Adsorbent dosage (0.1–0.6 g); **C**: Contact time (1–6 h) and **D**: pH (3–9). The optimization response was quantified as the percentage of dye removal from aqueous solutions. The coded variables and their corresponding optimization parameters are given in **Table 5.4.** Each independent variable was evaluated at five coded levels in the CCD-RSM experimental design. The RSM-CCD generated 29 runs for each set and the experimental and predicted values from trial runs are presented in **Table 5.5**.

Table 5.4 Factor ranges in CCD-RSM for dye removal optimization.

Factor	Name	Level of factors	I	II	III	IV	V
A	**Initial dye concentration (mg/mL)**	5	15.00	26.25	37.50	48.75	60.00
B	**Adsorbent dose (g)**	5	0.10	0.22	0.35	0.47	0.60
C	**Contact time (h)**	5	1.00	2.25	3.50	4.75	6.00
D	**pH**	5	3.00	4.50	7.50	6.00	9.00

Table 5.5 Comparative analysis of predicted and experimental values of percentage removal of CR and MB by EN-NC-CMC-Na hydrogel.

					CR- removal efficiency (%)		MB- removal efficiency (%)	
Run	**A**	**B**	**C**	**D**	**Predicted**	**Experimental**	**Predicted**	**Experimental**
1	26.25	0.225	2.25	4.5	80.5	79.72	78.63	79.46
2	48.75	0.225	2.25	4.5	75.54	76.91	80.34	80.14
3	26.25	0.475	2.25	4.5	77.16	77.73	88.73	87.06
4	48.75	0.475	2.25	4.5	79.1	78.21	90.6	91.61
5	26.25	0.225	4.75	4.5	84.14	82.45	82.34	82.17
6	48.75	0.225	4.75	4.5	76.42	76.84	82.26	80.96
7	26.25	0.475	4.75	4.5	85.05	84.27	92.08	91.32
8	48.75	0.475	4.75	4.5	84.24	83.76	92.17	92.64
9	26.25	0.225	2.25	7.5	80.99	81.54	87.84	86.13
10	48.75	0.225	2.25	7.5	76.9	77.84	84.27	85.23
11	26.25	0.475	2.25	7.5	80.17	79.91	88.73	90.23
12	48.75	0.475	2.25	7.5	82.98	84.73	85.32	84.25
13	26.25	0.225	4.75	7.5	85.03	86.08	93.05	92.24
14	48.75	0.225	4.75	7.5	78.18	77.68	87.69	88.12
15	26.25	0.475	4.75	7.5	88.47	87.18	93.58	92.54
16	48.75	0.475	4.75	7.5	**88.53**	**89.47**	88.4	87.77
17	15	0.35	3.5	6	87.75	89.18	93.39	94.78
18	60	0.35	3.5	6	82.84	81.18	89.9	89.55
19	37.5	0.1	3.5	6	78.73	78.17	83.88	84.34
20	37.5	0.6	3.5	6	85.73	86.06	**94.67**	**95.25**
21	37.5	0.35	1	6	77.12	75.62	83.89	83.55
22	37.5	0.35	6	6	86.3	87.58	90.67	92.05
23	37.5	0.35	3.5	3	74.24	75.49	78.09	78.46
24	37.5	0.35	3.5	9	79.01	77.53	83.52	84.19
25	37.5	0.35	3.5	6	84.4	84.4	82.75	83.55
26	37.5	0.35	3.5	6	83.4	83.12	80.19	81.25
27	37.5	0.35	3.5	6	82.23	80.42	82.23	86.67
28	37.5	0.35	3.5	6	83.63	82.53	82.85	82.45
29	37.5	0.35	3.5	6	84.4	85.21	83.55	81.43

A – Initial dye concentration (mg/L); **B** – EN-NC-CMC-Na, Adsorbent dosage (g);
C –Contact time (h); **D** – pH

Experimental analysis of adsorption studies indicated maximum removal rate of Congo Red (CR) at 89.5% under response conditions of pH 7.5, a time duration of 4.75 h, a concentration of 95.3 mg/L dye, and a dosage of 0.47 g of EN-NC-CMC-Na hydrogel. For Methylene Blue (MB), the removal rate reached 94.7% at pH 6.0, 3.5 h, a concentration of 37.5 mg/L of dye and a dosage of 0.6 g of EN-NC-CMC-Na hydrogel.

5.3.6.2 Interactive plot analysis

Multivariate optimization necessitates a comprehensive evaluation of independent variables on system responses. Three-dimensional quadratic surface plots were generated by systematically varying pairs of parameters across their experimental ranges while maintaining other factors at constant levels. These plots not only visualize parameter interactions but also reveal their relative intensities through characteristic curvature patterns. The following analysis examines 3-D response surfaces for critical process variables - including initial dye concentration, adsorbent (EN-NC-CMC-Na hydrogel) dosage, contact time and pH - in relation to Congo red (CR) and Methylene blue (MB) removal efficiency from aqueous solutions.

5.3.6.2.1 3D plot for Congo red

The response surface methodology (RSM) plots provided a comprehensive understanding of the interactive effect of different operational parameters on CR dye removal efficiency. The combined effect of initial dye concentration (A) and adsorbent (EN-NC-CMC-Na hydrogel) dose (B) is shown in **Fig. 5.10a**. As the hydrogel adsorbent dose increases, the Dye removal efficiency also increases, attributed to the greater availability of active sites for adsorption. In contrast, higher initial dye concentrations tend to slightly reduce the efficiency, likely due to saturation of the available adsorption sites. The curvature of the plot indicates a strong interaction

between these parameters, with optimal efficiency observed at low dye concentrations and high adsorbent doses. While increasing the adsorbent dose can somewhat counteract the effect of higher dye concentrations, the most efficient removal is achieved when low dye concentration is paired with a high adsorbent dose. The effect of initial dye concentration (A) and contact time (C) in shown in **Fig. 5.10b**. Longer contact time led to improved dye removal efficiency, as more time allows the system to reach adsorption equilibrium. However, higher dye concentrations again reduced the efficiency. The convex nature of the plot suggests that after a certain contact time, the benefits taper off, indicating that further extension of contact time yields minimal additional removal, especially post-equilibrium. **Fig. 5.10c** explores the interaction between initial dye concentration (A) and pH (D). Efficiency increases with rising pH up to a certain point, possibly due to the deprotonation of the adsorbent's surface functional groups, enhancing dye binding *via* electrostatic interaction. At very high pH values, however, the efficiency either plateaus or slightly drops, likely due to changes in dye solubility or reduced adsorbent affinity. Optimal removal is achieved near pH 6–7 at moderate dye concentrations. The contact time (C) and adsorbent dose (B) was examined in **Fig. 5.10d**. Both variables positively influenced the dye removal. **Fig. 5.10e** shows that higher pH and higher adsorbent doses synergistically improve dye removal, with the best performance near pH 7 and elevated adsorbent gel doses. **Fig. 5.10f** demonstrates a similar interaction between pH (D) and contact time (C), with peak efficiency around pH 6.5–7 and contact time near 4.5 h. Collectively, these plots underscore the importance of optimizing parameter interactions. The best operational conditions for CR dye removal were found to be: 26–32 mg/L dye concentration, 0.41–

0.47 g/100 mL adsorbent (EN-NC-CMC-Na hydrogel) dose, 4.5 h contact time and pH 6.5–7.0.

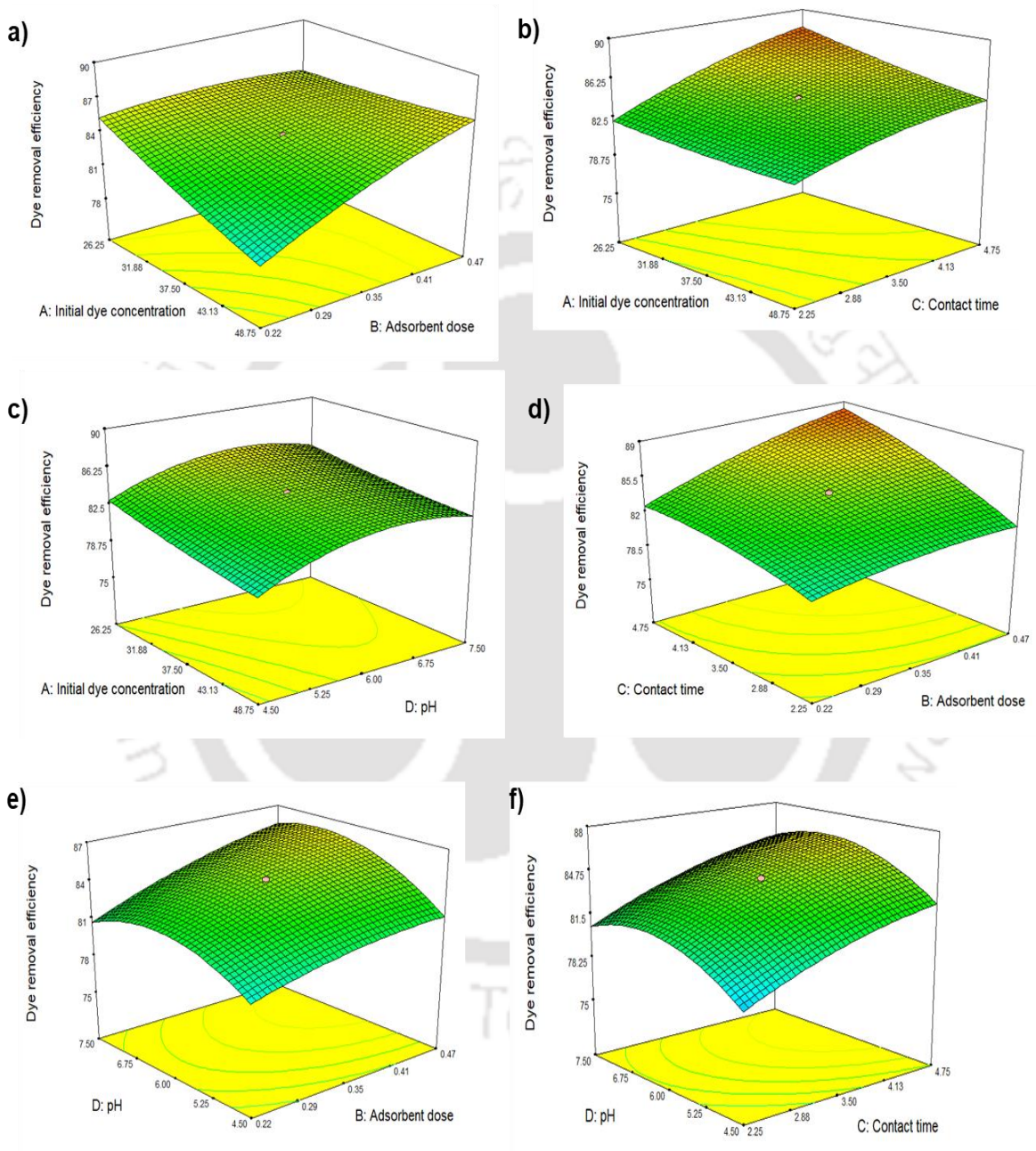


Fig. 5.10 RSM 3-D plots for CR dye removal (a) adsorbent dose and initial dye concentration (b) contact time and initial dye concentration (c) pH and initial dye concentration (d) contact time and adsorbent (EN-NC-CMC-Na hydrogel) dose (e) adsorbent (EN-NC-CMC-Na hydrogel) dose and pH (f) pH and contact time.

5.3.6.2.2 3D plot for Methylene blue

The interaction between initial MB dye concentration (A) and adsorbent (EN-NC-CMC-Na hydrogel) dose (B) on dye removal efficiency is displayed in **Fig. 5.11a**. As the adsorbent dose increased, a marked improvement in dye removal efficiency is observed, owing to the greater availability of active adsorption sites. Conversely, increasing the initial dye concentration resulted in a noticeable reduction in efficiency, likely due to competitive adsorption and eventual saturation of the adsorption sites. The curvature of the response surface suggested a significant interaction between these two variables. The optimal condition for MB dye removal was observed at low dye concentration and high adsorbent dose, where the abundance of adsorption sites maximizes dye uptake. The combined effect of initial dye concentration (A) and contact time (C) is displayed in **Fig. 5.11b**. A longer contact time enhanced the dye removal significantly, highlighting the necessity of adequate time to achieve adsorption equilibrium. As with previous observations, higher dye concentrations reduced the efficiency due to oversaturation of adsorption sites. The curved surface again implies an interaction effect, with optimal dye removal occurring at low dye concentrations and extended contact time. **Fig. 5.11c** shows the effect of initial MB dye concentration (A) and solution pH (D). Dye removal efficiency was higher at lower dye concentrations and increased with rising pH, especially in the near-neutral pH range. This trend was likely due to enhanced deprotonation of functional groups on the adsorbent, which strengthens electrostatic interactions with dye molecules. However, at both very low and very high pH values the dye removal efficiency slightly declines. The best results are achieved at moderate dye concentrations (26 mg/L to 40 mg/L) and a pH of around 6.5–7.0.

The interaction between contact time (C) and adsorbent dose (B) is shown in **Fig. 5.11d** Both variables positively influenced the dye removal. Longer contact times allowed for attaining the equilibrium, while higher adsorbent doses increase the number of active adsorption sites. The response surface displays a clear synergistic effect, with the best results occurring when both parameters are at high levels. **Fig. 5.11e** demonstrates the influence of pH (D) and adsorbent dose (B). As both pH and adsorbent dose increase, the dye removal improves. The optimal efficiency is observed at near-neutral pH and high adsorbent doses, emphasizing the combined role of surface charge and available binding capacity. **Fig. 5.11f** shows the interaction between pH (D) and contact time (C) at constant adsorbent dose. The dye removal efficiency increased with both variables, though the response levels off at higher values, indicating an equilibrium. Overall, the optimal conditions identified are: low initial dye concentration (26–32 mg/L), high adsorbent dose (0.41–0.47 g/100 mL), contact time of 4.5 h and pH near 7.0. These findings are critical for maximizing adsorption in dye removal processes.

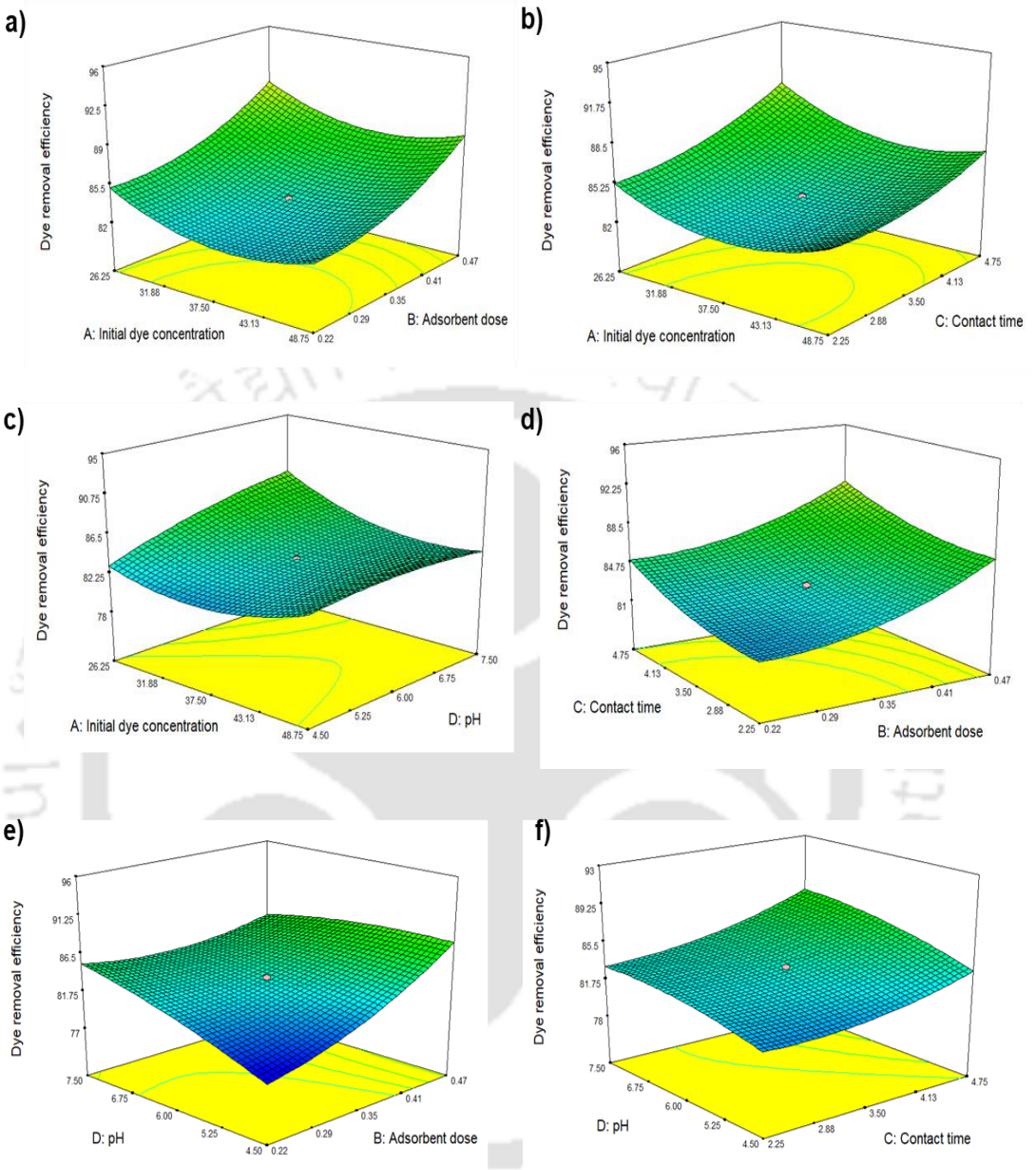


Fig. 5.11 RSM 3-D plots for MB dye removal (a) adsorbent (EN-NC-CMC-Na hydrogel) dose and initial dye concentration (b) contact time and initial dye concentration (c) pH and initial dye concentration (d) contact time and adsorbent dose (e) adsorbent (EN-NC-CMC-Na hydrogel) dose and pH (f) pH and contact time.

5.3.6.3 Analysis of variance

The statistical analyses for both Congo Red (CR) and Methylene Blue (MB) removal processes are presented in **Table 5.6**. The p-values associated with the model variables serve to assess the statistical significance ($p < 0.05$) of both individual parameters and their interaction effects. As indicated in the results, the p-values for both removal processes were found to be highly significant ($p < 0.0001$), demonstrating the robustness of the model in describing the system behavior. The F-value, determined as the ratio of the model mean square to the residual error mean square, provides further validation of model significance. The quadratic model proposed by the RSM was employed to evaluate the experimental data. For both Congo Red (CR) and Methylene Blue (MB) removal processes, the exceptionally high F-values of 16.23 and 27.99, respectively, coupled with highly significant p-values ($p < 0.0001$), collectively demonstrate that the quadratic model optimally describes the dye adsorption behavior by the adsorbent, EN-NC-CMC-Na hydrogel. The quadratic regression models derived for CR and MB removal processes using adsorbent, EN-NC-CMC-Na hydrogel, were expressed in terms of coded factors as follows:

CR removal (%) = $+84.40 - 1.23*A + 1.75*B + 2.29*C + 0.19*D + 1.73*AB - 0.69*AC + 0.22*AD + 1.07*BC + 0.63*BD + 0.10*CD + 0.22*A^2 - 0.54*B^2 - 0.67*C^2 - 1.94*D^2$

MB removal (%)= $+83.55 - 0.87*A + 2.70*B + 1.69*C + 1.36*D + 0.042*AB - 0.45*AC - 1.32*AD - 0.088*BC - 2.30*BD + 0.38*CD + 2.02*A^2 + 1.43*B^2 + 0.93*C^2 - 0.69*D^2$

In the developed model, terms exhibiting positive coefficients signify synergistic interactions between variables, while negative coefficients denote antagonistic effects on the response. The non-significant lack-of-fit ($F = 2.86$ for CR;

2.18 for MB) relative to pure error confirms the adequacy of the quadratic model, consistent with repeated center-point measurements. For model validation, the adjusted R^2 (0.88) and predicted R^2 (0.75) values for CR removal demonstrated good agreement, differing by less than 20%. Similarly, the MB removal process showed comparable consistency, with adjusted R^2 (0.94) and predicted R^2 (0.82) values maintaining the recommended <20% difference threshold.

Table 5.6 ANOVA and regression analysis of dye removal efficiency.

ANOVA	CR	MB
Model		
F-value	16.23	27.99
p-value	<0.0001	<0.0001
Significance	S	S
Lack of fit		
F-value	2.86	2.18
p-value	0.123	0.012
Significance	NS	NS
Coefficient of variation	1.74 %	1.44
R^2	0.94	0.97
Adjusted R^2	0.88	0.94
Predicted R^2	0.75	0.82
Adeq. precision	21.20	37.60

S – significant; NS – non-significant.

5.3.6.4 Validation of the model

Application of the desirability function yielded optimal removal efficiencies of 90.3% for Congo Red (CR) and 96.0% for Methylene Blue (MB) under specified conditions. For CR, maximum removal was achieved at 2.94 h contact time, 26.7 mg/L initial concentration, 0.45 g adsorbent (EN-NC-CMC-Na hydrogel) dose and pH 4.8. MB removal was optimized at 3.49 h, 44.95 mg/L concentration, 0.23 g adsorbent dose, and pH 6.2. Experimental validation under these optimized conditions demonstrated average removal efficiencies of 88.1% (CR) and 93.2% (MB), closely matching the

predicted values (CR: 90.3%; MB: 96.0%). This close agreement between experimental and predicted results confirmed the model's validity and predictive accuracy for dye removal optimization.

5.3.7 Adsorption interaction analysis of dyes with EN-NC-CMC-Na hydrogel

Isotherm studies characterizes the dye adsorption capacity and surface properties of EN-NC-CMC-Na hydrogel at equilibrium, while kinetic analysis identifies the rate-determining processes. The integration of these approaches (examined through Langmuir, Freundlich, and pseudo-order models) provides fundamental insights for optimizing the dye removal efficiency of hydrogel and system scalability.

5.3.7.1 Isotherm analysis

The adsorptive removal of Congo Red (CR) and Methylene Blue (MB) dyes by EN-NC-CMC-Na hydrogel was examined through isotherm modeling. The Langmuir model, which presumes uniform monolayer adsorption with equivalent energy across all binding sites, was applied. In contrast, the Freundlich model suggested multilayer adsorption for CR and MB. The Temkin isotherm provided insights into the influence of adsorbate-adsorbent interactions on adsorption enthalpy. The Sips isotherm, integrating aspects of both Langmuir and Freundlich models, was utilized to describe adsorption behavior in heterogeneous systems. Non-linear isotherm plots for CR and MB are presented in **Fig. 5.12**. The adsorption capacity (q_e) exhibited a positive correlation with increasing initial dye concentration, indicative of progressive occupancy of available binding sites on the adsorbent surface. The removal efficiency of CR dye exhibits a concentration-dependent decline, decreasing from 87.2% at 25 mg/L to 51.8% at 60 mg/L. A comparable trend is observed for MB, with its removal

efficiency diminishing from 93.3% at 25 mg/L to 67.8% at 60 mg/L. The EN-NC-CMC-Na hydrogel showed improved adsorption capacitiy (q_e) for CR and MB dyes as the initial dye concentration increased from 25 to 60 mg/L. For CR, q_e increased from 21.8 mg/g at 25 mg/mL to a peak of 35.6 mg/g at 55 mg/mL before declining to 31.1 mg/g at 60 mg/mL. Similarly, for MB, q_e rose from 23.3 mg/g at 25 mg/mL to a maximum of 40.7 mg/g at 60 mg/mL. This concentration-dependent behavior confirms the effective utilization of adsorption sites by dye molecules. At low concentrations, high removal efficiency is achieved as abundant adsorption sites facilitate dye uptake. However, increasing dye concentration leads to adsorption sites saturation, reducing the dye removal efficiency. Meanwhile, q_e increases because more dye is present and the adsorbent retains more dye mass per gram, until a saturation point is reached.

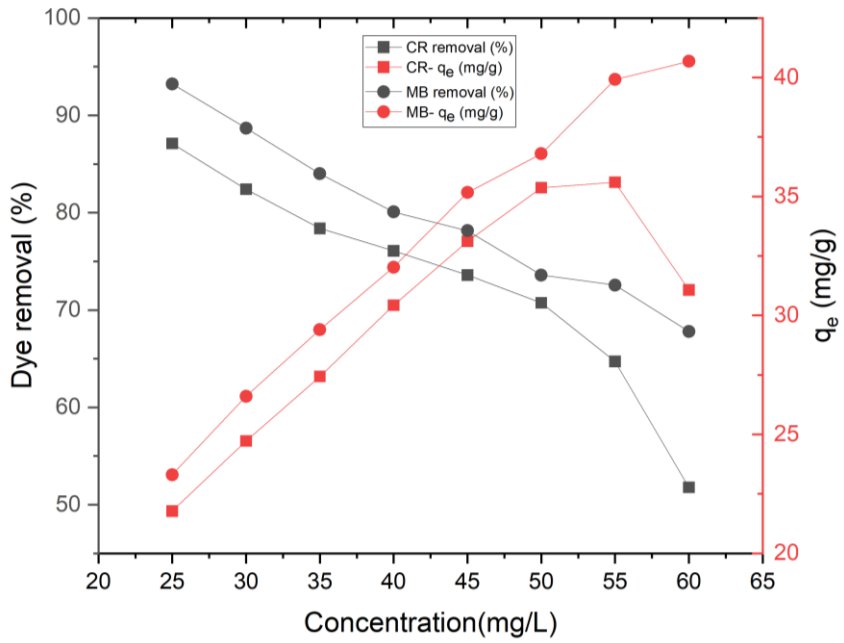


Fig. 5.12 Influence of initial dye concentration on dye removal efficiency and adsorption capacity of adsorbent EN-NC-CMC-Na hydrogel. Dye removal (%) and q_e were measured, keeping the adsorbent dose at 0.1 g and time 3 h for both CR and MB.

Table 5.7 summarizes the regression alignment with experimental data compared with the different models used for adsorption of dyes on EN-NC-CMC-Na hydrogel surface. The appropriateness of the model was assessed based on the coefficient of determination (R^2) and the chi-square (χ^2) values.

Table 5.7 Adsorption isotherm models and their respective parameters.

Adsorption isotherm											
Langmuir			**Freundlich**			**Temkin**			**Sips**		
	CR	**MB**		**CR**	**MB**		**CR**	**MB**		**CR**	**MB**
q_m (mg /g)	32.64	40.72	**1/n**	0.32	0.22	**B**	0.10	0.12	**K_s**	0.594	38.58
K_L (L/mg)	0.41	0.53	**K_F (mg/ g)**	0.33	0.35	**K_T (L/mg)**	15.19	10.89	**n**	0.719	4.01
R^2	0.81	0.88	**R^2**	0.68	0.99	**R^2**	0.71	0.96	**R^2**	0.82	0.98
X^2	0.016	0.029	**X^2**	0.029	0.002	**X^2**	0.025	0.005	**X^2**	0.015	0.002

*The Langmuir isotherm is defined by q_m (maximum capacity, mg/g) and K_L (constant, L/mg), while the Freundlich model uses K_F (capacity constant) and n (heterogeneity factor). The Temkin isotherm involves B (heat of adsorption, J/mol), K_T (binding constant, L/g), and b (adsorption energy, mol^2/J^2). In Sips, K_s represent Sips equilibrium constant and n as the heterogeneity factor.The Polanyi potential (ε, kJ/mol) incorporates the gas constant (R = 8.314 J/mol/K) and temperature (T, K).

Analysis of R^2 and χ^2 of the fitting from **Table 5.7** demonstrated for CR adsorption near-Langmuir behavior, though slight site heterogeneity was evident (best described by the Sips model) Conversely, MB adsorption followed Freundlich behavior, suggesting a heterogeneous surface. The theoretical maximum adsorption capacities (q_m (maximum capacity, mg/g)) derived from the Langmuir model were 32.64 mg/g for CR and 40.72 mg/g for MB. These values are comparable to those reported for other nanocellulose-based adsorbents. For instance, unmodified cellulose nanocrystals (CNCs) showed a capacity of ~35 mg/g for MB, while CNF-reinforced alginate–PVA hydrogels and CMC/CNC/zeolite composites achieved MB adsorption capacities of 36.3 mg/g and 38.2 mg/g, respectively [40,41]. In contrast, significantly higher adsorption capacities were reported for more complex systems such as

chitosan/CMC/PEG porous hydrogels (1053.9 mg/g for CR and 331.7 mg/g for MB) and MXene–carbon foam aerogels (647.8 mg/g for CR and 356.0 mg/g for MB) [42,43]. While these advanced materials demonstrate superior adsorption performance, they typically require intricate synthesis routes, costly reagents, or non-biodegradable components. In contrast, the CMC–NC hydrogel developed in this study offers a simpler, greener, and more sustainable alternative, with sufficient adsorption performance for practical dye removal applications. Its performance may be further enhanced through surface modifications or incorporation of functional fillers. The Temkin model also provided a strong fit for MB, implying a linear decrease in adsorption energy with increasing surface coverage. Additionally, the Sips isotherm exhibited excellent agreement with experimental data, with the Sips constant (n) approaching unity, further supporting monolayer adsorption as described by the Langmuir model**.**

5.3.7.2 Kinetics analysis

The adsorption kinetics of CR and MB onto EN-NC-CMC-Na hydrogel was investigated using pseudo-first-order, pseudo-second-order, Elovich and intraparticle diffusion models. The pseudo-first-order model describes adsorption kinetics based on the adsorbed quantity, while the pseudo-second-order model assumes the rate-limiting step is governed by adsorption capacity. The non-linear kinetic plots for for CR and MB are presented in **Fig. 5.13**. It showed rapid initial dye removal kinetics followed by equilibrium attainment. The dye removal efficiency (%) for both dyes increases rapidly at first, then gradually levels off and finally gives plateau after100 min for CR (76.4%) and 120 min for MB (93.4%). Similarly, the adsorption capacities (q_t) for Congo Red (CR), reached equilibrium at 100 min, while Methylene Blue (MB) required 120 min.

The EN-NC-CMC-Na hydrogel exhibited maximum adsorption capacities of 45.8 mg/g for CR and 55.8 mg/g for MB. Both dye removal efficiency and q_t follow the same trend because they fundamentally reflect the same process and both metrics are directly related to the amount of dye adsorbed from solution.

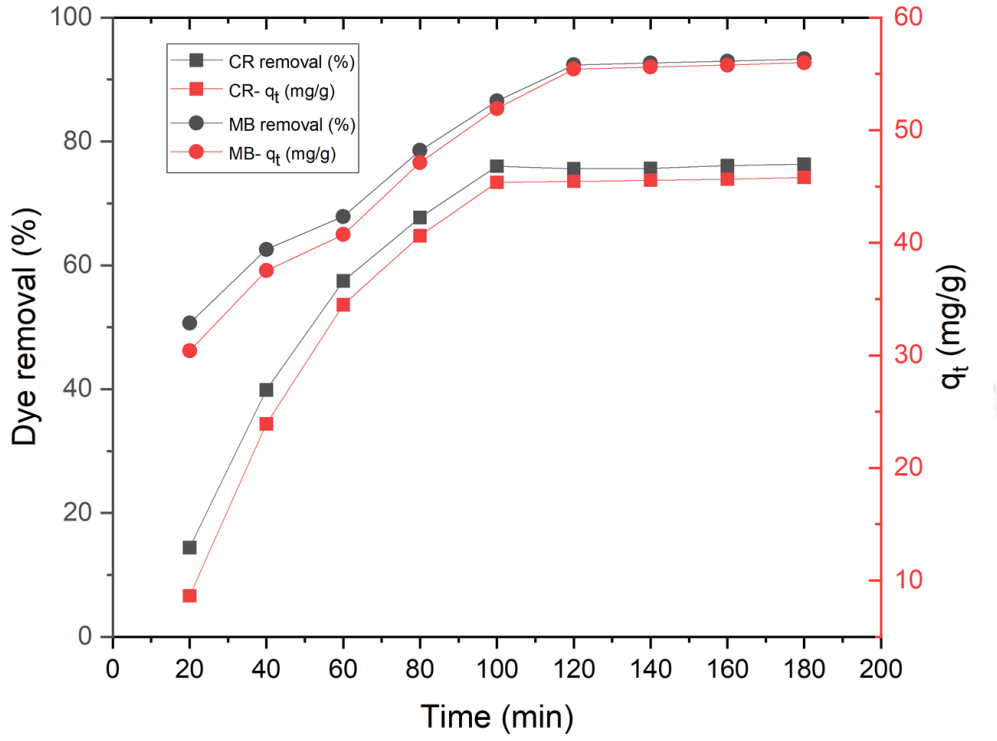


Fig. 5.13 Kinetic analysis of dye removal performance of EN-NC-CMC-Na hydrogel. Dye removal (%) and q_t were measured at varying time from 20 min to 180 min, keeping the adsorbent dose at 0.1 g and initial dye concentration at 60 mg/L.

Table 5.8 summarizes the kinetic parameters for different models. Both dyes exhibited higher correlation coefficients (R^2) for the pseudo-second-order model. This suggests adsorption was primarily controlled by the availability of adsorption sites on the EN-NC-CMC-Na hydrogel, consistent with chemisorption mechanisms involving electron sharing or exchange between the adsorbent and adsorbate. The close agreement between experimental (q_e) and theoretical (q_t) adsorption capacities further validates the robustness of the obtained results.

Table 5.8 Adsoption isotherm kinetics models and their respective parameters.

Adsorption kinetics										
Pseudo-first order			**Pseudo-second order**			**Temkin**			**Sips**		
	CR	**MB**		**CR**	**MB**		**CR**	**MB**		**CR**	**MB**
k_1 (1/min)	0.023	0.035	**k_2 (g/mg/min)**	0.0021	0.0045	**α (g/mg/min)**	12.45	18.72	**K_{id} (mg/g)**	4.82	7.35
$q_{e,cal}$ (mg/g)	48.15	55.62	**$q_{e,cal}$ (mg/g)**	46.92	54.34	**β (g/mg)**	0.085	0.062	**C**	6.15	12.42
$q_{e,\ exp}$ (mg/g)	47.13	51.16	**$q_{e,\ exp}$ (mg/g)**	44.54	52.61						
R^2	0.97	0.98	**R^2**	0.99	0.98	**R^2**	0.96	0.96	**R^2**	0.91	0.90
X^2	2.85	2.12	**X^2**	0.85	1.12	**X^2**	3.42	2.87	**X^2**	0.101	0.052

* k_1 (min^{-1}) and k_2 (g/mg/min) represent the rate constants for pseudo-first-order and pseudo-second-order kinetic models, respectively. The terms q_e and q_t denote the adsorption capacities (mg/g) at equilibrium and time t, respectively, for dye uptake by EN-NC-CMC-Na hydrogel beads. The initial adsorption rate (a, mg/g/min) and desorption constant (b, g/mg) characterize the adsorption dynamics, while k_{ia} (g mg^{-1} $min^{1/2}$) is the intraparticle diffusion rate constant and C reflects the boundary layer thickness effect.

5.3.8 Effect of various chemicals on hydrogel regeneration by desorption of dyes

The dye-laden EN-NC-CMC-Na hydrogel was subjected to desorption studies utilizing multiple chemical agents to evaluate their efficacy of the adsorbent recovery. This regeneration approach significantly reduces both material costs and resource requirements by enabling adsorbent reuse. **Fig. 5.14** shows desorption efficiency variation with various chemical agents. The highest desorption of CR from the hydrogel occurred with 1 M NaOH (79.6%) and 50% ethanol (82.3%), while 0.1 M HCl (93.6%) and 0.1 M HNO_3 (92.7%) proved most effective for MB desorption in 1 h. Notably, competing adsorbents exhibited reduced performance (<70% efficiency) after regeneration, likely due to saturation of the pores of the hydrogel matrix [44].

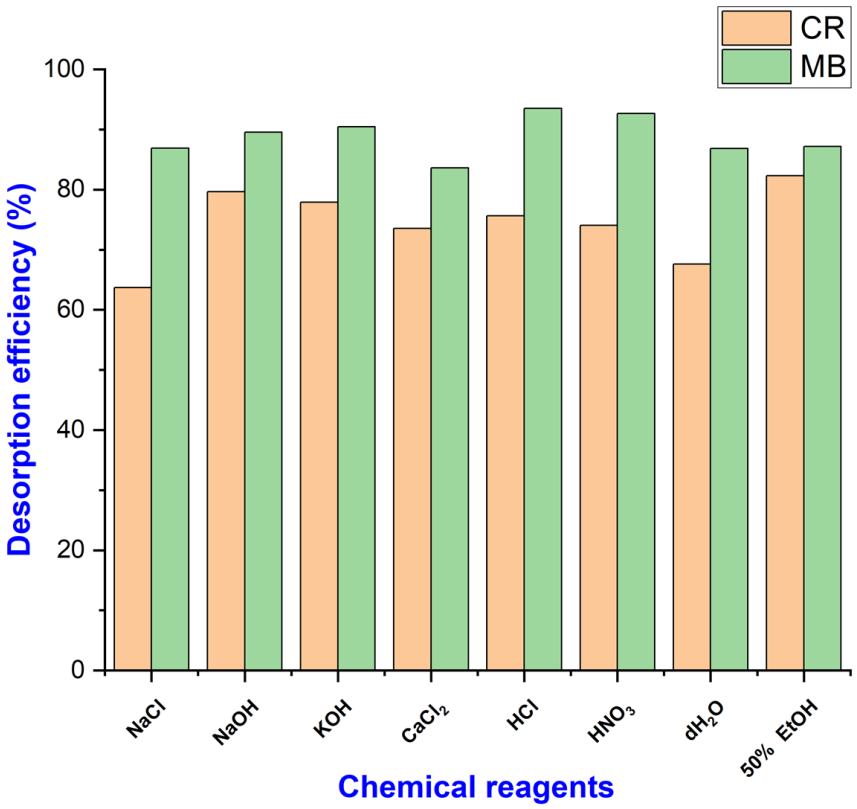


Fig. 5.13 Systematic investigation of chemical reagent effects on EN-NC-CMC-Na hydrogel regeneration capacity. Each trial used 0.1 g of adsorbent (EN-NC-CMC-Na) hydrogel with 10 mL of reagent solution for 1 h at 25°C.

The superior CR desorption in alkaline conditions (1 M NaOH) results from attenuated electrostatic interactions between the hydrogel and dye molecules. The effectiveness of 50% ethanol arises from its dual functionality: hydrophobic interactions with aromatic carbon structures and hydrogen bonding with surface functional groups (–OH, –COOH) on the EN-NC-CMC-Na hydrogel, which competitively inhibit dye adsorption [45]. For MB removal, 0.1 M HCl demonstrated exceptional regeneration capacity through proton-mediated disruption of ionic dye-adsorbent bonds. While both HCl and HNO_3 are strong acids, HCl monoprotic nature

ensures efficient proton delivery at low concentrations (0.1 M) without secondary oxidative effects that might compromise adsorbent integrity [46].

5.4 Conclusion

Enzymatic production of nanocellulose offers a sustainable alternative to conventional methods, which are energy-intensive and low-yield. The recombinant endoglucanase, *At*GH9C-CBM3A-CBM3B (97.2 kDa) and cellobiohydrolase, *At*CBH5A (62 kDa) were expressed and purified, showing specific activities of 58.8 ± 0.8 U/mg and 96.8 ± 1.2 U/mg, respectively, against 1.0% (w/v) CMC-Na. Cellulose was extracted from sugarcane trash (SCT) by using optimized alkali pretreatment method. The extracted cellulose was hydrolyzed by both purified enzymes (3.0 mg total enzyme/g.$_{\text{SCT cellulose}}$, 6 h) to produce enzymatic nanocellulose (EN-NC) and compared with chemical (TEMPO) treated SCT cellulose that gave TEMPO-nanocellulose (TO-NC). FTIR and XRD analyses of SCT cellulose, EN-NC and TO-NC showed similar spectra, but FESEM revealed morphological differences. SCT cellulose showed micron-scale structures (~10 μm), while EN-NC and TO-NC both showed nanoscale, 10–100 nm structures. However, EN-NC showed narrower fibrils (≤20 nm), reflecting enzymatic specificity. EN-NC-CMC-Na hydrogel was synthesized by dispersing EN-NC (0.5 g/50 mL) in 2% acetic acid and added to a solution of CMC-Na (1.0 g/50 mL) in deionized water at pH 6.0 and incubating at 35°C. Swelling analysis of EN-NC-CMC-Na hydrogel showed hydration equilibrium at 6 h. The hydrogel showed a pHpzc of 6.6 and surface area of 4.81 m^2/g. The removal of dyes by EN-NC-CMC-Na hydrogel optimized by RSM showed maximum Congo Red (CR) removal (88.5%) 48.75 mg/L dye, at pH 7.5 and 0.47 g hydrogel dosage in 4.75 h. Methylene Blue (MB) removal was 94.7% (37.5 mg/L dye) at pH 6.0 and 0.6 g hydrogel dose in 3.5 h. CR adsorption isotherm fitted both Freundlich and Langmuir models, whereas MB adsorption followed the Freundlich model (multilayer

adsorption). Kinetic analysis indicated pseudo-second-order behavior, suggesting chemisorption *via* hydrogen bonding, electrostatic and n–π interactions. Hydrogels showed excellent regeneration using 0.1 M NaOH with 50% ethanol for CR and 0.1 M HCl/HNO_3 for MB. These results highlighted the EN-NC-CMC-Na hydrogel as a promising, eco-friendly tertiary treatment for dye-laden industrial effluents.

.

5.5 References

[1] R. Jamee, R. Siddique, Biodegradation of synthetic dyes of textile effluent by microorganisms: An environmentally and economically sustainable approach, Eur J Microbiol Immunol (Bp) 9 (2019) 114–118. https://doi.org/10.1556/1886.2019.00018

[2] B. Lellis, C.Z. Fávaro-Polonio, J.A. Pamphile, J.C. Polonio, Effects of textile dyes on health and the environment and bioremediation potential of living organisms, Biotechnology Research and Innovation 3 (2019) 275–290. https://doi.org/10.1016/J.BIORI.2019.09.001

[3] P.O. Oladoye, M.O. Bamigboye, O.D. Ogunbiyi, M.T. Akano, Toxicity and decontamination strategies of Congo red dye, Groundw Sustain Dev 19 (2022) 100844. https://doi.org/10.1016/J.GSD.2022.100844

[4] P.O. Oladoye, T.O. Ajiboye, E.O. Omotola, O.J. Oyewola, Methylene blue dye: Toxicity and potential elimination technology from wastewater, Results in Engineering 16 (2022) 100678. https://doi.org/10.1016/J.RINENG.2022.100678

[5] H. Ben Slama, A.C. Bouket, Z. Pourhassan, F.N. Alenezi, A. Silini, H. Cherif-Silini, T. Oszako, L. Luptakova, P. Golińska, L. Belbahri, Diversity of Synthetic Dyes from Textile Industries, Discharge Impacts and Treatment Methods, Applied Sciences 2021, Vol. 11, Page 6255 11 (2021) 6255. https://doi.org/10.3390/APP11146255

[6] D. Klemm, E.D. Cranston, D. Fischer, M. Gama, S.A. Kedzior, D. Kralisch, F. Kramer, T. Kondo, T. Lindström, S. Nietzsche, K. Petzold-Welcke, F. Rauchfuß, Nanocellulose as a natural source for groundbreaking applications in materials science: Today's state, Materials Today 21 (2018) 720–748. https://doi.org/10.1016/J.MATTOD.2018.02.001

[7] Y. Du, G. Feng, When nanocellulose meets hydrogels: the exciting story of nanocellulose hydrogels taking flight, Green Chemistry 25 (2023) 8349–8384. https://doi.org/10.1039/D3GC01829F

[8] A. Blanco, M.C. Monte, C. Campano, A. Balea, N. Merayo, C. Negro, Nanocellulose for Industrial Use: Cellulose Nanofibers (CNF), Cellulose Nanocrystals (CNC), and Bacterial Cellulose (BC), Handbook of Nanomaterials for Industrial Applications (2018) 74–126. https://doi.org/10.1016/B978-0-12-813351-4.00005-5

[9] P. Thomas, T. Duolikun, N.P. Rumjit, S. Moosavi, C.W. Lai, M.R. Bin Johan, L.B. Fen, Comprehensive review on nanocellulose: Recent developments, challenges and future prospects, J Mech Behav Biomed Mater 110 (2020) 103884. https://doi.org/10.1016/J.JMBBM.2020.103884

[10] Y. Habibi, L.A. Lucia, O.J. Rojas, Cellulose nanocrystals: Chemistry, self-assembly, and applications, Chem Rev 110 (2010) 3479–3500. https://doi.org/10.1021/cr900339w

[11] T.C. Mokhena, M.J. John, Cellulose nanomaterials: new generation materials for solving global issues, Cellulose 2019 27:3 27 (2019) 1149–1194. https://doi.org/10.1007/S10570-019-02889-W

[12] K.C. Khaire, V.S. Moholkar, A. Goyal, Bioconversion of sugarcane tops to bioethanol and other value added products: An overview, Mater Sci Energy Technol 4 (2021) 54–68. https://doi.org/10.1016/J.MSET.2020.12.004

[13] B. Ram, K. Ramaiyan, Current status of sugarcane agriculture and sugar industry in the World with special reference to India, in: 2021: pp. 1–21

[14] S. R, S. CM, Integrated Sugarcane Trash Management: A Novel Technology for Sustaining Soil Health and Sugarcane Yield, Advances in Crop Science and Technology 3 (2015) 1–4. https://doi.org/10.4172/2329-8863.1000160

[15] S. Quereshi, T.K. Naiya, A. Mandal, S. Dutta, Residual sugarcane bagasse conversion in India: current status, technologies, and policies, Biomass Conversion and Biorefinery 2020 12:9 12 (2020) 3687–3709. https://doi.org/10.1007/S13399-020-00871-2

[16] N.K. Bhardwaj, D. Kaur, S. Chaudhry, M. Sharma, S. Arya, Approaches for converting sugarcane trash, a promising agro residue, into pulp and paper using soda pulping and elemental chlorine-free bleaching, J Clean Prod 217 (2019) 225–233. https://doi.org/10.1016/J.JCLEPRO.2019.01.223

[17] O. Nechyporchuk, M.N. Belgacem, J. Bras, Production of cellulose nanofibrils: A review of recent advances, Ind Crops Prod 93 (2016) 2–25. https://doi.org/10.1016/J.INDCROP.2016.02.016

[18] L. Solhi, V. Guccini, K. Heise, I. Solala, E. Niinivaara, W. Xu, K. Mihhels, M. Kröger, Z. Meng, J. Wohlert, H. Tao, E.D. Cranston, E. Kontturi, Understanding Nanocellulose–Water Interactions: Turning a Detriment into an Asset, Chem Rev 123 (2023) 1925. https://doi.org/10.1021/ACS.CHEMREV.2C00611

[19] A. Errokh, A. Magnin, J.L. Putaux, S. Boufi, Morphology of the nanocellulose produced by periodate oxidation and reductive treatment of cellulose fibers, Cellulose 25 (2018) 3899–3911. https://doi.org/10.1007/s10570-018-1871-7

[20] A. Tejado, M.N. Alam, M. Antal, H. Yang, T.G.M. van de Ven, Energy requirements for the disintegration of cellulose fibers into cellulose nanofibers, Cellulose 19 (2012) 831–842. https://doi.org/10.1007/s10570-012-9694-4

[21] S. Koskela, S. Wang, D. Xu, X. Yang, K. Li, L.A. Berglund, L.S. McKee, V. Bulone, Q. Zhou, Lytic polysaccharide monooxygenase (LPMO) mediated production of ultra-fine cellulose nanofibres from delignified softwood fibres, Green Chemistry 21 (2019) 5924–5933. https://doi.org/10.1039/C9GC02808K

[22] F. Rol, M.N. Belgacem, A. Gandini, J. Bras, Recent advances in surface-modified cellulose nanofibrils, Prog Polym Sci 88 (2019) 241–264. https://doi.org/10.1016/J.PROGPOLYMSCI.2018.09.002.

[23] R.S.A. Ribeiro, B.C. Pohlmann, V. Calado, N. Bojorge, N. Pereira, Production of nanocellulose by enzymatic hydrolysis: Trends and challenges, Eng Life Sci 19 (2019) 279–291. https://doi.org/10.1002/elsc.201800158

[24] K. Gourlay, T. van der Zwan, M. Shourav, J. Saddler, The potential of endoglucanases to rapidly and specifically enhance the rheological properties of micro/nanofibrillated cellulose, Cellulose 25 (2018) 977–986. https://doi.org/10.1007/s10570-017-1637-7

[25] X.Q. Chen, X.Y. Deng, W.H. Shen, M.Y. Jia, Preparation and characterization of the spherical nanosized cellulose by the enzymatic hydrolysis of pulp fibers, Carbohydr Polym 181 (2018) 879–884. https://doi.org/ 10.1016/j.carbpol.2017.11.064

[26] J. Hu, D. Tian, S. Renneckar, J.N. Saddler, Enzyme mediated nanofibrillation of cellulose by the synergistic actions of an endoglucanase, lytic polysaccharide monooxygenase (LPMO) and xylanase, Scientific Reports 2018 8:1 8 (2018) 1–8. https://doi.org/10.1038/s41598-018-21016-6

[27] A. de Campos, A.C. Correa, D. Cannella, E. de M Teixeira, J.M. Marconcini, A. Dufresne, L.H.C. Mattoso, P. Cassland, A.R. Sanadi, Obtaining nanofibers from curauá and sugarcane bagasse fibers using enzymatic hydrolysis followed by sonication, Cellulose 20 (2013) 1491–1500. https://doi.org/10.1007/s10570-013-9909-3

[28] K. Saelee, N. Yingkamhaeng, T. Nimchua, P. Sukyai, An environmentally friendly xylanase-assisted pretreatment for cellulose nanofibrils isolation from sugarcane bagasse by high-pressure homogenization, Ind Crops Prod 82 (2016) 149–160. https://doi.org/10.1016/J.INDCROP.2015.11.064

[29] B.R. Rossi, V.O.A. Pellegrini, A.A. Cortez, E.M.S. Chiromito, A.J.F. Carvalho, L.O. Pinto, C.A. Rezende, V.R. Mastelaro, I. Polikarpov, Cellulose nanofibers production using a set of recombinant enzymes, Carbohydr Polym 256 (2021) 117510. https://doi.org/10.1016/J.CARBPOL.2020.117510

[30] A. Mandal, A. Thakur, A. Goyal, Role of carbohydrate binding modules, CBM3A and CBM3B in stability and catalysis by a β-1,4 endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus* ATCC 27405, Int J Biol Macromol 242 (2023). https://doi.org/10.1016/J.IJBIOMAC.2023.125164

[31] P.D. Maibam, A. Goyal, Designing of recombinant hydrolytic enzymes cocktail for effective saccharification of delignified rice straw, Ind Crops Prod 206 (2023) 117727. https://doi.org/10.1016/J.INDCROP.2023.117727

[32] K.C. Khaire, V.S. Moholkar, A. Goyal, Separation and characterization of cellulose from sugarcane tops and its saccharification by recombinant cellulolytic enzymes, Prep

Biochem Biotechnol 51 (2021) 811–820. https://doi.org/10.1080/10826068.2020.1861011

[33] N. Nelson, A photometric adaptation of the somogyi method for the determination of glucose, Journal of Biological Chemistry 153 (1944) 375–380. https://doi.org/10.1016/S0021-9258(18)71980-7

[34] M. Somogyi, Determination of blood sugar, Journal of Biological Chemistry 160 (1945) 69–73. https://doi.org/10.1016/S0021-9258(18)43098-0

[35] H. Fukuzumi, T. Saito, Y. Okita, A. Isogai, Thermal stabilization of TEMPO-oxidized cellulose, Polym Degrad Stab 95 (2010) 1502–1508. https://doi.org/10.1016/J.POLYMDEGRADSTAB.2010.06.015

[36] L. Segal, J.J. Creely, A.E. Martin, C.M. Conrad, An Empirical Method for Estimating the Degree of Crystallinity of Native Cellulose Using the X-Ray Diffractometer, Textile Research Journal 29 (1959) 786–794. https://doi.org/10.1177/004051755902901003

[37] C. Poornachandhra, R.M. Jayabalakrishnan, M. Prasanthrajan, G. Balasubramanian, A. Lakshmanan, S. Selvakumar, J.E. John, Cellulose-based hydrogel for adsorptive removal of cationic dyes from aqueous solution: isotherms and kinetics, RSC Adv 13 (2023) 4757–4774. https://doi.org/10.1039/D2RA08283G

[38] D. Allouss, Y. Essamlali, O. Amadine, A. Chakir, M. Zahouily, Response surface methodology for optimization of methylene blue adsorption onto carboxymethyl cellulose-based hydrogel beads: adsorption kinetics, isotherm, thermodynamics and reusability studies, RSC Adv 9 (2019) 37858–37869. https://doi.org/10.1039/C9RA06450H

[39] H. Karyab, Z. Hamidi, F. Ghotbinia, Z. Mosa Khani, N. Nazeri, Application of the central composite design and response surface methodology for optimization of reactive color removal from aqueous solutions using dicyandiamide-formaldehyde resin modified by ammonium chloride, Desalination Water Treat 308 (2023) 217–228. https://doi.org/10.5004/DWT.2023.29829

[40] Y. Zhou, L. Zhang, J. Zhang, Preparation of nanocellulose-based hydrogel for efficient adsorption of methylene blue from aqueous solution, Int J Biol Macromol 183 (2021) 1545–1553. https://doi.org/10.1016/j.ijbiomac.2021.05.056

[41] R. Singh, S. Tiwari, P.K. Mishra, Composite hydrogel of CMC/CNC/zeolite for efficient adsorption of dyes from aqueous solution, J Environ Chem Eng 8 (2020) 104230. https://doi.org/10.1016/j.jece.2020.104230

[42] J. Wang, C. Li, Y. Gao, Y. Liu, Superadsorbent hydrogel based on chitosan/CMC/PEG for efficient removal of anionic and cationic dyes, Carbohydr Polym 275 (2022) 118692. https://doi.org/10.1016/j.carbpol.2021.118692

[43] M. Hu, F. Li, C. Wang, Y. Wang, MXene–carbon foam aerogel composite for high-performance dye adsorption, Chemical Engineering Journal 384 (2020) 123282. https://doi.org/10.1016/j.cej.2019.123282

[44] Y. Yang, Q. Zhu, X. Peng, J. Sun, C. Li, X. Zhang, H. Zhang, J. Chen, X. Zhou, H. Zeng, Y. Zhang, Hydrogels for the removal of the methylene blue dye from wastewater: a review, Environmental Chemistry Letters 2022 20:4 20 (2022) 2665–2685. https://doi.org/10.1007/S10311-022-01414-Z

[45] X. Xing, H. Qu, R. Shao, Q. Wang, H. Xie, Mechanism and kinetics of dye desorption from dye-loaded carbon (XC-72) with alcohol-water system as desorbent, Water Science and Technology 76 (2017) 1243–1250. https://doi.org/10.2166/WST.2017.268

[46] H. Patel, Review on solvent desorption study from exhausted adsorbent, Journal of Saudi Chemical Society 25 (2021) 101302. https://doi.org/10.1016/J.JSCS.2021.101302

List of articles published/submitted from Ph.D. Thesis Work

Research articles

1. **Mandal, A.,** Thakur, A., & Goyal, A. (2023). Role of carbohydrate binding modules, CBM3A and CBM3B in stability and catalysis by a β-1,4 endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus* ATCC 27405. **International Journal of Biological Macromolecules**, 242, 125164. https://doi.org/10.1016/j.ijbiomac.2023.125164 (JIF 7.7).
2. **Mandal, A**., Ahmed, J., Singh, S., & Goyal, A. (2024). Structure elucidation of a multi-modular recombinant endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus* ATCC 27405 and its substrate binding analysis. **International Journal of Biological Macromolecules**, 273, 133212. https://doi.org/10.1016/J.IJBIOMAC.2024.133212 (JIF 7.7).
3. **Mandal, A.,** Goyal, A. (2025). Synthesis of nanocellulose from sugarcane trash using recombinant cellulases and its application in efficient dye removal through nanocellulose-based hydrogel. C**ellulose**, (Under review).

List of research articles in collaboration with others

1. Sherya Biswas, **Ardhendu Mandal**, Carlos M.G.A. Fontes and Arun Goyal, 2025. A unique highly efficient, thermostable and multi-substrate specific galactanase (*At*GH53) from *Acetivibrio thermocellus* cleaving both β (1,4)- and β (1,6)- linked galactans. **Current Research in Biotechnology,** (Submitted). (JIF 3.6)

List of Conferences

1. **Ardhendu Mandal**, Jebin Ahmed, Abhijeet Thakur and Arun Goyal (2021). Cloning, expression, biochemical and in-silico characterization of an endoglucanase *Ct*GH9C from *Clostridium thermocellum* ATCC 27405. **International Conference on Biotechnology for Resource Efficiency, Energy, Environment, Chemical and Health** 2021, Dec 01-04, 2021, BREEECH 2021 Dehradun.
2. **Ardhendu Mandal,** Jebin Ahmed and Arun Goyal (2022) Structure elucidation of recombinant endoglucanase (*Ct*GH9C-CBM3A-CBM3B) from Clostridium thermocellum ATCC 27405 and its substrate binding analysis. **Biotechnology for Sustainable Bioresources and Bioeconomy (BSBB)**, IIT Guwahati, Assam, India. 7-11 December, 2022.
3. **Ardhendu Mandal**, Abhijeet Thakur and Arun Goyal (2023) Impact of associated carbohydrate binding modules, CBM3A and CBM3B in stability and catalysis by an β-1,4 endoglucanase, *At*GH9C-CBM3A-CBM3B from *Acetivibrio thermocellus* ATCC 27405. **Research and Industrial Conclave, An amalgamation of Academia, Industry & Strat-ups**, May 14-16, 2023, IIT Guwahati, Guwahati, Assam.
4. **Ardhendu Mandal,** Jebin Ahmed and Arun Goyal (2023) Structure elucidation of recombinant endoglucanase (*At*GH9C-CBM3A-CBM3B) from *Acetivibrio thermocellus* ATCC 27405 and its substrate binding analysis at **92nd Annual meeting of The Society of Biological Chemists (India): Biological Chemistry, Opportunities and Way forward (SBC 2023)** held from 18 Dec to 20 Dec, 2023 at BITS-Goa, India (Best poster presentation awarded).
5. **Ardhendu Mandal** and Arun Goyal (2024) Synthesis of nanocellulose from biomass and paper pulp using a synergistic action of recombinant cellulases. **Research and Industrial Conclave, An amalgamation of Academia, Industry & Strat-ups**, August 09-11, 2024, IIT Guwahati, Guwahati, Assam.
6. **Ardhendu Mandal** and Arun Goyal (2024) Synthesis of nanocellulose from biomass cellulose using a synergistic action of recombinant cellulases and its application as hydrogel. **International Conference on Environmental Challenges, Opportunities and Sustainable Solutions, 9-11 December, 2024,** IIT Guwahati, Guwahati, Assam.

VITAE

The author was born on November 14, 1994 in the small village of Kolsur (West Bengal, India). He passed Secondary Examination (10th Class) conducted by Central Board of Secondary Examination, New Delhi in 2011 and Higher Secondary Examination (12th Class) conducted by Central Board of Secondary Examination, New Delhi in 2013. He completed BS-MS dual degree (Biological Science) from Indian Institute of Science Education and Research Kolkata, West Bengal, India in July, 2019 with a Second-Best Thesis award for final year MS project presentation.

Mr. Ardhendu Mandal joined the PhD program in July, 2020 at Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India. He successfully completed the coursework with 9.14/10 CPI. He received Institute Fellowship (IIT Guwahati) from Sept 2020 to August 2025, under the scheme run by the Ministry of Education, New Delhi. He delivered the open (PhD Synopsis) Seminar on 12th June 2025 and presented his thesis work before the Doctoral Committee and his performance was satisfactory. He submitted the PhD thesis in July 2025.