Protein engineering of β-1,4-endoglucanase and Chimera construction with β-glucosidase from *Clostridium thermocellum* for improving ligno-cellulosic biomass saccharification

PhD Thesis

by

Priyanka Nath

November 2019

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI – 781039, ASSAM, INDIA

Protein engineering of β-1,4-endoglucanase and Chimera construction with β-glucosidase from *Clostridium thermocellum* for improving ligno-cellulosic biomass saccharification

A Thesis

Submitted in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy

by

Priyanka Nath

Under supervision of

Professor Arun Goyal

November 2019

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

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

STATEMENT

I do hereby declare that the content embodied in this thesis entitled as **"Protein engineering of β-1,4-endoglucanase and Chimera construction with β-glucosidase from *Clostridium thermocellum* for improving lingo-cellulosic biomass saccharification"** 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.

November, 2019 *Priyanka Nath*

(146106042)

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES & BIOENGINEERING

CERTIFICATE

It is certified that the work described in this thesis entitled **"Protein engineering of β-1,4-endoglucanase and Chimera construction with β-glucosidase from *Clostridium thermocellum* for improving lingo-cellulosic biomass saccharification"** by **Priyanka Nath (Roll No. 146106042)** for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision at the Department of Biosciences & Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India and this work has not been submitted elsewhere for a degree.

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

ACKNOWLEDGEMENTS

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

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

I would also like to express my sincere gratitude to all my doctoral committee members Dr. Ranjan Tamuli, Dr. Manish Kumar and Dr. Sachin Kumar for their valuable suggestions and constructive criticism that has led to the successful completion of my thesis.

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

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

I am also thankful to my seniors Dr. Arun Dhillon, Dr. Aruna Rani and Dr. Rwivoo Baruah for their help and suggestions. I am immensely thankful to my research group members Vikky, Sumitha, Kedar, Shweta, Krishan, Abhijeet, Kaustubh, Dishant, Parmeshwar, and Maibam. I am grateful to all the people with whom I have worked in the lab at the Department of Biosciences and Bioengineering for their cooperation and support. I am grateful to my parents (Mr. Subhas Chandra Nath and Mrs Rina Nath), my Sister Mampi Nath and my husband Debjyoti Nath for their immense love, support and care.

I wish to acknowledge the support received from other teaching and non-teaching staff of the Department of Biosciences and Bioengineering, IIT Guwahati. I duly acknowledge the financial support of DST Inspire fellowship from Department of Science and Technology, India.

My PhD endeavor would not have been successful without the love, trust, support and blessings of my parents. I owe my achievements to my family and friends.

Priyanka Nath
November, 2019

Synopsis

Introduction

The rise in the human population on earth has been tremendous in the last 70 years. The substantial increase in population from 2.7 billion in 1955 to 7.6 billion in 2018 has created major burden over non-renewable resources and their utilization. Similarly, the urban population in 2018 was 4.2 billion which was only 1.0 billion in 1955 (www.worldometers.info). This expansion of urbanization has led to increase in energy consumption upto 200%, including non-renewable and renewable resources from 1990 to 2016 (Global Energy Statistical Yearbook, 2017). Petroleum and Oil are still serving as the world's main supply of fuel. The petroleum and oil shares 22.9% in world energy consumption whereas, coal provides around 40% of the world's electricity (World Energy Report, 2016). Petroleum and Oil is presently serving as the main energy source, but it is being exhausted rapidly from the world (Ragauskas *et al.*, 2006; Tong *et al.*, 2012). On contrary to the oil and petroleum, renewable sources of energy have a good future to support our ever-demanding civilization. Towards the early 19th century biomass was known to be used as an energy source. But nowadays energy generated from biomass shares only a small part in the total energy supply of the world (REN21, 2006). Currently among the total energy consumed, 18% is from renewable sources for example forest biomass (mostly, around 90% from woody biomass). This is the primary source of energy supply, that is generally utilized in

smelting of iron, other metal ores and cooking across the world (WER, 2016). Biodiesel and Ethanol production from biomass is another growing area. The production bioethanol and biodiesel is predicted to grow by 2024 to an amount of approximately 134.5 and 39 billion litres respectively (REN21,2006). However, the increase in energy demand due to several developments for example urbanization, power driven technology and transportation etc. were responsible for depletion in the non-renewable reserve crude oil. Estimated data suggested that by 2025, there will be approximately 50% increase in the energy demand from several developed and developing nations (Tong *et al.,* 2012). Producing bioenergy by utilizing biomass through cost-effective and sustainable approach is vital alternative of non-renewable energy resource.

Biofuels, like bioethanol and biodiesel, are derived from optimized conversion of lignocellulosic biomass. Bioethanol is produced by fermentation from carbohydrate of lingocellulosic biomass while biodiesel is produced from any fatty acids from vegetable or animal origin using transesterification reaction (Claassen *et al.,* 1999; Meher *et al.,* 2006; Galbe and Zacchi, 2007; Bhatia *et al.,* 2012; Saravanakumar *et al.,* 2013; Aulitto *et al.,* 2019). Several agricultural waste biomass having cellulosic component can be used for production of bioethanol. Several plant components like corn stover, sugarcane, grasses, and sorghum etc. are used around the world as lingocellulosic biomass for production of bioethanol. This biomass were further processed for breakdown of their cell wall using different pretreatment technologies. The pretreated biomass are then enzymatically hydrolysed into fermentable sugars (Lin and Tanaka, 2006 & Kazi *et al.,* 2010). The efficient conversion and usage of

lignocellulosic waste from plants to bioethanol can lead to clean energy production and reduce burden on the non-renewable energy sources. Bioethanol is an environmentally friendly product, Bioethanol releases less carbon monoxide (around 12%) and is mainly produced from sugars, cellulose and starch (Priya *et al*., 2010; Zainab and Fakhra, 2014). The lesser emission of carbon monoxide will prevent the greenhouse gases that were important for maintenance of the climate and global temperature of earth (Priya *et al.*, 2010; Zainab and Fakhra, 2014). The cellulosic biomass from agricultural and industrial wastes are abundant in nature which can be used for the production of biofuel. The prime components of the plant lignocellulosic biomass are made up of cellulose, hemicellulose and lignin. The cellulose and hemicellulose can be hydrolysed to monomeric sugars. These monomeric sugars can be utilized by the yeast for production of bioethanol (Zhang *et al.*, 2004). For complete degradation of this complex polymer *i.e* cellulose and hemicellulose requires multiple, related enzyme activities acting synergistically. Further, the degradation of cellulose polymer itself is an elaborate multistep process and the last step of this multistep process is the action of β-glucosidase on the cellobiose for converting it into glucose (Lynd *et al*., 2002). Cellobiose being an inhibitor of both endoglucanase and cellobiohydrolase can be acted upon by β-glucosidase enzyme to reduce its effect. Similarly glucose at high concentration can block the active site of β-glucosidase for the substrate or prevent the hydrolyzed substrate from leaving (Krogh *et al.*, 2010). To address this problem glucose tolerance were also incorporated into these enzymes by protein engineering technologies (Riou *et al.*, 1998, Gunata and Vallier, 1999). In the present situation, commercially available cellulases for biomass hydrolysis contain β-

glucosidase enzyme in the cocktail of enzymes in significant amount (Bidard-Michelot et al., 2019). This commercial enzyme claims to have improved β-glucosidase activity with higher titers, but the cost of these enzymes is very high. From this, it can be assumed that engineering the enzyme for improving the β-glucosidase activity and its production can overcome the limiting factors for bioethanol production.

Present investigation is carried out on **Site-directed mutagenesis of endoglucanase *(Ct*GH5) of family 5 glycoside hydrolase from *Clostridium thermocellum* for enhancing endoglucanase activity. Further, the mutant endoglucanase was fused with β-glucosidase (*Ct*GH1) family 1 glycoside hydrolase from *Clostridium thermocellum* for the development of an enzyme with bi-functional β-glucosidase and endoglucanase activity. The developed Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was biochemically and structurally characterised. The Chimera along with cellobiohydrolase (CBH5A) used as a cocktail of cellulases for enzymatic saccharification of dual alkali soaking and organosolv pretreated *Sugarcane bagasse*. This saccharified sugars obtained can be utilized for the production of bioethanol.**

The thesis work comprises 5 chapters **Chapter 1** is the general introduction that starts with the requirement of alternative energy source due to the depletion of fossil fuel. It also states that bioethanol can be a potential alternative energy source to overcome the demands of fuel and reduce burden on the non-renewable energy sources. For the production of bioethanol the cellulosic biomass from agricultural and industrial wastes can be used. The prime components of the plant lignocellulosic

biomass are made up of cellulose, hemicellulose and lignin. The cellulose and hemicellulose can be hydrolysed to monomeric sugars. These monomeric sugars for example glucose and xylose can be utilized by the yeast for production of bioethanol. In the plant cell wall cellulose is the key component and is known to be the largest biopolymer in nature. Cellulose can be depolymerized to give glucose by the combined action of endoglucanase, cellobiohydrolase and β-glucosidase. But the enzymatic hydrolysis by this enzymes require plant lignocellulosic biomass structure to be partially broken down by using pretreatment process, making them accessible to enzymatic hydrolysis. This chapter also includes the importance of different cellulases on the depolarization of plant cell wall with special emphasis on endoglucanases from family 5 glycoside hydrolase and β-glucosidase of family 1 glycoside hydrolase. It also demonstrates the use of protein engineering technologies used to produce multifunctional enzymes to reduce the cost of enzyme production. This chapter defines the significance of investigation undertaken and then lists the specific objectives.

Chapter 2 The site-directed mutagenesis was performed for the wild-type *Ct*GH5 to develop mutant *Ct*GH5-F194A. The mutant plasmid was transformed in *E. coli* DH5α competent cells. The plasmid was extracted and sequenced to confirm the mutation at the specific residue. This recombinant plasmid was transformed in *E. coli* BL21 cells for *Ct*GH5-F194A expression. The mutant *Ct*GH5-F194A was further engineered for development of Chimera by fusing *Ct*GH1 (β-glucosidase) from *Clostridium thermocellum* to the N-terminal of mutant *Ct*GH5-F194A using the natural linker (Fig. 1 and Fig. 2). The PCR amplified fragment of Chimera (*Ct*GH1-

L1-*Ct*GH5-F194A) showed a band of ~2.3 kb. The restriction enzyme digested fragment of the Chimera was ligated to the linearized pET28a(+) vector. The ligated mixture was transformed into the DH5α cells. The positive clone was confirmed by the double digestion using the restriction enzymes *NheI* and *XhoI* and confirmed by sequencing. The gel electrophoresis of the digested product showed a band of ~ 5.4 kb was produced for the pET-28a(+) vector and the band ~2.3 kb produced for the corresponding for the chimeric gene. The recombinant plasmid was then transformed in *E. coli* BL21 cells for Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) expression. The recombinant protein Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) along with its individual module *Ct*GH1, wild-type *Ct*GH5 and *Ct*GH5-F194A were purified using Immobilized Metal Ion Affinity Chromatography (IMAC). The purified recombinant proteins displayed a band of approximately, 89 kDa for Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), 52 kDa for *Ct*GH1 and a band of approximately, 35 kDa for wild-type *Ct*GH5 and mutant *Ct*GH5-F194A on SDS-PAGE gels. The amount of purified recombinant Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), *Ct*GH1, wild-type *Ct*GH5 and *Ct*GH5-F194A obtained from 100 ml *E. coli* BL21 cell cultures were 3.24 mg, 1.62 mg,1.8 mg and 1.2 mg, respectively.

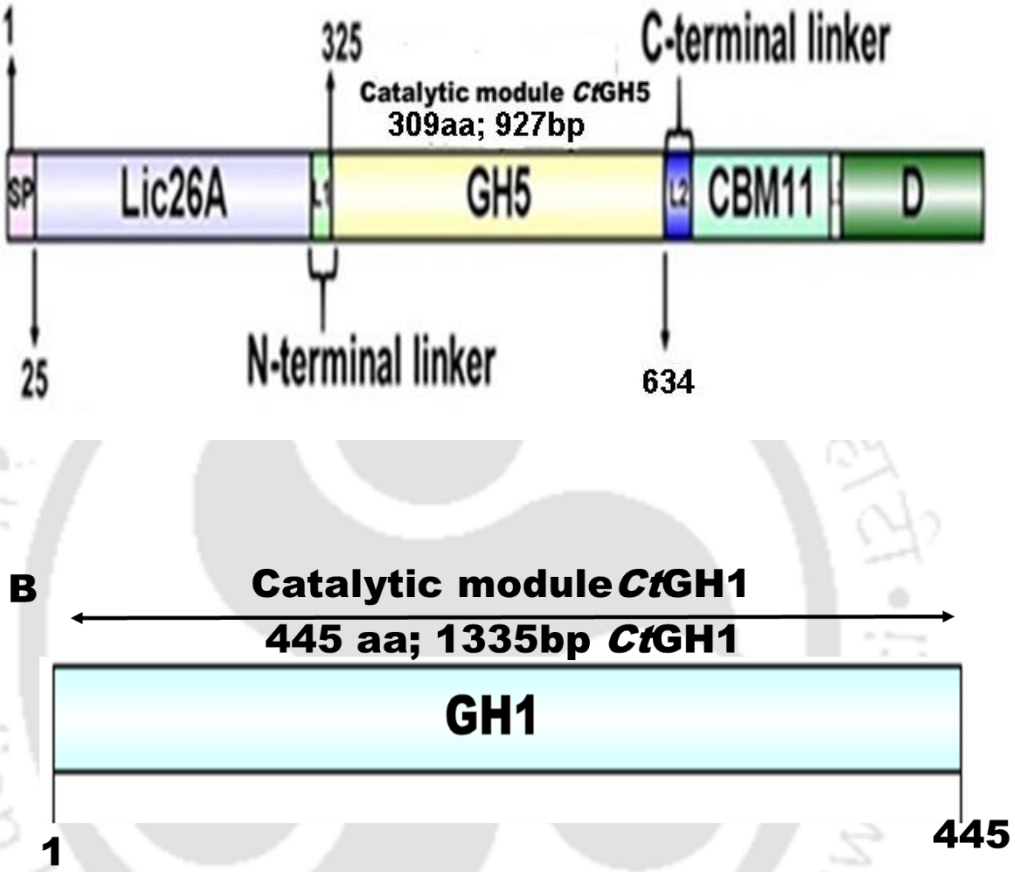


Fig. 1 A) Molecular architechture *Ct*CelH from *Clostridium thermocellum,* The domain *Ct*GH5 used for site-directed mutagenesis of amino acid residue Phe194 and the N-terminal (57 bp) used for Chimera construction by fusion of *Ct*GH1 and *Ct*GH5-F194A B) β-glucosidase (*Ct*GH1) from *C. thermocellum* was fused at the N-terminal of *Ct*GH5-F194A.

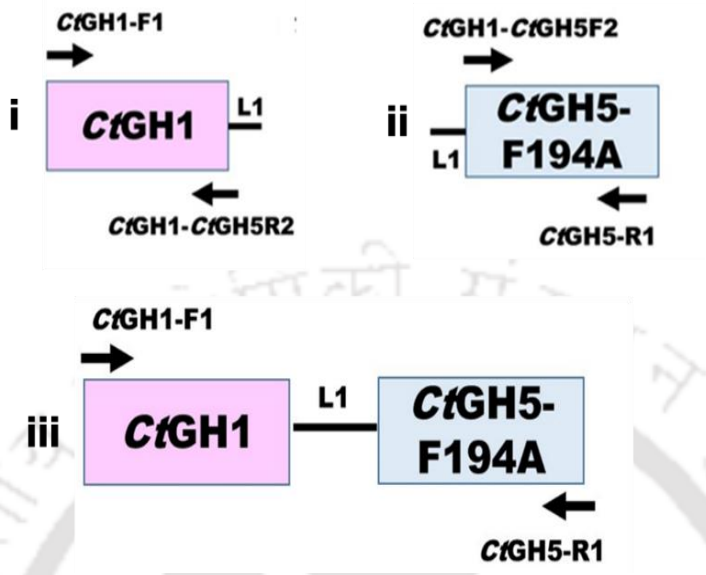


Fig. 2 Schematic presentation for development of chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A). The arrow (→) represents the primers for different gene constructs,
i) *Ct*GH1-L1 and ii) L1-*Ct*GH5-F194A
iii) Full length Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)

Chapter 3 describes the comparative biochemical characterization of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual catalytic modules- β-glucosidase (*Ct*GH1) and mutant endoglucanase (*Ct*GH5-F194A). The mutant endoglucanase (*Ct*GH5-F194A) gave 40 U/mg specific activity against carboxymethyl cellulose, resulting 2-fold higher activity than wild-type *Ct*GH5. *Ct*GH5-F194A was fused with a β-1,4-glucosidase, *Ct*GH1 from *Clostridium thermocellum* to develop a chimeric enzyme with bifunctional β-glucosidase and endoglucanase activity. The Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) expressed as a soluble protein using *E. coli* BL-21cells. For endoglucanase, the Chimera and the wild-type *Ct*GH5 showed temperature maxima at 60°C, whereas the mutant *Ct*GH5-F194A showed temperature maxima at 50°C. This showed that mutant is temperature sensitive and gets inactivated. The temperature sensitivity and inactivation of mutant *Ct*GH5-F194A is counteracted by

the linker or the adjacent more thermotolerant module *Ct*GH1 in the N-terminal domain conferring the Chimera higher temperature resistance. The Chimera also displayed improved temperature optima at 70°C for β-glucosidase activity than its individual counterpart *Ct*GH1 which is 65°C. The Chimera showed maximum endoglucanase activity at pH 5.0. The pH optima of wild-type *Ct*GH5 is 4.2 which shifted to pH 5.5 in mutant *Ct*GH5-F194A. A shifting in pH optimum towards acidic pH of 5.5 was also observed in the β-glucosidase activity of the Chimera, because the β-glucosidase has a pH optimum of pH 6.5 for individual module *Ct*GH1. The pH stability analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) for endoglucanase activity showed that is stable between pH 4.0-5.0. The individual mutant *Ct*GH5-F194A was stable at higher pH between 5.0-6.0. For β-glucosidase activity, the Chimera showed pH stability between pH 5.0 to 5.5, while the individual enzyme *Ct*GH1 was stable between pH 5.5 to 6.0. The advantageous outcome was that the Chimera displayed pH stability between pH 4.0-5.5 for both the enzyme activities. This may prove beneficial in the synergistic functioning of both enzyme modules in the chimeric form. The thermostability analysis showed that the wild-type *Ct*GH5 is stable up to 60ºC and the individual mutant enzyme *Ct*GH5-F194A is stable up to 55ºC. The thermostability analysis of Chimera showed that its endoglucanase activity is stable at 55ºC up to 1 h. The thermal stability of both, *Ct*GH1 and β-glucosidase domain of Chimera was found up to 60ºC for 1 h. Protein melting studies of Chimera and its individual enzyme modules showed that *Ct*GH5-F194A melting temperature (T_m) is 68ºC and of *Ct*GH1 is 79ºC, whereas, Chimera showed melting temperature (T_m) of 78ºC. This showed improved structural integrity, thermostability of Chimera. The K_m and k_{cat} of Chimera

were compared against their individual modules for substrates Lichenan, β-Glucan, CMC and *p*NPG. For endoglucanase activity Chimera showed higher catalytic efficiency (k_{cat}/K_m) against Lichenan, β-Glucan, and CMC i.e.$2.9\pm0.8\times10^5$ ml $min^{-1}mg^{-1}$, $1.6\pm4.7\times10^5$ ml $min^{-1}mg^{-1}$, $1.2\pm0.4\times10^4$ ml $min^{-1}mg^{-1}$ which was 3-5 fold higher than the individual mutant *Ct*GH5-F194A. For β-glucosidase activity Chimera showed $4.1\pm0.2\times10^4$ $mM^{-1}min^{-1}$ catalytic efficiency (k_{cat}/K_m) against *p*NPG which was 5.5 fold higher than the individual module *Ct*GH1. The TLC analysis of hydrolysed products of CMC by Chimera revealed glucose as final product that confirmed its both, β-1,4-endoglucanase and β-1,4-glucosidase activities. The hydrolysed products of CMC by *Ct*GH5-F194A were cellobiose and cello-oligosaccharides and no glucose. The Chimera displayed higher accumulation of glucose after 48h than the mixture of *Ct*GH1 and *Ct*GH5-F194A from the pretreated Sorghum stalk. This study demonstrated the developed bi-functional Chimera with β-glucosidase and endoglucanase activities has potential applications for bioethanol, biobutanol or lactic acid production.

Chapter 4 elaborates the SAXS and computational approaches for structure and bindingcharacteristics of chimera (*Ct*GH-L1-*Ct*GH5-F194A). The chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A) developed by the fusion of *Ct*GH1 (β-glucosidase) at the N-terminal and *Ct*GH5-F194A (endoglucanase) at C-terminal was structurally characterized. The amino acid sequence of the chimera (*Ct*GH1-L1-*Ct*GH5-F194A) composed of β-glucosidase (*Ct*GH1) and endoglucanase (*Ct*GH5-F194A) was evaluated for secondary structure analysis by using Psi-Pred and RaptorX prediction tools and compared with the CD data. The CD analysis revealed that the secondary

structure of chimera consists of 38% of α-helix, 9.32% of β-sheet and 52.68% of the loops corroborating with the predicted secondary structure elements. The 3-dimensional modeled structure of chimera was generated by RaptorX, which showed close homology with β-glucosidase, 3AHX from *Clostridium cellulovorans* and an endoglucanase, 5BYW from *Clostridium thermocellum.* The modeled structure generated was validated by Ramachandran plot that showed approximately, 99.9% residues in the allowed region. The modeled structure of chimera represented a modular structure consisting of N-terminal β-glucosidase (*Ct*GH1) and C-terminal endoglucanase (*Ct*GH5-F194A) modules connected by the linker. The overall moduler structure of chimera represented a classical (α/β)-TIM barrel fold. The molecular dynamics study of the chimera showed stable RMSD till 70 ns with 0.01 oscillation suggesting the stable conformation of the two modules, *Ct*GH1 (β-glucosidase) and *Ct*GH5-F194A (endoglucanase) in the chimeric form. The average SASA of the chimera remained stable with an average value of 320 nm^2. The minor variation in the SASA also suggested that the accessibility of the substrate in the catalytic sites were not affected in the designed chimera. This indicated that the binding sites were not obstructed by deformation or steric effect between the two modules of chimeric enzyme. The docking studies of the chimera with cello-oligosaccharides suggested that both the modules *Ct*GH1(β-glucosidase) and *Ct*GH5-F194A (endoglucanase) are active in chimeric form and involved in the catalysis. The SAXS analysis revealed that the chimera displayed elongated structure with two modules showing fully folded form in the solution. Guinier estimation analysis showed that the radius of gyration (Rg) for chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was 3.15±0.10 nm for globular shape

and 0.57±0.05 nm for rod-shape. The Kratky plot analysis confirmed a compact and folded structure of chimera. The Gasbor modeled structure of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) displayed an elongated structure with two modules having an overall shape similar to a bean-bag contour.

Chapter 5 reports the optimization of pretreatment of Sugarcane bagasse using alkali followed by organosolv pretreatment method for efficient enzymatic saccharification in context to get maximum TRS yield was carried out. In single stage alkali pretreatment, Sugarcane bagasse by 1% (w/v) NaOH for 1h, 2h, 3h and 4h gave a TRS loss of 20.3, 13.5, 9.4 and 5.9 mg/g, of raw biomass respectively, in the pretreated hydrolysate The loss of TRS was because the sodium hydroxide disrupts the sugarcane bagasse (scb) cell wall by solubilizing the hemicellulose and lignin. The dual stage pretreatments of Sugarcane bagasse by alkali pretreatment followed by organosolv pretreatment method resulted in a minimum loss in TRS yield in the pretreated hydrolysate. The less total reducing sugar (TRS) loss in the hydrolysate during phosphoric acid/acetone pretreatment method was because the phosphoric acid pretreatment can reduce the crystallinity of cellulose and avoid its further degradation into water-soluble sugar. The enzymatic saccharification using Chimera and Cellobiohydrolase of all 5 pretreated sugarcane bagasse which were single stage organosolv pretreted and dual stage pretreted by alkali pretreatment for different time interval of 1h, 2h, 3h and 4h followed by organosolv pretreatment were analysed for hexose sugars by HPLC. The enzymatic hydrolysates of the dual pretreated sugarcane bagasse using phosphoric acid-acetone method resulted in maximum yield of 137 mg/g of glucose from the pretreated biomass which was 1.9-fold higher as compared

to the enzymatic hydrolysates of single organosolv pretreated sugarcane bagasse which gave a glucose yield of 71.5 mg/g from pretreated biomass. The FESEM analysis showed Alkali using 1% (w/v) NaOH for 1h and 2h pretreated Sugarcane bagasse showed visible voids. These changes in the surface morphology of biomass can be attributed because of the delignification of biomass. The single organosolv pretreated Sugarcane bagasse by Phosphoric acid-acetone pretreatment method and dual pretreated Sugarcane bagasse from Alkali using 1%(w/v) NaOH for 1h and 2h followed by organosolv pretreatment showed rough surfaces. This indicates the pretreatment has removed external fibers which in turn increased the surface area of pretreated biomass in which the cellulase enzymes are more accessible to cellulose fibers. The FT-IR analysis showed that the Phosphoric acid-acetone pretreatment on 2h alkali soaked pretreated sugarcane bagasse the intensity of peak in 1062 cm^{-1} and 1171 cm^{-1} has increased. This suggested that the cellulose contend in the solid residue during the dual pretreated sample has increased since hemicellulose fraction was removed by the phosphoric acid pretreatment. This result showed that alkaline NaOH soaking followed by organosolv pretreatment using phosphoric acid-acetone method is an efficient pretreatment strategy than the single organosolv pretreatment. This dual stage pretreatment strategy aids the cocktail cellulase enzymes comprising of Chimera and CBH5A to act efficiently on the degradation of cellulose into glucose which can be further utilized for bioethanol production

CONTENTS

Statement……..... i
Certificate……..... iii
Acknowledgements……..... v
Synopsis……..... vii
Contents…… xxi

Chapter 1 General interoduction
1.Bioethanol as an alternative energy source…….. 1
1.1 Plant cell wall and its carbohydrates……. 3
1.1.1 Cellulose…….. 4
1.1.2 Hemicellulose……. 6
1.1.3 Lignin……. 7
1.2 Cellulases……. 8
1.2.1 Endoglucanase……. 9
1.2.2 Cellobiohydrolase……... 10
1.2.3 β-Glucosidase…….. 11
1.3 Importance of cellulases in bioethanol production……... 12
1.4 Industrial production and global demand of cellulases……. 13
1.5 Protein engineering of cellulases……. 14
1.5.1 Directed Evolution……... 15
1.5.2 Rational designing……... 15
1.5.3 Multifunctional Chimera…… 16
1.6 The microorganism…….. 18
1.6.1 Cellulosome structure…….. 19
1.7 Significance and objectives of the present study……. 20
1.7.1 Significance of this study……. 20
1.7.2 Specific objectives of the study…… 23
References…….. 24

Chapter 2 Site-directed mutagenesis of wild-type endoglucanase (*Ct*GH5) to develop mutant *Ct*GH5-F194A and construction of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) using mutant *Ct*GH5-F194A and β-glucosidase (*Ct*GH1)

2.1 Introduction…… 35
2.2 Materials and Methods…….. 38
2.2.1 Chemicals, reagents and kits……. 38
2.2.2 Microorganisms……. 39
2.2.3 Site-directed mutagenesis of wild-type endoglucanase (*Ct*GH5)……. 39
2.2.4 Cloning of Chimeric gene (*Ct*GH1-L1-*Ct*GH5-F194A) by fusing the genes encoding mutant endoglucanase (*Ct*GH5-F194A) and glucosidase (*Ct*GH1)……. 41
2.2.4.1 Strategy for Chimera construction…….. 41
2.2.4.2 PCR amplification and cloning of Chimeric gene *(Ct*GH1-L1- *Ct*GH5-F194A)……. 44
2.2.5 Agarose gel electrophoresis of PCR amplified products…… 46
2.2.5.1 DNA loading buffer ……. 47

2.2.6 Extraction of DNA from agarose gel……. 47
2.2.6.1 Protocol for extraction of DNA from agarose gel…….. 47
2.2.7 Preparation of the culture medium……. 49
2.2.8 Preparation of *E. coli* DH5α and *E. coli* BL21 competent cells……... 50
2.2.9 Cloning of gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) into pET28a(+) vector…….. 51
2.2.9.1 Restriction digestion of PCR amplified genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and pET-28a(+) vector DNA…….. 53
2.2.9.2 Ligation of restriction digested genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) into pET-28a(+) vector……. 53
2.2.9.3Transformation of mutagenic plasmid DNA or ligated recombinant DNA into *E. coli* DH5α cells…….. 54
2.2.9.4 Isolation of plasmid DNA from transformed colonies by plasmid miniprep kit…….. 56
2.2.9.4.1Plasmid isolation protocol by miniprep kit protocol……. 56
2.2.9.5 Screening of recombinant plasmid DNA for positive clone of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) by restriction digestion…… 58
2.2.10 Transformation of recombinant plasmids containing genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and *Ct*GH5-F194A in *E. coli* BL21 (DE3)……. 58
2.2.11 Expression of recombinant plasmids containing genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and *Ct*GH5-F194A in *E. coli* BL21 (DE3)……... 58
2.2.11.1Purification of recombinant proteins by affinity chromatography…….. 59
2.2.11.2 IMAC purification protocol for recombinant Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual modules (*Ct*GH5-F194A and *Ct*GH1)……... 61
2.2.11.3 Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis (SDS-PAGE) analysis of recombinant proteins……. 62
2.2.11.4 Preparation of sample buffer for protein……. 64
2.2.11.5 Preparation of protein samples for SDS-PAGE…….. 65
2.2.11.6 Preparation of SDS-PAGE running buffer…….. 65
2.2.11.7 Preparation of staining and destaining solutions…….. 66
2.2.12 Protein concentration determination of purified recombinant proteins…… 66

2.3 Results and Discussion…… 67
2.3.1 Site-directed mutagenesis of wild-type *Ct*GH5 for the development of mutant *Ct*GH5-F194A…….. 67

2.3.2 PCR amplification of genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)…….. 69
2.3.3 Cloning of gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) into pET-28a(+) vector……. 70
2.3.3.1 Isolation of recombinant plasmid DNA……. 70
2.3.3.2 Restriction digestion of isolated plasmid DNA for confirmation of positive clone……... 70
2.3.4 Expression of recombinant *Ct*GH5-F194A and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)……. 74
2.3.5 Purification of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and individual enzymes β-glucosidase (*Ct*GH1) and endoglucanase (*Ct*GH5-F194A)…… 76
2.3.6 Production of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and individual modules *Ct*GH1 (β-glucosidase), *Ct*GH5-F194A and wild-type *Ct*GH5 (endoglucanase) using LB medium ……. 78
2.4 Conclusions…… 79
References…… 81

Chapter 3 Comparative biochemical characterization of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual modules catalytic modules β-glucosidase (*Ct*GH1) and mutant endoglucanase (*Ct*GH5-F194A)

3.1 Introduction ……. 85
3.2 Materials and Methods…… 88
3.2.1 Substrates and reagents…… 88
3.2.2 Enzyme assay……... 88
3.2.2.1 Preparation of reagents for reducing sugar estimation……. 89
3.2.2.2 Generation of standard plot of D-glucose…… 90
3.2.2.3 Calculation of enzyme activity for wild-type endoglucanase (*Ct*GH5), mutant *Ct*GH5-F194A and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)…….. 91
3.2.2.4Calculation of enzyme activity for β-glucosidase (CtGH1) and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)…….. 92
3.2.3 Determination of optimum pH and temperature of wild-type *Ct*GH5 endoglucanase, mutant *Ct*GH5-F194A and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)……... 92
3.2.4 pH stability analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) by its individual modules……. 93
3.2.5 Thermal stability analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) from its individual modules…….. 94
3.2.6 Substrate specificities and kinetic parameters of wild-type *Ct*GH5, mutant *Ct*GH5-F194A and Chimera…….. 94
3.2.7 Protein melting studies…… 95
3.2.8 End-product determination by Thin Layer Chromatography…….. 95
3.2.9 Analysis of hydrolysed products from pretreated Sorghum stalk by chimera and individual enzyme…….. 96
3.3 Results and Discussion…… 98

3.3.1 Effect of temperature and pH on different enzymatic construct……. 98
3.3.2 pH stability of Chimera and its individual construct…….. 102
3.3.3 Thermal stability of Chimera and its individual constructs……. 104
3.3.4 Comparison of kinetic parameters of enzyme constructs……. 106
3.3.5 Melting temperature of Chimera and its individual modules……... 108
3.3.6 Analysis of hydrolysed products by purified Chimera using TLC…... 110
3.3.7 Analysis of hydrolysed products from pretreated biomass (Sorghumstalk) by purified Chimera using TLC……. 112
3.4 Conclusions…….. 113
References…….. 116

Chapter 4 SAXS and computational approaches for structure and binding characteristics of chimera (*Ct*GH-L1-*Ct*GH5-F194A)

4.1 Introduction…….. 119
4.2. Materials and Methods…… 122
4.2.1 Experimental and *in silico* secondary structure analysis of chimera……. 122
4.2.2 Computer-based structural prediction of chimera and its validation…….. 123
4.2.3 Molecular dynamics simulation of modeled chimera structure…… 124
4.2.4 Binding interaction analysis of chimera…… 125
4.2.5 Molecular dynamics of ligand bound chimera……. 126
4.2.6 Small angle x-ray scattering analysis of chimera……... 127
4.3. Results and Discussion…… 129
4.3.1 Secondary structure analysis by circular dichroism and web server's prediction…….. 129
4.3.2 Three-Dimensional structure prediction of chimera and its validation…….. 131 133
4.3.3 Molecular dynamics simulation of chimera……. 134
4.3.4 Molecular Docking analysis of chimera…… 137
4.3.5 Molecular dynamics of ligand-bound chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and corresponding individual enzymes……. 141
4.3.6 Small angle x-ray scattering analysis of chimera……... 147
4.4 Conclusion……... 151
References…….. 153

Chapter 5 Saccharification of alkali and organosolv pretreated sugarcane bagasse by cocktail of cellulases, Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and Cellobiohydrolase (*Ct*CBH5A)

5.1 Introduction……. 159
5.2 Materials and Methods……... 162
5.2.1 Chemical and Reagents……. 162
5.2.2 Pretreatment of sugarcane bagasse…… 162
5.2.2.1 Stage 1 Alkali pretreatment……. 162
5.2.2.2 Stage 2 Organosolv (phosphoric acid-acetone) pretreatment of alkali pretreated sugarcane bagasse…… 163

5.2.3 Production of recombinant enzymes Chimera and Cellobiohydrolase (*Ct*CBH5A)…… 164
5.2.4 Determination of enzyme concentration of recombinant enzymes (Chimera and Cellobiohydrolase)…… 165
5.2.5 Determination of enzyme activity of recombinant enzymes (Chimera and Cellobiohydrolase)…… 165
5.2.6 Determination of enzyme stability of recombinant enzymes (Chimera and Cellobiohydrolase)…… 165
5.2.7 Saccharification of pretreated sugarcane bagasse by Chimera (*Ct*GH1-L1-*Ct*GH5- F194A) and Cellobiohydrolase (*Ct*CBH5A)…… 166
5.2.8 Fourier-transform infrared spectroscopy (FTIR) analysis of untreated and pretreated sugarcane bagasse ……. 167
5.2.9 Field Emission Scanning Electron Microscope (FESEM) analysis of untreated and pretreated sugarcane bagasse…….. 167

5.3 Results and Discussion…… 167
5.3.1 Reducing sugar analysis of hydrolysates from pretreatment of sugarcane bagasse…….. 167
5.3.2 Saccharification of pretreated sugarcane bagasse by Chimera (*Ct*GH1-L1-*Ct*GH5- F194A) and Cellobiohydrolase (*Ct*CBH5A)…….. 168
5.3.3 Fermentable hexose sugar analysis in enzymatic hydrolysates from saccharification of pretreated sugarcane bagasse…… 169
5.3.4 FESEM analysis of untreated and pretreated sugarcane bagasse…… 173
5.3.5 FT-IR analysis of untreated and pretreated sugarcane bagasse…… 174
5.4 Conclusion…….. 177
References……. 179

Future prospects……... 183
List of publication…… xxvii
List of conferences……... xxix
Vitae……. xxxi

Chapter 1

General Introduction

1. Bioethanol as an alternative energy source

The rise in the human population on earth has been tremendous in the last 70 years. The substantial increase in population from 2.7 billion in 1955 to 7.6 billion in 2018 has created major burden over non-renewable resources and their utilization. Similarly, the urban population in 2018 was 4.2 billion which was only 1.0 billion in 1955 (www.worldometers.info). This expansion of urbanization has led to increase in energy consumption upto 200%, including non-renewable and renewable resources from 1990 to 2016 (Global Energy Statistical Yearbook, 2017). Petroleum and Oil are still serving as the world's main supply of fuel. The petroleum and oil share 22.9% in world energy consumption whereas, coal provides around 40% of the world's electricity (World Energy Report, 2016). Petroleum together with Oil is presently serving as the main energy source, but it is being exhausted rapidly from the world (Ragauskas *et al.*, 2006; Tong *et al.*, 2012). On contrary to the oil and petroleum,

renewable sources of energy have a good future to support our ever-demanding civilization. Towards the early 19th century biomass was known to be used as an energy source. But nowadays energy generated from biomass shares only a small part in the total energy supply of the world (REN21, 2006). Currently among the total energy consumed, 18% is from renewable sources, for example forest biomass (mostly, around 90% from woody biomass). This is the primary source of energy supply, that is generally utilized in smelting of iron, other metal ores and cooking across the world (WER, 2016). Biodiesel and ethanol production from biomass is another growing area. The production of bioethanol and biodiesel is predicted to grow by 2024 to approximately, 134.5 and 39 billion litres respectively (REN21, 2006). However, the increase in energy demand due to several developments for example urbanization, power driven technology and transportation etc. are responsible for depletion in the non-renewable reserve crude oil. Estimated data suggested that by 2025, there will be approximately 50% increase in the energy demand from several developed and developing nations (Tong *et al*., 2012). Producing bioenergy by utilizing biomass through cost-effective and sustainable approach is vital alternative of non-renewable energy resource.

Biofuels, like bioethanol is derived from optimized conversion of lignocellulosic biomass whereas, biodiesel is derived from vegetable oil and animal fat. Bioethanol is produced by fermentation from carbohydrate of lingocellulosic biomass while biodiesel is produced from any fatty acids from vegetable or animal origin using transesterification reaction (Claassen *et al.,* 1999; Meher *et al.,* 2006; Galbe and Zacchi, 2007; Bhatia *et al.,* 2002; Saravanakumar *et al.,* 2013; Aulitto *et al.,* 2019). Several agricultural waste biomasses having cellulosic component can be

used for production of bioethanol. Several plant components like corn stover, sugarcane, grasses, and sorghum etc. are used around the world as lingocellulosic biomass for production of bioethanol (Lynd *et al.*, 2008). These biomasses as are further processed to breakdown their cell wall using different pretreatment technologies.

The pretreated biomasses are then enzymatically hydrolysed into fermentable sugars (Lin and Tanaka, 2006 & Kazi *et al.*, 2010). The efficient conversion and usage of lignocellulosic waste from plants to bioethanol can lead to clean energy production and reduce burden on the non-renewable energy sources. Bioethanol is an environmentally friendly product. Bioethanol releases less carbon monoxide (around 12%) and is mainly produced from sugars, cellulose and starch (Mathapati *et al*., 2010; Zainab and Fakhra, 2014). The lesser emission of carbon monoxide will prevent the greenhouse gases that were important for maintenance of the climate and global temperature of earth (Mathapati *et al.*, 2010; Zainab and Fakhra, 2014). The cellulosic biomass from agricultural and industrial wastes are abundant in nature which can be used for the production of biofuel. The prime components of the plant lignocellulosic biomass are made up of cellulose, hemicellulose and lignin. The cellulose and hemicellulose can be hydrolysed to monomeric sugars. These monomeric sugars can be utilized by the yeast for production of bioethanol (Zhang *et al.*, 2006).

1.1 Plant cell wall and its carbohydrates

The plant lignocellulosic biomass is made up of cellulose, hemicellulose and lignin as shown in Fig. 1.1. Plant lignocellulosic biomass is one of the most abundant

components present on earth. The following sections 1.1.1, 1.1.2 and 1.1.3 illustrates these components and their importance in bioethanol production.

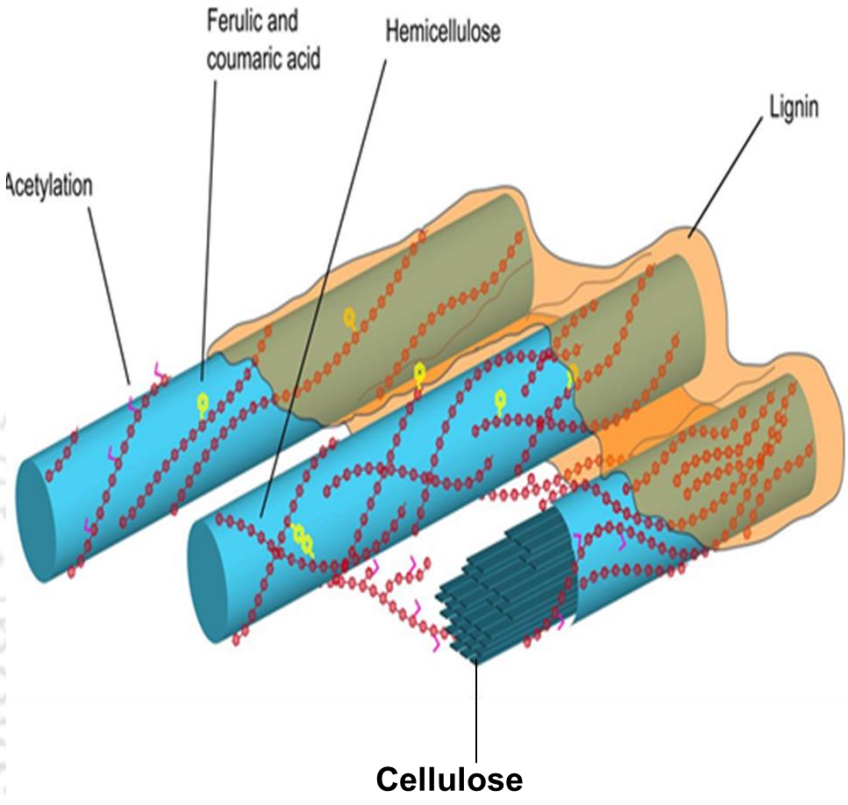


Fig 1.1 The primary structural component of plant cell wall (Sorek *et al.*, 2014)

1.1.1 Cellulose

Cellulose is the key component of plant cell walls and is known to be the largest biopolymer in nature (Somerville 2006). The insoluble microfibrils of the cellulose forms cable-like structures (Fig. 1.1) which are composed of approximately 24 hydrogen-bonded chains containing β-(1,4)-linked glucose molecules as shown in

Fig. 1.2 (Guerriero *et al.*. 2010, Fernandes *et al.*, 2011, Nsor *et al.*, 2017). The glucan chains which are polysaccharides made up of glucose units forms parallel and consecutive chain of glucose molecules. The repeating unit of cellulose is cellobiose, a disaccharide of glucose in which there is a 180° rotation of each monomer in relation to its neighbour glucose molecule as shown in Fig. 1.3 (McDonald *et al.*, 2012). The hydrolysis of cellobiose is an measure of β-glucosidase activity. This arrangement helps the glucan chains to form a flat, inflexible and ribbon-like crystalline structure merged together by Van der Waals forces and hydrogen bonds to form microfibrils. Hydroxyl groups present in cellulose macromolecules are connected with large number of intra- and intermolecular hydrogen bonds, which resulted in various ordered crystalline arrangements such as flat ribbon like conformations of cellulose chain (Sorek *et al.*, 2014).

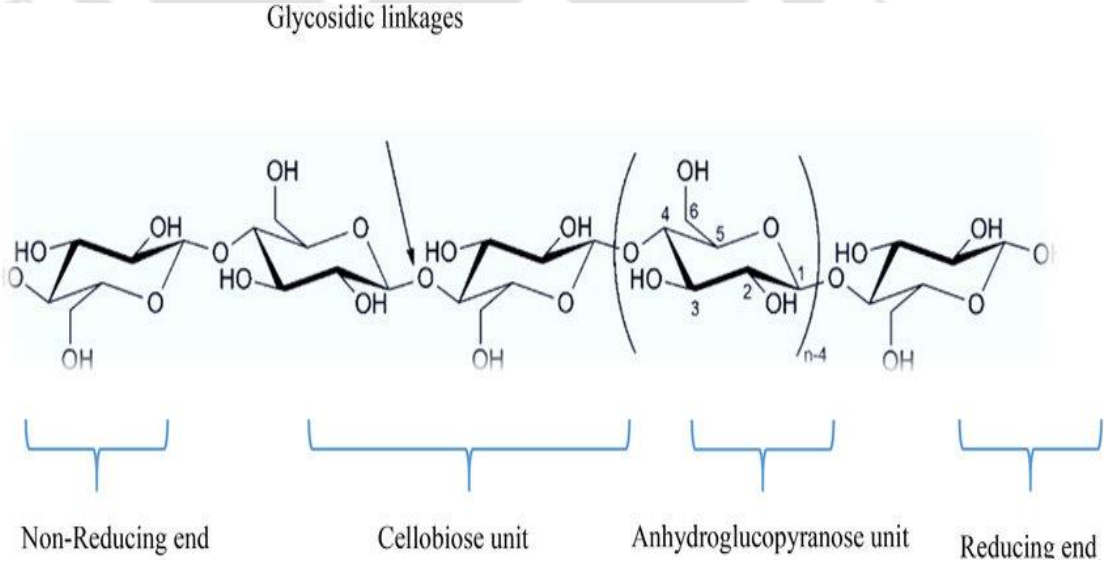


Fig. 1.2 Chemical structure of microcrystalline cellulose (Nsor *et al.*, 2017)

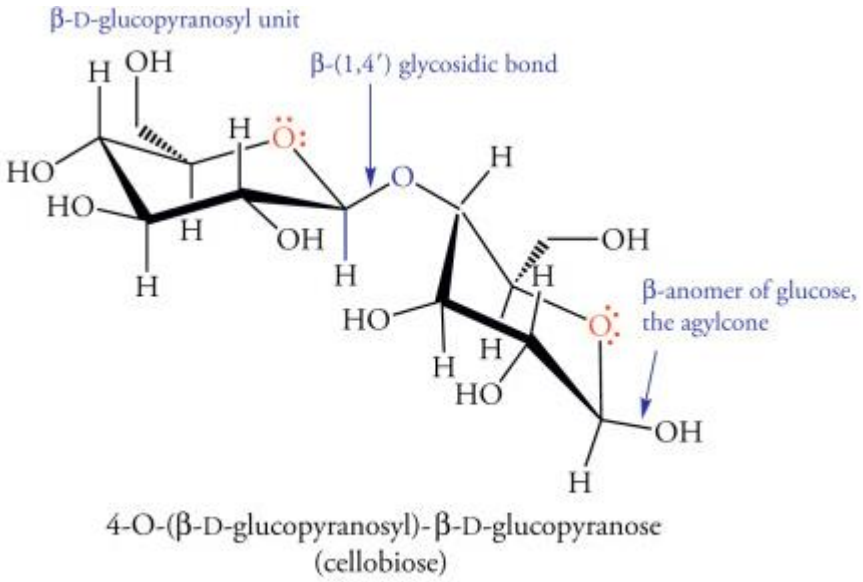


Fig. 1.3 Structure of cellobiose (McDonald *et al.*, 2012)

1.1.2 Hemicellulose

The microfibriller structure of cellulose are mostly covered with hemicellulose, which forms hydrogen bond in the microfibril surface of cellulose (Fig. 1.1). Since the hemicellulose is usually branched in nature, it may prevent the aggregation of cellulose microfibrils. Hemicelluloses are of different types according to their backbone structure such as xylan, galactan, mannan, or glucomannan which are branched with longer or single glycosyl residues (Pauly *et al.*, 2013). Secondary cell wall hemicelluloses form approximately half of the carbohydrates and one-third of the overall biomass available in the stem and woody tissues can be considered for biofuel production. Although hemicelluloses are considered as a potentially large fermentable pool of sugar, they also develop challenges in production of biofuel.

Hemicelluloses are a diverse catagory of linear and branched polysaccharides that vary largely in composition among species and plant tissues and that change significantly during plant development (Sorek *et al.*, 2014). Hemicellulose plays as cross-linking agents in the cell wall, associating the cellulose bundles, cell wall proteins, lignins, pectins, and nonstructural polysaccharides by the help of noncovalent and covalent interactions (Sorek *et al.*, 2014).

1.1.3 Lignin

Lignin is abundantly found in the cell walls of plant (contributing around 20-30% in lignocellulosic matter) (Fisher and Fong, 2014; Gall *et al.*, 2018). Lignin is a alkyl-aromatic, non-carbohydrate heteropolymer (Sorek *et al.*, 2014). These are mostly syringyl and guaiacyl monoaromatic phenyl propanoid units (Sorek *et al.*, 2014).. The main building blocks of lignin are the hydroxycinnamyl alcohols (or monolignols) which are coniferyl alcohol, sinapyl alcohol and *p*-coumaryl alcohol (Fisher and Fong, 2014; Cragg *et al.*, 2015). The oligomers of lignin are oxidatively coupled to this monolignols forming phenylpropanoid subunits: guaiacyl (G), *p*-hydroxyphenol (H), and syringyl (S) respectively, that differs in individual plants (Fisher and Fong, 2014). Lignin polymers are deposited mainly in the plant cell wall making them rigid and impervious (Fig. 1.1). The biosynthesis of the lignin can also occur even upon various biotic and abiotic stress and perturbations in cell wall structure (Can o-Delgado *et al.*, 2003; Tronchet *et al.*, 2010). Since lignin protects cell wall polysaccharides of plant from microbial degradation, thus it becomes an important limiting factor for conversion of plant polysacharides into monomeric sugars for production of biofuel. The process for lignin removal is a cost intensive process, therefore the research efforts are now aimed at designing plants that either

deposit less lignin or produce lignins that are more amenable to chemical degradation (Sticklen, 2008; Weng *et al.*, 2008a; Mansfield, 2009).

1.2 Cellulases

Cellulose can be depolymerized to give glucose by the combined action of endoglucanase, cellobiohydrolase and β-glucosidase. These Cellulase enzymes involved in hyrolysis of cellulose polymers into cellooligosacharides and glucose belong to different glycoside hydrolase families (GH) (www.cazy.org). Cellulase enzymes consist of two separate modules, the catalytic module, which contains the active site and the non-catalytic carbohydrate binding module (CBM). Both modules are connected by flexible linkers, containing mostly serine and threoninine (Zhang *et al.*, 2006). Cellulolytic enzymes act synergistically, in which endo-β-1,4-glucanase acts randomly on the cellulose chain and produces cellodextrins which are the larger cellooligosacharides as a hydrolyzed products (Urbanowicz *et al.*, 2007) (Fig. 1.4). Cellobiohydrolase acts at the end of the cellodextrin and releases cellobiose as the main product (Barr *et al.*, 1996) (Fig. 1.4). Finally, β-glucosidase hydrolyzes the cellobiose to form two molecules of glucose (Grabnitz *et al.*, 1991) (Fig.1.4). The following sections describe different cellulases and their mode of actions.

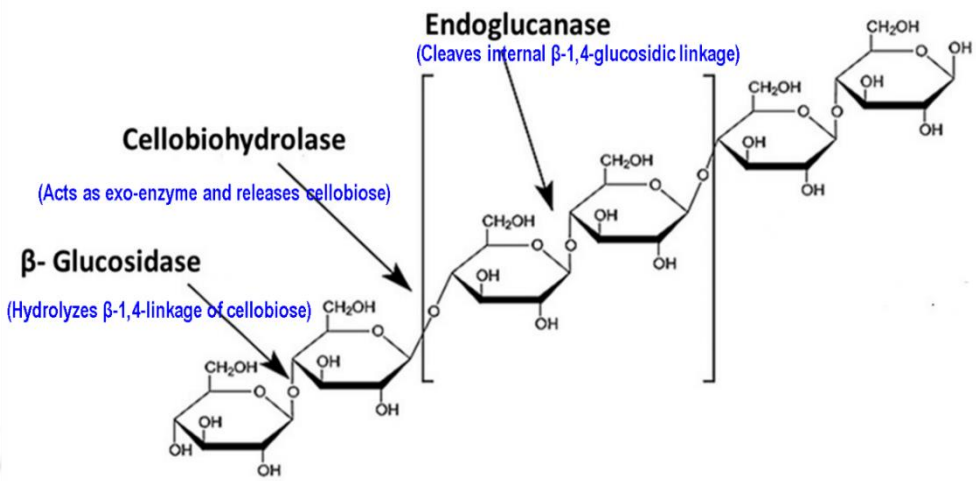


Fig. 1.4 Action of different cellulases in depolimerization of cellulose chain.

1.2.1 Endoglucanase

Endoglucanases (3.2.1.4) play a major role in hydrolysis of cellulose polymer to cellodextrins. They belong to glycoside hydrolases (GH) in a systemof classification based on amino acid sequence and three-dimensional structure. (www.cazy.org). Endoglucanases are categorized in 13 GH families *viz.* GH5, GH6, GH7, GH8, GH9, GH12, GH44, GH45, GH48, GH51, GH61, GH74 and GH124 (www.cazy.org). Endoglucanase display both inverting and retaining type catalytic mechanisms (Davies & Henrissat, 1995). Most of the thermostable endoglucanases are found in GH families 5, 8 and 12 and 45 (Rye *et al.*, 2000 & Yeoman *et al.*, 2010). Endo β-1,4-glucanase from family 5 glycoside hydrolase is one of the largest families found in CAZy database (http://www.cazy.org/GH5.html). It is known to show variety of specificities, notably endoglucanase (cellulase), endomannase, exoglucanases, exomannases and β-mannosidase activities**.** One of the members of family 5 GH,

Cel5E from *C. thermocellum,* is a part of cellulosomal *celH,* comprising two N-terminal catalytic modules (Lic26A and Cel5E) and a C-terminal family 11 carbohydrate binding module (CBM11) and two Type I dockerins at C-terminal connected by C-terminal linker of CBM11 (Taylor *et al.,* 2005). This family 5 GH (Cel5E) was shown to hydrolyse both soluble and insoluble cellulolosic substrates (Bharali *et al.,* 2005). The enzyme activity of the family 5 GH, (Cel5E) endoglucanase (Gen Bank Acc No. ABN52701.1) was enhanced by attaching the linker at both N- and C-terminal of GH5 and by mutating Phe267 to Ala (Yuan *et al.,* 2015). The TLC analysis of Cel5E hydrolysed carboxymethyl cellulose (CMC) showed the production of cellodextrin as the final product (Yuan *et al.,* 2015).

1.2.2 Cellobiohydrolase

Cellobiohydrolases (CBH) (EC 3.2.1.91) release cellobiose from the Cellodextrins (Barr *et al.,* 1996). The CBH are of two types i.e. CBH-I and CBH-II. CBH-I and CBH-II produced cellobiose from the reducing and non-reducing ends respectively, of cellulose chains in the crystalline regions (Arantes *et al.,* 2010). Cellobiohydrolases consist of two conserve module, the catalytic module and carbohydrate binding module (Arantes *et al.,* 2010). Both modules are linked by peptide linker (Zhang *et al.,* 2006). CBHs bind the cellulose chain in a more or less closed tunnel and cleave off cellobiose units processively from one end of the cellulosic polymer (Sandgren *et al.,* 2013). CBH-I has N- terminal CBM1 domain (Hilden *et al.,* 2004). CBH-I attacks the cellulose chain at the reducing end while the CBH-II with C-terminal CBM1 domain acts at the non-reducing end (Liu *et al.,* 2011). The CBH-I has inverting mechanism while CBH-II has retaining mechanism (Barr *et*

al., 1996, Mertz *et al.*, 2007). CBH-I belongs to GH7 family while CBH-II belongs to GH6 family (Arantes *et al.*, 2010).

.

1.2.3 β-Glucosidase

β-Glucosidase (EC 3.2.1.21) synergistically plays a significant role as a cellulase enzyme. It converts cellobiose into monomeric glucose (Rani *et al.*, 2014). β-Glucosidases are reported in glycoside hydrolase (GH) families 1, 3, 5, 9, 30 and 116 (http://www.cazy.org/glycoside-hydrolases.html). β-Glucosidases mostly belong to either family 1 or family 3 of GH, while these are also found in GH families 5, 9, 30 and 116 (Henrissat, 1991, Opassiri *et al.*, 2007, Cantarel *et al.*, 2008). GH family 1 contains approximately, 62 β-glucosidases from archaebacteria, plants, mammals and also includes 6-phosphoglycosidases and thioglucosidases. Predominantly, GH1 enzymes, also show significant β-galactosidase activity (Yang *et al.*, 2015). β-Glucosidase from family 1 GH was found to hydrolyse cellodextrin up to five glucose chain length (Adlakha *et al.*, 2012). The family GH1 β-glucosidases are also classified as members of the 4/7 super family with a common $(\beta/\alpha)_8$ barrel fold. Family 3 of glycoside hydrolases contains around 44 β-glucosidases and hexosaminidases of bacterial, mold and yeast origin (Singhania *et al.*, 2013). The most studied fungal β-glucosidases belong to the family 3 of glycoside hydrolases (Singhania *et al.*, 2013). Structural data of GH3 enzymes is still limited. Structural data on GH1 showed canonical TIM-barrel $(\beta/\alpha)_8$ fold (Jones *et al.*, 2008). The active site analysis of β-Glucosidase of family 1 glycoside hydrolase hrom *Clostridium thermocellum* showed that E166 acts as acid/base residue and the E355 act as a nucleophile (Sharma *et al.*, 2018).

1.3 Importance of cellulases in bioethanol production

The plant lignocellulosic biomass structure can be partially broken down by using pretreatment process, making them accessible to enzymatic hydrolysis (Wyman *et al.*, 1999). Therefore, extensive research has been carried out on enzymatic degradation of cellulose and hemicellulose with significant progress (Thomsen *et al.*, 2008, Liu *et al.*, 2009). Hemicellulose can be hydrolyzed by mild-acid conditions. The cellulose being recalcitrant semi-crystalline polymer has been the target of most biomass degradation enzyme research (Paech 1994 and Himmel *et al.*, 2018). Historically, cellulases were first to be included in a sequential process of pretreatment of biomass followed by cellulase hydrolysis and ethanol fermentation (Ogier *et al.*, 1999). However, though there has been extensive research on cellulases (Sheehan, 2001, Lantz *et al.*, 2010, Khoshnevisan *et al.*, 2012), the cost of production of these enzymes have remained high, in comparison to the costs of proteases and amylases, for a few important reasons. These are mainly due to the complex and insoluble semi-crystalline nature of cellulose and low catalytic activity of cellulases as catalysts (Esteghlalian *et al.*,2001). Moreover, the complete biomass degradation requires multiple, related enzyme activities acting synergistically.

The degradation of cellulose polymer is an elaborate multistep process and the last step of this multistep process is the action of β-glucosidase on the cellobiose for converting it into glucose (Lynd *et al.*, 2002). Cellobiose being an inhibitor of both endoglucanase and cellobiohydrolase can be acted upon by β-glucosidase enzyme to reduce its effect. Similarly, glucose at high concentration can block the active site of β-glucosidase for the substrate or prevent the hydrolyzed substrate from leaving (Krogh *et al.*, 2010). To address this problem glucose tolerance were also incorporated

into these enzymes by protein engineering technologies (Riou *et al.*, 1998, Gunata and Vallier, 1999). In the present situation, commercially available cellulases for biomass hydrolysis contain β-glucosidase enzyme in the cocktail of enzymes in significant amount (Bidard *et al.*, 2019). This commercial enzyme claims to have improved β-glucosidase activity with higher titers, but the cost of these enzymes is very high. From this, it can be assumed that engineering the enzyme for improving the β-glucosidase activity and its production can overcome the limiting factors for bioethanol production.

1.4 Industrial production and global demand of cellulases

Enzymes are one of the biomolecules that have applications in different industries, such as brewing, dairy products, detergent, food and feed, pharmaceutical, paper and pulp industry. One such enzyme is cellulase. According to the latest global market analysis, the demand for the production of cellulases is increasing exponentially. For the production of cellulases using economic processes for biofuel industries, many factors are significant such as high substrate loading, low enzyme loading and a short hydrolysis period. In the previous report by Klein *et al*. 2012, it was stated that these factors could significantly reduce the production cost of enzymes at the commercial platform. Two companies, namely Genencor, USA and Novozymes, Denmark are known to produce cellulases. These companies have significantly contributed towards the development of efficient cellulases. Genencor has developed a cellulase complex known as Accelerase®1500, mainly for lignocellulosic biomass processing industries (Singhania *et al*., 2007) which is considered to be more cost-effective than previously developed, predecessor Accelerase®. Moreover, the Accelerase®1500 carries higher levels of β-glucosidase

enzyme activity as compared to other commercial cellulases and thus it acts efficiently for almost conversion of cellobiose into glucose (Penttila *et al*., 1998; Singhania *et al*., 2007). There are many other cellulases which can hydrolyze biomass along with β-glucosidase (Chandra *et al.*, 2010). Moreover, Novozymes has also developed various ranges of cellulase preparations depending upon the application. Novozymes has developed cellulases mixture or cocktails of multiple enzymes for hydrolysis of biomass. Moreover, Amano Enzyme Inc., Japan and MAP's Enzyme Limited India are the other enzymes developing industries which are actively engaged in the production of cellulase. Although, the number of cellulase producing companies around the globe are contributing towards the production and marketing of enzymes, however, solely few of them have developed cellulase for the conversion of biomass in a cost effective manner.

1.5 Protein engineering of cellulases

Cellulases are extensively used for a large number of industrial purposes and consequently, significant efforts have been made for improving their production and performance. Protein engineering has been used as a tool to study the underlying catalytic mechanism of these enzymes, as well as improving their activity. Protein engineering includes the mutagenesis of potential active residues and their kinetic analysis. Inactive mutants are often used to study the protein-ligand complexes at the three-dimensional level. The protein can be engineered by three strategies namely, directed evolution, rational designing and developing multifunctional chimeras. In directed evolution random mutagenesis of the target genes were performed and the variant with the improved activity were selected (Zhang *et al.*, 2006). In rational designing under protein engineering, the three-dimensional complexes of enzymes

have been used to design new strategies for modification and exploitation of the glycoside hydrolases for engineering the enzymes and modifying their functions (Huang *et al.,* 2012). The multifunctional chimeras were constructed by using molecular biology techniques for fusing two or more modules in a single polypeptide chains (Morais *et al.*, 2012). The multifunctional chimera developed can reduce the cost of production of several hydrolytic enzymes required for complete degradation of lignocellulosic biomass (McKee *et al.,* 2012). The detailed strategies for enzyme improvement were given in section 1.5.1, 1.5.2 and 1.5.3 and Fig. 1.5.

1.5.1 Directed evolution

Directed evolution or irrational design is an approach to non-informatic protein engineering technique. This technique uses natural selection processes to evolve proteins and selection of desired traits. It utilizes the DNA techniques for example, error-prone PCR (epPCR) and DNA shuffling to generate an extensive library of gene variants randomly (Fig. 1.5). Improved variants are identified by screening, that accurately reflects the desired properties (Maki *et al.,* 2009).

1.5.2 Rational designing

The advancement in the process of rational design was introduced after the development in recombinant DNA technology and site-directed mutagenesis. The approach considers the choice of a suitable enzyme, its crystallographic studies to identify the amino acid sites to be mutated and further characterization of this developed mutant(s) (Zhang *et al.,* 2006). Rational design considers a very intensive knowledge of the catalytic mechanism of proteins and additional information on function of proteins and their structures. Computational techniques are utilized to

visualize the protein structure and its active site. The amino acid residues can be identified that play significant role in docking of the substrate molecule to the active site and confer the overall protein stability. The site-directed mutagenesis at these sites gives a strategy to manipulate the enzyme performance (Mohanram *et al.*, 2013) (Fig. 1.5). The protein engineering technique by structure-based rational designing of *Ct*Cel5E from *Clostridium thermocellum* was performed by Yuan *et al.*, 2015. The technology developed an efficient mutant by site-directed mutagenesis of Phe267 to Ala, displaying improvement in the enzyme activity. Moreover, site-directed mutagenesis data confirmed that His168 of *Ct*Cel5E is essential for xylanase activity (Yuan *et al.*, 2015). The structure and function studies on *Ct*Cel5E revealed the substrate recognition mechanism and resulted in enhancing the performance of *Ct*Cel5E.

1.5.3 Multifunctional Chimera

Engineering multimodular enzymes can reduce the number of enzymes required for the complete conversion of plant biomass. Multifunctional hydrolases can act synergistically in hydrolysing complex carbohydrates. In nature, many organisms have multifunctional enzymatic properties for the degradation of polysaccharides. Natural, multifunctional enzymatic complexes are known as cellulosomes (Doi and Kosugi 2004). The enzymes may have broad substrate specificity (Wagschal *et al.*, 2009). The single polypeptide may contain multiple functional domains (Blum *et al.*, 2000). In protein engineering, these natural strategies can be utilized for the development of multifunctional enzymes (Fig. 1.5). These variants can be used for catalyzing two or more reactions for depolymerization

of the lignocellulosic biomass. Using the recombinant DNA technology development of chimeric enzyme with endoglucanase (Endo5A) and β-glucosidase (Gluc1C) modules from *Paenibacillus sp.* having higher catalytic efficiencies was achieved (Adlakha *et al.*, 2012). They observed that Gluc1C improves the efficiency of Endo5A and releases the reducing sugar from carboxymethyl cellulose. Therefore, the fusion of Gluc1C with Endo5A was performed. This developed bifunctional enzyme resulted in 3.3- and 2- fold higher activity towards β-glucosidase and endoglucanase, respectively. Lee *et al.*, 2011 developed chimeric enzymes using a β-glucosidase (*Cc*BG) from *Clostridium cellulovorans* fused it with three different types of cellulases from *Clostridium thermocellum* which were a cellulosomal endoglucanase (*Ct*CD), a cellulosomal exoglucanase CBHA (*Ct*CA) and a non-cellulosomal endoglucanase *Ct*C9I. Out of all the chimeras the *Ct*CD-*Cc*BG a chimera with cellulosomal endoglucanase (*Ct*CD) in the N-terminal and β-glucosidase (*Cc*BG) in the C-terminal showed 2-fold higher activity with phosphoric acid swollen cellulose (PASC) while retaining similar activity for β-glucosidase. Efforts were made to develop chimeric cellulases by attaching a Carbohydrate Binding Module (CBM3) to a family 9 endoglucanase, Cel9A (Telke *et al.*, 2012) and to a family 5 endoglucanase, Cel5H (Shi *et al.*, 2013) resulting in 8- to 12-fold higher catalytic efficiency. All these results showed that developing multifunctional enzymes by protein engineering is a cost-effective process that lead to efficient hydrolysis of biomass. Therefore, research on the development of multifunctional chimera needs continued efforts. The synergistic action of different enzymatic modules can be combined to enhance the overall efficiency of plant cell wall deconstruction.

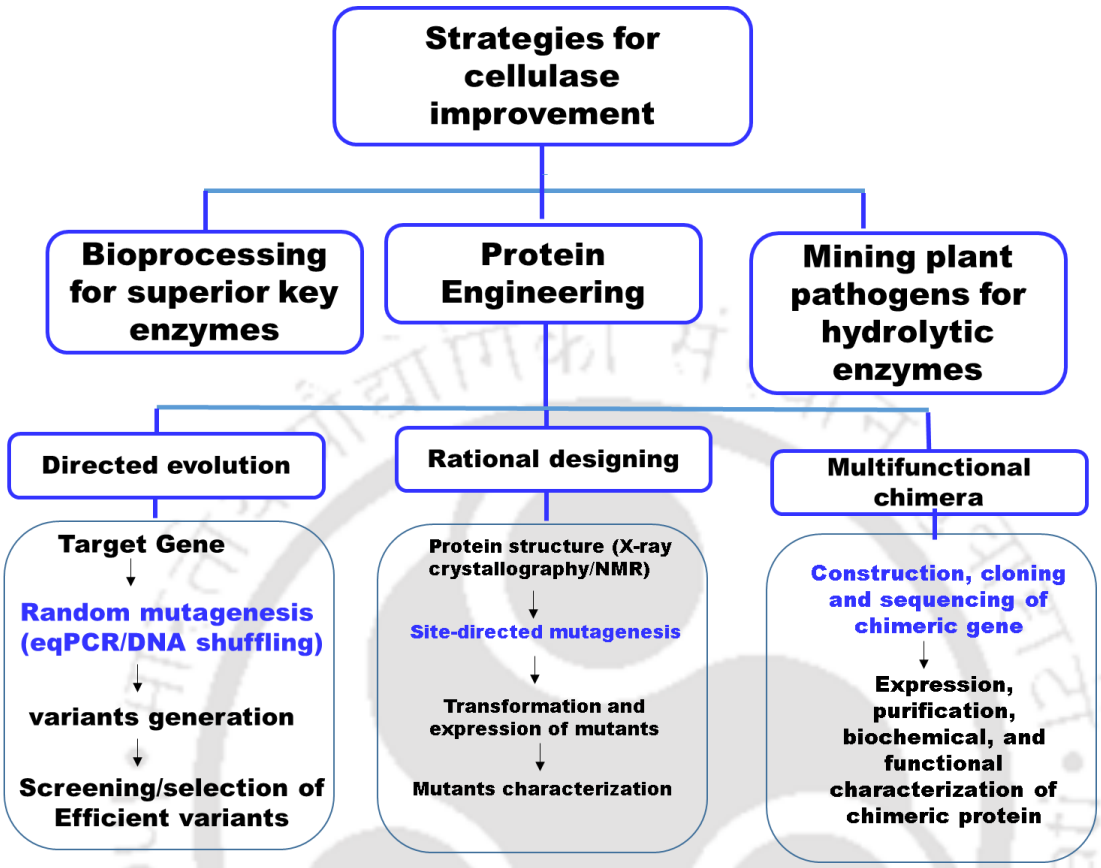


Fig. 1.5. Steps for advancement in enzyme technology (Mohanram *et al.*, 2013)

1.6 The microorganism

Clostridium thermocellum is an anaerobic, cellulolytic, ethanologenic and thermophilic Gram-positive bacterium. This bacterium can directly convert cellulose biomass into ethanol (Tachaapaikoon *et al*., 2012). The cellular structure of *Clostridium thermocellum* is similar to most rod-shaped bacteria as shown in Scanning Electron Microscopic images in Fig. 1.6 (Tachaapaikoon *et al*., 2012).

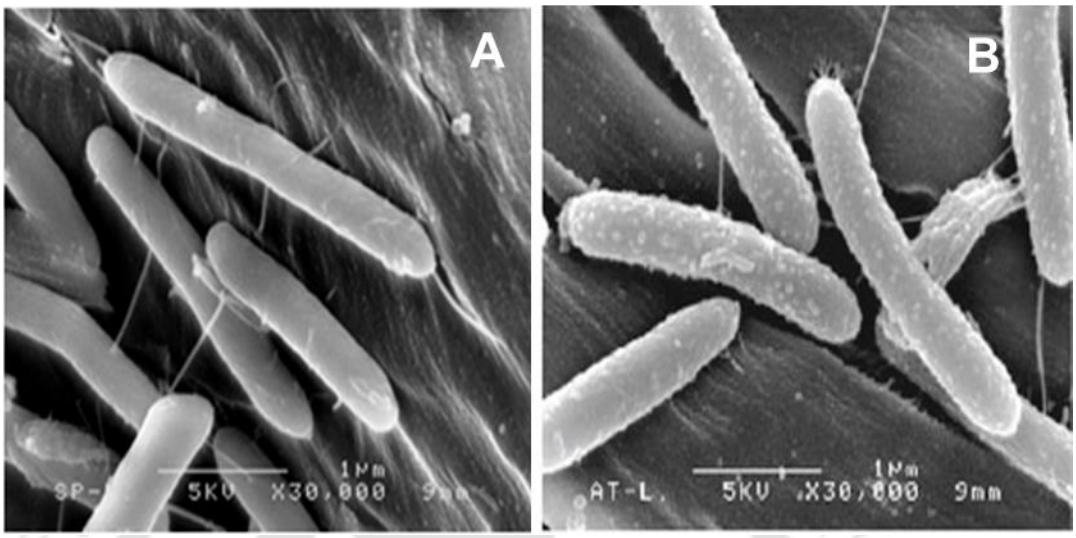


Fig. 1.6 Scanning electron micrographs of (A) *C. thermocellum* S14 (Tachaapaikoon *et al*., 2012) and (B) *C. thermocellum* ATCC 27405 (Tachaapaikoon *et al*., 2012)

1.6.1 Cellulosome structure

The cellulosome is a macromolecular complex. The cellulosome complex consists catalytic and noncatalytic modules, which are associated into the complex by forming a unique type of scaffolding subunit known as scaffoldin (Bayer *et al*., 2004). The components of the cellulosome shows intermodular interaction in a synergistic manner for the catalysis of the cellulose in the efficient manner (Bayer *et al.*, 2007). In 1983, the cellulosome concept was established in *Clostridium thermocellum* (Lamed *et al.*, 1983) wherein cellulases were found to be organized into high molecular weight cellulolytic complex (Bayer *et al.*, 2007). The molecular size of the Clostridial cellulosomal enzymes varies from about 40 to 180 kDa (Fontes and Gilbert, 2010). After the discoveries of the cellulosome from *C. thermocellum* various genes from this bacterium was cloned, sequenced and their modular structures were determined (Doi *et al*., 1994; Belaich *et al.*, 1997; Doi and Kosugi, 2004). In *C. thermocellum,* the cellulosome complex contains many different types of glycosyl hydrolases, including

cellulases, hemicellulases and even carbohydrate esterases, all of which are bound to a major polypeptide called scaffoldin (also known as the cellulosome-integrating protein, CipA). The roles of scaffoldin is mainly for cellulose-binding and cell-anchoring functions (Shoham *et al.*, 1999).

1.7 Significance and objectives of the present study

1.7.1 Significance of this study

The lignocellulosic biomass or plant biomass is composed of cellulose, hemicellulose and lignin. It is abundant in nature and can serve as the alternative energy source. There is an increasing effort for converting the cellulosic component to alcohols, which can serve as biofuels. The enzymatic hydrolysis is the critical step in the process for releasing fermentable sugar. For enzymatic hydrolysis, efficient enzymes are required for the complete hydrolysis of the lignocellulosic biomass. The high cost of commercially available *Trichoderma reesei* cellulolytic enzymes used for enzymatic hydrolysis is of prime concern (Chowdary *et al.*, 2001). The cellulosome of *Clostridium thermocellum* is reported to have 50-fold higher specific activity against crystalline cellulose than the corresponding fungal system (Gilbert and Fontes, 2010). The existing recombinant cellulolytic enzymes of *Clostridium thermocellum* with substrate specificities and catalytic efficiency against cellulosic substrates were screened. The three dimensional structures can help in identifying the active site and substrate binding residues involved in binding and catalysis of different substrates. The structural study will help in identifying the underlying mechanism of action these enzymes. This will help in designing variants of this enzymes which may show improved catalytic efficiency.

In this study, an endoglucanse *(Ct*GH5) of family 5 glycoside hydrolase from *Clostridium thermocellum* will be mutated by site-directed mutagenesis for possibility of enhancing endoglucanase activity. The mutagenesis will be performed on a substrate binding residue Phe194 to develop a mutant Ala194. The mutation was adapted from the previous mutagenesis study performed by Yuan *et al.*, 2015 where, enzyme activity of the family 5 GH, (Cel5E) endoglucanase (Gen Bank Acc No. ABN52701.1) was enhanced by attaching the linker at both N- and C-terminal of GH5 and by mutating Phe267 to Ala.

Another aspect of complete hydrolysis of lignocellulose biomass is the requirement of the multiple enzymes. The production of multiple enzymes for complete hydrolysis of lignocellulose biomass will increase the overall cost of biofuel production. This cost can be minimized by engineering enzymes for improving its catalytic efficiency and decreasing the number of enzymes by fusing two or more enzymes. Therefore, for development of an enzyme with different activity, the mutant *Ct*GH5-F194A will be fused with β-1,4-glucosidase (*Ct*GH1) from family 1 glycoside hydrolase from *Clostridium thermocellum* to construct a bi-functional enzyme. The *Ct*GH1 was previously cloned, expressed and biochemically characterised by Sharma *et al.*, 2019. The fusion of the two constructs *Ct*GH1 and *Ct*GH5-F194A to develop a Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) will be performed using a natural N-terminal linker from the cellulosomal gene *cel*H from *C. thermocellum* (Taylor *et al.*, 2005). The catalytic efficiency of the developed Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) will be compared with its individual modules using biochemical and functional characterization. The structural characterization of chimeric enzyme will be performed

to analyze the arrangement of domains in the chimeric form using *in silico* molecular modeling approach. The molecular dynamics simulation on the chimeric enzyme will be performed for validation of the protein model to explore the protein conformation. Moreover, the SAXS (small angle X-ray scattering) analysis of the chimeric enzyme will be performed to study the arrangement of the two modules of chimera in the solution form. The action of the developed bi-functional Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) with Cellobiohydrolase (CBH5A) from *Clostridium thermocellum* on the alkali followed by organosolv pretreated Sugarcane bagasse will be performed to estimate the reducing sugars released from the pretreated biomass. The glucose released from the action of the cocktail of cellulases enzymes consisting of Chimera and CBH5A will be quantified by HPLC. The released sugars will taken for further fermentation by fermentative microbes at shake flask level and analytical assays will be performed.

1.7.2 Specific objectives of the study

1. Site-directed mutagenesis of wild-type endoglucanase (*Ct*GH5) to develop mutant *Ct*GH5-F194A and construction of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) using mutant *Ct*GH5-F194A and β-glucosidase (*Ct*GH1).
2. Comparative biochemical characterization of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual catalytic modules - β-glucosidase (*Ct*GH1) and mutant endoglucanase (*Ct*GH5-F194A).
3. SAXS and computational approaches for determining structure and binding characteristics of chimera (*Ct*GH-L1-*Ct*GH5-F194A).
4. Saccharification of alkali and organosolv pretreated Sugarcane bagasse by cocktail of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and Cellobiohydrolase (CBH5A).

References

Adlakha, N., Sawant, S., Anil, A., Lali, A., & Yazdani, S. S. (2012). Specific fusion of β-1, 4-endoglucanase and β-1, 4-glucosidase enhances cellulolytic activity and helps in channeling of intermediates. *Applied and Environmental Microbiology*, *78*, 7447-7454

Arantes, V., & Saddler, J. N. (2010). Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnology for Biofuels*, *3*, 4.

Aulitto, M., Fusco, S., Nickel, D. B., Bartolucci, S., Contursi, P., & Franzén, C. J. (2019). Seed culture pre-adaptation of *Bacillus coagulans* MA-13 improves lactic acid production in simultaneous saccharification and fermentation. *Biotechnology for Biofuels*, *12*, 45.

Bayer, E.A., Belaich, J.P., Shoham, Y. & Lamed, R. (2004) The cellulosomes: multi-enzyme machines for degradation of plant cell wall polysaccharides. *The Annual Review of Microbiology*, *58*, 521-554.

Bayer, E.A., Lamed, R. & Himmel, M.E. (2007) The potential of cellulases and cellulosomes for cellulosic waste management. *Current Opinion in Biotechnology*, *18*, 237-245.

Belaich, J.P., Tardif, C., Belaich, A. & Gaudin, C. (1997) The cellulolytic system of *Clostridium cellulolyticum*, *Journal of Biotechnology*, *57*, 3–14.

Bhatia, Y., Mishra, S., & Bisaria, V. S. (2002). Microbial β-glucosidases: cloning, properties, and applications. *Critical Reviews in Biotechnology*, *22*, 375-407.

Bidard-Michelot, F., Tong, L. C. H., Margeot, A., & Cohen, C. (2019). *U.S. Patent Application No. 16/142,929*.

Blum, D. L., Kataeva, I. A., Li, X. L., & Ljungdahl, L. G. (2000). Feruloyl esterase activity of the *Clostridium thermocellum* cellulosome can be attributed to previously unknown domains of XynY and XynZ. *Journal of Bacteriology*, *182*, 1346-1351.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2008). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research, 37*, D233-D238.

Chandra M., Kalra A., Sharma, P.K., Kumar H. & Sangwan, R.S. (2010) Optimization of cellulases production by *Trichoderma citrinoviride* on marc of *Artemisia annua* and its application for bioconversion process. *Biomass Bioenergy, 34*, 805–811.

Claassen, P.A.M., Van Lier, J.B., Lopez Contreras, A.M., VanNiel, E.W.J., Sijtsma, L., & Stams, A.J.M. (1999). Utilization of biomass for the supply of energy carriers. *Applied Microbiology* and *Biotechnology*, *52*, 741-755.

Rye, C.S., & Withers, S.G. (2000), *Current Opinion in Chemical Biology*. 4, 573-580.

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

Doi, R. H., & Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nature Reviews Microbiology*, *2*, 541-551.

Doi, R.H., Goldstein, M., Hashida, S., Park, J.S. & Takagi, M. (1994) The *Clostridium cellulovorans* cellulosome. *Critical Reviews in Microbiology, 20,* 87-93.

Taylor, E.J., Goyal, A., Guerreiro, C.I., Prates, J.A., Money, V., Ferry, N., Gilbert H.J., How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides. (2005). Structure and activity of a *Clostridium thermocellum* lichenase, *Ct*Lic26A. *Journal of Biological Chemistry*. *280*, 32761–32767.

Esteghlalian, A. R., Srivastava, V., Gilkes, N., Gregg, D. J., & Saddler, J. N. (2001). An overview of factors influencing the enzymatic hydrolysis of lignocellulosic feedstocks. *ACS Symposium Series.769*, 100-111.

Fisher, A. B., & Fong, S. S. (2014). Lignin biodegradation and industrial implications. *AIMS Bioengineering*, *1*, 92-112.

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

Galbe, M., & Zacchi, G. (2007). Pretreatment of lignocellulosic materials for efficient bioethanol production. *Biofuels*. Springer, Berlin, Heidelberg. 41-65.

GESY (Global Energy Statistical Yearbook), 2017. Source: https://yearbook.enerdata.net/. Paech, Christian. "Approaches to Cellulase Purification. 1994. 130-178.

Gunata, Z., & Vallier, M. J. (1999). Production of a highly glucose-tolerant extracellular β-glucosidase by three *Aspergillus strains*. *Biotechnology Letters*, *21*, 219-223.

Hilden, L., & Johansson, G. (2004). Recent developments on cellulases and carbohydrate-binding modules with cellulose affinity. *Biotechnology Letters*, *26*, 1683-1693.

Himmel, M. E., Adney, W. S., Baker, J. O., Nieves, R. A., & Thomas, S. R. (2018). Cellulases: structure, function, and applications. In *Handbook on Bioethanol*, 143-161.

Huang, J. W., Cheng, Y. S., Ko, T. P., Lin, C. Y., Lai, H. L., Chen, C. C., & Liu, J. R. (2012). Rational design to improve thermostability and specific activity of the truncated *Fibrobacter succinogenes* 1, 3-1, 4-β-D-glucanase. *Applied Microbiology and Biotechnology*, *94*, 111-121.

Jones, B. V., Begley, M., Hill, C., Gahan, C. G., & Marchesi, J. R. (2008). Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proceedings of the National Academy of Sciences USA*, *105*, 13580-13585.

Kazi, F. K., Fortman, J. A., Anex, R. P., Hsu, D. D., Aden, A., Dutta, A., & Kothandaraman, G. (2010). Techno-economic comparison of process technologies for biochemical ethanol production from corn stover. *Fuel*, *89*, S20-S28.

Khoshnevisan M.D., Oleskowicz P.P, Simmons B.A. & Blanch H.W. (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology* & *Bioengineering*,*109*,1083–1087.

Krogh, K. B., Harris, P. V., Olsen, C. L., Johansen, K. S., Hojer-Pedersen, J., Borjesson, J., & Olsson, L. (2010). Characterization and kinetic analysis of a thermostable GH3 β-glucosidase from *Penicillium brasilianum*. *Applied Microbiology and Biotechnology*, *86*, 143-154.

Lamed, R., Setter, E., Kenig, R., & Bayer, E. A. (1983). The cellulosome-a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate

antigenic, cellulose-binding and various cellulolytic activities. *Biotechnology* and *Bioengineering*, *13*, 163-181.

Lantz, S. E., Goedegebuur, F., Hommes, R., Kaper, T., Kelemen, B. R., Mitchinson, C., & Larenas, E. A. (2010). *Hypocrea jecorina* CEL6A protein engineering. *Biotechnology for Biofuels*, *3*, 20.

Lee, H. L., Chang, C. K., Teng, K. H., & Liang, P. H. (2011). Construction and characterization of different fusion proteins between cellulases and β-glucosidase to improve glucose production and thermostability. *Bioresource Technology*, *102*, 3973-3976.

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

Liu, Y. S., Baker, J. O., Zeng, Y., Himmel, M. E., Haas, T., & Ding, S. Y. (2011). Cellobiohydrolase hydrolyzes crystalline cellulose on hydrophobic faces. *Journal of Biological Chemistry*, *286*, 11195-11201.

Liu, L., Sun, J., Cai, C., Wang, S., Pei, H., & Zhang, J. (2009). Corn stover pretreatment by inorganic salts and its effects on hemicellulose and cellulose degradation. *Bioresource Technology*, *100*, 5865-5871.

Lynd, L. R., Weimer, P. J., Van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, *66*, 506-577.

Lynd, L.R., Mark, S. L, David, B., Bruce E. D, Brian, D., Richard H., & Michael. H. (2008). How biotech can transform biofuels. *Nature Biotechnology*, *26*, 169.

McDonald, J. E., Rooks, D. J., & McCarthy, A. J. (2012). Methods for the isolation of cellulose-degrading microorganisms. In *Methods in Enzymology*, *510*, 349-374.

McKee, L. S., Peña, M. J., Rogowski, A., Jackson, A., Lewis, R. J., York, W. S., & Marles-Wright, J. (2012). Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proceedings of the National Academy of Sciences USA*, *109*, 6537-6542.

Maki, M., Leung, K. T., & Qin, W. (2009). The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International Journal of Biological Sciences*, *5*, 500.

Mathapati, P. R., Ghasghase, N. V., & Kulkarni, M. K. (2010). Study of *Saccharomyces cerevisiae* 3282 for the production of tomato wine. *International Journal* of *Chemical* Sciences, *1*, 5-15.

Meher, L. C., Sagar, D. V., & Naik, S. N. (2006). Technical aspects of biodiesel production by transesterification-a review. *Renewable and Sustainable Energy Reviews*, *10*, 248-268.

Mertz, B., Hill, A. D., Mulakala, C., & Reilly, P. J. (2007). Automated docking to explore subsite binding by glycoside hydrolase family 6 cellobiohydrolases and endoglucanases. *Biopolymers: Original Research on Biomolecules*, *87*, 249-260.

Mohanram, S., Dolamani A., Jairam C., Arora, A., & Nain, L., (2013). Novel perspectives for evolving enzyme cocktails for lignocellulose hydrolysis in biorefineries. *Sustainable Chemical Processes, 1*, 15.

Morais S, Barak Y, Lamed R, Wilson DB, Xu Q, Himmel ME, & Bayer EA:, (2012). Paradigmatic status of an endo- and exo-glucanases and its effect on crystalline cellulose degradation. *Biotechnology* for *Biofuels*, *5*, 78

Nsor-Atindana, J., Chen, M., Goff, H. D., Zhong, F., Sharif, H. R., & Li, Y. (2017). Functionality and nutritional aspects of microcrystalline cellulose in food. *Carbohydrate Polymers*, *172*, 159-174.

Ogier, JC, Leygue, JP, Ballerini, D., Rigal, L., & Pourquie, J. (1999). Ethanol production from lignocellulosic biomass; *Revue de l`Institut Francais du Petrole*, *54*, 67-94.

Opassiri, R., Pomthong, B., Akiyama, T., Nakphaichit, M., Onkoksoong, T., Cairns, M. K., & Cairns, J. R. K. (2007). A stress-induced rice (*Oryza sativa L.*) β-glucosidase represents a new subfamily of glycosyl hydrolase family 5 containing a fascin-like domain. *Biochemical Journal*, *408*, 241-249.

Penttila, M. (1998). Heterologous protein production in Trichoderma. *Trichoderma & Gliocladium-Enzymes*, *2*, 365-382.

Rani, V., Mohanram, S., Tiwari, R., Nain, L., & Arora, A. (2014). Beta-glucosidase: key enzyme in determining efficiency of cellulase and biomass hydrolysis. *Journal of Bioprocessing and Biotechniques*, *5*, 2.

Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., & Eckert, C. A. (2006). WJF Jr., JP Hallett, DJ Leak, CL Liotta, JR Mielenz, R. Murphy, R. Templer and T. Tschaplinski. *Science*, *311*, 484.

REN21, 2006. Renewable Energy Policy Network for the 21stcentury. Renewables Global Status Report 2006 (Source: https://web.archive.org/web/

20110718181410/http://www.ren21.net/globalstatusreport/download/RE_GSR_2006_Update.pdf.

Sandgren, M., Wu, M., Karkehabadi, S., Mitchinson, C., Kelemen, B. R., Larenas, E. A., & Hansson, H. (2013). The structure of a bacterial cellobiohydrolase: the catalytic core of the *Thermobifida fusca* family GH6 cellobiohydrolase Cel6B. *Journal of Molecular Biology*, *425*, 622-635.

Saravanakumar, K., Senthilraja, P., & Kathiresan, K. (2013). Bioethanol production by mangrove-derived marine yeast, *Sacchromyces cerevisiae*. *Journal of King Saud University-Science*, *25*, 121-127.

Singhania R.R., Sukumaran R.K. and Pandey A. (2007) Improved cellulase production by *Trichoderma reesei* RUT C30 under SSF through process optimization. *Biotechnology and Applied Biochemistry, 142*, 60–70.

Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C., & Pandey, A. (2013). Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresource Technology*, *127*, 500-507.

Sheehan, J. (2001). The road to bioethanol: a strategic perspective of the US Department of Energy's national ethanol program.

Shoham, Y., Lamed, R., & Bayer, E. A. (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends in Microbiology*, *7*, 275-281.

Sorek, N., Yeats, T. H., Szemenyei, H., Youngs, H., & Somerville, C. R. (2014). The implications of lignocellulosic biomass chemical composition for the production of advanced biofuels. *BioScience*, *64*, 192-201.

Tachaapaikoon, C., Kosugi, A., Pason, P., Waeonukul, R., Ratanakhanokchai, K., Kyu, K..L. Arai, T., Murata, Y. and Mori, Y. (2012) Isolation and characterization of a new cellulosome-producing *Clostridium thermocellum* strain. *Biodegradation*, *23*, 57-68.

Thomsen, M. H., Thygesen, A., & Thomsen, A. B. (2008). Hydrothermal treatment of wheat straw at pilot plant scale using a three-step reactor system aiming at high hemicellulose recovery, high cellulose digestibility and low lignin hydrolysis. *Bioresource Technology*, *99*, 4221-4228.

Tong, Z., Pullammanappallil, P., & Teixeira, A. A. (2012). How ethanol is made from cellulosic biomass. *University of Florida IFAS Extension*, 1-4.

Wagschal, K., Heng, C., Lee, C. C., & Wong, D. W. (2009). Biochemical characterization of a novel dual-function arabinofuranosidase/xylosidase isolated from a compost starter mixture. *Applied Microbiology and Biotechnology*, *81*, 855-863.

Wyman, C. E. (1999). Biomass ethanol: technical progress, opportunities, and commercial challenges. *Annual Review of Energy and the Environment*, *24*, 189-226.

WER (World Energy Report), 2016. World Energy Council, ISBN 9780946121588.; Source: https://www.worldenergy.org.

Yang, F., Yang, X., Li, Z., Du, C., Wang, J., & Li, S. (2015). Overexpression and characterization of a glucose-tolerant β-glucosidase from Thermoanaerobacterium *aotearoense* with high specific activity for cellobiose. *Applied Microbiology and Biotechnology*, *99*, 8903-8915.

Yeoman, C. J., Han, Y., Dodd, D., Schroeder, C. M., Mackie, R. I., & Cann, I. K. (2010). Thermostable enzymes as biocatalysts in the biofuel industry. In *Advances in Applied Microbiology*. Academic Press*, 70,* 1-55

Yuan, S. F., Wu, T. H., Lee, H. L., Hsieh, H. Y., Lin, W. L., Yang, B., & Ho, M. C. (2015). Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum. Journal of Biological Chemistry*, *290*, 5739-5748.

Zainab, B., & Fakhra, A. (2014). Production of Ethanol by fermentation process by using Yeast *Saccharomyces cerevisae*. *International Research Journal of Environmental Sciences*, *3*, 24-32.

Zhang, Y. H. P., Himmel, M. E., & Mielenz, J. R. (2006). Outlook for cellulase improvement: screening and selection strategies. *Biotechnology Advances*, *24*, 452-481.

Chapter 2

Site-directed mutagenesis of wild-type endoglucanase (*Ct*GH5) to develop mutant *Ct*GH5-F194A and construction of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) using mutant *Ct*GH5-F194A and β-glucosidase (*Ct*GH1)

2.1 Introduction

The risk of depletion of fossil fuels as an energy source has led to the development of alternative strategies to overcome this challenge. Production of bio-ethanol from cheap lignocellulosic biomass is one of the most promising approaches to address this problem. Plant cell walls contain cellulose, hemicellulose and lignin which can be converted into alcohol, thus plant biomass can serve as alternative energy source. Cellulose is the most abundant polymer (polysaccharide) and for its complete break down three enzymes are required i.e. endoglucanase, cellobiohydrolase and β-glucosidase (Beguin *et al.,* 1990, Yennamalli *et al.,* 2013). Endo-β-1,4-glucanase hydrolyses cellulose backbone randomly and produces cellodextrin as final product (Urbanowicz *et al.,* 2007). Cellobiohydrolase binds at the end of the cellodextrin and releases cellobiose as the major product (Barr *et al.,* 1996), while β-glucosidase cleaves cellobiose to form two molecules of glucose (Grabnitz *et al.,* 1991).

For efficient hydrolysis of cellulosic biomass, it is important to engineer the enzymes with improved catalytic efficiency. Moreover, the requirement of multiple

enzymes for complete degradation of lignocellulosic biomass is a cost intensive process (McKee *et al.,* 2012). Therefore, to reduce the cost of production of multiple enzymes, the development of multifunctional Chimera is beneficial that can catalyze two or more reactions for efficient hydrolysis of biomass (Lu *et al.,* 2008, Fan, *et al.,* 2009). Efforts were made to develop chimeric cellulases by attaching a Carbohydrate Binding Module (CBM3) to a family 9 endoglucanase, Cel9A (Telke *et al.,* 2012) and to a family 5 endoglucanase, Cel5H (Shi *et al.,* 2013) resulting in 8- to 12-fold higher catalytic efficiency. Telke *et al.,* 2012 also proved from the computational protein modelling that the active site configuration of native catalytic module does not get affected upon fusing the two modules. Therefore, other catalytic modules with different substrate specificities and mode of actions can be introduced in a single polypeptide chain for development of chimeric cellulases with increased catalytic efficiency.

The family 5 glycoside hydrolase, GH5 is one of the largest and well characterized families (http://www.cazy.org/GH5.html). One of the members of family 5 GH, Cel5E from *C. thermocellum,* is a part of cellulosomal *celH,* comprising two N-terminal catalytic modules (Lic26A and Cel5E) and a C-terminal family 11 carbohydrate binding module (CBM11) and two Type I dockerins at C-terminal connected by C-terminal linker of CBM11 (Taylor *et al*., 2005). This family 5 GH (Cel5E) was shown to hydrolyse both soluble and insoluble cellulolosic substrates (Bharali *et al.,* 2005). The enzyme activity of the family 5 GH, (Cel5E) endoglucanase (Gen Bank Acc No. ABN52701.1) was enhanced by attaching the linker at both N- and C-terminal of GH5 and by mutating Phe267 to Ala (Yuan *et al.,* 2015). The TLC analysis of Cel5E hydrolysed carboxymethyl cellulose (CMC) showed the production of cellodextrin as the final product (Yuan *et al.,* 2015). β-Glucosidase from family 1

GH was found to hydrolyse cellodextrin up to a five glucose chain length (Adlakha *et al.*, 2012). The fusion of catalytic domains of family 5 GH and family 1 GH will produce bifunctional enzyme that could alone convert cellulolosic biomass to glucose as a major product.

In the present study the site-directed mutagenesis of wild-type endoglucanase, *Ct*GH5 (Gen Bank Acc No. ABN52701.1) from *Clostridium thermocellum* was performed to develop its mutant, *Ct*GH5-F194A by following the method as mentioned by Yuan *et al.*,2015. The wild-type endoglucanase (CtGH5) was previously cloned, expressed and characterised by Bharali *et al*., 2005. The catalytic efficiency of mutant, *Ct*GH5-F194A was enhanced by 2 fold as compared with the wild-type *Ct*GH5 (Nath *et al*., 2019). The mutant of endoglucanase, *Ct*GH5-F194A was fused with β-glucosidase, *Ct*GH1 (Genbank Acc No. CAA42814.1) from *Clostridium thermocellum* to construct a Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) for development of a bifunctional enzyme. The β-glucosidase (*Ct*GH1) was previously cloned and characterised (Sharma *et al.*, 2019). All the proteins were expressed in *Escherichia coli* BL21 cells and purified by immobilized metal ion affinity chromatography (IMAC) for further biochemical, functional and structural characterization.

2.2 Materials and Methods

2.2.1 Chemicals, reagents and kits

The mutagenic primer for the mutagenesis of wild-type *Ct*GH5 and the oligonucleotides for the PCR amplification of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) were procured from Eurofins, India. The site-directed mutagenesis kit was obtained from Agilent Technologies India Pvt. Ltd. The DNA polymerase was procured from Thermo Fisher Scientific, USA. dNTPs and $MgCl_2$ were obtained from MP Biomedicals, India Pvt. Ltd. PCR tubes (0.2 ml) were supplied by Axygen, Germany. Restriction enzymes *Nhe*I, and *Xho*I were obtained from NEB, UK. The expression vector, pET28a(+) was procured from Novagen, Germany. T_4 DNA ligase and 10x ligase buffers were supplied by Takara Bio Inc, UK. RNAse solution (20 mg/ml), glacial acetic acid (99.9% pure) Trizma base (Tris free base), ethidium bromide, Bradford reagent, nuclease free water (pH 8.0) and components of polyacrylamide gel electrophoresis were purchased from Sigma-Aldrich, India. The GenElute miniprep plasmid isolation kit and GenElute gel-extraction kit was obtained Sigma-Aldrich, India. DNA was electrophoresed on agarose gels prepared using Agarose, with low EEO purchased from Sigma-Aldrich, India. DNA marker, 1kb plus DNA ledder was purchased from NEB, UK. Protein markers were procured from Biobharti, India and GeNei, India. Disodium ethylenediamine tetra acetate salt (EDTA), glucose, sodium hydroxide, sodium dodecyl sulphate (SDS), LB medium were supplied by Himedia Pvt. Ltd., India. The antibiotic, kanamycin and ampicilin was procured from Himedia Pvt. Ltd., India. SDS-PAGE was performed using Mini-PROTEAN Tetra Cell purchased from Bio-Rad Laboratories (India) Private Limited. The protein staining dye Coomassie Brilliant Blue R250 was procured from Himedia Pvt. Ltd., India and methanol from Merck, India.

2.2.2 Microorganisms

The commercially available *E. coli* DH5α and *E. coli* BL-21 (DE3) cells were purchased from Novagen, Germany.

2.2.3 Site-directed mutagenesis of wild-type endoglucanase (*Ct*GH5)

The wild-type endoglucanase (*Ct*GH5) was previously cloned, expressed and characterised by Bharali *et al.,* 2005. The Phe194 residue of wild type endoglucanase (*Ct*GH5) was mutated to alanine by site-directed mutagenesis. The wild-type plasmid *Ct*GH5-pET21a(+) was used as the template and the mutagenic forward and reverse primers were used for PCR. The details of primers are provided in Table 2.1. Components of 50 µl PCR amplification was performed in a thermal cycler (Applied Biosystems, GeneAmp PCR System 9700) given in Table 2.2 and the PCR cycles for amplification are given in Table 2.3. The total PCR product was digested by *Dp*nI using 1µl (0.6U/µl) by incubating the reaction at 37ºC for 2 hours to digest the parental DNA. The *Dpn*I digested product was transformed in *E. coli* DH5α competent cells as mentioned in section 2.2.9.3. The plasmid was isolated using sigma plasmid isolation kit mentioned in section 2.2.9.4. The mutation was confirmed by sequencing. The mutant plasmid was then transformed in *E. coli* BL21 cells as mentioned in section 2.2.10 and its expression was achieved. The mutant *Ct*GH5-F194A was further engineered for development of Chimera by fusing *Ct*GH1 (β-glucosidase) from *C. thermocellum* to its N-terminal using the natural linker.

Table. 2.1 Mutagenic primers for the site-directed mutagenesis of wild-type *Ct*GH5 and oligonucleotide primers used for PCR amplification of genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

Construct	Primers	Sequence
Mutant F194A	Mutagenic forward	TGAGACCCATACTTAATCGGAACT**GCC**CATTACTATGACCCATATG
	Mutagenic reverse	CATATGGGTCATAGTAATG**GGC**AGTTCCGATTAAGTATGGGTCATCA
*Ct*GH1-L1-*Ct*GH5-F194A	*Ct*GH1-F1	CG**GCTAGC**ATGTCAAAGATAACTTTCCC
	*Ct*GH1-*Ct*GH5R2	ACTAAAAGCCTTTTTTTCATTGGTGAGCTTGATGGTGGAGTAGAGGTCCA AGTTGGAGTAGGGGTAGGATTTGATGAAAAACCGTTGTTTTTGATTA
	*Ct*GH1-*Ct*GH5F2	TAATCAAAAACAACGGTTTTTCATCAAATCCTACCCCTACTCCAACTTGG ACCTCTACTCCACCATCAAGCTCACCAAAGGCTGTCGACCCCTTTGA
	*Ct*GH5-R1	CC**CTCGAG** ATTAAATAGTGCATTGAGGA

Nucleotides shown in bold represent mutagenic site and restriction enzyme sites in the development of mutant CtGH5-F194A and Chimera (CtGH1-L1-CtGH5-F194A) respectively.

Table. 2.2 Site-directed mutagenesis of wild-type *Ct*GH5 for generation of mutant *Ct*GH5-F194A.

Reaction Component	Final Concentration	Volume Used (µl)
Nuclease free water	-	28
10x reaction buffer	1x	5
Template wild-type *Ct*GH5(33ng/µl)	3.3 ng/µl	5
dNTPs (2.5mM)	0.06 mM	1
Forward primer (15µM)	1.5 µM	5
Reverse primer (15 µM)	1.5 µM	5
Pfu DNA polymerase(2.5U/ µl)	0.05U/µl	1
Total Reaction Volume		50 µl

Table. 2.3 PCR conditions for generation of mutant *Ct*GH5- F194A

Steps		Time
I. Denaturation at 95°C		60 s
II. 18 cycles of	i) Denaturation at 95°C	30 s
	ii) Annealing at 55°C	60 s
	iii) Extension at 68°C	360 s
III. Final extension at 68°C		9 min

2.2.4 Cloning of Chimeric gene (*Ct*GH1-L1-*Ct*GH5-F194A) by fusing the genes encoding mutant endoglucanase (*Ct*GH5-F194A) and glucosidase (*Ct*GH1)

2.2.4.1 Strategy for Chimera construction

The molecular architecture of *Ct*CelH from *Clostridium thermocellum* is shown in Fig. 2.1A, where wild-type endoglucanase, *Ct*GH5 was used for site-directed mutagenesis of amino acid residue Phe194 to Ala for developing mutant, *Ct*GH5-F194A by following the method as reported earlier (Yuan *et al.*,2015). The wild-type endoglucanase (*Ct*GH5) was previously cloned, expressed and characterised by Bharali *et al.*, 2005. The molecular architecture of *Ct*GH1 from *Clostridium thermocellum* is shown in Fig. 2.1B. The β-glucosidase (*Ct*GH1) was previously cloned in pET28a(+) vector, expressed and biochemically characterised (Sharma *et al.*, 2019). The N-terminal linker, L1 (57 bp) of *Ct*GH5 (Fig. 2.1A) was used for Chimera construction by fusing β-glucosidase (*Ct*GH1) and mutant endoglucanase (*Ct*GH5-F194A) modules from *Clostridium thermocellum* as shown in Fig. 2.2. The N-terminal β-glucosidase (*Ct*GH1) module of Chimera consists of 445 amino acids and the C-terminal endoglucanase (*Ct*GH5-F194A) module consists of 309 amino acids (Fig. 2.2). The natural linker consists of 19 amino acids as shown in Fig. 2.2.

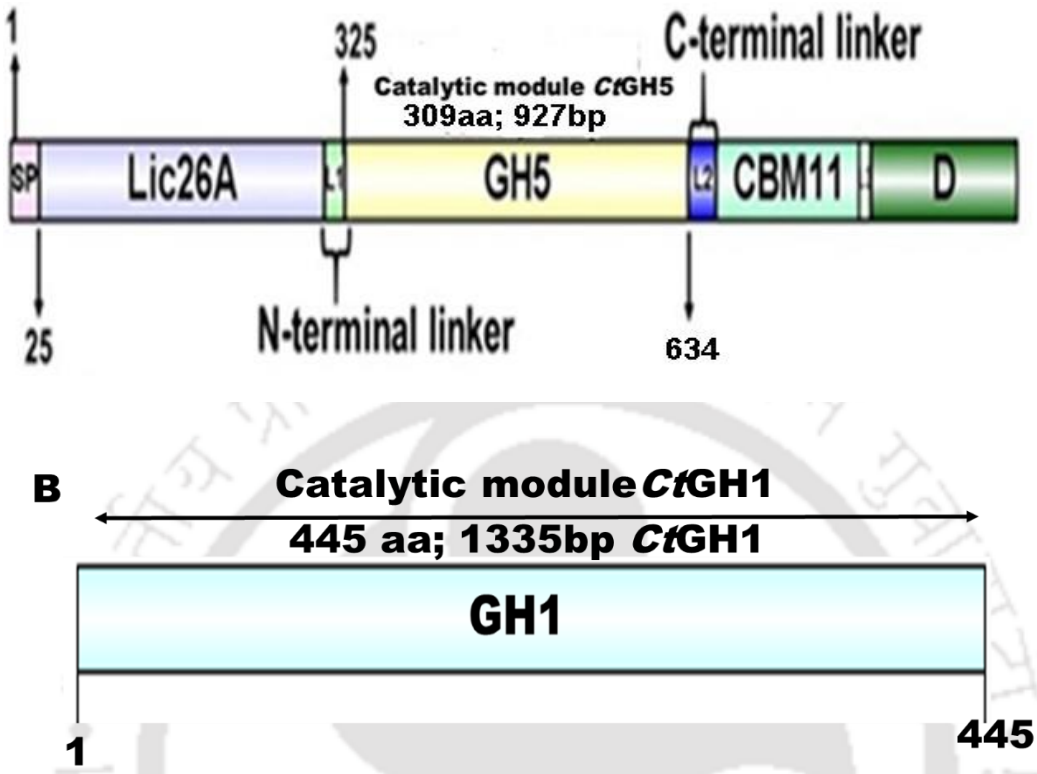


Fig. 2.1 A) Molecular architechture *Ct*CelH from *Clostridium thermocellum,* The domain *Ct*GH5 used for site-directed mutagenesis of amino acid residue Phe194 and the N-terminal (57 bp) used for Chimera construction by fusion of *Ct*GH1 and *Ct*GH5-F194A B) β-glucosidase (*Ct*GH1) from *C. thermocellum* was fused at the N- terminal of *Ct*GH5-F194A.

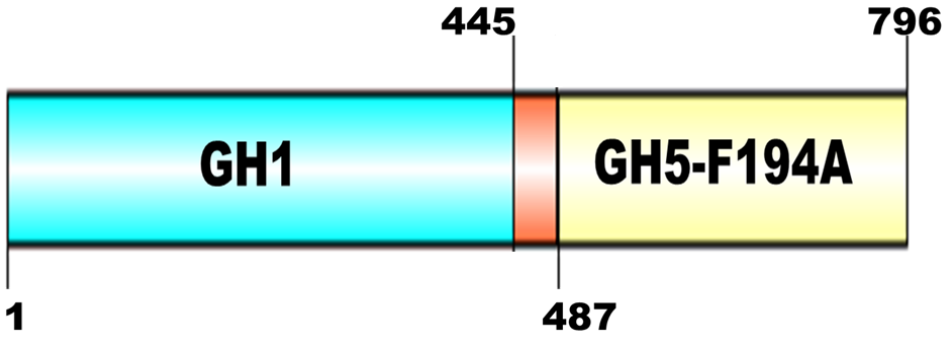


Fig. 2.2 Molecular architecture of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) showing The Chimera is made up of two modules, *viz.* N-terminal module 1 (*Ct*GH1) comprising 445 amino acids (cyan), the C-terminal module 2 (*Ct*GH1-F194A) comprising 309 amino acids (yellow) and the two modules are connected by a natural linker made up of 19 amino acids (red). The difference in the number of amino acid residues between the endoglucanase module of Chimera and the individual endoglucanase is of 487 (pET28a(+)-His-tag region, 23 + *Ct*GH1, 445 + Natural Linker L1, 19 = 487)

The schematic presentation of Chimera construction by fusion PCR approach is shown in Fig. 2.3 i, ii and iii. Chimera consisting of N-terminal β-glucosidase (*Ct*GH1) module and C-terminal endoglucanase (*Ct*GH5-F194A) module were fused with a natural linker sequence of modular *Ct*GH5-F194A. The 2322 bp gene encoding the Chimera was cloned in expression vector, pET28a(+). The Chimera contains 796 amino acids including 23 amino acids of the vector sequence. The Chimera with mutant *Ct*GFH5-F194A at N-terminal and *Ct*GH5-F194A at C-terminal using the C-terminal natural linker of *Ct*GH5 was also constructed by following the same strategy, but the construct did not show any activity.

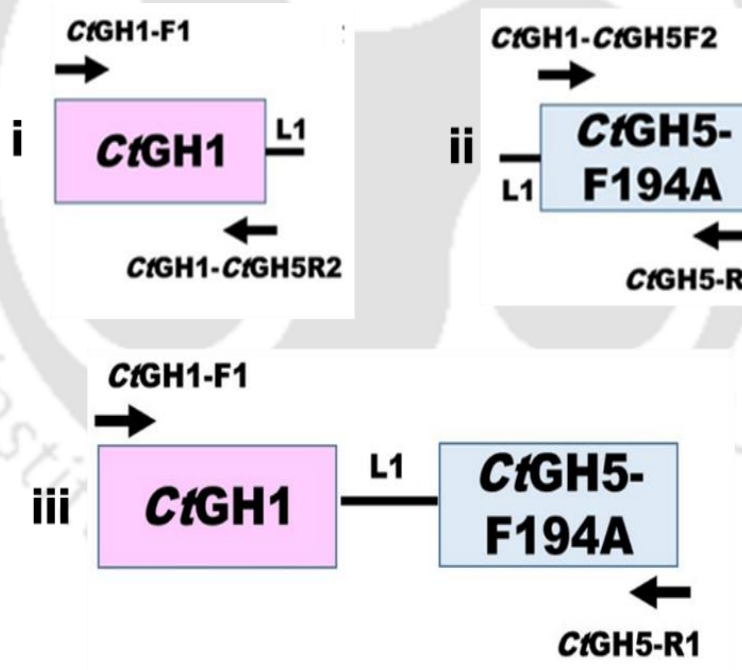


Fig. 2.3 Schematic presentation for development of chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A). The arrow (→) represents the primers for different gene constructs,
i) *Ct*GH1-L1 and ii) L1-*Ct*GH5-F194A
ii) Full length Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)

2.2.4.2 PCR amplification and cloning of Chimeric gene (CtGH1-L1-CtGH5-F194A)

The Chimeric gene was amplified by using the primers given in Table 2.1. The Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was constructed by using the gene encoding *Ct*GH1 and *Ct*GH5-F194A as a template by two step PCR. In the first step, two rounds of PCR were performed, first the *Ct*GH1 was used as template and amplified using primers *Ct*GH1-F1 and *CtG*H1-*Ct*GH5R2 as shown in Fig. 2.3 i. In the second round of PCR, *Ct*GH5-F194A was used as template and two primers used were *Ct*GH1-*Ct*GH5F2 and *Ct*GH5-R1 as shown in Fig. 2.3 ii. In the second step of PCR, the full length Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was amplified by using the primers *Ct*GH1-F1 and *Ct*GH5-R1 as shown in Fig. 2.3 iii while the templates were *Ct*GH1-L1 and L1-*Ct*GH5-F194A encoding genes obtained from the first step PCR as shown in Fig. 2.3 i & ii. The components of 30 µl PCR reaction mixture and the PCR cycles for amplification *Ct*GH1-L1 are given in Tables 2.4.1 & 2.4.2, respectively. The components of 30 µl PCR reaction mixture and the PCR cycles for amplification L1-*Ct*GH5-F194A are given in Tables 2.4.3 & 2.4.4, respectively. The components of 30 µl PCR reaction mixture and the PCR cycles for amplification full Chimera construct *Ct*GH1 L1- *Ct*GH5-F194A are given in Tables 2.4.5 & 2.4.6, respectively. The PCR amplification was performed on a thermal cycler (Applied Biosystems, GeneAmp PCR System 9700). The PCR amplicons were electrophoresed on a 0.8% (w/v) agarose gel along with a DNA marker (1kb plus DNA ladder) as mentioned in Section 2.2.5.

Table 2.4.1 PCR amplification of *Ct*GH-L1.

Reaction Component	Final Concentration	Volume Used (µl)
Nuclease Free Water	-	19.46
5x Phusion HF buffer (Thermo-Scientific)	1x	6.0
25 mM dNTPs (MP Biomedicals)	200 µM	0.24
Forward Primer (15 mM)	0.5 µM	1.0
Reverse Primer (15 mM)	0.5 µM	1.0
Template (Plasmid *Ct*GH1, 81 ng/ µl)	1.35 ng/µl	0.5
DMSO	3%	0.9
Phusion DNA Polymerase (2U/ µl)	0.02 U/µl	0.3
Total Reaction Volume		30

Table 2.4.2 PCR conditions for amplification of *Ct*GH-L1.

Steps		Time
I. Denaturation at 98°C		30 s
II. 30 cycles of	i) Denaturation at 98°C	20 s
	ii) Annealing at 58°C	30 s
	iii) Extension at 72°C	90 s
III. Final extension at 72°C		10 min

Table 2.4.3 PCR amplification of L1-*Ct*GH5-F194A.

Reaction Component	Final Concentration	Volume Used (µl)
Nuclease Free Water	-	19.46
5x Phusion HF buffer (Thermo-Scientific)	1x	6.0
25 mM dNTPs (MP Biomedicals)	200 µM	0.24
Forward Primer (15 mM)	0.5 µM	1.0
Reverse Primer (15 mM)	0.5 µM	1.0
Template (Plasmid *Ct*GH5-F194A, 53 ng/ µl)	0.8 ng/µl	0.5
DMSO	3%	0.9
Phusion DNA Polymerase (2U/ µl)	0.02 U/µl	0.3
Total Reaction Volume		30

Table 2.4.4 PCR conditions for amplification of L1-*Ct*GH5-F194A.

Steps		Time
I. Denaturation at 98°C		30 s
II. 30 cycles of	i) Denaturation at 98°C	20 s
	ii) Annealing at 58°C	30 s
	iii) Extension at 72°C	90 s
III. Final extension at 72°C		10 min

Table 2.4.5 PCR set up for amplification of full length Chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

Reaction Component	Final Concentration	Volume Used (µl)
Nuclease Free Water	-	19.46
5x Phusion HF buffer (Thermo-Scientific)	1x	6.0
25 mM dNTPs (MP Biomedicals)	200 µM	0.24
Forward Primer (15 mM)	0.5 µM	1.0
Reverse Primer (15 mM)	0.5 µM	1.0
Template (*Ct*GH1-L1+L1-*Ct*GH5-F194A, 34.8 ng/µl +36 ng/µl)	(0.4 +0.42) ng/µl	(0.35+0.35)
DMSO	3%	0.9
Phusion DNA Polymerase (2 U/µl)	0.02 U/µl	0.3
Total Reaction Volume		30

Table 2.4.6 PCR condtions for amplification of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

Steps		Time
I. Denaturation at 98°C		60 s
II. 30 cycles of	i) Denaturation at 98°C	20 s
	ii) Annealing at 60°C	60 s
	iii) Extension at 72°C	150 s
III. Final extension at 72°C		10 min

2.2.5 Agarose gel electrophoresis of PCR amplified products

The amplified PCR products were electrophoresed on 0.8% (w/v) agarose gel prepared in 1x TAE buffer. A stock solution of TAE buffer was prepared according to Sambrook and Russell (2001) maintaining the concentrations of components to 10x (400 mM Tris-acetate, 10 mM EDTA, pH 8.0). A gel was prepared using agarose (400 mg for 0.8% (w/v) in 50 ml of 1x TAE buffer and then heating in a microwave oven to get a clear solution. Then 5.0 µl of ethidium bromide (5.0 mg/ml) was added when the solution temperature was 50°C. The solution was mixed well and poured down on the casting apparatus and a comb was placed on the gel and was allowed to solidify. 1x TAE (Tris-acetate-EDTA) buffer was used as an electrophoresis buffer (Sambrook and Russel, 2001). The DNA sample and DNA loading dye as mentioned in section 2.2.5.1 were mixed in a ratio 4:1 and the gel was run at constant 40 Volt till the dye

migrated over 70% of the gel length. The DNA bands were then visualized under UV illumination in a gel documentation system (BioRad XR).

2.2.5.1 DNA loading buffer

The DNA sample loading buffer (5x) was prepared by using the components mentioned below in Table 2.5. A 4 volume of DNA was mixed with 1 volume of 5x stock solution of DNA loading buffer to make 1x, before loading on to the agarose gels. The final pH of the DNA loading dye was maintained at pH 8.0.

Table 2.5 Composition of 5x DNA loading buffer.

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

2.2.6 Extraction of DNA from agarose gel

The PCR amplified DNA or other plasmids were purified from agarose gel using a kit (Sigma GenElute, Sigma-Aldrich, USA). The protocol was followed as given by the manufacturer is mentioned in Section 2.2.6.1. The extracted DNA was eluted in 30 µl elution buffer provided with the kit.

2.2.6.1 Protocol for extraction of DNA from agarose gel

1. 1.5 ml sterile microcentrifuge tube was weighed and the weight was noted.
2. The PCR amplified DNA or plasmid was excised from gel using sharp sterile scalpel and transferred to the micro-centrifuge tube. The tube was weighed again and the weight of excised gel was determined by subtracting the weight of the empty tube.

3. Now, 3 volumes of Gel solubilisation solution were added to every 1 volume of gel (100 mg ~ 100 µl).
4. The micro-centrifuge tube containing excised gel was incubated at 50°C for 10 min (or until the gel slice dissolved completely)
5. 1 gel volume of isopropanol was added to this solution.
6. GenElute binding column G was placed in a 2 ml collection tube provided with the kit. 500 µl of column preparation solution was added over column membrane and centrifuged at 16,000*g* for 1 min. The flow through was discarded before the next step.
7. The solution containing PCR-amplified DNA or plasmid (~700 µl) were added to DNA binding columns and centrifuged at 16,000*g* for 1 min at room temperature discarding the flow through. If the volume was more than 700 µl, the remaining solution was centrifuged similarly and again the flow through was discarded.
8. 700 µl of Wash solution was added to each the DNA bound spin column and the mixture centrifuged at 16000g for 1 min at room temperature, discarding the flow through. The column was given an additional spin of 1 min at 16000*g* to completely remove the residual ethanol.
9. Now the column containing bound DNA was placed on a fresh 1.5 ml sterile microcentrifuge tube. 30 µl of DNAse free water (Sigma-Aldrich, USA) or eluent solution (10 mM Tris-Cl, pH 8.5) was added at the centre of the column. The column was incubated for 2 min at room temperature and then centrifuged at 16000*g* for 1 min. For efficient recovery of DNA, the elution solution was

preheated to 65°C prior to adding it to the membrane. Eluting at 65°C improves the DNA recovery by 2 to 3-fold.

10. The PCR-amplified DNA or plasmid DNA were eluted from GenElute spin columns and collected in 1.5 ml sterile microcentrifuge tube and stored at -20°C for further use.

2.2.7 Preparation of the culture medium

The most commonly used medium called Luria Bertani was used for growing the *E. coli* cells containing recombinant plasmid. The LB medium was prepared by dissolving the ingredients (Table 2.6) in 800 ml of deionized water. The pH of the medium was adjusted to 7.2 and final volume was made up to 1 litre. 100 ml of LB medium was then transferred to 250 ml conical flask and autoclaved at 121°C at 15 psi for 20 min. The filter sterilized antibiotic (Kanamycin; 50 µg/ml or Ampicilin; 100 µg/ml) were added separately to the cooled and autoclaved LB medium before the inoculations.

Table 2.6 Composition of Luria-Bertani medium (Sambrook *et al*., 1989).

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

LB agar medium was prepared by adding 2% (w/v) Agar Agar type I in LB liquid medium. The medium was then autoclaved as described earlier and cooled to around 50-55°C and the antibiotics (Kanamycin; 50 µg/ml or Ampicilin; 100 µg/ml)) was added separately under the laminar air flow. 25 ml of medium supplemented with appropriate amount of antibiotics were poured in the sterile petri plates and allowed to solidify for 15- 20 min.

2.2.8 Preparation of *E. coli* DH5α and *E. coli* BL21 competent cells

Day 1

1. 50 µl of culture of *E. coli* DH5α cells or *E. coli* BL21 from glycerol stock was inoculated into 5.0 ml LB medium (Sambrook *et al*., 1989) contained in a test tube and grown overnight at 37ºC and 180 rpm.
2. Filter-sterilized 0.1 M $CaCl_2$ solution was pass through 0.22 µm filter in laminar air flow and kept in refrigerator.

Day 2

3. The Day 1 1.0 ml culture was inoculated into 100 ml LB medium kept in 250 ml conical flask and incubated at 37ºC with 180 rpm till cell Optical Density reached 0.4-0.6 at 550 nm.
4. 50 ml centrifuge tubes (round bottom), Micro-centrifuge tubes and micro tips were autoclaved and kept on ice and placed in a laminar air flow hood.
5. 40 ml culture was transferred aseptically to round bottom centrifuge tubes.
6. The tubes were centrifuged for 10 min at 4ºC with 4000*g*.
7. The entire 100 ml culture was centrifuge by repeating the previous step.
8. The cell pellet was re-suspended in 3-4 ml ice-chilled and sterile 0.1 M $CaCl_2$ solution followed by making up the final volume to 20 ml. The centrifuge tubes containing the cell suspension was kept on ice for 10 min.
9. The tube was centrifuged again at 4000*g* at 4ºC for 10 min.
10. The supernatant was carefully separated and the pellet re-suspension was performed in 3.0 ml of sterile ice chilled 0.1 M $CaCl_2$ solution.

11. 200 µl of competent cells were aliquoted into each 1.5 ml microcentrifuge tube containing 10 (%, v/v) glycerol (final concentration) and kept at -80ºC for further use.

2.2.9 Cloning of gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) into pET28a(+) vector

The pET-28a(+) is a modified form of pBR322 plasmid. It is a regularly used vector for cloning and expression of recombinant proteins in *E. coli*. pET-28a(+) vector has a strong T7 promoter system in the beginning developed by Studier and colleagues (Studier and Moffatt, 1986; Studier *et al.*, 1990). The expression of genes cloned in pET plasmids is under the control of T7 bacteriophage promoter. The cloned genes are transcribed by T7 RNA polymerase of the host cell. The genes cloned in pET vectors continue to be transcriptionally silent in the uninduced state. The proteins encoded by the cloned genes contain a His_6-Tag which can be purified by single step purification by affinity chromatography. The pET-28a(+) vector allows for incorporation of expressed protein with an N-terminal His_6-Tag/thrombin/T7-Tag in addition to an optional C-terminal His_6-Tag sequence (Fig. 2.4). The region of sequence encoding His-Tag, T7 promoter, T7 terminator, kanamycin resistance and f1 origin are indicated in the Fig. 2.4.

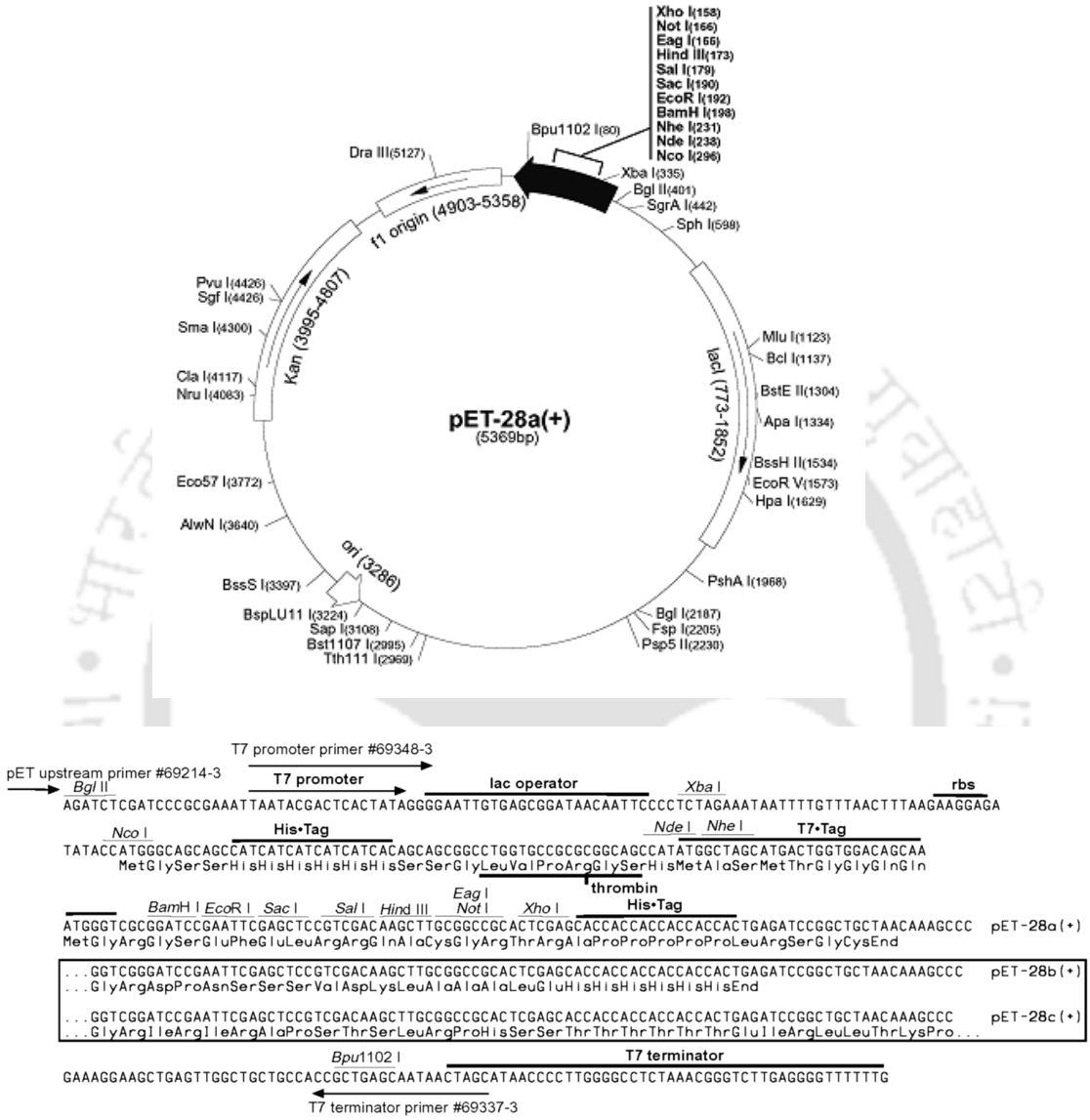


Fig. 2.4 Restriction map of the pET-28a(+) expression vector showing multiple cloning site (158-203 bp), restriction enzyme sites, N-terminal His_6-Tag coding sequence (270-287 bp), C-terminal His_6-Tag coding sequence (140-157 bp), T7 promoter (370-386), T7 terminator (26-72 bp), pBR322 origin (3286 bp), kanamycin marker (3995-4807 bp) and a f1 origin (4903-5358). *Nhe*I cuts at 231 and *Xho*I at 158.

2.2.9.1 Restriction digestion of PCR amplified genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and pET-28a(+) vector DNA

The pET-28a(+) vector DNA was digested with *Nhe*I and *Xho*I restriction enzymes (Table 2.7). The PCR amplified gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was also digested with *Nhe*I and *Xho*I enzymes to prepare it for ligation with restriction digested pET-28a(+) vector (Table 2.8). The digestion reactions were incubated in a water bath at 37°C for 90 min. The *Nhe*I-*Xho*I digested pET vector and PCR amplified genes were purified from agarose gel as described in Section 2.2.6.

Table 2.7 Restriction enzyme digestion of pET-28a(+) vector DNA.

Reaction components	Volume (µl)
10x buffer	2.0
Nuclease free water	1.5
Plasmid DNA (approx. 100 ng/ µl)	15.0
*Nhe*I (10 U/µl)	0.5
*Xho*I (20 U/µl)	1.0
Total	20.0

Table 2.8 Restriction enzyme digestion of PCR amplified gene encoding *Ct*GH1-L1-*Ct*GH5-F194A.

Reaction components	Volume (µl)
10x buffer	1.7
Nuclease free water	0.75
Amplified DNA (approx. 20 ng/ µl)	15.0
*Nhe*I (10 U/µl)	0.3
*Xho*I (20 U/µl)	0.15
Total	18.0

2.2.9.2 Ligation of restriction digested genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) into pET-28a(+) vector

The *Nhe*I-*Xho*I digested gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was ligated into pET-28a(+) vector, which was also digested with same restriction enzymes

as described in Section 2.2.9.1. The ligation reactions were setup using the reaction components as mentioned in Table 2.9 and incubated at 16°C overnight to get maximum number of transformants. pET-28a(+) vector. The insert: vector molar ratio was kept at 3:1 for all the fragments. The amounts of *Nhe*I-*Xho*I digested insert of *Ct*GH1-L1-*Ct*GH5-F194A required for cloning were calculated by using the following formula (Engler and Richardson, 1982). The reactions were setup at an insert: vector molar ratio of 3:1, where the amount of insert is calculated as mentioned below:

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

$$\frac{70\ (\text{ng}) \times 2.1\ (\text{kb})}{5.36\ (\text{kb})} \times \frac{3}{1} = 82.7\ \text{ng of Chimera (}\textit{Ct}\text{GH1-L1-}\textit{Ct}\text{GH5-F184A)}$$

Table 2.9 Components of reaction for ligating gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) into pET-28a(+) expression vector.

Reaction component	**(µl)**
10x Rapid Ligation Buffer	1.1
pET-28a(+) Vector	2.0 (70 ng)
Restriction digested DNA Insert	6.0 (82.7 ng)
T4 DNA Ligase (Takara Bio Inc, UK, 350 Units/µl)	1.0
Nuclease-free water	1.0
Total	11.0

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

The mutagenic plasmid DNA of mutant endoglucanase (*Ct*GH5-F194A) in pET-21a(+) or ligation reaction of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) were transformed in the *E. coli* DH5α competent cells, after overnight ligation. Preparation

of *E. coli* competent cell preparation has been described in Section 2.2.8. The step-wise transformation protocol is described below:

1. The micro-centrifuge tube containing competent cell (200 µl) was taken out from -80°C and kept on ice for 5 min.
2. 10 µl of ligation mixture was added to cells and the tube was gently tapped 4-5 times and kept on ice for 30 min. The cells were occasionally tapped gently during 30 min incubation.
3. The cells were given a heat shock at 42°C for 42 s.
4. The cells were immediately transferred back to the ice and kept for 5 min.
5. 800 µl of LB medium was added to the transformed cells.
6. The transformed cells were incubated at 37°C in a shaking incubator at 220 rpm for 1h.
7. The cells were centrifuged at 2000*g*, 25°C for 5 min.
8. 800 µl supernatant was discarded and the cell pellet was re-suspended in remaining 200 µl supernatant.
9. The 200 µl cells were spread plated on LB agar plates as described in Section 2.2.7.1 supplemented with antibiotics. The LB agar plates were incubated overnight at 37°C.
10. The transformation efficiency was calculated using the following formula,

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

The 10 µl of mutagenic plasmid DNA or ligation mixture was added to 200 µl *E. coli* DH5α competent cells, following the above transformation protocol. The transformed DH5α cells were plated on LB plates supplemented with appropriate antibiotics and grown overnight at 37°C, 180 rpm.

2.2.9.4 Isolation of plasmid DNA from transformed colonies by plasmid miniprep kit

Overnight incubated petri plates from section 2.2.9.3 were observed for colonies. Colonies for mutant endoglucanase (*Ct*GH5-F194A) and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) preferred from the center of the plate and were randomly picked under a laminar air flow. For growing DH5α cells containing *Ct*GH5-F194A in pET-21a(+) vector the 5 ml LB medium supplemented with ampicillin (100 µg/ml) was used. For growing DH5α cells containing *Ct*GH1-L1-*Ct*GH5-F194A in pET-28a(+) vector Kanamycin (50 µg/ml) was added in the 5 ml LB medium. Both the cultures were grown by incubating at 37°C under shaker incubator at 180 for 12 h. The plasmid DNA of *Ct*GH5-F194A and *Ct*GH1-L1-*Ct*GH5-F194A from their respective 5 ml cultures was isolated by plasmid miniprep kit (Sigma-Aldrich, USA) following the protocol mentioned in Section 2.2.9.4.1.

2.2.9.4.1 Plasmid isolation protocol by miniprep kit protocol

1. 9 ml from each of the grown culture containing recombinant plasmid were pelleted in 1.5 ml microcentrifuge tube aseptically.
2. The cells were then centrifuged at 14000*g* for 1 min and the process was repeated six times with 1.5 ml culture (Total 9ml culture).
3. The resulting cell pellet of each recombinant derivative was re-suspended in 200 µl resuspension solution and vortexed. RNAse at final concentration of 0.3 mg/ml was added to the re-suspension solution prior to use.
4. 200 µl of lysis solution was added to each tube and the tubes were inverted gently 5-6 times to ensure mixing and allowed to stand for 2-5 min.

5. 350 µl of neutralization solution was added to the mixture and the tubes were inverted again for 4–6 times to mix properly.
6. The mixture was centrifuged at 16,000*g* for 10 min.
7. The DNA binding columns were prepared and activated by adding 500 µl of column preparation solution to binding column and centrifuging at 14,000*g* for 1 min. The flow through accumulated in collection tube was discarded.
8. The clear lysate was then transferred to activated DNA binding column, centrifuged at 14,000*g* for 1 min and the flow through in the collection tube was discarded again.
9. The plasmid DNA bound to the column was washed with wash solution and spun at 14,000*g* for 1 min. The flow through was discarded and the column was given another 1 min spin at 14,000*g* for removing the wash solution completely.
10. The DNA binding column was transferred to a fresh sterile microcentrifuge tube and 30 µl of TE buffer solution or DNAse free water was added at the centre of binding column. The microcentrifuge tube was allowed to stand for 10 min at room temperature and then plasmid DNA was eluted by centrifugation at 14,000*g* for 1 min.
11. The eluted plasmid DNA in sterile microcentrifuge tube was stored at -20°C.

2.2.9.5 Screening of recombinant plasmid DNA for positive clone of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) by restriction digestion

3 µl of recombinant plasmids from pET-28a(+) clones of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) that were isolated in the last step, were taken in separate fresh sterile micro-centrifuge tubes for restriction enzyme digestion analysis. The recombinant plasmid DNA was digested with restriction enzymes, *Nhe*I and *Xho*I, to verify the positive clones following a 5 µl reaction mixture set up as mentioned in Table 2.10.

Table 2.10 Restriction enzyme digestion of cloned plasmid of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

Reaction components	Volume (µl)
10x buffer	0.5
Nuclease free water	1.2
Plasmid DNA (approx. 100 ng/ µl)	3.0
*Nhe*I (10 U/µl)	0.2
*Xho*I (20 U/µl)	0.15
Total	5.0

2.2.10 Transformation of recombinant plasmids containing genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and *Ct*GH5-F194A in *E. coli* BL21 (DE3)

The recombinant plasmids pET-21a(+) plasmid containing gene encoding mutant *Ct*GH5-F194A construct and recombinant pET-28a(+) plasmid containing gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) were transformed in *E. coli* BL21 (DE3) competent cells as mentioned in section 2.2.8.

2.2.11 Expression of recombinant plasmids containing genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and *Ct*GH5-F194A in *E. coli* BL21 (DE3)

The colonies were picked randomly and grown in 5 ml LB medium supplemented with ampicillin (100 µg/m) for mutant F194A enzymes and kanamycin (50 µg/ml) for Chimera. The cells were induced for expression of the recombinant

protein by adding 1 mM final concentration of isopropyl-1-thio-β-D-galactopyranoside (IPTG) at the mid-exponential phase (A_{550} 0.6). After addition of IPTG, the culture was incubated at 37°C and at 180 rpm for 6h for *Ct*GH5-F194A whereas, the culture of chimeric protein was kept at 24°C and 180 rpm for 16h. Protein expression was analysed by using SDS-PAGE gels as mentioned in Section 2.2.11.3. The recombinant wild-type endoglucanase *Ct*GH5 and β-glucosidase (*Ct*GH1) were expressed and purified by using the protocol mentioned in Bharali *et al.,* 2005 and Sharma *et al.,* 2019, respectively.

2.2.11.1 Purification of recombinant proteins by affinity chromatography

For the purification of recombinant proteins, the cells were cultured in the 100 ml LB medium supplemented with 100 µg/ml of ampicillin for wild-type *Ct*GH5 and mutant F194A enzymes (Bharali *et al.,* 2005). The enzyme, *Ct*GH1 was expressed using *E. coli* BL21 cells in 100 ml LB medium supplemented with 50 µg/ml of kanamycin as mentioned earlier (Sharma *et al.,* 2019). The chimeric gene was cloned in pET-28a(+) vector, therefore, 50 µg/ml of kanamycin was used for growing chimeric construct containing *E. coli* cells in 100 ml LB medium at 37°C and 180 rpm for 4h. The cells were induced with 1 mM final concentration of isopropyl-1-thio-β-D-galactopyranoside (IPTG) at the mid-exponential phase (A_{550} 0.6) for over-expression of recombinant proteins. After addition of IPTG, the cells of wild-type *Ct*GH5 and mutant F194A were kept at 37°C and at 180 rpm for 6h whereas, the cells of chimeric protein and *Ct*GH1 were kept at 24°C and 180 rpm for 16h. The cells were centrifuged at 10,000g at 4°C for 10 min. The cells were sonicated for 10s on and 10s off pulse; with 33% amplitude and were centrifuged for 40 min at 15000g at 4°C to separate the cell free extract. The recombinant Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)

and its individual modules (*Ct*GH5-F194A, wild-type *Ct*GH5 and *Ct*GH1) proteins containing a His_6-tag at the N-terminal were purified through a single step purification method based on immobilized metal-ion affinity chromatography (IMAC) as described in Section 2.2.11.2. The purification of these recombinant proteins was carried out by using 1.0 ml sepharose columns (GE Healthcare, HiTrap chelating HP). The composition of binding as well as elution buffers used for affinity column purification is mentioned in Table 2.11. All the proteins were then dialysed for removal of imidazole and sodium chloride using buffer as mentioned in Table 2.11. The recombinant Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual modules *Ct*GH1, *Ct*GH5-F194A and wild-type *Ct*GH5 expressed as soluble proteins. After their purification they were analysed for purity and molecular mass using SDS-PAGE as mentioned in section 2.2.11.3**.**

Table 2.11 Composition of buffers required for purification of recombinant proteins by affinity purification (IMAC).

Buffer	Composition
Equilibration buffer	50 mM Sodium phosphate, pH 7.0 300 mM NaCl, 50 mM Imidazole
Elution buffer	50 mM Sodium phosphate, pH 7.0 300 mM NaCl, 300 mM Imidazole
Cleaning buffer	50 mM Sodium phosphate, pH 7.5 500 mM NaCl, 50 mM EDTA
Dialysis Buffer	50 mM Sodium phosphate, pH 7.0

2.2.11.2 IMAC purification protocol for recombinant Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual modules (*Ct*GH5-F194A and *Ct*GH1)

1. The bacterial cells (100 ml culture) were harvested by centrifugation at 10,000*g* at 4°C. The cell pellet was re-suspended in 5 ml of 50 mM Sodium phosphate buffer, pH 7.0.
2. The cells were sonicated on ice for 10 min (10s on and 10s off pulse; with 33% amplitude) and centrifuged at 15,000*g* at 4ºC for 40 min to get the crude cell free extract.
3. The cell free extract was passed through a 0.45 µm filter membrane before loading onto 1ml HiTrap chelating HP column. The column was pre-washed with 5 volumes of filtered and degassed water to remove the alcohol.
4. Column was charged using 2.0 ml of 0.1 M $NiSO_4$ solution and the unbound Ni^{2+} ions were washed away with 2-5 volumes of water.
5. Then the column was equilibrated with 10 volumes of equilibration buffer (Table 2.11).
6. The filtered cell free extract of recombinant protein was loaded on to the column at a flow rate of 0.5 ml/min.

7. The column was then washed with 50 column volumes of equilibration buffer to remove the unbound proteins.
8. The retained protein of interest was then eluted with elution buffer under a gradient of 0-100% imidazole concentration and 1 ml fractions were collected (Carvalho *et al.*, 2004).
9. The column was cleaned using cleaning buffer as mentioned in Table 2.11, washed with 5 volumes of water and incubated in 1N NaOH at 4°C for 2h. The column was then washed with 20 volumes of water to remove NaOH, and finally stored in 20% (v/v) ethanol at 4°C.

The purified recombinant proteins Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual modules *Ct*GH1, *Ct*GH5-F194A and wild-type *Ct*GH5 were dialyzed against 50 mM Sodium phosphate, pH 7.0. The purity and molecular mass of recombinant proteins were verified by SDS-PAGE as described in Section 2.2.10.3.

2.2.11.3 Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis (SDS-PAGE) analysis of recombinant proteins

The recombinant proteins were separated on SDS-PAGE gel on the basis of their respective molecular size. SDS-PAGE was used to separate components of a protein mixture based on their size (Laemmli, 1970; Sambrook *et al*., 1989). Analysis of expression and purification of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), *Ct*GH1, *Ct*GH5-F194A and wild-type *Ct*GH5 was done on 12% (w/v) SDS-PAGE gels. The ingredients for making the gel are mentioned in Tables 2.12 and 2.13. The SDS-polyacrylamide gel electrophoresis was conducted in a vertical slab mini gel unit (Mini-PROTEAN®Tetra cell, BioRad, USA) with 1.0 mm thick gels.

The protein sample buffer was prepared by following the protocol mentioned in section 2.2.11.4. The protein samples were prepared as mentioned in section 2.2.11.5. SDS-PAGE was run using 1x Tris-Glycine (pH 8.3-8.5) running buffer at constant 40 mA current. The preparation of the running buffer is mentioned in section 2.2.11.6. The expressed and purified protein samples were visualized after staining the gel with staining solution. The staining solution contained (0.25% w/v) Coomassie Brilliant Blue (CBB) R-250 dye 100 ml solution of deionized water, methanol and glacial acetic acid in 5:4:1 ratio as mentioned in section 2.2.11.7. The gels were de-stained by immersing the gel in de-staining solution containing deionized water, methanol and glacial acetic acid in 5:4:1 ratio as mentioned in section 2.2.11.7. The gels were destained with periodic change of de-staining solution was performed by using rocker platform until the protein bands were clearly visible.

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

Component	12% gel volume (ml)
Acrylamide solution *(30%,w/v)	4.0
Deionized water	0.7
SDS (10%,w/v)	1.0
Glycerol (50%,v/v)	1.0
1.5 M Tris-HCl (pH 8.8)	3.3
APS (10%,w/v)	0.1
TEMED	0.01
Total	10.0

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

Table 2.13 Composition of SDS-PAGE components for preparation of stacking gel.

Components	5% gel volume (ml)
Acrylamide solution* (30%, w/v)	0.7
Deionized water	2.8
SDS (10%, w/v)	0.5
0.5 M Tris-HCl (pH 6.8)	1.0
APS (10%, w/v)	0.05
TEMED	0.005
Total	5.0

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

2.2.11.4 Preparation of sample buffer for protein

5x sample loading buffer for protein was prepared by dissolving the components while keeping the concentration of components as described in Table 2.14 and the pH of the buffer was adjusted to 6.8. The components were dissolved in the order as mentioned in Table 2.2.14 to make 5x sample buffer. However, the final concentration while loading to a SDS-PAGE gel was always kept to 1x by mixing 4 volumes of sample (protein) with 1 volume of 5x sample buffer.

Table 2.14 Composition of 5x sample loading buffer (Laemmli, 1970).

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

2.2.11.5 Preparation of protein samples for SDS-PAGE

The protein samples were mixed with 5x protein sample buffer in the ratio of 4:1. The mixture was heating in boiling water bath for 5 min, then cooled to 25°C and loaded on the gel. The electrophoresis was carried out by using 1x Tris-Glycine (pH 8.3-8.5) running buffer as mentioned in section 2.2.11.6. Molecular mass marker (10-200 kDa), purchased from Biobharti India Private Ltd. and Genei India Private Ltd. were used as standard for SDS-PAGE.

2.2.11.6 Preparation of SDS-PAGE running buffer

The SDS-PAGE gels were run using a 1x running or tank buffer prepared from the 5x stock solution as described in Table 2.15. 15.14 g of Tris free base and 94 g of glycine were dissolved in 800 ml of deionized water. To this 50 ml of 10% (w/v) SDS was added and the final volume was adjusted to 1 litre. The final pH of the buffer was adjusted to 8.3. The 5x buffer was filtered (Whatman, Filter No. 1) and stored at 4°C.

Table 2.15 Composition of 5x Tris-Glycine, running or tank buffer.

Components	Final concentration (5x buffer)
Tris base	0.125 M
Glycine	1.25 M
SDS	0.5 % (w/v)

2.2.11.7 Preparation of staining and destaining solutions

The staining solution (100 ml) was prepared by dissolving 250 mg or 0.25% (w/v), of Coomassie Brilliant Blue (CBB R-250) dye in 50 ml of deionized water in an amber colour bottle by keeping on a magnetic stirrer for overnight. The solution was filtered (Whatman, Filter No. 1), then 40 ml of methanol and 10 ml of glacial acetic acid were added to finally make the ratio 5:4:1 (deionized water:methanol:glacial acetic acid). The destaining solution was prepared by dissolving deionized water: methanol: glacial acetic in 5:4:1 ratio. The gels were destained by immersing it in destaining solution with gentle shaking condition and changing the destaining solution every 30 min, until the protein bands were clearly visible.

2.2.12 Protein concentration determination of purified recombinant proteins

The concentration of purified proteins was determined from their corresponding absorbance at 280 nm using the equation below (Layne, 1957; Stoscheck, 1990). Absorbance was measured after appropriate dilution of the proteins using a spectrophotometer (Gene Quant, GE healthcare, USA) having a path length of 1 cm. The molar extinction co-efficient 192185 $M^{-1}cm^{-1}$ for Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), 100730 $M^{-1}cm^{-1}$ for *Ct*GH1 and 85830 $M^{-1}cm^{-1}$ for wild-type *Ct*GH5 and mutant *Ct*GH5-F194A were used.

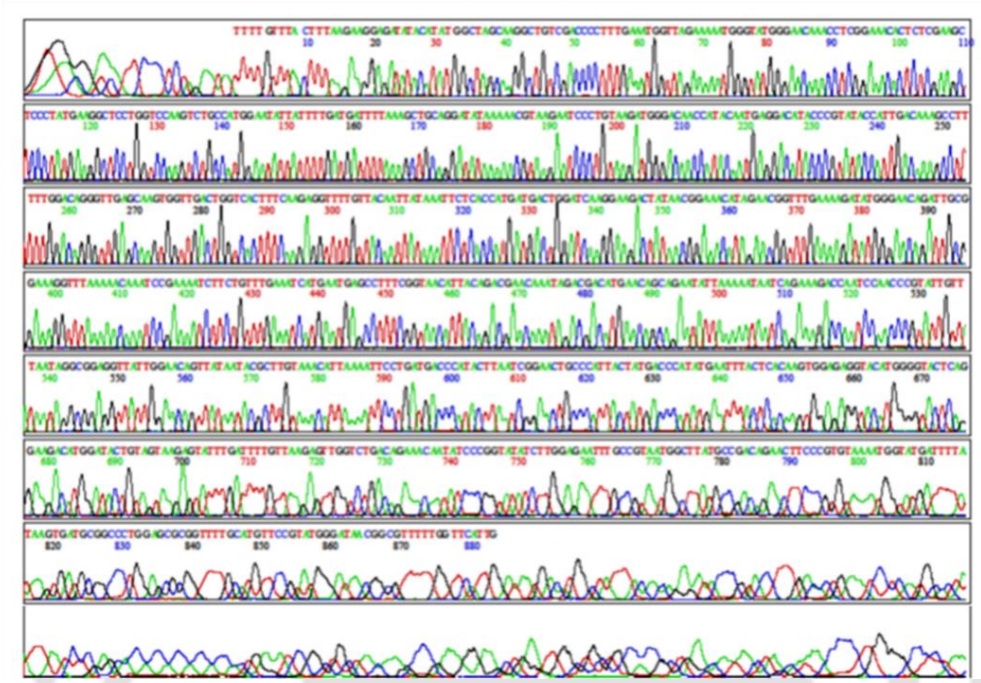
$$\text{Concentration of protein (mg/ml)} = \frac{\text{Absorbance at 280 nm x Mol. weight (Da)}}{\text{Extinction coefficient } (M^{-1}cm^{-1}) \text{ x Path length (1 cm)}}$$

2.3 Results and Discussion

2.3.1 Site-directed mutagenesis of wild-type *Ct*GH5 for the development of mutant *Ct*GH5-F194A

The site-directed mutagenesis was performed for the wild-type endoglucanase *Ct*GH5 as described in the section 2.2.3. The mutant plasmid was transformed in *E. coli* DH5α competent cells. The plasmid was extracted and sequenced to confirm the mutation at the specific residue shown in Fig. 2.5. This recombinant plasmid was transformed in *E. coli* BL21 cells for mutant endoglucanase *Ct*GH5-F194A expression.

A



B

TTTGTTTACTTTAAGAAGGAGATATACATATGGCTAGCAAGGCTGTCGACCCCTTTGAAA
TGGTTAGAAAAATGGGTATGGGAACAAACCTCGGAAACACTCTCGAAGCTCCCTATGAAG
GCTCCTGGTCCAAGTCTGCCATGGAATATTATTTTGATGATTTTAAAGCTGCAGGATATA
AAAACGTAAGAATCCCTGTAAGATGGGACAACCATACAATGAGGACATACCCGTATACCA
TTGACAAAGCCTTTTTGGACAGGGTTGAGCAAGTGGTTGACTGGTCACTTTCAAGAGGTT
TTGTTACAATTATAAATTCTCACCATGATGACTGGATCAAGGAAGACTATAACGGAAACA
TAGAACGGTTTGAAAAGATATGGGAACAGATTGCGGAAAGGTTTAAAAACAAATCCGAAA
ATCTTCTGTTTGAAATCATGAATGAGCCTTTCGGTAACATTACAGACGAACAAATAGACG
ACATGAACAGCAGAATATTAAAAATAATCAGAAAGACCAATCCAACCCGTATTGTTATAA
TAGGCGGAGGTTATTGGAACAGTTATAATACGCTTGTAAACATTAAAATTCCTGATGACC
CATACTTAATCGGAACT**GCC**CATTACTATGACCCATATGAATTTACTCACAAGTGGAGAG
GTACATGGGGTACTCAGGAAGACATGGATACTGTAGTAAGAGTATTTGATTTTGTTAAGA
GTTGGTCTGACAGAAACAATATCCCGGTATATCTTGGAGAATTTGCCGTAATGGCTTATG
CCGACAGAACTTCCCGTGTAAAATGGTATGATTTTATAAGTGATGCGGCCCTGGAGCGCG
GTTTTGCATGTTCCGTATGGGATAACGGCGTTTTTGGTTCATT

Fig. 2.5 DNA Sequencing of *Ct*GH5-F194A (A) Electropherogram showing the mutagenesis result for mutant plasmid. Sequencing was done using T7 forward primer. (B) the deduced sequence.

2.3.2 PCR amplification of genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)

The genes encoding *Ct*GH1-L1 and L1-*Ct*GH5 were amplified using *Ct*GH1 and *Ct*GH5-F194A plasmids, respectively, as templates. The full length Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was amplified using *Ct*GH1-L1 and L1-*Ct*GH5 amplified products as a template. The conditions are mentioned in Section 2.2.4. The PCR amplified fragments *Ct*GH1-L1, L1-*Ct*GH5-F194A and *Ct*GH1-L1-*Ct*GH5-F194A were detected on 0.8% (w/v) agarose gel and are displayed in Fig. 2.6 A, B and C respectively. The PCR products were purified from gel using gel extraction kit as mentioned in Section 2.2.6 and stored at - 20°C.

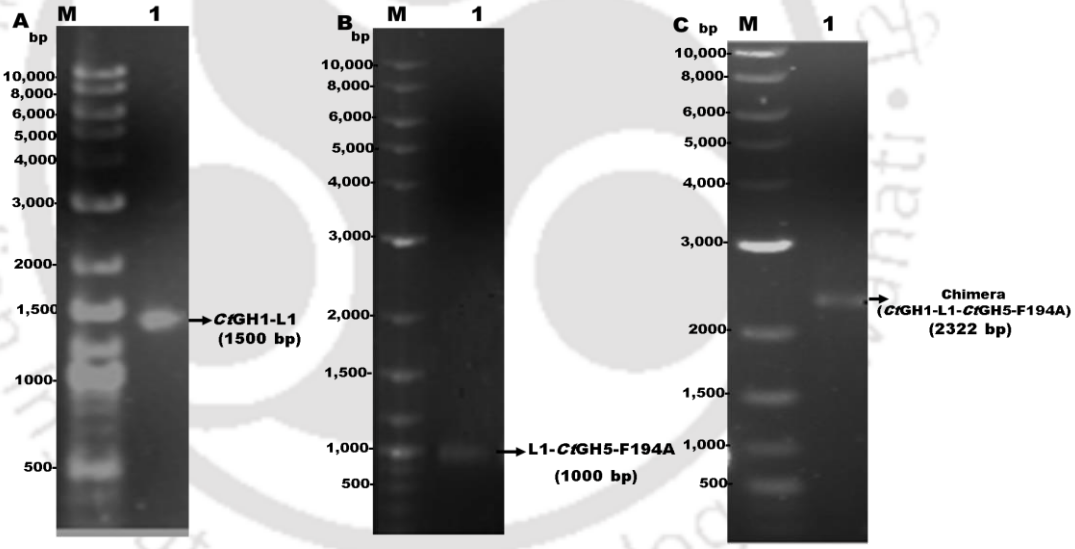


Fig. 2.6 Agarose Gel (0.8%, w/v) showing amplification of A) *Ct*GH1-L1: Lane M, DNA marker; Lane 1, *Ct*GH1-L1 extracted purified product; B) L1-*Ct*GH5: Lane M, DNA marker; Lane 1, L1-*Ct*GH5 extracted product; C) Lane M, DNA marker; Lane 1, Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) extracted product.

2.3.3 Cloning of gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) into pET-28a(+) vector

The restriction enzyme digested gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was ligated with the linearized fragments of pET-28a(+) vector following the protocol mentioned in Section 2.2.9.2. The ligated product was transformed into *E. coli* DH5α competent cells and grown overnight on LB agar plates grown at 37°C under stationary condition. The transformation efficiency of *E. coli* DH5α competent cells was 4.1 x10^6 cfu/µg.

2.3.3.1 Isolation of recombinant plasmid DNA

Plasmid DNA from after cloning into pET-28a(+) grown colonies was isolated using Plasmid miniprep kit following the protocol mentioned in Section 2.2.9.4.1. The isolated plasmids were visualized after electrophoresis on 0.8% (w/v) agarose gel. The positive clone was confirmed by restriction digestion of this isolated plasmid DNA.

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

The isolated plasmid was digested with *Nhe*I and *Xho*I restriction enzymes for confirming the positive clone. The plasmid after restriction digestion was electrophoresed on 0.8% (w/v) agarose gel. *NheI* and *XhoI* digested fragment of gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) (Fig. 2.7; Lane 1) was visualized on agarose gels around 2.3 kb. Linearized pET-28a (+) vector was seen at around 5.4 kb after restriction digestion. The positive clones were sequenced (Scigenom Labs Pvt. Ltd, India) and no undesired mutations were detected (Fig. 2.8 and Fig. 2.9).

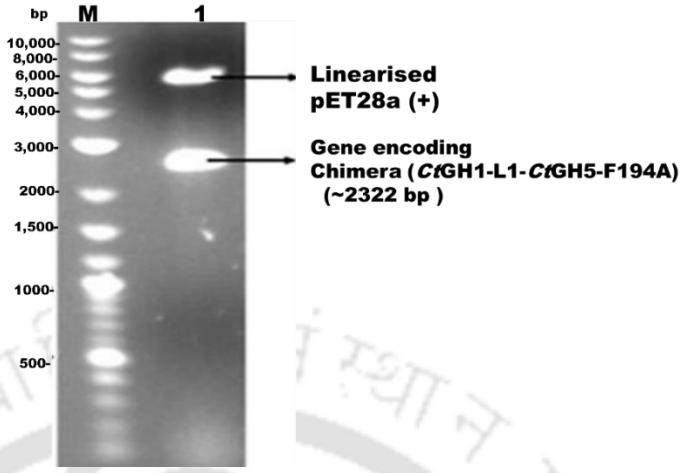


Fig. 2.7 Agarose (0.8%, w/v) showing *Nhe*I-*Xho*I digested recombinant plasmid. Lane 1 shows the genes encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), (2.3 kb); Lane M: DNA marker (1kb DNA ladder, New England Biolabs).

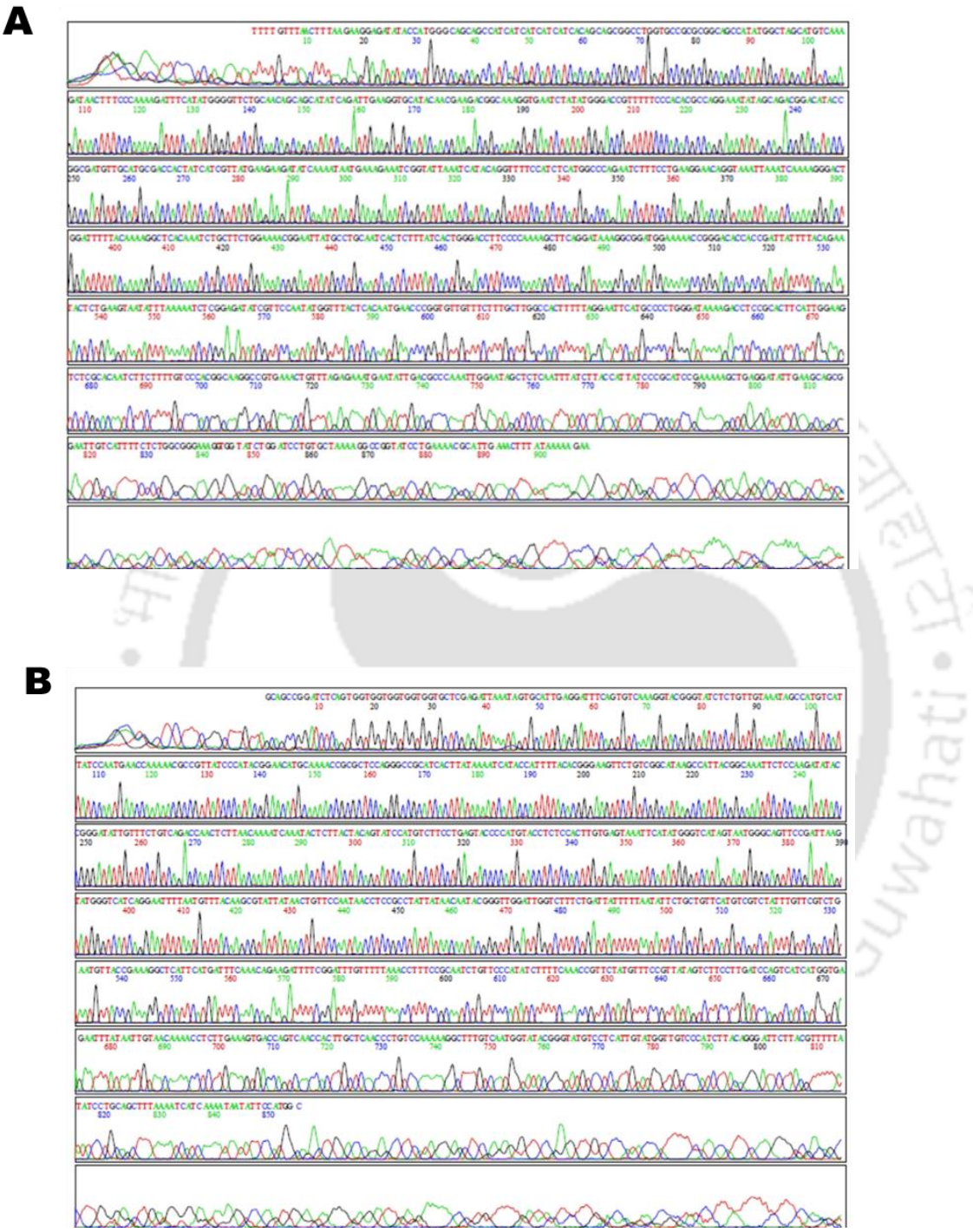


Fig. 2.8 DNA Sequencing of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), Electropherogram showing the DNA sequencing result for cloned gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) (A) N-terminal Sequencing was done using T7 forward primer. (B) C-terminal Sequencing was done using T7 reverse primer.

A

GAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGC
CGCGCGGCAGCCATATGGCTAGCATGTCAAAGATAACTTTCCCAAAAGATTTCATATGGG
GTTCTGCAACAGCAGCATATCAGATTGAAGGTGCATACAACGAAGACGGCAAAGGTGAAT
CTATATGGGACCGTTTTTCCCACACGCCAGGAAATATAGCAGACGGACATACCGGCGATG
TTGCATGCGACCACTATCATCGTTATGAAGAAGATATCAAAATAATGAAAGAAATCGGTA
TTAAATCATACAGGTTTTCCATCTCATGGCCCAGAATCTTTCCTGAAGGAACAGGTAAAT
TAAATCAAAAGGGACTGGATTTTTACAAAAGGCTCACAAATCTGCTTCTGGAAAACGGAA
TTATGCCTGCAATCACTCTTTATCACTGGGACCTTCCCCAAAAGCTTCAGGATAAAGGCG
GATGGAAAAACCGGGACACCACCGATTATTTTACAGAATACTCTGAAGTAATATTTAAAA
ATCTCGGAGATATCGTTCCAATATGGTTTACTCACAATGAACCCGGTGTTGTTTCTTTGC
TTGGCCACTTTTTAGGAATTCATGCCCCTGGGATAAAAGACCTCCGCACTTCATTGGAAG
TCTCGCACAATCTTCTTTTGTCCCACGGCAAGGCCGTGAAACTGTTTAGAGAAATGAATA
TTGACGCCCAAATTGGAATAGCTCTCAATTTATCTTACCATTATCCCGCATCCGAAAAAG
CTGAGGATATTGAAGCAGCGGAATTGTCATTTTCTCTGGCGGGAAAGGTGGTATCTGGAT
CCTGTGCTAAAAGGCCGGTATCCTGAAAACGCATTGAAACTTTATAAAAAGA

B

CAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGATTAAATAGTGCATTGAGGATTT
CAGTGTCAAAGGTACGGGTATCTCTGTTGTAAATAGCCATGTCATTATCCAATGAACCAA
AAACGCCGTTATCCCATACGGAACATGCAAAACCGCGCTCCAGGGCCGCATCACTTATAA
AATCATACCATTTTACACGGGAAGTTCTGTCGGCATAAGCCATTACGGCAAATTCTCCAA
GATATACCGGGATATTGTTTCTGTCAGACCAACTCTTAACAAAATCAAATACTCTTACTA
CAGTATCCATGTCTTCCTGAGTACCCCATGTACCTCTCCACTTGTGAGTAAATTCATATG
GGTCATAGTAATGGGCAGTTCCGATTAAGTATGGGTCATCAGGAATTTTAATGTTTACAA
GCGTATTATAACTGTTCCAATAACCTCCGCCTATTATAACAATACGGGTTGGATTGGTCT
TTCTGATTATTTTTAATATTCTGCTGTTCATGTCGTCTATTTGTTCGTCTGTAATGTTAC
CGAAAGGCTCATTCATGATTTCAAACAGAAGATTTTCGGATTTGTTTTTAAACCTTTCCG
CAATCTGTTCCCATATCTTTTCAAACCGTTCTATGTTTCCGTTATAGTCTTCCTTGATCC
AGTCATCATGGTGAGAATTTATAATTGTAACAAAACCTCTTGAAAGTGACCAGTCAACCA
CTTGCTCAACCCTGTCCAAAAAGGCTTTGTCAATGGTATACGGGTATGTCCTCATTGTAT
GGTTGTCCCATCTTACAGGGATTCTTACGTTTTTATATCCTGCAGCTTTAAAATCATCAA
AATAATATTCCATGG

Fig. 2.9 DNA Sequencing of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), the deduced sequence. A) N-terminal Sequencing was done using T7 forward primer. B) C-terminal Sequencing was done using T7 reverse primer.

2.3.4 Expression of recombinant *Ct*GH5-F194A and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)

The *E. coli* BL21 (DE3) competent cells were used for the transformation of recombinant pET-28a(+) plasmid containing gene encoding Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and pET-21a(+) plasmid containing gene encoding mutant *Ct*GH5-F194A construct. The colonies were picked randomly and grown in 5 ml LB medium supplemented with kanamycin (50 µg/ml) or ampicillin (100 µg/m) as described in Section 2.2.10. The cells were induced for expression of protein at mid exponential stage as described in Section 2.2.10. Protein expression was analysed using SDS-PAGE gels by loading uninduced as well as the induced cells in adjacent wells. The expression of *Ct*GH1 (β-glucosidase) was achieved by following the protocol mentioned in Sharma *et al.*, 2019. The protein expression for mutant endoglucanase *Ct*GH5-F194A and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) are depicted in Fig(s). 2.10 and 2.11, respectively.

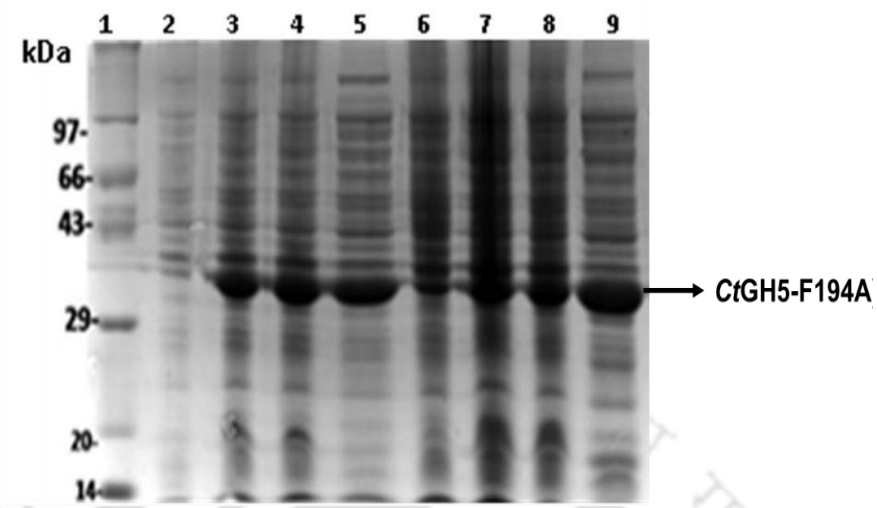


Fig. 2.10 Expression profile of *Ct*GH5 mutant F194A at 37°C and 24°C. SDS-PAGE (12%, w/v). Lanes 1, Mol wt. marker (Genei); 2, Un-induced cells at 37°C; 3,4 Induced cells at 37°C; 5, Cell free extract; 6, Un-induced cells at 24°C; 7,8 Induced cells at 24°C; 9, Cell free extract.

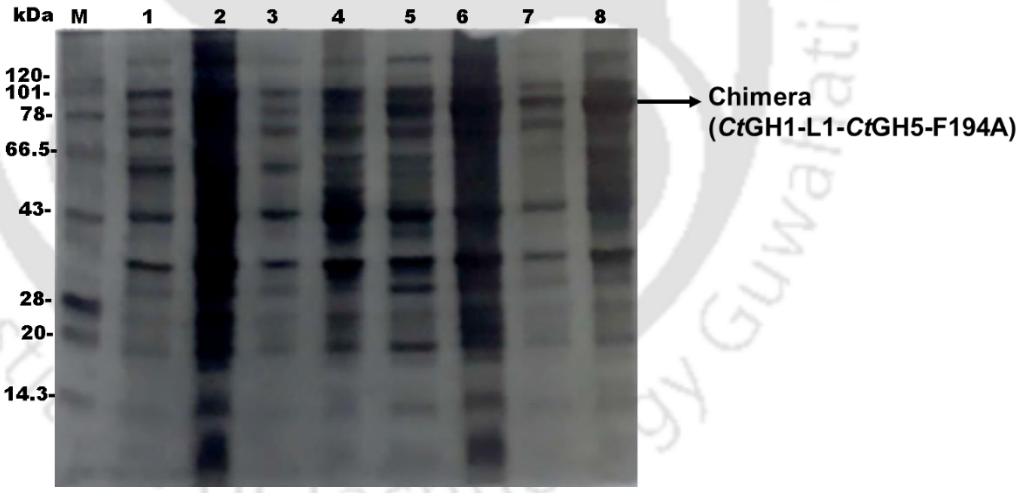


Fig. 2.11 Expression profile of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) at 37°C and 24°C by SDS-PAGE (12%, w/v). Lanes M, Mol wt. marker (Biobharti); 1, Un-induced cells at 37°C; 2, Induced cells at 37°C; 3, Sonicated cell pellet at 37°C; 4, Cell free extract; 5, Un-induced cells at 24°C; 6, Induced cells at 24°C; 7, Sonicated cell pellet at 24°C; 8, Cell free extract at 24°C.

2.3.5 Purification of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and individual enzymes β-glucosidase (*Ct*GH1) and endoglucanase (*Ct*GH5-F194A)

The recombinant Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and the individual enzyme modules, *Ct*GH5-F194A and *Ct*GH1 were purified by immobilized metal affinity chromatography (IMAC) as mentioned in section 2.2.11.2. The mutant *Ct*GH5-F194A and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) were expressed as soluble proteins and after purification displayed homogeneous single bands on SDS-PAGE gel shown in Fig. 2.12 and Fig. 2.13, respectively. The expression and purification of wild-type *Ct*GH5 and *Ct*GH1 (β-glucosidase) were achieved by following the protocol mentioned in Bharali *et al.,* 2005 and Sharma *et al.,* 2019, respectively. The comparative analysis of molecular masses of the recombinant Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual modules *Ct*GH1, *Ct*GH5-F194A along with wild-type *Ct*GH5 including the N-terminal histidine tag or C-terminal histidine tag were theoretically calculated to be 89 kDa, 52 kDa and 35 kDa, respectively, which are in close agreement with those observed by SDS-PAGE gel as shown in Fig. 2.14.

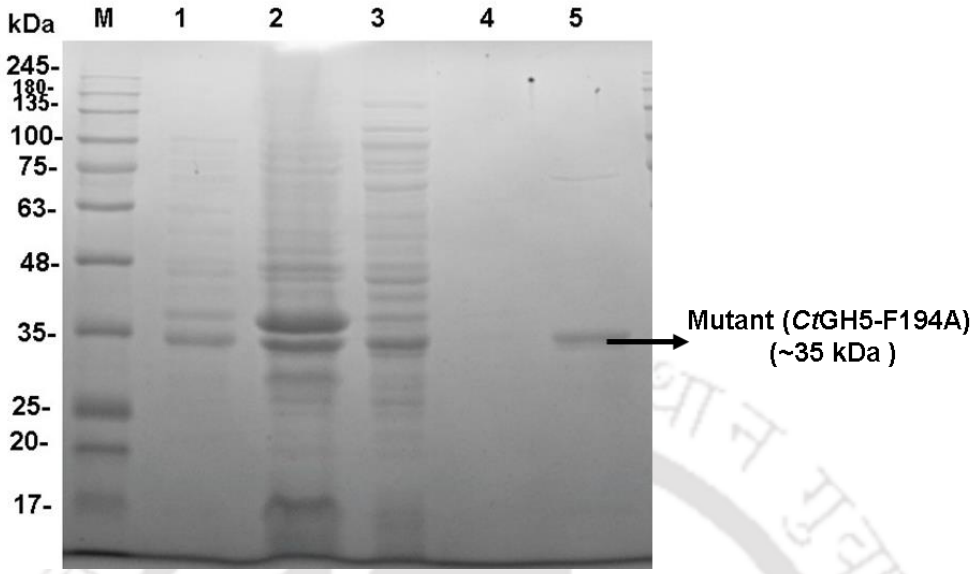


Fig. 2.12 Expression and purification analysis of mutant *Ct*GH5-F194A by SDS-PAGE (12%, w/v). Lanes M: Protein marker (Biobharti pre-stained), 1: Un-induced cells, 2: Induced cell pellet after sonication, 3: Cell free extract, 4: Last column wash, 5: *Ct*GH5-F194A after column purification.

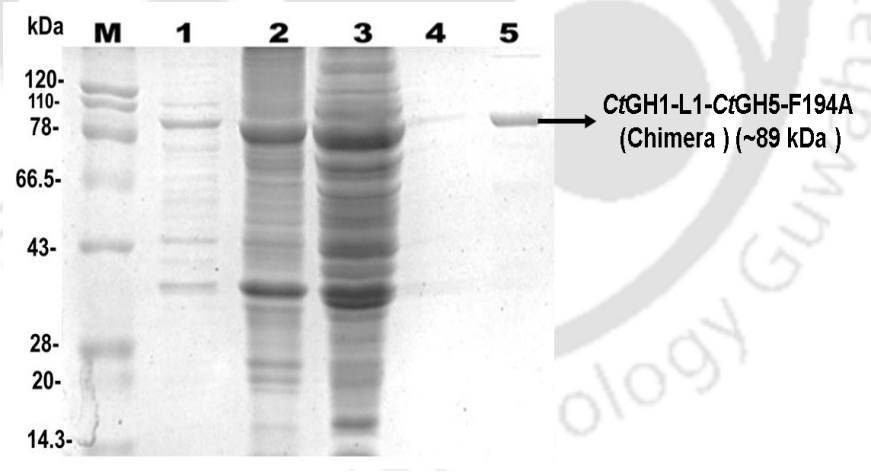


Fig. 2.13 Expression and purification analysis of *Ct*GH1-L1-*Ct*GH5-F194A by SDS-PAGE (12%, w/v). Lanes M: Protein marker (Biobharti), 1: Un-induced cells, 2: Induced cell pellet after sonication, 3: Cell free extract, 4: Last column wash, 5: *Ct*GH1-*Ct*GH5-F194A after column purification.

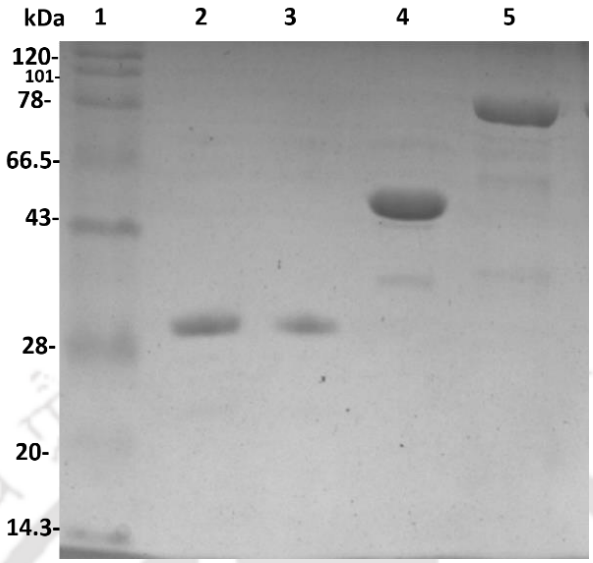


Fig. 2.14 SDS-PAGE (12%, w/v) analysis of the different purified enzymatic constructs. Lanes: 1, Protein marker (Biobharti); 2, wild-type *Ct*GH5; 3, mutant *Ct*GH5-F194A; 4, *Ct*GH1 5, Chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

2.3.6 Production of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and individual modules *Ct*GH1 (β-glucosidase), *Ct*GH5-F194A and wild-type *Ct*GH5 (endoglucanase) using LB medium

The recombinant proteins wild-type *Ct*GH5 and mutant *Ct*GH5-F194A were eluted in 2x1 ml fractions obtained from 100 ml culture. *Ct*GH1 and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) were eluted in 3x1 ml fractions obtained from 100 ml culture. After dialysis the concentration of wild-type *Ct*GH5, *Ct*GH5-F194A, *Ct*GH1 and *Ct*GH1-L1-*Ct*GH5-F194A were 0.9 mg/ml, 0.6 mg/ml 0.54 mg/ml and 1.08 mg/ml, respectively, as determined by UV method (Table 2.16).

Table 2.16 Purified recombinant proteins obtained from 100 ml cultures.

Recombinant Protein	Protein concentration (mg/ml)	Purified protein (ml)	Total amount of purified protein (mg)
wild-type *Ct*GH5	0.9	2.0	1.8
*Ct*GH5-F194A	0.6	2.0	1.2
*Ct*GH1	0.54	3.0	1.62
*Ct*GH1-L1-*Ct*GH5-F194A	1.08	3.0	3.24

2.4 Conclusions

The site-directed mutagenesis was performed for the wild-type *Ct*GH5 to develop mutant *Ct*GH5-F194A. The mutant plasmid was transformed in *E. coli* DH5α competent cells. The plasmid was extracted and sequenced to confirm the mutation at the specific residue. This recombinant plasmid was transformed in *E. coli* BL21 cells for *Ct*GH5-F194A expression. The mutant *Ct*GH5-F194A was further engineered for development of Chimera by fusing *Ct*GH1 (β-glucosidase) from *Clostridium thermocellum* to the N-terminal of mutant *Ct*GH5-F194A using the natural linker. The PCR amplified fragment of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) showed a band of ~2.3 kb. The restriction enzyme digested fragment of the Chimera was ligated to the linearized pET28a(+) vector. The ligated mixture was transformed into the DH5α cells. The positive clone was confirmed by the double digestion using the restriction enzymes *NheI* and *XhoI*. The gel electrophoresis of the digested product showed a band of ~ 5.4 kb was produced for the pET-28a(+) vector and the band ~2.3 kb produced for the corresponding for the chimeric gene. The recombinant plasmid was then transformed in *E. coli* BL21 cells for Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) expression. The recombinant protein Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) along with its individual

module *Ct*GH1, wild-type *Ct*GH5 and *Ct*GH5-F194A were purified using Immobilized Metal Ion Affinity Chromatography (IMAC). The purified recombinant proteins displayed a band of approximately, 89 kDa for Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), 52 kDa for *Ct*GH1 and a band of approximately, 35 kDa for wild-type *Ct*GH5 and mutant *Ct*GH5-F194A on SDS-PAGE gels. The amount of purified recombinant Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), *Ct*GH1, wild-type *Ct*GH5 and *Ct*GH5-F194A obtained from 100 ml *E. coli* BL21 cell cultures were 3.24 mg, 1.62 mg,1.8 mg and 1.2 mg, respectively.

References

Adlakha, N., Sawant, S., Anil, A., Lali, A., & Yazdani, S. S. (2012). Specific fusion of β-1,4-endoglucanase and β-1,4-glucosidase enhances cellulolytic activity and helps in channeling of intermediates. *Applied and Environmental Microbiology*, *78*, 7447-7454.

Beguin, P., (1990). Molecular biology of cellulose degradation. *Annual Reviews in Microbiology*, *44*, 219-248.

Bharali S., Purama R.K., Majumder A., Fontes, C. M., & Goyal, A. (2005). Molecular cloning and biochemical properties of family 5 glycoside hydrolase of bi-functional cellulase from *Clostridium thermocellum*. *Indian Journal of Microbiology*, *45*, 317-321.

Engler, M.J. and Richardson, D.C. (1982) DNA ligases. In P.D. Boyer (ed.), *The Enzymes*. Academic Press, San Diego. *15*, 3-30.

Grabnitz, F., Seiss, M., RucknageL, K. P., & Staudenbauer, W. L. (1991). Structure of the β-glucosidase gene bglA of *Clostridium thermocellum*: Sequence analysis reveals a superfamily of cellulases and β-glycosidases including human lactase/phlorizin hydrolase. *European Journal of Biochemistry*, *200*, 301-309.

Jamaldheen, S. B., Sharma, K., Rani, A., Moholkar, V. S., & Goyal, A. (2018). Comparative analysis of pretreatment methods on Sorghum (*Sorghum durra*) stalk agrowaste for holocellulose content. *Preparative Biochemistry and Biotechnology*, *48*, 1-27.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*, 680.

Layne, Ennis. (1957) Spectrophotometric and turbidimetric methods for measuring proteins.447-454.

Lu, P., & Feng, M. G. (2008). Bifunctional enhancement of a β-glucanase-xylanase fusion enzyme by optimization of peptide linkers. *Applied Microbiology and Biotechnology*, *79*, 579-587.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (No. Ed. 2). Cold Spring Harbor Laboratory Press.

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

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

Stoscheck, C. M. (1990). Quantitation of protein. In *Methods in Enzymology*, *182*, 50-68. Academic Press.

Studier, F. W., Davanloo, P., Rosenberg, A. H., Moffatt, B. A., & Dunn, J. J. (1990). *U.S. Patent No. 4,952,496*. Washington, DC: U.S. Patent and Trademark Office.

Telke, A. A., Ghatge, S. S., Kang, S. H., Thangapandian, S., Lee, K. W., Shin, H. D., .& Kim, S. W. (2012). Construction and characterization of chimeric cellulases with enhanced catalytic activity towards insoluble cellulosic substrates. *Bioresource Technology*, *112*, 10-17.

Urbanowicz, B. R., Bennett, A. B., del Campillo, E., Catala, C., Hayashi, T., Henrissat, B., & Teeri, T. T. (2007). Structural organization and a standardized nomenclature for plant endo-1,4-β-glucanases (cellulases) of glycosyl hydrolase family 9. *Plant Physiology*, *144*, 1693-1696.

Yennamalli, R. M., Rader, A. J., Kenny, A. J., Wolt, J. D., & Sen, T. Z. (2013). Endoglucanases: insights into thermostability for biofuel applications. *Biotechnology for Biofuels*, *6*, 136.

Yuan, S. F., Wu, T. H., Lee, H. L., Hsieh, H. Y., Lin, W. L., Yang, B. & Ho, M. C. (2015). Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*. *Journal of Biological Chemistry*, *290*, 5739-5748.

.

Chapter 3

Comparative biochemical characterization of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual catalytic modules - β-glucosidase (*Ct*GH1) and mutant endoglucanase (*Ct*GH5-F194A)

3.1 Introduction

The decline in the fossil fuels had shifted the global focus towards the alternative renewable energy sources. The production of Bioethanol from cheap plant lignocellulosic wastes is one of the alternative strategies to produce more economical and eco-friendly biofuels. The lignocellulosic plant cell walls are typically made up of cellulose (40-60%), hemicelluloses (20-40%), and lignin (10-25%) and other components such as proteins, ash and other extractives. The cellulose and hemicellulose can be degraded to simple sugars by cellulases and hemicellulases. This simple sugars can be utilized by the yeast to produce bioethanol. The major portion of the plant lignocellulose is made up of cellulose. For the complete degradation of the cellulosic part three different types of enzymes were required *viz*. endoglucanase, cellobiohydrolase and β-glucosidase (Beguin *et al.,* 1990). Endo-β-1,4-glucanase acts randomly on the cellulose chain and produces cellodextrin which are the larger cellooligosacharide as a hydrolyzed product (Urbanowicz *et al.,* 2007). Cellobiohydrolase acts on the end of the cellodextrin and releases cellobiose as the key

product (Barr *et al.,* 1996), while β-glucosidase hydrolyzes the cellobiose to form two molecules of glucose (Grabnitz *et al.*, 1991). The production of the three enzymes is a cost intensive process (McKee *et al.,* 2012). Also, these enzymes were required to function synergistically for the efficient and complete degradation of the polymer (Lynd *et al.*, 2002). An interesting approach to minimize the production cost of different enzymes is to fuse one or more enzymes in a single polypeptide chain (Fan *et al.,* 2009). The most feasible process for minimizing the production of one or more enzymes is to produce Chimeras using the existing functionalities present in their gene sequence (Nixon *et al.*, 1998). There are many ways for the production of Chimera by fusion enzyme technology which comprised end-to-end fusion of different domains. This fusion technology of end-to end fusion may lead to misfolding of the two domains in the polypeptide chain (Liu *et al.,* 2012). This obstacle led to the development of linkers, which were incorporated between different domains. These linkers provide extended conformation and stability to the different domains in the polypeptide chain (Liu *et al.,* 2012). Efforts were made to develop chimeric cellulases by attaching a Carbohydrate Binding Module (CBM3) to a family 9 endoglucanase, Cel9A (Telke *et al.,* 2012) and to a family 5 endoglucanase, Cel5H (Shi *et al.,* 2013) resulting in 8- to 12-fold higher catalytic efficiency. Telke *et al.,* 2012 also proved from the computational protein modeling that the active site configuration of native catalytic module does not get affected upon fusing the two modules. Therefore, other catalytic modules with different substrate specificities and mode of actions can be introduced in a single polypeptide chain for development of chimeric cellulases with increased catalytic efficiency. Adlakha *et al.*,2012 used similar approaches for the development of new endoglucanase (Endo5A) and β-glucosidase (Gluc1C) from *Paenibacillus sp.* having higher catalytic

efficiencies. They observed that Gluc1C improves the efficiency of Endo5A and releases the reducing sugar from carboxymethyl cellulose. Therefore, the fusion of Gluc1C with Endo5A resulted in 3.3- and 2- fold higher activity towards β-glucosidase and endoglucanase, respectively. In this study, the developed Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) by fusion of β-glucosidase (*Ct*GH1) at the N-terminal and mutant endoglucanase (*Ct*GH5-F194A) at C-terminal as described in chapter 2 was biochemically and functionally characterized and compared with its individual native enzymes *viz.* β-glucosidase (*Ct*GH1) and mutant endoglucanase (*Ct*GH5-F194A).

3.2 Materials and Methods

3.2.1 Substrates and reagents

Barley *β*-D-glucan was purchased from Megazyme International (Ireland). Carboxy methylcellulose (CMC), Lichenan, cellobiose and glucose were purchased from Sigma-Aldrich, USA Synthetic *p*NP-β-D-glucopyranoside was purchased from SRL Pvt. Ltd., India. Sodium carbonate, sodium potassium tartarate, sodium bicarbonate, sodium sulphate, sodium phosphate (monobasic), sodium phosphate (dibasic) were procured from HiMedia Laboratories Pvt. Ltd., India. Sodium arsenate, ammonium molybdate, sulphuric acid, hydrochloric acid, acetone, acetonitrile and acetic acid were purchased from Merck Limited, India. The TLC plate was purchased from Merck KGaA, Darmstadt, Germany.

3.2.2 Enzyme assay

The enzyme assays for wild-type *Ct*GH5 and mutant *Ct*GH5-F194A were performed with 20 mM citrate phosphate buffer, pH 4.2 and 5.5, respectively. For enzymatic assay of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), 20 mM citrate phosphate buffer (pH 5.0) was used. 100 µl reaction mixture contained 1.0%, w/v carboxymethyl cellulose as substrate and 10 µl of enzyme (wild-type *Ct*GH5, 0.1 mg/ml), (mutant *Ct*GH5-F194A, 0.05 mg/ml) and (*Ct*GH1-L-*Ct*GH5-F194A, 0.05 mg/ml) and incubated at 60ºC for wild-type *Ct*GH5 and Chimera and at 50ºC for the mutant *Ct*GH5-F194A for 4 min. The enzyme activity was calculated by measuring the reducing sugar concentration by method as described earlier (Nelson, 1944 and Somogyi,1945). The activity calculated using equation mentioned in section 3.2.2.3. The concentration of the reducing sugar was estimated by using glucose as standard. One unit of activity was defined as the amount of enzyme which produced 1 µmole of glucose per min. For the

assay of β-glucosidase activity of Chimera and its individual module *Ct*GH1, the substrate *p*-Nitrophenyl β-D-glucopyranoside (*p*NPG) was used. The enzymatic reaction was performed in a 1 ml reaction mixture in 20 mM citrate phosphate buffer, pH 5.5 containing 1 mM of the substrate and 20 µL *Ct*GH1-L1-*Ct*GH5-F194A or *Ct*GH1 (0.05 mg/ml). The reaction was performed at 70ºC for 5 min for Chimera and 65ºC for *Ct*GH1 for 5 min on a UV visible spectrophotometer (Varian, Cary 100 Bio) with a peltier temperature controller. The reactions were performed for 5 min and continuous absorbance at 405 nm was monitored. The quantification was done by the release of *p*-nitrophenol using 24150 M^{-1} cm^{-1} as an extinction coefficient described by Cartmell *et al.,* 2011. The activity calculated using equation mentioned in section 3.2.2.4.

3.2.2.1 Preparation of reagents for reducing sugar estimation

Reagent A

Sodium carbonate anhydrous	6.25 g
Sodium potassium tatarate	6.25 g
Sodium bicarbonate	5.0 g
Sodium sulphate anhydrous	50.0 g

The above mentioned components were dissolved in 100 ml of deionized water and the final volume was adjusted to 250 ml. The solution was filtered (Whatman No. 1) and stored at 30°C.

Reagent B

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

Reagent C

Reagent C was prepared in two steps under dark. First, 2.5 g of ammonium molybdate was dissolved in 45 ml of deionized water in 100 ml beaker and 2.1 ml of concentrated sulphuric acid was added to it. In another beaker 0.3 g of sodium arsenate was dissolved in 2.5 ml of deionized water. Now, this solution was added to ammonium molybdate solution and the contents were mixed (total volume was around 50 ml). The solution was filtered (Whatman No. 1) under dark conditions and stored at 37°C. The solution was used after 24h incubation.

Reagent D

Reagent D was prepared by mixing reagent A and reagent B in the ratio 25:1. Reagent D was freshly prepared for the assay.

3.2.2.2 Generation of standard plot of D-glucose

The standard plot using of D-glucose was prepared by varying the concentration from 5-500 µg/ml. 100 µl reaction mixture contained 50 mM sodium phosphate buffer pH 7.0 and in a 1.5 ml microcentrifuge tube D-glucose was incubated at 50°C for 15 min and then 100 µl of solution D (Section 3.2.2.1) was added to it. The reaction mixture was then heated in boiling water bath for 20 min and cooled. 100 µl of solution C (Section 3.2.2.1) was added and the contents were mixed. Then 700 µl of deionized water was added to make the final volume to 1 ml. The absorbance at 500 nm (A_{500}) was measured using UV-Visible spectrophotometer (GeneQuant, GE Healthcare, USA) against a buffer blank. A standard plot of A_{500} versus D-glucose concentration (µg/ml) was generated and 1 A_{500} equivalent of D-glucose (µg/ml) was calculated. 1 A_{500} equivalent of D-glucose (µg/ml) was converted into mg/ml for calculation of enzyme activity.

3.2.2.3 Calculation of enzyme activity for wild-type endoglucanase (CtGH5), mutant CtGH5-F194A and Chimera (CtGH1-L1-CtGH5-F194A)

The enzyme activity of the enzyme was expressed as U/ml, the specific activity as U/mg of protein and the k_{cat} or turn over number (activity) was expressed as U/µmol of protein. One unit (U) of enzyme activity is defined as the amount of enzyme that liberates 1 µmole of reducing sugar per min. The enzyme activities of endoglucanase (wild-type *Ct*GH5, mutant *Ct*GH5-F194A) and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) were calculated as described below,

$$\text{Enzyme activity (U/ml)} = \frac{\Delta A_{500} \times C \times V}{180 \times t \times v} = (\mu\text{mole/min/ml})$$

where,

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

C = A_{500} equivalent of D-glucose normal form concentration from the standard plot

V = volume of the reaction mixture (ml)

t = time of reaction (min)

180 = molecular mass of D-glucose

v = volume of the enzyme taken in assay (ml) for reducing sugar estimation.

$$k_{cat} \text{ or Activity (U/µmol)} = \frac{\text{Enzyme activity (U/ml)}}{\text{Concentration of protein used (µmol)}}$$

3.2.2.4 Calculation of enzyme activity for β-glucosidase (CtGH1) and Chimera (CtGH1-L1-CtGH5-F194A)

Enzyme activity was expressed as U/ml and k_{cat} or turn over number (activity) was expressed as U/μmol. One unit of enzyme activity was defined as the amount of enzyme required to produce 1 μmol *para*-Nitrophenol (product) per minute. The enzyme activities of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its individual module *Ct*GH1 were calculated as described below,

$$\text{Enzyme activity (U/ml)} = \frac{\Delta A_{405} \text{ x } 1000}{\varepsilon \text{ x t x v}}$$

Where,

ΔA_{405} = Change in absorbance of *para*-Nitrophenol

ε = Molar Extinction coefficient (24150 M^{-1} cm^{-1}) of *para*-Nitrophenol

t = Time of reaction (min)

v = Total volume of reaction (ml)

$$k_{cat} \text{ or Activity (U/μmol)} = \frac{\text{Enzyme activity (U/ml)}}{\text{Concentration of protein used (μmol)}}$$

3.2.3 Determination of optimum pH and temperature of wild-type *Ct*GH5 endoglucanase, mutant *Ct*GH5-F194A and Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)

The pH profiles of the enzymes were performed under wide range of pH. The buffers 20 mM citrate phosphate, for pH (3-7) and 20 mM sodium phosphate, for pH (6.5- 8) were used. The pH profile was generated by using substrate 1% (w/v) CMC using temperature 60ºC for wild-type *Ct*GH5 and *Ct*GH1-L1-*Ct*GH5-F194A and 50ºC

for mutant *Ct*GH5-F194A. Similarly, the optimum temperatures of enzymes were determined by incubating the enzymes at varying temperatures between 30-80ºC, at their respective optimised pH and optimised time period. For the β-glucosidase activity of Chimera the optimum pH was obtained by using *p*NPG (1 mM) as a substrate in 1 ml of reaction mixture under different pH ranges, of 20 mM citrate phosphate, pH (3-7) and sodium phosphate, pH (6.5-8). The temperature optima for β-glucosidase activity of Chimera was performed by assaying it by *p*NPG as substrate at different temperature ranging from 30 to 80ºC for 5 min.

3.2.4 pH stability analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) by its individual modules

The pH stability analysis for Chimera and its individual enzyme, *Ct*GH5-F194A for endoglucanase activity was performed by incubating 0.05-0.1 mg/ml of enzymes at different pH, using 50 mM citrate phosphate, pH (3-7) and 50 mM sodium phosphate, ranging from pH (6.5-8.0) for 1 h at 25ºC. An aliquot of 10 µl was taken and the assay for endoglucanase activity was performed at their respective optimum pH and temperature, for Chimera at pH 5.0 and 60°C and for *Ct*GH1 at pH 5.5 and 50°C for 4 min. The pH stability analysis for Chimera and its individual enzyme, *Ct*GH1 for β-glucosidase activity was performed by incubating 0.05 mg/ml of enzymes (Chimera and *Ct*GH1) at different pH, using 50 mM citrate phosphate, pH (3-7) and 50 mM sodium phosphate, ranging from pH (6.5-8) for 1 h at 25ºC. An aliquot of 20 µl was taken and the assay was performed in triplicates for Chimera at pH 5.5 and 70°C and for *Ct*GH1 at pH 6.5 and 65°C for 5 min.

3.2.5 Thermal stability analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) from its individual modules

The thermal stability analysis of Chimera and the individual enzymes, wild-type *Ct*GH5 and *Ct*GH5-F194A for endoglucanase activity was performed by incubation of 0.05-0.1 mg/ml of (wild-type *Ct*GH5, *Ct*GH5-F194A and Chimera) at varying temperature from 20-80ºC, for 60 min. An aliquot was taken for assay of activity by incubating the wild-type *Ct*GH5 at pH 4.2 and 60°C, *Ct*GH5-F194A at pH 5.5 and 50°C and Chimera at pH 5.0 and 60°C for 4 min. The thermal stability analysis of Chimera and its individual enzyme, *Ct*GH1 for β-glucosidase activity was performed by incubation of 0.05 mg/ml of enzyme (Chimera and *Ct*GH1) at varying temperatures from 20-80ºC for 60 min and the assay was performed by taking an aliquot from each and incubating the Chimera at pH 5.5 at 70°C for 5 min and incubating *Ct*GH1 at 65°C and pH 6.5 for 5 min.

3.2.6 Substrate specificities and kinetic parameters of wild-type *Ct*GH5, mutant *Ct*GH5-F194A, *Ct*GH1 and Chimera

The kinetic parameters of wild-type *Ct*GH5, mutant *Ct*GH5-F194A and Chimera for endoglucanase activity against various substrates i.e. CMC, Lichenan and β-Glucan was determined. The optimum conditions used for wild-type *Ct*GH5 were 20 mM citrate phosphate buffer, pH 4.2 at 60ºC, for mutant *Ct*GH5-F194A 20 mM citrate phosphate buffer, pH 5.5 at 50ºC and for Chimera, 20 mM citrate phosphate buffer, pH 5.0 at 60ºC. The reaction was performed in 100 µl reaction mixture for 4 min containing varying substrate concentration ranging from 0.01% (w/v) to 2% (w/v) and using enzyme 10 µl (0.005-0.1 mg/ml). The kinetic parameters were determined by taking the initial velocities in Michaelis-Menten equation. The kinetic parameters of *Ct*GH1 and

Chimera for β-glucosidase activity against substrate *p*NPG was determined by varying its concentration from 0.01 mM to 1 mM. The reactions were performed at their respective optimized conditions, for *Ct*GH1, 20 mM citrate phosphate buffer, pH 6.5 at 65ºC and for Chimera, 20 mM citrate phosphate buffer, pH 5.5 at 70ºC. The reaction was monitored by taking the absorbance at 405 nm (A_{405}) for 5 min. The quantification was done by calculating the concentration of *p*-nitrophenol released and by using its molar extinction coefficient as 24150 $M^{-1}cm^{-1}$ as described by Cartmell *et al.*, 2011.

3.2.7 Protein melting studies

The protein melting studies for Chimera and the individual enzymes, wild-type *Ct*GH5, mutant *Ct*GH5-F194A and *Ct*GH1 were performed at different temperatures (25ºC to 100ºC) by monitoring the change in absorbance at 280 nm using an UV-Visible spectrophotometer (Varian, Cary 100-Bio) coupled with a peltier temperature controller. The concentration of each enzyme used was 0.05 mg/ml Chimera and wild-type *Ct*GH5 were dissolved in 50 mM sodium acetate buffer, pH 5.0 while the mutant *Ct*GH5-F194A and *Ct*GH1 were dissolved in 50 mM MES buffer, pH 5.5. The 1 ml protein sample were kept at 30°C for 10 min to reach the equilibrium. A curve was generated by plotting the change in the absorbance at 280 nm due to protein unfolding with respect to the change in temperature.

3.2.8 End-product determination by Thin Layer Chromatography

The mode of action of Chimera with respect to its individual enzymes was determined by the TLC. The wild-type *Ct*GH5, *Ct*GH5-F194A, *Ct*GH1, *Ct*GH1 + *Ct*GH5-F194A and Chimera were used with substrates 1% (w/v) CMC and/or 1% (w/v) cellobiose. The 100 µl reaction mixtures of these enzymes were set up by using 20 mM citrate phosphate buffer incubated under their respective optimum conditions for 1 h

except in the case of Chimera. The enzyme concentration 0.2 µM (was kept constant) for each case. The time dependent hydrolysis of substrate by Chimera was performed from 5 min -12 h using 1% (w/v) CMC under optimum condition of 20 mM citrate phosphate buffer, pH 5.0 at 60ºC using 0.2 µM concentration of enzyme. The reactions were stopped by adding two volumes absolute ethanol. The 0.3 ml reaction mixtures were centrifuged at 13000*g* for 10 min. The supernatants were separated and dried under oven at 80ºC and concentrated up to a volume of 25 µl. From these concentrated hydrolysed products, 0.8 µl samples of each and 1.0 µl of standards (glucose and cellobiose, 2 mg/ml) were applied on silica gel coated aluminium TLC plates (TLC Silica gel 60 F254, Merck India Ltd.). The mobile phase used was butanol/acetic acid/MiliQ water in the ratio of 2:1:1, for separation. The visualization of the migrated sugars was achieved by immersing the TLC plate in the solution containing sulphuric acid: methanol 5:95, (v/v); and α-naphthol 0.5% (w/v). The plate was dried under oven at 80°C for 15 min to visualize the spots of the standard and hydrolysed products.

3.2.9 Analysis of hydrolysed products from pretreated Sorghum stalk by chimera and individual enzyme

The TLC analysis of hydrolysed products released from pretreated Sorghum stalk biomass by Chimera and its individual enzymes was also performed. The enzymatic hydrolysis of 1% (w/v) Sorghum stalk pretreated by 1% (w/v) NaOH and autoclaving was performed in 20 mM citrate phosphate buffer, pH 5.0 by Chimera, 20 mM citrate phosphate buffer, pH 5.5 by *Ct*GH5-F194A and 20 mM citrate phosphate buffer, pH 5.5, by mixture of *Ct*GH1 + *Ct*GH5-F194A. One ml reactions were carried out by maintaining equimolar concentration of each enzyme (1.1 µM) and incubating in a shaker incubator at 50°C and 180 rpm for 48h. The reactions were stopped by

adding two volumes absolute ethanol. The 1 ml reaction mixtures were centrifuged at 13000*g* for 10 min. The supernatants were separated and dried under oven at 80ºC and concentrated up to a volume of 25 µl. 0.8 µl of concentrated hydrolysed products and 1.0 µl of standards were applied and analysed on silica gel coated aluminium TLC plates (TLC Silica gel 60 F254, Merck India Ltd.). The standard used were glucose and cellobiose at concentrations, 2 mg/ml. The mobile phase used was butanol/acetic acid/MiliQ water in the ratio of 2:1:1, for separation. The visualization of the migrated sugars was achieved by immersing the TLC plate in the solution containing sulphuric acid: methanol 5:95, (v/v); and α-naphthol 0.5% (w/v). The plate was dried under oven at 80°C for 15 min to visualize the spots of the standard and hydrolysed products.

3.3 Results and Discussion

3.3.1 Effect of temperature and pH on different enzymatic modules

The Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) showed maximum endoglucanase activity using 1% (w/v) CMC at temperature 60°C (Fig. 3.1A) and at pH 5.0 (Fig. 3.1B), while the mutant *Ct*GH5-F194A showed maximum activity at 50°C (Fig. 3.1A) and pH 5.5 (Fig. 3.1B). The wild-type *Ct*GH5 enzyme showed maximum activity at 60°C (Fig. 3.1A) and pH 4.2 (Fig. 3.1B). The wild-type and the Chimera showed temperature maxima at 60°C whereas the mutant *Ct*GH5-F194A showed temperature maxima at 50°C. This showed that mutant *Ct*GH5-F194A is temperature sensitive and gets inactivated. The temperature sensitivity and inactivation of mutant *Ct*GH5-F194A is counteracted by the linker or the adjacent more thermotolerant module *Ct*GH1 at the N-terminal conferring the Chimera higher temperature resistance. The pH optima of wild-type *Ct*GH5 is 4.2 which shifted to 5.5 in mutant *Ct*GH5-F194A while in the Chimera the maximum endoglucanase activity was observed at pH 5.0. This pH optimum shift has been associated to either the change of pH in the microenvironment created by the mutant in the Chimera or their altered tertiary structures as also reported earlier (Bulow, 1987).

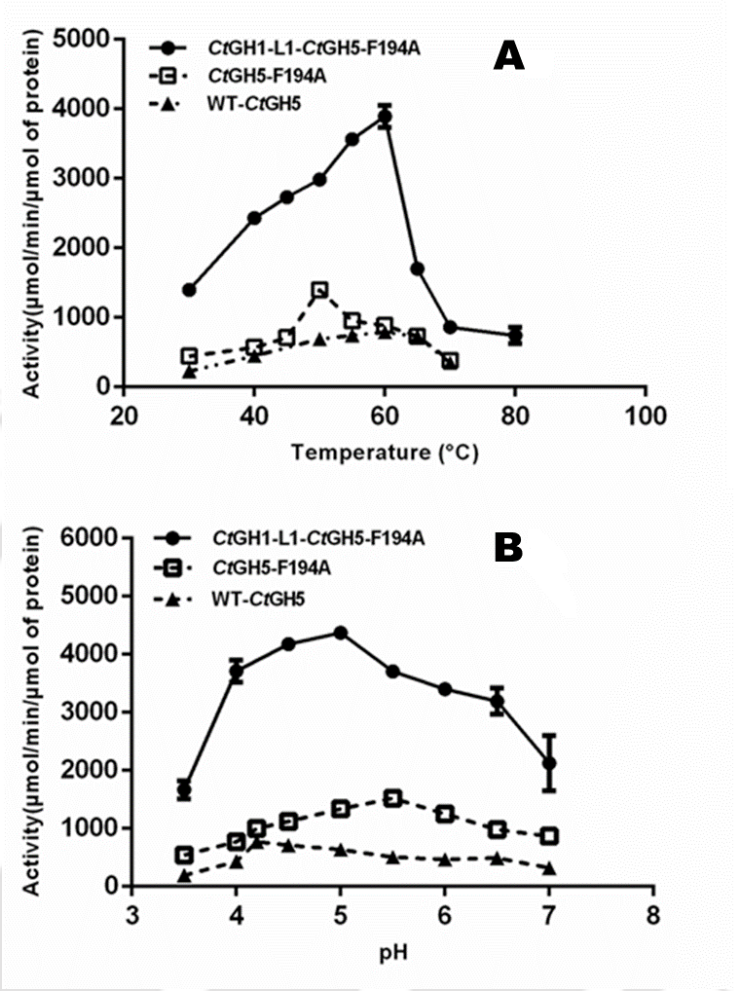
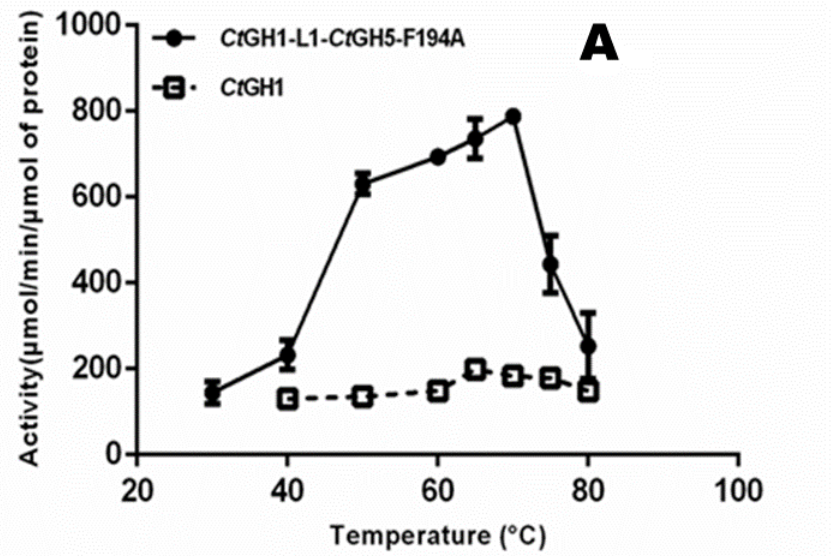


Fig. 3.1 Endoglucanase activity of Chimera and individual enzymes (wild-type *Ct*GH5 and *Ct*GH5-F194 A) was measured a) by incubating the enzymes with 1% (w/v) CMC at between temperature range, 30-80°C and B) by incubating the enzymes with 1% (w/v) CMC at different pH. The experimental details are mentioned in the methods. The experiment was performed in triplicate. Each data point represents the mean ± SE (n=3).

The Chimera showed maximum β-glucosidase activity using substrate, *p*NPG at 70°C (Fig. 3.2A) and at pH 5.5 (Fig. 3.2B), while, the individual enzyme, *Ct*GH1 showed maximum activity at 65ºC (Fig. 3.2A) and pH 6.5 (Fig. 3.2B). The Chimera displayed improved temperature optima for β-glucosidase activity than its individual counterpart *Ct*GH1 which may be due to the role played by the linker sequence between the two modules. A shift in optimum pH towards acidic pH, 5.5 was also observed in the β-glucosidase activity of the Chimera (similar to endoglucanase activity), from 6.5 for individual enzyme, *Ct*GH1.



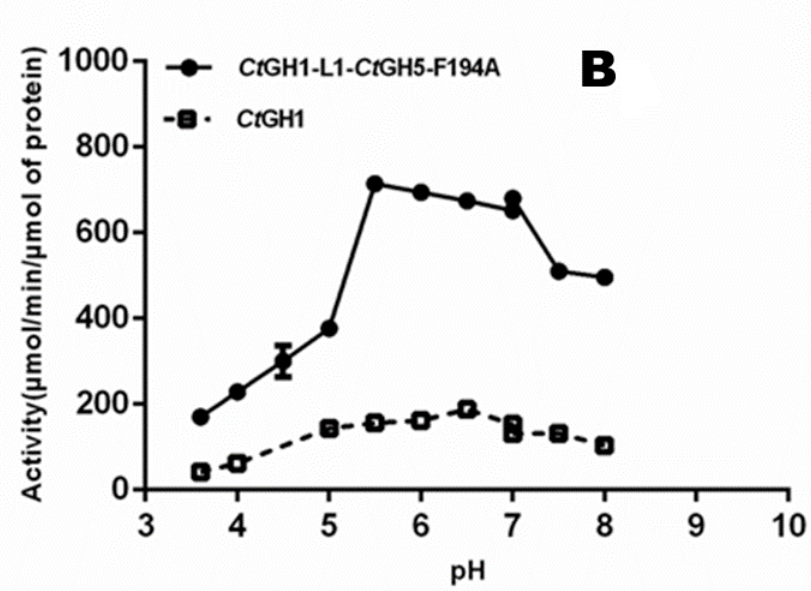


Fig. 3.2 β-Glucosidase activity of Chimera and individual enzyme (*Ct*GH1) was measured, A) by incubating the enzymes with 1 mM *p*NPG at different temperatures ranging between 30-80°C and B) by incubating the enzymes with 1 mM *p*NPG at different pH. The experimental details are mentioned in the methods. The experiment was performed in triplicate. Each data point represents the mean ± SE (n=3).

3.3.2 pH stability of Chimera and its individual modules

The pH stability analysis of Chimera for endoglucanase activity showed that the enzyme is stable between pH 4.0 and 5.0 and retained 80% activity at pH 5.5 (Fig. 3.3A). The individual mutant *Ct*GH5-F194A was stable between pH 5.0 to 6.0 (Fig. 3.3A), while its activity declines at higher and lower pHs. For β-glucosidase activity the Chimera showed pH stability between pH 5.0 to 5.5 and retained 80% activity at pH 4.0 (Fig. 3.3B), while the individual enzyme *Ct*GH1 was stable between pH 5.5 to 6.0 (Fig. 3.3B). The interesting outcome was that the Chimera displayed pH stability between pH 4.0-5.5 for both the enzyme activities. The closer range of pH stability of two modules in the Chimera may prove beneficial for the synergistic functioning of both enzyme modules in the chimeric form. For wider difference in pH stability of the two, may affect the substrate binding and may result in slower catalysis (Somero 1995). Therefore, the Chimera may play a potential role as a better substitute of individual enzymes, β-glucosidase and endoglucanase, in the applications such as lignocellulosic biomass saccharification for glucose production for bioethanol, biobutanol or lactic acid production.

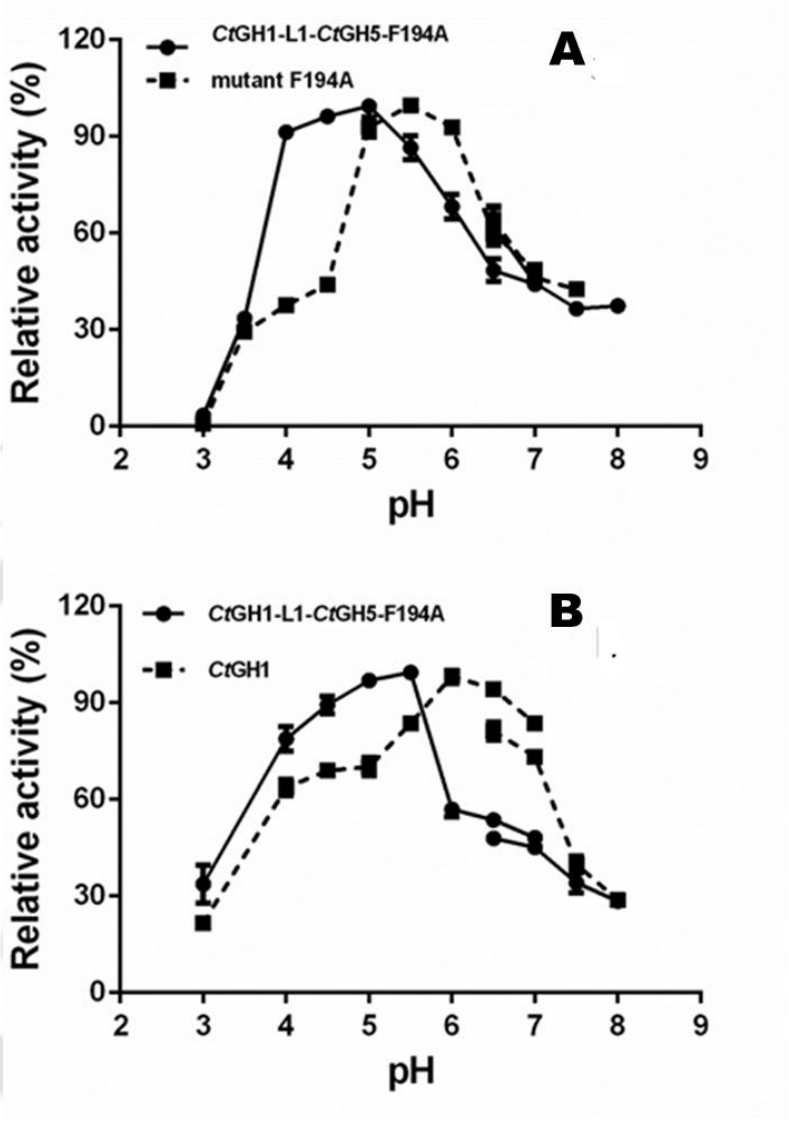


Fig. 3.3 pH stability analysis of different enzymatic modules. A) endoglucanase activity of Chimera and *Ct*GH5-F194A and B) β-glucosidase activity of Chimera and *Ct*GH1. The pH stability was assayed for endoglucanase and β-glucosidase activities by incubating the enzymes at different pH for 60 min. In panel A) for endoglucanase activity, the reactions were performed in 100 µl reaction mixture containing 1% (w/v) CMC under optimized conditions. In panel B) for β-glucosidase activity, the assays were performed in 1 ml reaction containing 1 mM *p*NPG under optimized conditions. The assays were performed in triplicates and each data point represents the mean ± SE (n=3).

3.3.3 Thermal stability of Chimera and its individual modules

The thermal stability analysis showed that the wild-type *Ct*GH5 is stable up to 60ºC (Fig. 3.4 A) and the individual mutant enzyme *Ct*GH5-F194A is stable up to 55ºC (Fig. 3.4 B). This decrease in thermostability of mutant endoglucanase *Ct*GH5-F194A could be because of the side chain flexibility incorporated through mutagenesis of Phe194 by Ala near the substrate binding site of the wild-type *Ct*GH5. This has been also reported earlier that the thermostability of an enzyme is related with the conformational flexibility (Anbar *et al.*, 2012). The thermostability analysis of *Ct*GH1 enzyme showed that it is stable at 60ºC up to 1 h (Fig. 3.4 C). The thermostability analysis of Chimera showed that its endoglucanase activity is stable at 55ºC up to 1h (Fig. 3.4 D) whereas, the thermal stability of β-glucosidase domain of Chimera was found up to 60ºC for 1h (Fig. 3.4 E). This showed that the Chimera developed is thermostable.

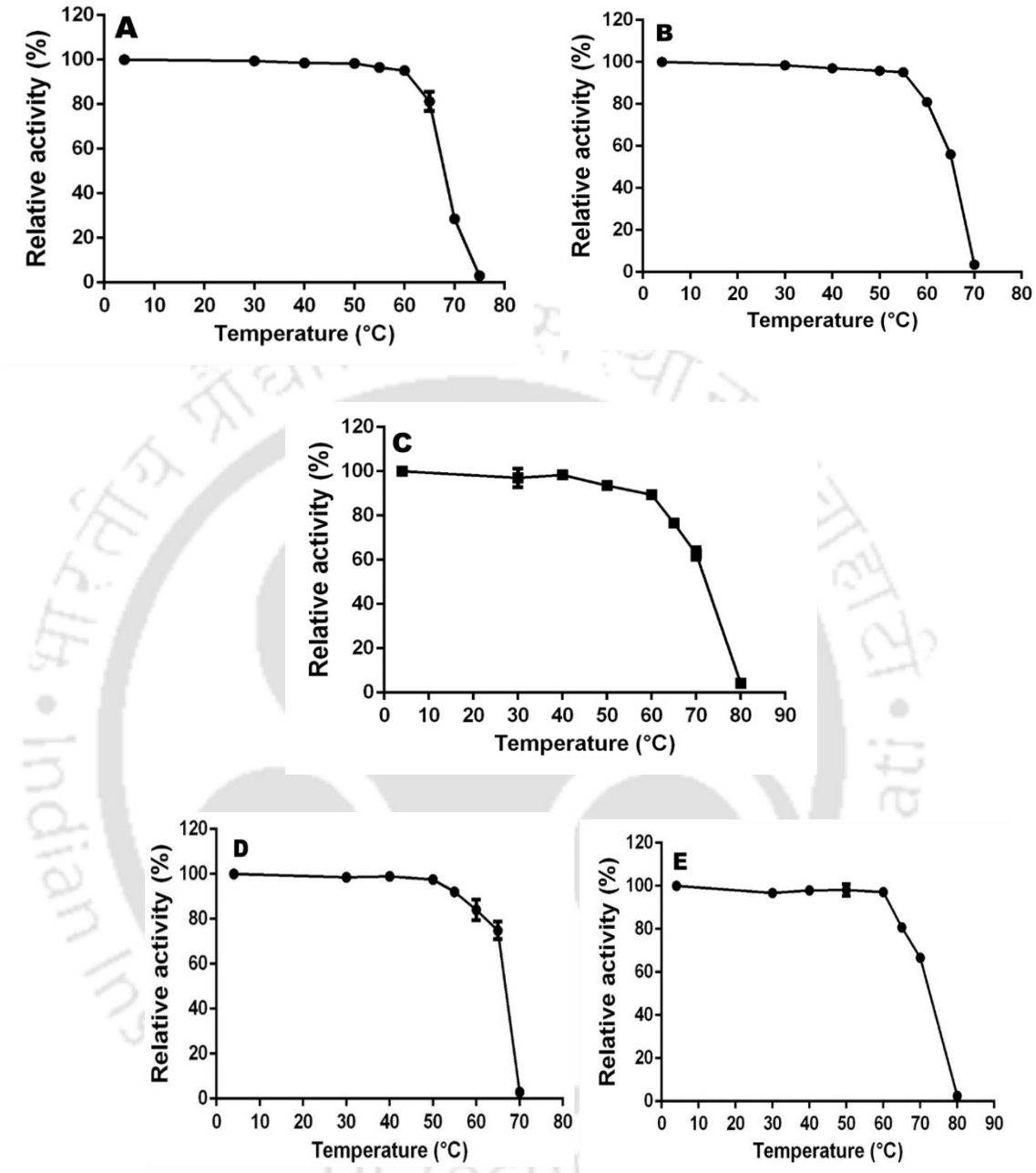


Fig. 3.4 Thermal stability analysis of different enzymatic modules: for A) wild-type *Ct*GH5, B) *Ct*GH5-F194A, C) *Ct*GH1 D) Chimera (endoglucanase) and E) Chimera (β-glucosidase). Each enzyme was incubated at different temperatures for 1h. The endoglucanase activity was determined by using 100 µl reaction mixture containing 1% (w/v) CMC. The β-glucosidase activity was determined by using 1 ml reaction containing 1 mM *p*NPG. The assays were performed in triplicate and each data point represents the mean ± SE (n=3).

3.3.4 Comparison of kinetic parameters of enzyme modules

The K_m and k_{cat} of Chimera were compared against their individual modules for substrates Lichenan, β-Glucan, CMC and *p*NPG (Table 3.1) under optimal conditions. For endoglucanase activity Chimera showed higher catalytic efficiency (k_{cat}/K_m) against Lichenan, β-Glucan, and CMC i.e.$2.9\pm0.8\times10^5$ ml $min^{-1}mg^{-1}$, $1.6\pm4.7\times10^5$ ml $min^{-1}mg^{-1}$, $1.2\pm0.4\times10^4$ ml $min^{-1}mg^{-1}$ which was 3-5 fold higher than the individual mutant *Ct*GH5-F194A. For β-glucosidase activity Chimera showed $4.1\pm0.2\times10^4$ $mM^{-1}min^{-1}$ catalytic efficiency (k_{cat}/K_m) against *p*NPG which was 5.5 folds higher than the individual module *Ct*GH1. Previous report on hybrid enzyme between exoglucanase (cellobiohydrolase and β-glucosidase) *Ct*CD-*Cc*BG showed 2-fold higher activity with phosphoric acid swollen cellulose (PASC) while retaining similar activity for β-glucosidase as (Lee *et al*., 2011). Another fusion enzyme between endoglucanase (Endo5A) and β-glucosidase (Gluc1C) from *Paenibacillus sp.* with synthetic Glycine-serine linkers (G_4S_3) in between, showed 3.2- and 2- fold higher activity against both β-glucosidase and endoglucanase activities, respectively (Adlakha, *et al*., 2012). It was also previously reported that inter-domain flexibility between two modules in a hybrid enzyme results in significant increase in activity (Lu and Feng, 2008). Thus, the results from present study revealed that the chimeric enzyme constructed by fusing β-glucosidase (*Ct*GH1) at N-terminal and endoglucanase (*Ct*GH5-F194A) at C-terminal using natural linker gave much higher catalytic efficiency for β-glucosidase and endoglucanase activities than the individual enzymes.

Table 3.1: Kinetic parameters of individual and chimeric enzyme against different substrates.

Substrate	Wild-type *Ct*GH5			Mutant *Ct*GH5-F194A			Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)			*Ct*GH1		
	K_m (mg.ml^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (ml.min^{-1}mg^{-1})	K_m (mg.ml^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (ml.min^{-1}mg^{-1})	K_m (mg.ml^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (ml.min^{-1}mg^{-1})	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m (mM^{-1}min^{-1})
Lichenan	0.17±0.08	6.0±.1×10^3	3.8±1.3×10^4	0.15±0.05	8.7±0.9 ×10^3	5.8±1.3×10^4	0.11±0.01	3.2±0.9×10^4	2.9.±0.8×10^5	--	--	--
β-Glucan	0.16±0.05	5.4±0.4×10^3	3.3±0.8×10^4	0.18±0.04	8.1±0.6×10^3	4.5±1.5×10^4	0.16±0.08	2.7.±3.5×10^4	1.6.±4.7×10^5	--	--	--
CMC[a]	0.35±0.10	9.0±0.1×10^2	2.5±0.9×10^3	0.38±0.11	1.57±0.13×10^3	4.1±1.2×10^3	0.54±0.15	6.9±0.7×10^3	1.2±0.4×10^4	--	--	--
*p*NPG[b]	--	--	--	--	--	--	[c]0.02±0.034	8.1±0.7×10^2	[d]4.1±0.2×10^4	0.04±..02	2.8±0.3×10^2	7.1±0.2×10^3

[a] CMC-Carboxymethylcellulose [b] *p*NPG- *p*-Nitrophenyl β-D-glucopyranoside [c]Unit of K_m in mM [d]Unit of k_{cat}/K_m in mM^{-1}min^{-1}

3.3.5 Melting temperature of Chimera and its individual modules

The melting temperature (T_m) of wild-type *Ct*GH5 was 79ºC and mutant *Ct*GH5-F194A was 68°C as shown in Fig. 3.5A. This decrease in T_m of the mutant *Ct*GH5-F194A can be correlated with the results of the decrease in thermostability of the mutant *Ct*GH5-F194A as compared with the wild-type *Ct*GH5 (Section 3.3.3). The melting temperature for the Chimera was 78ºC (Fig. 3.5B) and of *Ct*GH1 was 79ºC (Fig. 3.5B). The overall T_m of the Chimera was retained at similar to that of GH1. This might be due to the effect of the N-terminal *Ct*GH1 in the chimeric construct which has T_m of 79°C (Fig. 3.5B). This positive effect of N-terminal domain exerted on the enzyme has been also described in a previous study (Lee *et al.*,2011).

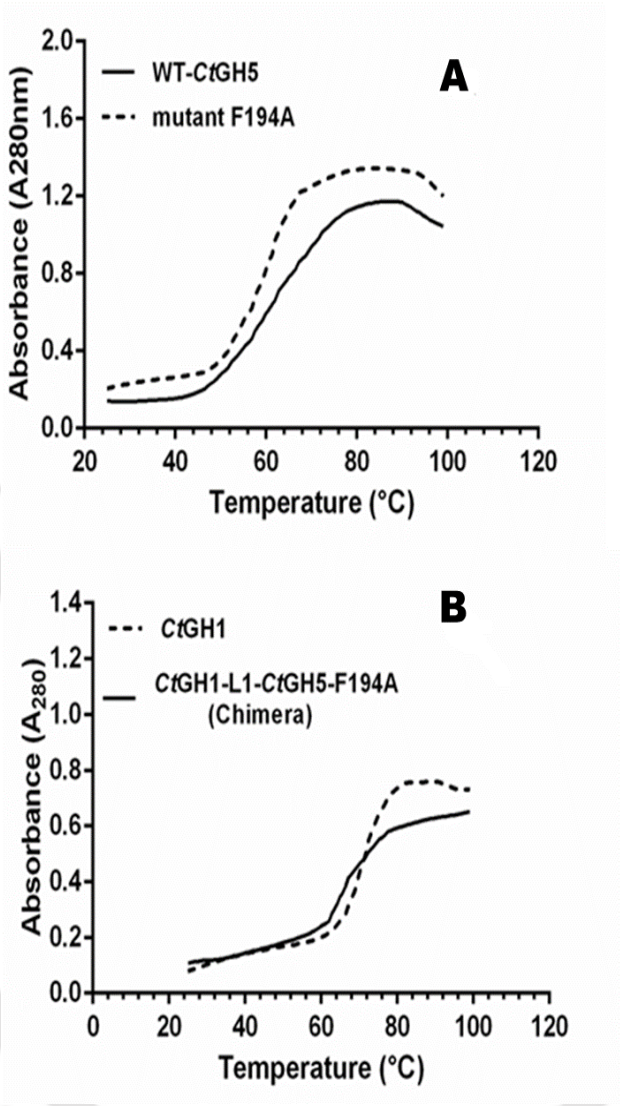


Fig. 3.5 Melting temperature analysis of Chimera and different enzymatic modules. A) wild-type *Ct*GH5 and mutant *Ct*GH5-F194A and B) *Ct*GH1 and Chimera. 50 µg/ml of each, Chimera or wild-type *Ct*GH5 was dissolved in 1 ml 50 mM sodium acetate buffer (pH 5.0), while 50 µg/ml of each, mutant *Ct*GH5-F194A and *Ct*GH1 was dissolved in 1 ml 50 mM MES buffer (pH 5.5) and the absorbance was measured a UV-Visible spectrophotometer at different temperatures ranging between 30-100°C.

3.3.6 Analysis of hydrolysed products by purified Chimera using TLC

The hydrolysed products of CMC by Chimera at different time interval was analysed by TLC (Fig. 3.6A). In the initial 5 min major cello-oligosaccharides were formed and after that the major-product was only glucose till 12h. This showed that both endo- and exo-activities are incorporated in the Chimera. The hydrolysed product of individual enzymes i.e. the wild-type *Ct*GH5, the mutant *Ct*GH5-F194A and the mixed enzymes at equimolar concentrations were compared with the Chimera (Fig. 3.6B). It was confirmed that the individual enzymes showed cellobiose and higher oligosaccharide as the major product but the Chimera alone forms glucose along with cello-oligosacharides as a final product in 1h of incubation time. Therefore, this showed that both β-glucosidase and endoglucanase activities were incorporated into the engineered chimeric construct and is more efficient than its individual enzymes.

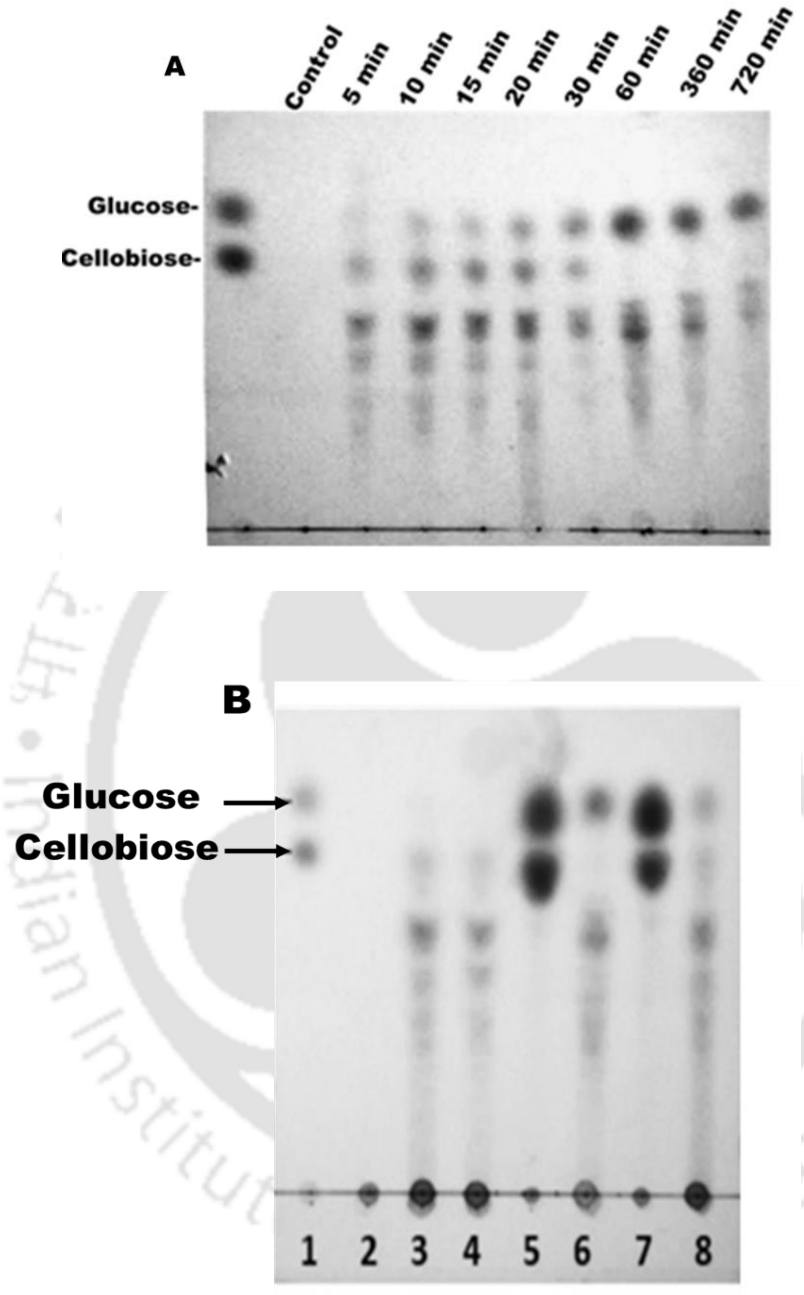


Fig. 3.6 A) Time dependent analysis of hydrolysed product of 1% (w/v) CMC by Chimera. The reaction was performed at 60°C using 20 mBM citrate phosphate buffer, pH 5.0 using 0.2 µM enzyme. B) TLC analysis of hydrolysed products of 1% (w/v) CMC or 1% (w/v) cellobiose by different enzymatic constructs. Lanes: 1, Standard (Glucose + cellobiose); 2, CMC as control; 3, CMC + wild-type *Ct*GH5; 4, CMC + mutant *Ct*GH5-F194A; 5, Cellobiose + C*t*GH1; 6, CMC + Chimera; 7, Cellobiose + Chimera; 8, CMC + mixture *Ct*GH1 + mutant *Ct*GH5-F194A. Each reaction was performed at optimal conditions using 0.2 µM of each enzyme for 1h.

3.3.7 Analysis of hydrolysed products from pretreated biomass (Sorghum stalk) by purified Chimera using TLC

The action of Chimera on the pretreated biomass (Sorghum stalk) was also analysed by TLC. The individual enzyme *Ct*GH5-F194A (endoglucanase) hydrolysed the pretreated biomass and produced cellobiose and the higher cello-oligosaccharides (Fig. 3.7). In the presence of mixture of *Ct*GH1 (β-glucosidase) and *Ct*GH5-F194A (endoglucanase) the glucose was formed as the final product from the pretreated biomass (Fig. 3.7). However, the Chimera displayed higher accumulation of glucose after 48h than the mixture of *Ct*GH1 and *Ct*GH5-F194A (Fig. 3.7). Therefore, Chimera acts on the pretreated Sorghum stalk biomass and efficiently releases the glucose as final product.

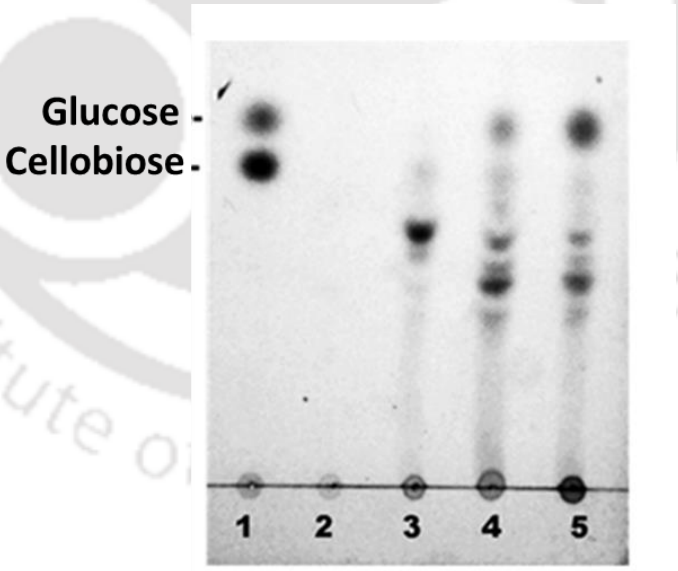


Fig. 3.7 Enzyme hydrolysed products of 1% (w/v) pretreated Sorghum stalk. Lanes: 1, Standard (Glucose + cellobiose); 2, pretreated biomass as control; 3, mutant *Ct*GH5-F194A; 4, mix (*Ct*GH1 + *Ct*GH5-F194A); 5, Chimera. The reactions were performed at 50°C for 48 h using 1.1 µM of each enzyme in 20 mM citrate phosphate buffer, pH 5.0 for Chimera and 20 mM citrate phosphate buffer, pH 5.5 for individual enzymes (*Ct*GH5-F194A and mixture of *Ct*GH1 + *Ct*GH5-F194A).

3.4 Conclusions

Site-directed mutagenesis of β-1,4-endoglucanase from family 5 glycoside hydrolase (*Ct*GH5) from *Clostridium thermocellum* was performed to develop a mutant *Ct*GH5-F194A that gave 40 U/mg specific activity against carboxymethyl cellulose, resulting 2-fold higher activity than wild-type *Ct*GH5. *Ct*GH5-F194A was fused with a β-1,4-glucosidase, *Ct*GH1 from *Clostridium thermocellum* to develop a chimeric enzyme with bifunctional β-glucosidase and endoglucanase activity. The Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) expressed as a soluble protein using *E. coli* BL-21cells. For endoglucanase, the Chimera and the wild-type *Ct*GH5 showed temperature maxima at 60°C, whereas the mutant *Ct*GH5-F194A showed temperature maxima at 50°C. This showed that mutant is temperature sensitive and gets inactivated. The temperature sensitivity and inactivation of mutant *Ct*GH5-F194A is counteracted by the linker or the adjacent more thermotolerant module *Ct*GH1 in the N-terminal domain conferring the Chimera higher temperature resistance. The Chimera also displayed improved temperature optima at 70°C for β-glucosidase activity than its individual counterpart *Ct*GH1 which is 65°C. The Chimera showed maximum endoglucanase activity at pH 5.0. The pH optima of wild-type *Ct*GH5 is 4.2 which shifted to pH 5.5 in mutant *Ct*GH5-F194A. A shifting in pH optimum towards acidic pH of 5.5 was also observed in the β-glucosidase activity of the Chimera, because the β-glucosidase has a pH optimum of pH 6.5 for individual module *Ct*GH1. The pH stability analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) for endoglucanase activity showed that is stable between pH 4.0-5.0. The individual mutant *Ct*GH5-F194A was stable at higher pH between 5.0-6.0. For β-glucosidase activity, the Chimera showed pH stability between pH 5.0 to 5.5, while the individual enzyme

*Ct*GH1 was stable between pH 5.5 to 6.0. The advantageous outcome was that the Chimera displayed pH stability between pH 4.0-5.5 for both the enzyme activities. This may prove beneficial in the synergistic functioning of both enzyme modules in the chimeric form. The thermostability analysis showed that the wild-type *Ct*GH5 is stable up to 60ºC and the individual mutant enzyme *Ct*GH5-F194A is stable up to 55ºC. The thermostability analysis of Chimera showed that its endoglucanase activity is stable at 55ºC up to 1 h. The thermal stability of both, *Ct*GH1 and β-glucosidase domain of Chimera was found up to 60ºC for 1 h. Protein melting studies of Chimera and its individual enzyme modules showed that *Ct*GH5-F194A melting temperature (T_m) is 68ºC and of *Ct*GH1 is 79ºC, whereas, Chimera showed melting temperature (T_m) of 78ºC. This showed improved structural integrity, thermostability of Chimera. The K_m and k_{cat} of Chimera were compared against their individual modules for substrates Lichenan, β-Glucan, CMC and *p*NPG. For endoglucanase activity Chimera showed higher catalytic efficiency (k_{cat}/K_m) against Lichenan, β-Glucan, and CMC i.e.$2.9\pm0.8\times10^5$ ml $min^{-1}mg^{-1}$, $1.6\pm4.7\times10^5$ ml $min^{-1}mg^{-1}$, $1.2\pm0.4\times10^4$ ml $min^{-1}mg^{-1}$ which was 3-5 fold higher than the individual mutant *Ct*GH5-F194A. For β-glucosidase activity Chimera showed $4.1\pm0.2\times10^4$ $mM^{-1}min^{-1}$ catalytic efficiency (k_{cat}/K_m) against *p*NPG which was 5.5 fold higher than the individual module *Ct*GH1. The TLC analysis of hydrolysed products of CMC by Chimera revealed glucose as final product that confirmed its both, β-1,4-endoglucanase and β-1,4-glucosidase activities. The hydrolysed products of CMC by *Ct*GH5-F194A were cellobiose and cello-oligosaccharides and no glucose. The Chimera displayed higher accumulation of glucose after 48h than the mixture of *Ct*GH1 and *Ct*GH5-F194A from the pretreated Sorghum stalk. This study demonstrated the developed bi-functional Chimera with β-

glucosidase and endoglucanase activities has potential applications for bioethanol, biobutanol or lactic acid production.

References

Adlakha, N., Sawant, S., Anil, A., Lali, A., & Yazdani, S. S. (2012). Specific fusion of β-1,4-endoglucanase and β-1,4-glucosidase enhances cellulolytic activity and helps in channeling of intermediates. *Applied and Environmental Microbiology*, *78*, 7447-7454.

Beguin, P., (1990). Molecular biology of cellulose degradation. *Annual Reviews in Microbiology*, *44*, 219-248.

Bulow L. 1987. Characterization of an artificial bifunctional enzyme, betagalactosidase/galactokinase, prepared by gene fusion. *European Journal of Biochemistry*, *163*, 443–448.

Cartmell, A., McKee, L., Pena, M. J., Larsbrink, J., Brumer, H., Kaneko, S., & Marles-Wright, J. (2011). The structure and function of an arabinan-specific alpha-1, 2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *Journal of Biological Chemistry*, *286*, 15483–15495.

Fan, Z., Werkman, J. R., & Yuan, L. (2009). Engineering of a multifunctional hemicellulase. *Biotechnology Letters*, *31*, 751-757.

Grabnitz, F., Seiss, M., RucknageL, K. P., & Staudenbauer, W. L. (1991). Structure of the β-glucosidase gene bglA of *Clostridium thermocellum*: Sequence analysis reveals a superfamily of cellulases and β-glycosidases including human lactase/phlorizin hydrolase. *European Journal of Biochemistry*, *200*, 301-309.

Lee, H. L., Chang, C. K., Teng, K. H., & Liang, P. H. (2011). Construction and characterization of different fusion proteins between cellulases and β-glucosidase to improve glucose production and thermostability. *Bioresource Technology*, *102*, 3973-3976

Lynd, L. R., Weimer, P. J., Van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews.*, *66*, 506-577.

Liu, L., Wang, L., Zhang, Z., Guo, X., Li, X., & Chen, H. (2012). Domain-swapping of mesophilic xylanase with hyper-thermophilic glucanase. *BMC Biotechnology*, *12*, 28.

McKee, L. S., Peña, M. J., Rogowski, A., Jackson, A., Lewis, R. J., York, W. S., & Marles-Wright, J. (2012). Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proceedings of the National Academy of Sciences*, *109*, 6537-6542.

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

Nixon, A. E., Ostermeier, M., & Benkovic, S. J. (1998). Hybrid enzymes: manipulating enzyme design. *Trends in Biotechnology*, *16*, 258-264.

Somero, G. N. (1995). Proteins and temperature. *Annual Review of Physiology*, *57*, 43-68.

Somogyi, M. (1945). Determination of blood sugar. *Journal of Biological Chemistry*, *160*, 69-73.

Telke, A. A., Ghatge, S. S., Kang, S. H., Thangapandian, S., Lee, K. W., Shin, H. D., & Kim, S. W. (2012). Construction and characterization of chimeric cellulases with enhanced catalytic activity towards insoluble cellulosic substrates. *Bioresource Technology*, *112*, 10-17.

Urbanowicz, B. R., Bennett, A. B., del Campillo, E., Catala, C., Hayashi, T., Henrissat, B., & Teeri, T. T. (2007). Structural organization and a standardized nomenclature

for plant endo-1,4-β-glucanases (cellulases) of glycosyl hydrolase family 9. *Plant Physiology*, *144*, 1693-1696.

Chapter 4

SAXS and computational approaches for structure and binding characteristics of chimera (*Ct*GH-L1-*Ct*GH5-F194A)

4.1 Introduction

The lignocellulosic plant biomass contains cellulose, hemicellulose and lignin. Cellulose can be hydrolysed to glucose while hemicellulose can be hydrolysed to various C5 sugars such as xylose, arabinose and C6 sugars such as galactose, mannose and glucose. The monomeric sugars glucose, xylose and arabinose can be utilized by the yeast to produce alcohol and the lignocellulosic plant biomass can thus serve as an alternative source of energy. The lignocellulosic plant biomass is rich in cellulose. The cellulose can be entirely hydrolyzed to glucose by treating with three enzymes *viz.* endoglucanase, cellobiohydrolase and β-glucosidase (Beguin, 1990). Endo-β-1,4-glucanase acts randomly on the cellulose chain and produces cellodextrin which are larger cellooligosaccharides as hydrolyzed products (Urbanowicz *et al.,* 2007). Cellobiohydrolase acts at the end of the cellodextrin and releases cellobiose as the product (Barr *et al.,* 1996), while β-glucosidase hydrolyzes the cellobiose to form two

molecules of glucose (Grabnitz *et al.,* 1991). The production of all three enzymes for cellulose hydrolysis is a cost-intensive method (McKee *et al.,* 2012). The production cost can be reduced by developing enzymes with improved catalytic efficiency and multifunctional activity (Lu *et al.*, 2008; Fan *et al.*, 2009). The multifunctional enzymes with two or more fused enzymes in a single polypeptide chain can act synergistically to hydrolyze the polymer into simple sugar (Fan *et al.,* 2009). An essential aspect for the construction of active multifunctional Chimera with synergistic catalytic activity and improved catalytic efficiency is to determine the structural organization of the modules in the chimeric construct and selection of the appropriate linker sequence (Crasto *et al.,* 2000). The structural organization of individual modules in the Chimera can be predicted by generating a homology model using computational approaches (Onuchic & Wolynes, 2004). The molecular dynamics can provide successful validation of the protein model to explore the protein conformation (Yang *et al.,* 2010). Efforts were made to create chimeric enzyme using a computational approach where bi-functional endo-xylanase (xynA) and lichenase (BglS) modules from *Bacillus subtilis* were fused (Cota *et al.*, 2013). For example, substrate accessibility to its active site and its dynamic behaviour in the solution were also determined (Cota *et al.,* 2013). The resulting Chimera of endo-xylanase (xynA) and lichenase (BglS) preserved the biochemical characteristics of the parental enzyme with slight variation in the catalytic efficiency (Cota *et al.,* 2013). The computational development for the prediction of stable conformation and the arrangement of the modules in a chimeric enzyme in the solution resulted in the development of multifunctional enzyme and predicted their extended conformation and validation of their experimental data (Jamros *et al.*, 2010).

The family, GH5 is one of the largest and well-characterized families of glycoside hydrolases (http://www.cazy.org/GH5.html). The catalytic activity of family 5 GH (Cel5E) (ABN52701.1) from *Clostridium thermocellum* was enhanced by attaching the linker at both N- and C-terminals of GH5 and by mutating Phe267 to Ala (Yuan *et al.,* 2015). In our previous study the catalytic efficiency of endoglucanase (*Ct*GH5) (ABN52701.1) of family 5 glycoside hydrolase from *Clostridium thermocellum* without the N- or C-terminal linker was cloned and its mutant F194A was generated by site-directed mutagenesis (Nath *et al.,* 2019), by following the method reported earlier (Yuan *et al.,* 2015). The mutant *Ct*GH5-F194A resulted in enhanced enzyme activity and was fused with β-glucosidase (C*tG*H1) to develop a Chimera, *Ct*GH1-L1-*Ct*GH5-F194A (Nath *et al.,* 2019). *Ct*GH1 (β-glucosidase) used in the chimeric construct was earlier biochemically and structurally characterized (Sharma *et al.,* 2019). The fusion of the two enzymatic modules *Ct*GH1 and *Ct*GH5-F194A to develop a Chimera was performed using a natural N-terminal linker from the cellulosomal gene *cel*H from *C. thermocellum* (Taylor *et al.,* 2005). The chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A) developed by using the module β-glucosidase (*Ct*GH1) at the N-terminal and endoglucanase (*Ct*GH5-F194A) at C-terminal showed significant enhancement in both the activities (Nath *et al.,* 2019). The Chimera showed 3- to 5-fold increase in the catalytic efficiency of β-glucosidase and endoglucanase activities (Nath *et al.,* 2019).

In the present study, the structural analysis of the developed Chimera, *Ct*GH1-L1-*Ct*GH5-F194A was undertaken to find out its organization, stability and the folding behaviour. The secondary structure composition of the Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), previously constructed by fusing β-glucosidase (*Ct*GH1) at N-terminal and

mutant endoglucanase (*Ct*GH5-F194A) at the C-terminal was determined (Nath *et al.*, 2019). The computer based prediction softwares were used to study its three-dimensional structure and the structural parameters. The binding interaction of the two modules of Chimera was performed by molecular docking against different ligands. The Chimera and its individual modules in the presence and absence of the ligands were MD simulated to elucidate the role of loops and the key residues involved in the catalysis or binding. The small angle X-ray scattering (SAXS) has been used to analyse the average particle size, molecular shape and other properties of various type of macromolecules such as biomacromolecules (Sharma *et al.*, 2019) and orthopaedics (Collins *et al.*,2012, Dalton *et al.*, 2014). SAXS analysis of the chimeric enzyme was performed that displayed the arrangement of the two modules of Chimera in the solution form.

4.2. Materials and Methods

4.2.1 Experimental and *in silico* prediction analysis of secondary structure of chimera

The chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was cloned, expressed and biochemically characterized earlier was used in the present study (Nath *et al.*, 2019). The secondary structure of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) i.e. the content α helices, β strands and random coils were predicted by different available webservers for prediction of secondary structure such as Psi-Pred v 3.3 web server (*http://bioinf.cs.ucl.ac.uk/psipred/*) and RaptorX (http://raptorx.uchicago.edu/). The secondary structure composition of chimera was also determined by circular dichroism (CD) analysis. The purified chimeric enzyme at a concentration of 5.3 µM in 50 mM sodium phosphate buffer, pH 7.0 was used for CD analysis. The CD spectrum was observed on a spectropolarimeter (JascoJ-815, Japan) at 25°C using 1 nm bandwidth

over far UV region between 190 to 250 nm at a scanning rate of 50 nm/min with an average of three scans. The CD experimental data was analyzed by the difference in molar extinction coefficient (Δε, deciliter mol^{-1} cm^{-1}) described as a function of wavelength (Kelly *et al.*, 2005). The percentage of α-helix and β-sheet were obtained from web-based K_2D_2 software (http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d2/) (Andrade et al., 1993).

4.2.2 Computer-based structural prediction of chimera and its validation

The 3D-structure prediction of the chimera was performed by using RaptorX web server (http://raptorx.uchicago.edu/). RaptorX is also an online server for protein structure and function prediction (Peng *et al.*, 2011). This program uses a statistical method for template-based modeling of protein. This methodology improves alignment accuracy by utilizing the structural data in single or multiple templates (Peng *et al.*, 2011). RaptorX also calculates *p*-value for the relative global quality, GDT (global distance test) and uGDT (un-normalized GDT) for the absolute global quality, and RMSD for the absolute local quality of each residue in the model. The prediction of the threading method is reliable when the p-value is less than 10^{-4} (Kallberg *et al.*, 2014). For generating the modeled structure of chimera, two modules for the input sequence of chimera, The chimera is made up of two modules, *viz.* N-terminal module 1 (*Ct*GH1) comprising 445 amino acids, the C-terminal module 2 (*Ct*GH1-F194A) comprising 309 amino acids and the two modules are connected by a natural linker made up of 19 amino acids. The difference in the number of amino acid residues between the endoglucanase module of Chimera and the individual endoglucanase is of 487 (pET28a(+)-His-tag region, 23 + *Ct*GH1, 445 + Natural Linker L1, 19 = 487). For module 1, amino acid

residues (1-474) p-value predicted was 7.14×10^{-12} and for module 2, amino acid residue (475-796) a p-value of 1.41×10^{-8} was obtained. The webserver RaptorX showed closest homology for the input sequence with 3AHX an β-glucosidase from *Clostridium cellulovorans* and 5BYW an endoglucanase from *Clostridium thermocellum.* The energy minimization of the modeled structure was performed using YASSARA energy minimization tool (http://www.yasara.org/minimizationserver.php). The modeled structure generated was analyzed by the software UCSF Chimera (Pettersen *et al.*, 2004). The energy minimized structure was validated by using the saves web server (http://servicesn.mbi.ucla.edu/SAVES/).

4.2.3 Molecular dynamics simulation of modeled chimera structure

Molecular dynamics simulation (MD) for the chimera modeled structure was accomplished by using Gromacs v 5.14 (Berendsen *et al.*, 1995) in a supercomputer facility Param-Ishan available in the Indian Institute of Technology Guwahati. GROMOS96 53a6 was used as a force field to calculate the protein forces. The modeled chimera was placed inside a cubic box with dimension 6.38× 7.73 × 10.12 and volume 1949.8 nm^3 with single point charge (SPC) along with the water molecules. The MD simulation contained 60088 water molecule and 22 Na^+ counter ions were added for the neutralization of the charges on the chimera. The chimeric structure was minimized using steepest descent method using 50,000 iteration steps with cut-off till 1000 $kJmol^{-1}$. Then the equilibration of the whole system was carried out for 500 ps using NVT ensemble in the constant number of particles, volume, and temperature. The next calibration was performed for 500 ps by the NPT ensemble in a constant number of particles, volume, temperature and pressure. The production run for 70 ns was

performed with NPT ensemble acquiring a 2 fs integration time. The linear constraint solver (LINC) algorithm (Hess *et al.*, 1997) was utilized to constrain the bonds linked with hydrogen atoms and radius of gyration. Throughout the simulation, the modeled chimera structure was analyzed as a time-dependent function to verify its stability in the solvent system. The root-mean-square deviation (RMSD) of the simulated system was analyzed by gmx rms command, and gmx rmsf was used for determining the fluctuation (RMSF). The radius of gyration (Rg), solvent accessible surface area (SASA) were determined by using the program gmx gyrate and gmx sasa, respectively.

4.2.4 Binding interaction analysis of chimera

Molecular docking of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) with the cello-oligosaccharides was performed by using SwisDock, the web based server (http://www.swissdock.ch/docking). Cello-oligosaccharides were downloaded using the GLYCAM server (Kirschner *et al.*, 2008). The modeled chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was saved in PDB format and the ligands were saved in Mol2 file format for docking analysis in the Swiss Dock web-based tool. Swiss Dock produced large number of ligand binding results. The enzyme-ligand docked complex showing the strongest binding with highest negative binding free energy were selected. This ligand bound structure was extracted and visualized in PyMol 2.0. The representation of ligand interaction with the amino acid residues of the protein was generated using the PDBsum Generate tool (*https://www.ebi.ac.uk/thornton srv/databases/pdbsum/Generate.html*).

4.2.5 Molecular dynamics of ligand bound chimera

Molecular dynamics (MD) simulation for determining the interaction of chimera with the ligands was performed in the presence of cellobiose and cellohexaose. The cellobiose was placed in the catalytic cleft of β-glucosidase module in the chimera, while the cellohexaose was kept in the catalytic cleft of endoglucanase module in the chimera. The docked complexes of Chimera with cellobiose and cellohexaose from section 2.4 were selected and were simulated by using GROMACS v 5.14. Similarly, the cellohexaose bound individual endoglucanase (*Ct*GH5-F194A) and cellobiose bound individual β-glucosidase (*Ct*GH1) were also simulated for the comparative analysis. Protein forces were calculated through GROMOS96 43A1 force field and ligand topologies for cellobiose and cellohexaose were generated by using PRODRG server. The docked complexes of the chimera or ligand bound chimera were placed in a cubic box with volume 1949.8 nm^3 and filled with water molecules. The MD simulation system containing 60088 water molecules were added with 23 Na^+ counter ions for the neutralization of the charges on ligand bound chimera for cellobiose in β-glucosidase module and 28 Na^+ counter ions for cellohexaose in endoglucanase module of chimera. Similarly, the docked complex of individual endoglucanase (*Ct*GH5-F194A) with cellohexaose and the only individual endoglucanase (*Ct*GH5-F194A) were placed inside a cubic box with a volume 489 nm^3 with single point charge (SPC) and 14509 water molecules were neutralized by 24 and 19 Na^+ counter ions. The docked complex of individual β-glucosidase (*Ct*GH1) with cellobiose and the only individual β-glucosidase (*Ct*GH1) were placed inside a cubic box with a volume 674 nm^3 with single point charge (SPC) and 25581 water molecules were neutralized by 11 and 10 Na^+ counter ions.

Initially, the equilibration of entire system was accomplished with restraints in NVT ensemble (constant number of particles, volume and temperature) for 500 ps. The entire system was equilibrated twice, first of all with restraints and secondly without restraints for 500 ps in NPT ensemble (constant number of particles, pressure and temperature). The 70 ns production run was setup along with 2 fs integration time of NPT ensemble. The docked complexes after 70 ns production were analysed with the help of PyMOL 2.0 software (Schrodinger et al., 2019). The root-mean-square deviation (RMSD) of the simulated ligand bound form of chimera and its individual endoglucanase (*Ct*GH5-F194A) was analyzed by gmx rms command. The gmx rmsf was used for determining the fluctuation (RMSF) and the gmx energy command was used to obtain the interaction energy between proteins and ligands. Average energies for short range interactions between protein and ligand was given in the terms of Lennard Jones and coulombic interaction energy. The representation of ligand interaction with the amino acid residues of the protein after molecular dynamics simulation was generated by using the PDBsum Generate tool (*https://www.ebi.ac.uk/thornton srv/databases/pdbsum/Generate.html*).

4.2.6 Small angle x-ray scattering analysis of chimera

The scattering profiles of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) at two different concentrations (3 mg/ml and 5 mg/ml) along with its matched buffer (50 mm Sodium phosphate, pH 7.0) were collected as per the method reported previously (Sharma *et al.*, 2018) by a home source small-angle x-ray scattering system (SAXSpace, Anton Paar GmbH, Graz, Austria) established at CSIR-Institute of Microbial Technology, Chandigarh, India. Incident x-rays were originated by using line collimation system and passed through the sample, filled in a thermostatic quartz capillary of a 1 mm diameter

and two independent exposures of 30 min each was collected at 10ºC. The SAXS analysis software was used for processing and subtraction of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its matched buffer SAXS profiles. The integrity and non-aggregation effect on the chimera scattering profiles were confirmed by visual inspection in the SAXS data analysis program of ATSAS 2.84 suite (Franke *et al.,* 2017). The radius of gyration (Rg) of globular and rod shape (Rc) of the chimera was calculated by applying the Guinier approximation equation in the primusqt (Fournet *et al.,* 1955; Konarev *et al.,* 2003). The length (L) of the chimera was calculated by implementing [$\sqrt{}$ 12{(Rg^2)-(Rc^2)}] formula. The maximum dimension (D_{max}) of the particle and the distance distribution function plot (P(R)) were evaluated by employing GNOM software integrated with ATSAS v2.84 (Svergun *et al.,* 1992). The molecular mass of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was calculated with the assistance of SAXSMoW program (Fischer *et al.,* 2010). Gasbor program was used for constructing 50 *ab initio* models by taking dummy aspartic acid residues equivalent to the molecular mass of *Ct*GH1-L1-*Ct*GH5-F194A. These 50 low-resolution shape *ab initio* models were averaged by applying DAMAVER program (Volkov *et al.,* 2003). The final averaged *ab initio* model was superposed with *Ct*GH1-L1-*Ct*GH5-F194A modeled structure using the SUPCOMB program (Kozin *et al.,* 2001). The CrYSOL program was used to generate theoretical scattering for *Ct*GH1-L1-*Ct*GH5-F194A modeled structure, and it was fitted with the experimental data generated from SAXS (Svergun *et al.,* 1995).

4.3. Results and Discussion

4.3.1 Secondary structure analysis by circular dichroism and web server's prediction

The circular dichroism (CD) spectrum of the chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was analyzed from the K_2D_2 web server, where the data points are measured with the already existing secondary structures of the known proteins (Louis *et al.*, 2012). The CD analysis displayed 38% of α-helix, 9.32% of β-sheets and 52.68% of random coils (Fig. 4.1A and Table 4.1). The secondary structure prediction analysis of the chimera by Psi-Pred web server showed 33% of α-helix, 10% of β-sheets and 54% of random coils (Fig. 4.1B and Table 4.1). The RaptorX prediction web server showed 34% of α-helix, 11% of β-sheets and 53% of random coils (Table 4.1). Thus, the predicted secondary structure result of the chimera corroborated with that obtained by CD analysis.

Table 4.1. Secondary structure analysis of chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

Prediction	α-helix (%)	β-sheet (%)	Random coils (%)
CD analysis (K_2D_2)	38	9.3	53
RaptorX	34	11	54
PSI-PRED	33	10	51

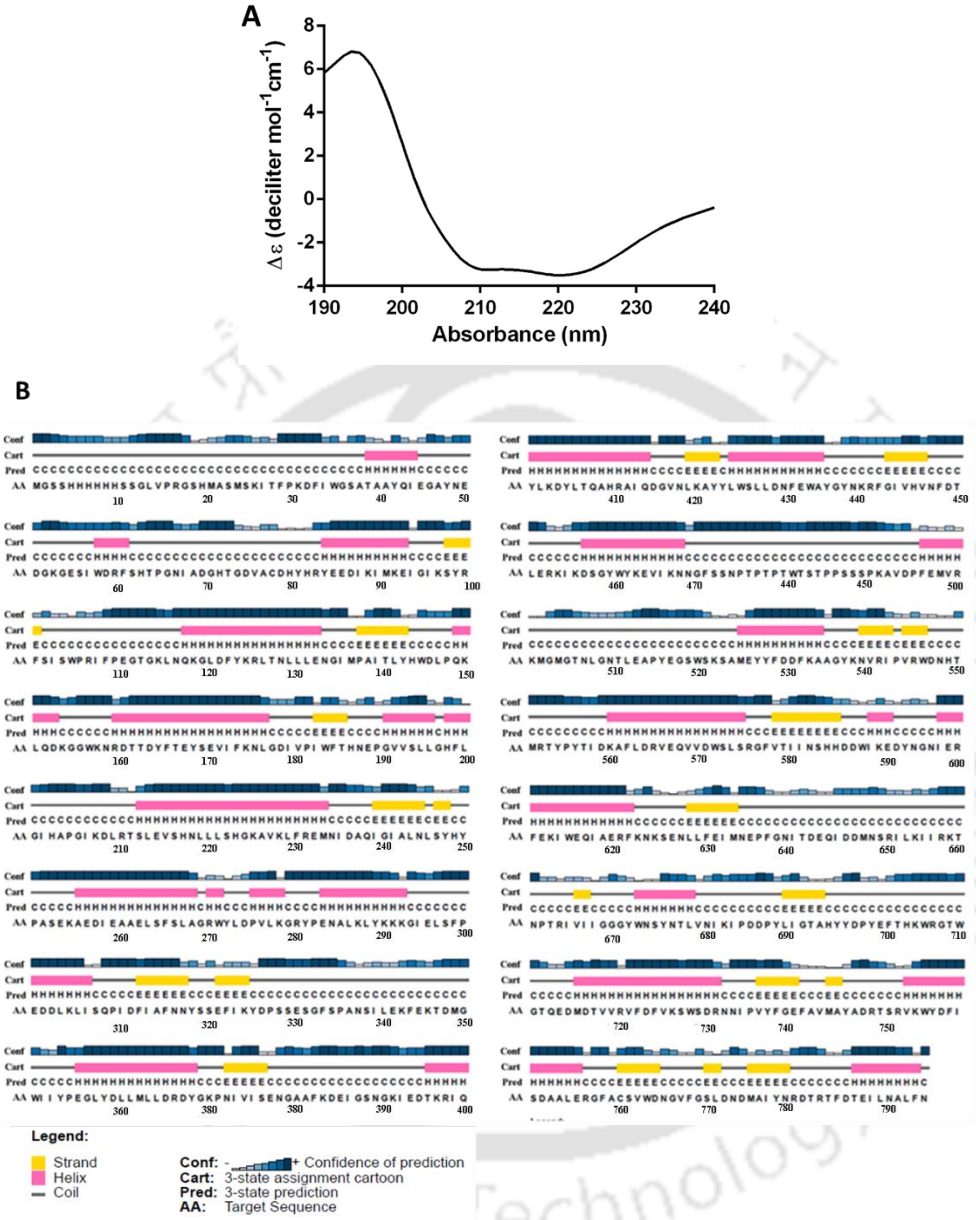


Fig. 4.1. A) Circular dichroism (CD) spectrum of chimera for estimating the percentage of secondary structure components. The CD data were explained by the difference in molar extinction coefficients (decilitre mol^{-1} cm^{-1}) as a function of wavelength. The purified chimera purified with a concentration of 5.3 µM in 50 mM sodium phosphate buffer, pH 7.0 was utilized for CD analysis. The CD spectrum was observed at 25°C using 1 nm bandwidth extending over far UV region between 190 to 250 nm at a scanning rate of 50 nm/min. B) Secondary structure prediction of chimera by Psi-Pred server tool demonstrating the amino acid residues involved in formation of α -helix (cylinders), β -sheet (arrow) and random coil (continuous line).

4.3.2 Three-Dimensional structure prediction of chimera and its validation

The homology based modeled 3D structure of chimera was generated by using RaptorX webserver. The 3D-model structure generated by the RaptorX server predicted a two-module protein structure with p-value of 7.14×10^{-12} for module 1, *Ct*GH1 (β-glucosidase), amino acid residues (1- 468) and 1.41×10^{-8} for module 2, *Ct*GH5-F194A (endoglucanase), amino acid residue (488-796). The 3D structure prediction of the chimera by RaptorX was considered for further analysis because the p-value was less than 10^{-4}. The modeled structure of chimera was energy minimized by using YASSARA energy minimization tool. The starting energy of the modeled structure was 1.1×10^{5} kJmol^{-1} while after minimization it showed energy of -4.36×10^{5} kJmol^{-1} for the end structure. The end structure of the chimera after minimization was extracted from the server and was used for further structural characterization. The modeled structure of the chimera showed N-terminal β-glucosidase (*Ct*GH1) module (grey colour) and C-terminal endoglucanase (*Ct*GH5-F194A) (green colour) connected by a linker (red colour) (Fig. 4.2A and 4.2B). The overall modeled structure of the chimera showed classical (α/β)-TIM barrel fold (Fig. 4.2A and 4.2B). The surface view of chimera shows the active site for both β-glucosidase (Glu189 and Glu378) (Fig. 4.2C) and endoglucanase (Glu626 and Glu731) (Fig. 4.2D).

This minimized structure was validated by using different validation tools available in SAVES v5.0 (http://servicesn.mbi.ucla.edu/SAVES/). The verify3D showed 91.4% of the residues gave average 3D-1D score >=0.2 (Fig. 4.2E). Quality assessment of the modeled structure of chimera by Ramachandran plot was carried out, before taking it further for MD simulation and other computational studies. It showed

90.9% non-glycine and non-proline amino acid residues in the most favourable region, 8.4% within the additional allowed region, 0.6% in the generously allowed region and only 0.1% amino acid residues in the disallowed region which were Ala77 and Trp433 (Fig. 4.2F). This result demonstrated that the amino acid residues in the modeled chimeric structure are involved in favourable phi (Φ) and psi (Ψ) backbone dihedral angles. The results were fairly in agreement with the individual modules even after a linker was introduced between them and the orientation of the amino acids residues were not disturbed. Quality assessment of the model from ProSA revealed that the C\chimera model matched X-Ray region of the plot with Z-score (−12.54) which is reliable to the Z-scores of the templates; (−10.79) for 3AHX an β-glucosidase from *Clostridium cellulovorans*, and (-10.21) for 5BYW an endoglucanase from *Clostridium thermocellum* (Fig. 4.2G). Similarly, local model quality (Fig. 4.2H) showed that absence of problematic or erroneous regions in the predicted structure of chimera as there were absence of positive energy values (thick green line indicates residue wise energy plot calculated based on smoothened average energy score over each-40 residues). From all these structure validation parameters it can be considered that the predicted protein structure is reliable and satisfactory.

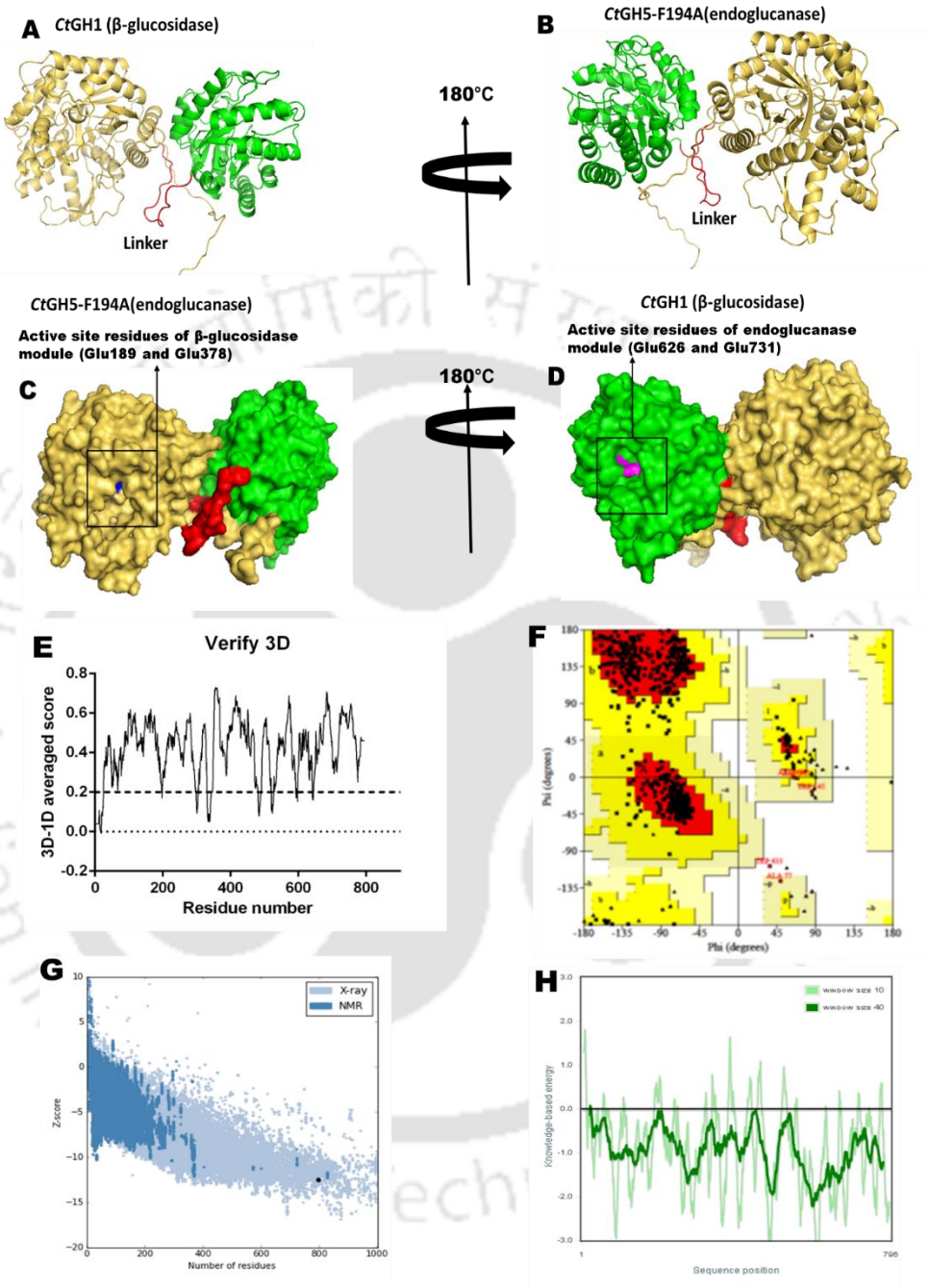


Fig. 4.2.A&B) Cartoon representation of 3D- modeled structure of chimera showing double domain with N- terminal *Ct*GH1 (β-glucosidase) module I (Yellow) and C-terminal *Ct*GH5-F194A (endoglucanase) module II (green) connected by a linker (red) Molecular surface view of chimera with active site highlighted, C) blue (β-glucosidase) and D) magenta (endoglucanase). Quality assessment of chimera by E) Verify-3D F) Ramachandran plot of modeled chimera structure, G) ProSA analysis and H) ProSA amino acid residue scores

4.3.3 Molecular dynamics simulation of chimera

The modeled structure of chimera was further subjected to molecular dynamics simulation for 70 ns. The dynamics study was carried out in order to determine the stability as well as the compactness of the modeled structure over the defined time scale. Molecular dynamics simulations along with the trajectory leads to the energetically more acceptable conformation for the protein in aqueous solution and can be represented by the RMSD between the starting structure and those obtained during the simulation. The RMSD profile shown in Fig. 4.3A demonstrated that the RMSD reached a plateau phase value of 0.7 nm at 40 ns and remained stable till 70 ns with ~0.01 nm oscillations in the RMSD profile after 40 ns (Fig. 4.3A). The stable RMSD with less oscillations represented the stable conformation of the two modules *Ct*GH1 (β-glucosidase) and *Ct*GH5-F194A endoglucanase in the chimeric form. The global compactness of chimera during the MD simulation was determined by using gmx gyrate program. The fluctuation in the R_g values (3.2-3.35 nm) for the chimera till the time scale of 20 ns was due to the flexibility, while after 20 ns the chimera achieved stability and compactness at 3.2 nm and remained stable till 70 ns (Fig. 4.3B). These results displayed that the chimeric modeled structure has a stable conformation. RMSF is measured by using gmx rmsf program. The RMSF calculates the displacement of a particular atom, or group of atoms, relative to the reference structure. The flexible residues were denoted as the residues showing the highest α- carbon flexibility among all the residues in a polypeptide chain, thus showing higher RMSF values. Rigid residues were denoted as the residues showing the lowest flexibility or lower RMSF values. In the chimera starting amino acid residues of N-terminal and end residues of

C-terminal showed highest flexibility The fluctuation in RMSF was also observed between residues between 470-480 this region comprising the linker connecting both N-terminal and C-terminal modules of the chimera signifying the flexibility of the linker in the chimeric structure. The amino acid residues within the protein active site core regions of the chimera for module 1 *Ct*GH1 (β-glucosidase) from amino acid residues 189-358 and module 2 endoglucanase (*Ct*GH5-F194A) 600-800 amino acid residues, respectively, showed comparatively lesser fluctuation than the starting N-terminal and end C-terminal residues (Fig. 4.3C). Previous studies also suggested that the protein core is the main determinant of protein stability and activity (Akasako *et al.,* 1997; Woolfson *et al.,* 2001). Therefore, the RMSF described that both catalytic cores of the chimera maintained their stability and activity in the chimeric form. The solvent accessible surface area was calculated using gmx sas programme. The average SASA of the chimera remained fairly stable with an average value of 320 nm^2 till 70 ns (Fig. 4.3D). This suggested that the overall exposure of the chimera towards the solvent area remains fairly unchanged. The minor variation in the SASA also suggested that the accessibility of the substrate in the catalytic sites were not affected in the designed chimera. This also suggested that the active sites of both the modules are not obstructed by steric effects between the modules due to the presence of the linker. The absence of inter-modular interaction between the two modules was also confirmed in a Chimera of xylanase and lichenase from *Bacillus subtilis* that resulted in the active form of chimera as reported earlier (Cota *et al.,* 2013). The superposition of initial structure of the chimera and structure obtained by MD simulation showed variation in the loop regions

and showed an RMSD of 1.244 Å (Fig. 4.3E). The MD simulated structure of chimera was stable and can be used for further analysis.

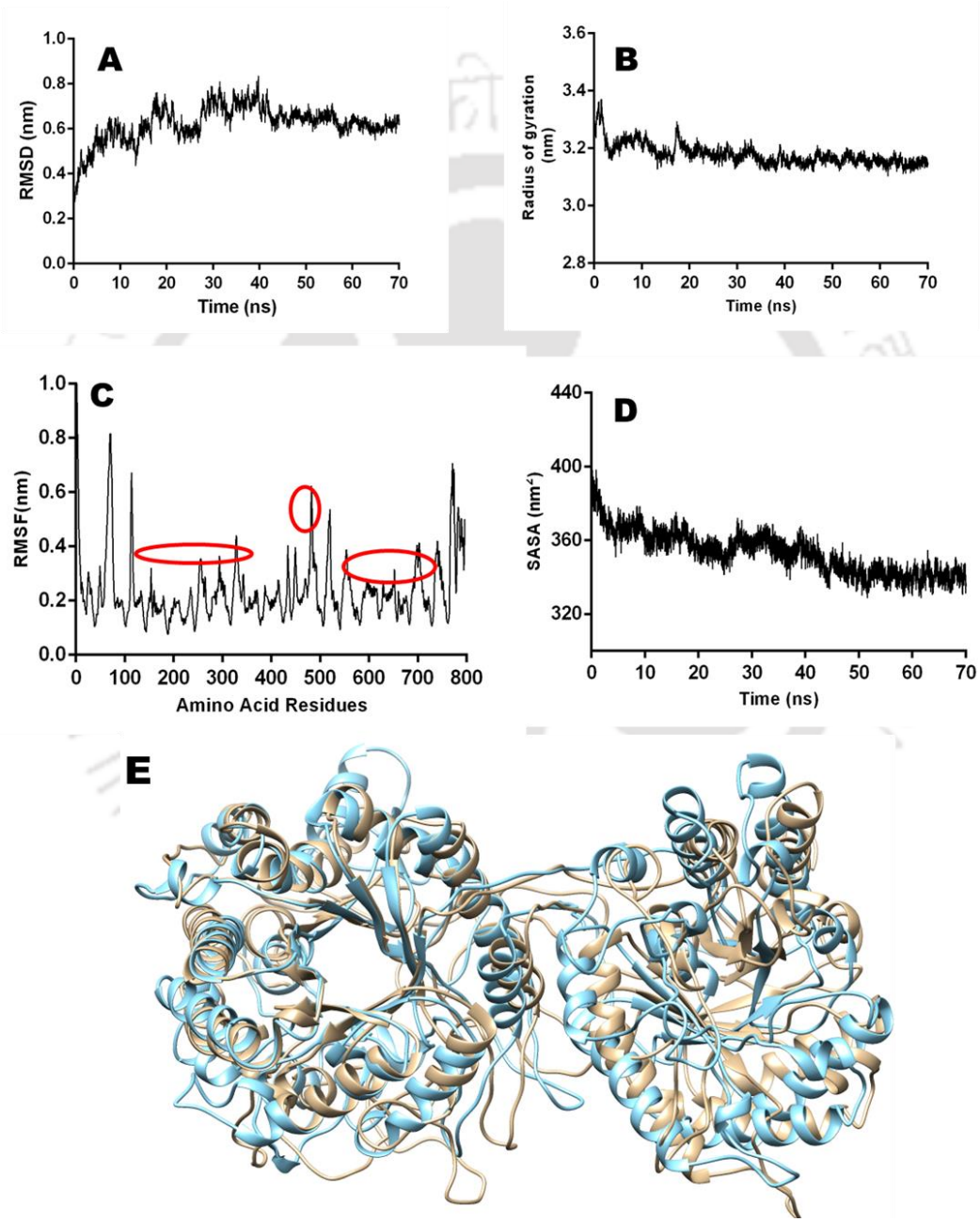


Fig. 4.3. Molecular dynamics simulation of chimera A) RMSD plot, B) Radius of gyration plot, C) RMSF plot, D) SASA plot and E) Superposition of simulated structure of chimera (cyan) with modeled chimera structure (grey).

4.3.4 Molecular Docking analysis of chimera

The molecular docking study of chimera was performed in order to determine the interaction of different cello-oligosaccharides along with the *para*-nitrophenyl derivative with the two catalytic clefts of the chimera which involved module 1, *Ct*GH1 (β-glucosidase) and module 2, *Ct*GH5-F194A (endoglucanase). The molecular docking in the SwissDock showed more than 20 possible docking solutions for both modules of the chimera. The binding interaction were observed in both catalytic sites of module 1, *Ct*GH1 and module 2, *Ct*GH5-F194A. This demonstrated the active involvement of both catalytic sites in the chimeric form. The module 1 showed binding with *p*NPG, cellobiose, cellotriose, cellotetraose, cellopentaose, cellohexaose and celloheptaose with free energy of binding (ΔG) of -6.92, -7.37, -7.62, -10.05, -10.86, -11.59 and -11.13 kcal/mol, respectively, as shown in Table 4.2. The binding studies demonstrated that the binding site of module 1, *Ct*GH1 (β-glucosidase) of the chimera can accommodate cello-oligosaccharides up to 7 degree of polymerization, DP (Table 4.2). Similar binding interactions have been reported for β-glucosidase (*Ht*Bgl) from *Hungateiclostridium thermocellum* which, showed that the active site of *Ht*Bgl displays strong binding affinities against cellobiose, cellotriose, cellotetraose up to cello-heptaose (Sharma *et al.,* 2019) and also for other glucosidases of family 1 glycoside hydrolases from different organisms such as *Humicola insolens* (Meleiro et al., 2013) and *Alicyclobacillus sp*. (Cao *et al.,* 2018). Similarly, module 2, *Ct*GH5-F194A (endoglucanase) in the chimera showed binding interaction with cellobiose, cellotriose, cellotetraose, cellopentaose, cellohexaose and celloheptaose with binding free energy (ΔG) of -7.62, -7.91, -10.05, -10.86, -11.59 and -11.54 kcal/mol, respectively as shown

in Table 4.2. This also indicated that the binding site of module 2, *Ct*GH5-F194A of chimera can accommodate cello-oligosaccharides with 7 DP (Table 4.2). The docked structure of moduled chimera and its non-covalent interactions were analysed. The binding interaction of the *p*NPG derivative and in the catalytic cleft of the module 1 *Ct*GH1 (β-glucosidase) of chimera showed involvement of amino acid residues Trp145, Glu189, Val192, Tyr319, Trp351 and Trp433 in hydrophobic interaction and Asn245, Leu246, Ser247, and Ser320 involved in hydrogen bond interaction (Table 4.2, Fig. 4.4A and Fig. 4.4B). Similarly, for the binding interaction of cellobiose with the catalytic cleft of the module 1 *Ct*GH1 (β-glucosidase) of chimera showed involvement of amino acid residues Trp145, Val192, Leu200, His203, Trp351, Trp433 and Trp435 in hydrophobic interaction and Glu189 and Asn245 in hydrogen bond interaction (Table 4.2, Fig. 4.4C and Fig. 4.4D). The binding interaction of the substrates cellobiose in the catalytic cleft of the module 2 *Ct*GH5-F194A (endoglucanase) of chimera showed amino acid residues His586, His585, Trp764, Met776, Met735, His694 involved in hydrophobic interaction, while Glu626, Glu731, Asn774 are involved in hydrogen bond formation (Table 4.2, Fig. 4.4E and Fig. 4.4F). The celloheptaose showed binding interaction in the catalytic cleft of the module 2, *Ct*GH5-F194A (endoglucanase) of chimera and amino acid residues Pro689, Trp663, Tyr690, His586, Trp520, Asn774, Val768, Asn510, Met776, Tyr687, His694 involved in hydrophobic interaction and Pro689, Trp663, Tyr690, His586, Trp520, Asn774, Val768, Asn510, Met776, Tyr687, His694 were involved in hydrogen bond interactions (Table 4.2, Fig. 4.4G and Fig. 4.4H). The docking studies of the chimera suggested that both the modules *Ct*GH1(β-

glucosidase) and *Ct*GH5-F194A (endoglucanase) are active in chimeric form and involved in the catalysis.

Table 4.2. Molecular docking analysis of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) with cello-oligosaccharides at catalytic sites of *Ct*GH1 and *Ct*GH5-F194A modules.

Cello-oligo-saccharides	β-Glucosidase (*Ct*GH1)			Endoglucanase (*Ct*GH5-F194A)		
	Binding free energy (kcal/mol)	Amino acids involved in hydrogen bond formation	Amino acids involved in hydrophobic bond formation	Binding free energy (kcal/mol)	Amino acids involved in hydrogen bond formation	Amino acids involved in hydrophobic bond formation
pNPG	-6.92	Ser247, Ser320, Leu246, Asn 245	Glu189, Val192, Trp351, Trp145, Trp433, Tyr 319	Nd	Nd	Nd
Cellobiose	-7.37	Glu189, Asn245	Leu200, Trp351, Val192, Tyr435, Trp145, Trp433, His203	-7.62	Glu626, Glu731, Asn774	His586, His585, Trp764, Met776, Met735, His694
Cellotriose	-7.62	Glu189,Glu378, Tyr319, Trp433	Glu432, His 203, Met349, Trp351, Leu193, Glu322, His249, Ser247, Asn 245	-7.91	His685, Asp668	Glu626, His694, Ser665, Tyr687, Trp663, Tyr690
Cellotetraose	-10.05	Glu331, Asn 245, Ser332, Glu432	Met349 Trp351, Glu189, Tyr319, Leu195, Ser335, Phe199, Leu200, Leu196, Trp433, Trp435	-10.05	His586, Glu731, His694, Asn766, Asn510	Glu626, Trp663, Tyr690, Tyr687, His694, Met776, Asn774, Trp520
Cellopentaose	-10.86	Trp433, Glu331, Ser335, Ser332, His 249	Glu189, Glu322, Ser247, Asn245, Tyr319, Val192, Glu432, Met349, Leu196, Trp351, Ala337, Leu268	-10.86	Asn766, Trp764, Asn688, Tyr690, Asn510	Asn774, Tyr687, Trp663, Tyr662, Glu626, Ser665, Trp520
Cellohexaose	-11.59	Glu331, Glu322, Glu350, Asn348	Tyr435, Ser265, Pro336, Leu200, Ser247, Ala269, Ile324, Leu268, His249, Trp351, Met349	-11.59	Glu691, Asn688, Glu626	Pro689, Trp663, Tyr690, His586, Trp520, Asn774, Val768, Asn510, Met776, Tyr687, His694
Celloheptaose	-11.13	Glu 342, Ser339, Glu32, Ala337, Asn 245	Glu189, Trp145, Tyr319, His203, Ser320, His249, Ser247, Ala267, Leu195, Leu268, Leu200, Glu331, Pro336	-11.54	Tyr690, Asp668, Glu691, Glu626, Trp764, Ser519	Asn766,Asn510, Asn774,Trp520, His586, Typ687, Trp663

nd- Not determined

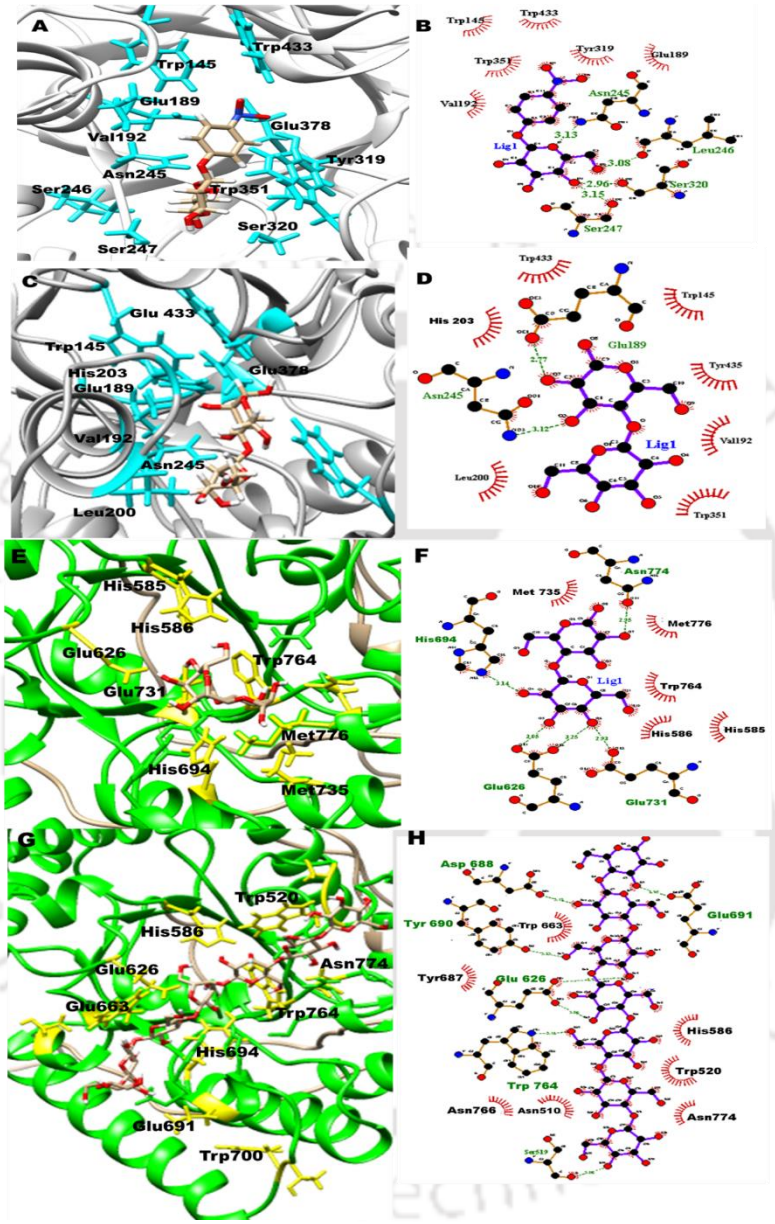


Fig. 4.4. Molecular docking analysis of chimera composed of module 1, *Ct*GH1(β-glucosidase) and module 2, *Ct*GH5-F194A (endoglucanase) with ligands. A) the active-site of *Ct*GH1 (β-glucosidase) showing the interaction of *p*NPG in the catalytic cleft, B) two-dimensional presentation of *p*NPG with the active site residues of *Ct*GH1 (β-glucosidase), C) the active-site of *Ct*GH1 (β-glucosidase) showing interaction of cellobiose in the catalytic cleft, D) two-dimensional presentation of cellobiose with active-site residues of *Ct*GH1 (β-glucosidase), E) the active-site of *Ct*GH5-F194A (endoglucanase) showing interaction of cellobiose in the catalytic cleft, F) two-dimensional presentation of cellobiose with the active-site residues of *Ct*GH5-F194A (endoglucanase), G) the active- site of *Ct*GH5-F194A (endoglucanase) showing the interaction of celloheptaose in the catalytic cleft and H) two-dimensional presentation of celloheptaose with the active-site residues of *Ct*GH5-F194A (endoglucanase) of the chimera.

4.3.5 Molecular dynamics of ligand-bound chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and corresponding individual enzymes

The molecular dynamics ofchimera in the presence of ligands cellobiose or cellohexaose was performed up to 70 ns. The structure stability of chimera in presence of the ligands was determined by analysing the time evolution of the C_α root-mean-square deviation (RMSD) and compared with the RMSD of only chimera. The chimeric enzyme containing cellobiose in the β-glucosidase module and cellohexaose in the endoglucanase module attained the stability after 15 ns and remained stable till 70 ns with a mean RMSD value of 0.5 nm (Fig. 4.5A). The MD simulation of the individual endoglucanase (*Ct*GH5-F194A) in presence of cellohexaose was also undertaken for 70 ns for comparative analysis with the cellohexaose bound endoglucanase module of Chimera. The ligand bound form of individual endoglucanase (*Ct*GH5-F194A) showed stable RMSD till 70 ns with mean RMSD value of 0.3 nm (Fig. 4.5B). The cellohexaose in the catalytic cleft of endoglucanase module of chimera interacts with the active site residues Glu626 and Glu731 (Fig. 4.5C). To determine the dynamics at the level of each residue the root mean square fluctuation (RMSF) was calculated. The active site residues Glu626 and Glu731 showed stability with RMSF ≤ 0.2 nm for both ligand bound and unbound chimera (Fig. 4.5D). The loop LA present in the endoglucanase module of the Chimera, is composed of amino acid residues ranging from 509-525 remains flexible during the MD simulation in both ligand bound and unbound forms of chimera (Fig. 4.5D). This loop (LA) is present near the flexible linker comprising amino acid residues (487-506), which is between β-glucosidase and endoglucanase modules (Fig. 4.5D). The loop LA is located near the -3 binding site of the catalytic cleft of endoglucanase module of the Chimera (Fig. 4.5C).

The structure superposition of ligand bound chimera from 0 ns (Blue) and after 70 ns MD simulation (Magenta), showed that the loop LA of MD simulated structure moves near the active site cleft of the endoglucanase module of Chimera and the distance of residue Asn510 between 0 ns and 70 ns structures is 4.7 Å (Fig. 4.5F). This relocation of the loop LA resulted in the stable hydrogen bond formation with a bond distance, ≤ 0.35 nm between the cellohexaose and the amino acid Asn510 in the catalytic site of endoglucanase module of Chimera as shown in 2D representation (Fig. 4.5H). The similar corresponding amino acid residue, Asn93 of a bifunctional cellulase/xylanase (*Ct*Cel5E) from C*lostridium thermocellum* was involved in the formation of hydrogen bond with the substrate for catalysis as earlier reported (Yuan *et al.,* 2015). Therefore, it can be concluded that Asn510 residue present in the binding site of the endoglucanase module in chimera is important for binding and catalysis of cellohexaose. Another loop LC present near the binding site of the endoglucanase module of chimeric enzyme comprising amino acid residues (685-689) displayed the flexibility upon ligand binding (Fig. 4.5D). In loop LC, the residue Tyr687 after 70 ns (Magenta) is displaced from the active site of the endoglucanase module of Chimera and the distance of the residue Tyr687 between 0 ns and 70 ns structures is 5.5 Å (Fig. 4.5F). The relocation of Tyr687 (corresponding residue Tyr270 for *Ct*Cel5E) was important for the catalysis, as it can facilitate the release of the product after the enzymatic cleavage by the acid/base catalysis carried out by Glu626/Glu731 (corresponding residues Glu209/Glu314 for *Ct*Cel5E) as also reported earlier (Yuan *et al.,* 2015). while the corresponding residues in individual endoglucanase are Glu139/Glu244. The RMSF for the loop LA for individual endoglucanase (*Ct*GH5-F194A) comprising amino acid residues (22-31) showed

minimal flexibility in the ligand bound and unbound forms (Fig. 4.5E). The structure superposition of the ligand bound individual endoglucanase (*Ct*GH5-F194A) (green) with its 70 ns MD simulated ligand bound structure (Red) showed that due to the lack of flexibility in the loop LA of individual endoglucanase (*Ct*GH5-F194A), the residue Asn23 (corresponding residue Asn510 in Chimera) cannot relocate towards the active site for hydrogen bond formation (Fig. 4.5G). Therefore, stable hydrogen bond formation was not initiated in the catalytic cleft of individual endoglucanase (*Ct*GH5-F194A) (Fig. 4.5I). The 2D representation of the ligand bound cellohexaose in the individual endoglucanase (*Ct*GH5-F194A) after 70 ns of simulation also confirmed that Asn23 (corresponding Asn510 in Chimera) is not involved in hydrogen bond formation with cellohexaose (Fig. 4.5I). The loop LC comprising the amino acid residues (197-201) in the individual endoglucanase (*Ct*GH5-F194A) showed the flexibility, when the cellohexaose is present in the structure (Fig. 4.5E). The superposition of cellohexaose bound individual endoglucanase (*Ct*GH5-F194A) at 0 ns (green) with 70 ns MD simulated structure (red) demonstrated the restriction in the movement of Tyr200 residue (corresponding residue Tyr 687 in chimera) from the active site (Fig. 4.5G). This restricted movement of residue Tyr200 in individual endoglucanase (*Ct*GH5-F194A) might be causing delayed product release from its active site. Therefore, its catalytic efficiency is lower than the endoglucanase module of the Chimera. The structural comparison after MD simulation concluded that the flexibility of the loops LA and LC and their relocation efficiency in the endoglucanase module of the chimera (*Ct*GH5-L1-*Ct*GH5-F194A) is higher than the individual endoglucanase (*Ct*GH5-F194A). This

favourable flexibility in the loops of endoglucanase module of chimera resulted in stable ligand binding and efficient release of product after catalysis from the catalytic cleft.

For further analysis, the average short range interaction energy was calculated for ligand bound chimera and its individual enzymes. The interaction energy of endoglucanase module of Chimera with cellohexaose was -383 kJ/mol, while the individual endoglucanase (CtGH5-F194A) showed interaction energy of -254 *kJ/mol* with cellohexaose after the MD simulation for 70 ns. Therefore, taking into consideration the MD simulation and the short range interaction energy calculation, a conclusion can be drawn that the endoglucanase module of chimera is more stable for interaction with the ligands in its catalytic cleft. Therefore, this may lead to efficient catalysis of the ligand. Moreover, the β-glucosidase module of chimera also showed sufficiently strong interaction with cellobiose giving an average short range interaction energy of -64 kJ/mol. Similarly, the individual β-glucosidase (*Ct*GH1) showed no substantial change in the interaction energy. Furthermore, no significant changes in the loops, near the binding site of both β-glucosidase module of chimera and its individual β-glucosidase (*Ct*GH1) after the MD simulation for 70 ns were found (data not shown). Although, the MD simulated ligand bound form of chimera showed minor fluctuations in the outer loops of β-glucosidase module after 70 ns (data not shown). This fluctuation in the outer loops might be responsible for the conformational change that might be causing improved catalysis in β-glucosidase module of the chimera. Therefore, from the MD simulation study of the interaction of chimera and its individual enzymes with ligands, it was found that endoglucanase module of chimera shows stable binding with cellohexaose and efficient release of the hydrolysed product (cellobiose) followed by the

action of fused β-glucosidase module of chimera with changed conformation leading to its improved catalytic efficiency. Therefore, all these favourable interactions made this Chimera a potent enzyme for hydrolysis of cellulosic substrates into glucose.

The overall MD simulation studies amply demonstrated that**,** the two modules in the Chimera, are present without any unfavourable inter-modular interactions and are in active form and independently functioning. The property of a natural linker for preventing unfavourable inter-modular interactions was also reported earlier (Chichili et al., 2013). Therefore, the independent functioning of the two modules might be because of the natural linker, that has been used for connecting the two modules. The MD simulation results of protein-ligand complex displayed that the relocation and orientation of the ligand binding site loops LA and LC in endoglucanase module in chimera plays a key role, that might be responsible for enhanced catalytic efficiency.

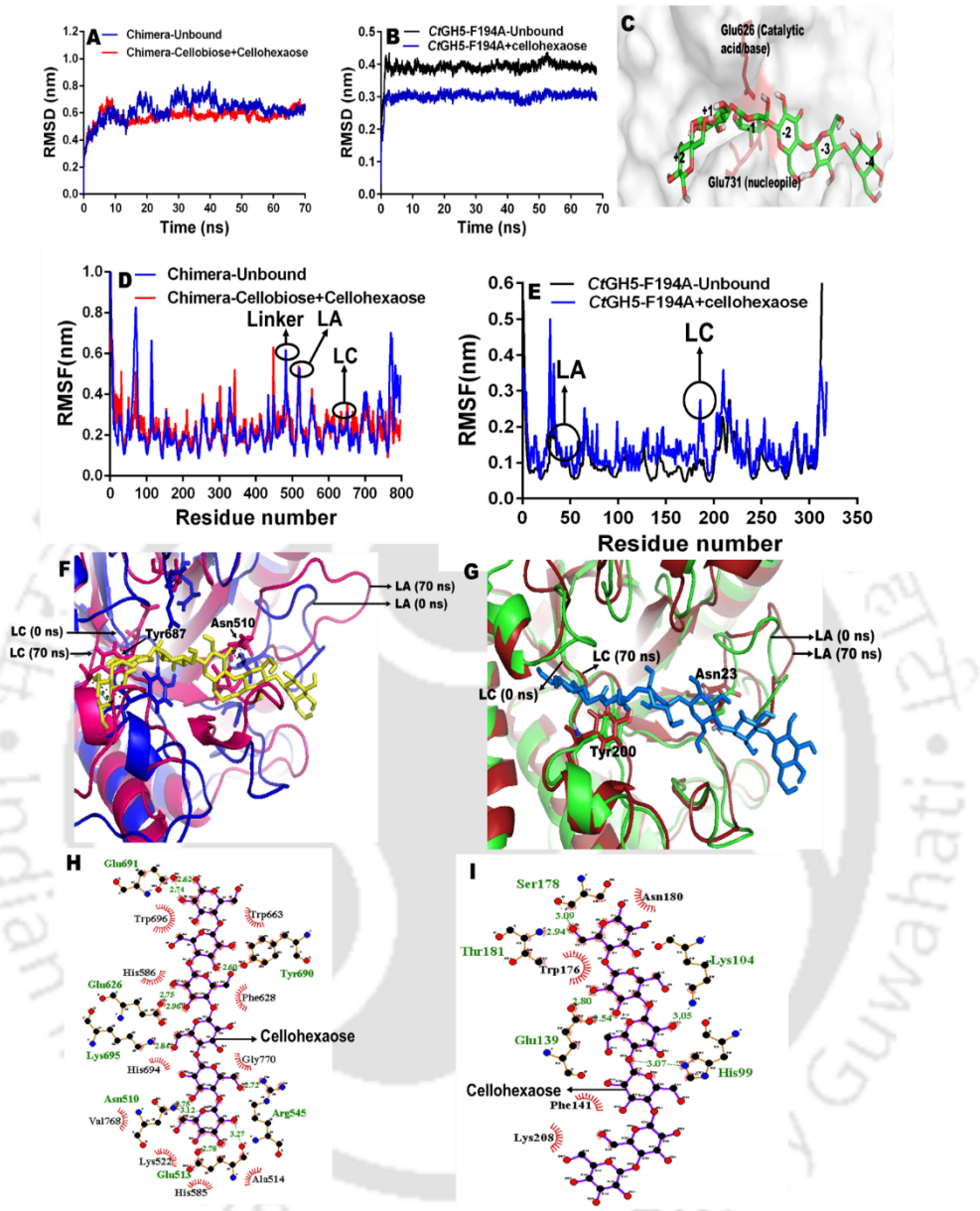


Fig. 4.5. Conformational dynamics for endoglucanase activity of substrate-bound Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and individual endoglucanase (*Ct*GH5-F194A) (A) RMSD of ligand bound and unbound forms of Chimera (B) RMSD of ligand bound and unbound forms of individual endoglucanase (*Ct*GH5-F194A) (C) position of cellohexaose (green) at the active site of endoglucanase module of Chimera (D) RMSF of ligand bound and unbound forms of Chimera (E) RMSF of ligand bound and unbound forms of individual endoglucanase (*Ct*GH5-F194A) (F) Active site residues of endoglucanase module of Chimera with cellohexaose (dark yellow) after 70 ns simulation (G) The interaction of active site residues of individual endoglucanase (*Ct*GH5-F194A) with cellohexaose (blue) after MD simulation of 70 ns (H) two-dimensional schematic presentation of cellohexaose with the active-site residues of endoglucanase module (*Ct*GH5-F194A) of Chimera after MD simulation for 70 ns (I) two-dimensional schematic presentation of cellohexaose with the active-site residues of individual endoglucanase (*Ct*GH5-F194A) after MD simulation for 70 ns.

4.3.6 Small angle x-ray scattering analysis of chimera

The small angle x-ray scattering analysis of chimera showed the conformational behaviour and molecular arrangement of both the linked modules in the solution form. The SAXS data of chimera at both concentrations (3 mg/ml and 5 mg/ml) were processed, and the results were given in Table 4.3. The initial processing and through visual examination of the scattering pattern generated by the chimera (*Ct*GH1-L1-*Ct*GH5-F194A) for different concentrations revealed monodisperse nature for the chimera in solution (Fig. 4.6A). The SAXS profiles at both concentrations were similar and lacked the aggregation or inter-particle interaction effect. The SAXS data collected at 5 mg/mL was used for further processing and analysis. Guinier estimation analysis demonstrated that the radius of gyration (Rg) for chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was 0.57±0.05 nm for rod-shape and 3.15±0.10 nm for globular shape. The length of the chimera (*Ct*GH1-L1-*Ct*GH5-F194A) molecule was 10.7 nm. The linearity of Guinier plot at low q region unveiled that the chimera exists in a monodispersed state in the solution form (Fig. 4.6B). The evaluation of P(R) function acquired by Fourier transformation of the chimera (*Ct*GH1-L1-*Ct*GH5-F194A) scattering profile displayed the symmetric profile (Fig. 4.6C). The P(R) plot in Fig. 4.6(C) displayed a shoulder signifying that there is an inter-modular separation between the two modules. The maximum dimension (D_{max}) and the radius of gyration (Rg) of chimera estimated from P(R) plot were found to be 10 nm and 3.17 nm, respectively (Fig. 4.6C). The global compactness and flexibility in solution for the chimera were determined by Kratky plot analysis (Fig. 4.6D). The analysis using Kratky plot for the chimera displayed bell-shape peaks at low Q region, confirming a compact and folded structure of chimera.

The molecular mass of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) procured from SAXSMow server employing SAXS scattering profile was 87 kDa, which is similar to the theoretical and experimental molecular mass (89 kDa) as described earlier (Nath *et al.*, 2019). The averaged Gasbor generated *ab initio* models for Chimera displayed two-module structure (Fig. 4.6E). The Gasbor modeled structure of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) displayed an elongated structure with two modules existing as independent units with no inter-modular contact, and having an overall shape similar to a bean-bag contour (Fig. 4.6F). The superposition of modeled structure of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) with Gasbor derived *ab initio* model from the collected SAXS data, displayed an excellent fitting with a minor deviation in the hexa-histidine tag region at the N-terminal (Fig. 4.6G & 4.6H). The phenomenon of independent existence of different modules in a multi-modular ubiquitin protein with no inter-modular contact by *ab initio* shape reconstruction using SAXS studies was also reported earlier (Bernado *et al.*, 2010). The CRYSOL fitting of the chimera (*Ct*GH1-L1-*Ct*GH5-F194A) modeled structure with the SAXS experimental data (Fig.4.6I) also corroborated with the results obtained by SAXS data processing. Therefore, the SAXS data supported the results of MD simulation as the two modules of chimera displayed no interaction between them. Therefore, the results clearly showed that the two modules of chimera are functionally independent and are in active form as also elucidated earlier by biochemical studies (Nath *et al.*, 2019.)

Table 4.3. SAXS data collection parameter and derived parameters of chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

Data-collection parameter	Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)
Instrument	SAXSpace Anton-Paar
Wavelength (Å)	1.54 Å
Q range (nm^{-1})	0.135-5.95
Exposure time (min.)	30x2
Temperature (°C)	10
Protein concentration (mg/ml)	5
Structural parameter	
Q range (nm^{-1}) used for Rg analysis	0.10-0.41
I(0) au from Guinier	57810±693
Rg nm from Guinier	3.13±0.1
I(0) au from P(r)	57810
Rg nm from P(r)	3.12±0.07
D_{max} (nm)	10.0
Porod volume estimate (nm^3)	120.82
Molecular Mass Determination	
Theoretical molecular mass (kDa)	89
Molecular Mass from SAXSMOW (kDa)	87
Modeling parameters	
Q range (nm^{-1}) used for structure modeling	0.12-2.53
Resolution (nm)	33±0.5
NSD	0.93
Software employed	
Data processing	Primus
P(r) function calculation	GNOM
Ab initio modeling	GASBOR
Validation and averaging	DAMAVER
Structure superposition	SUPCOMB
3-D graphical representation	PyMOL

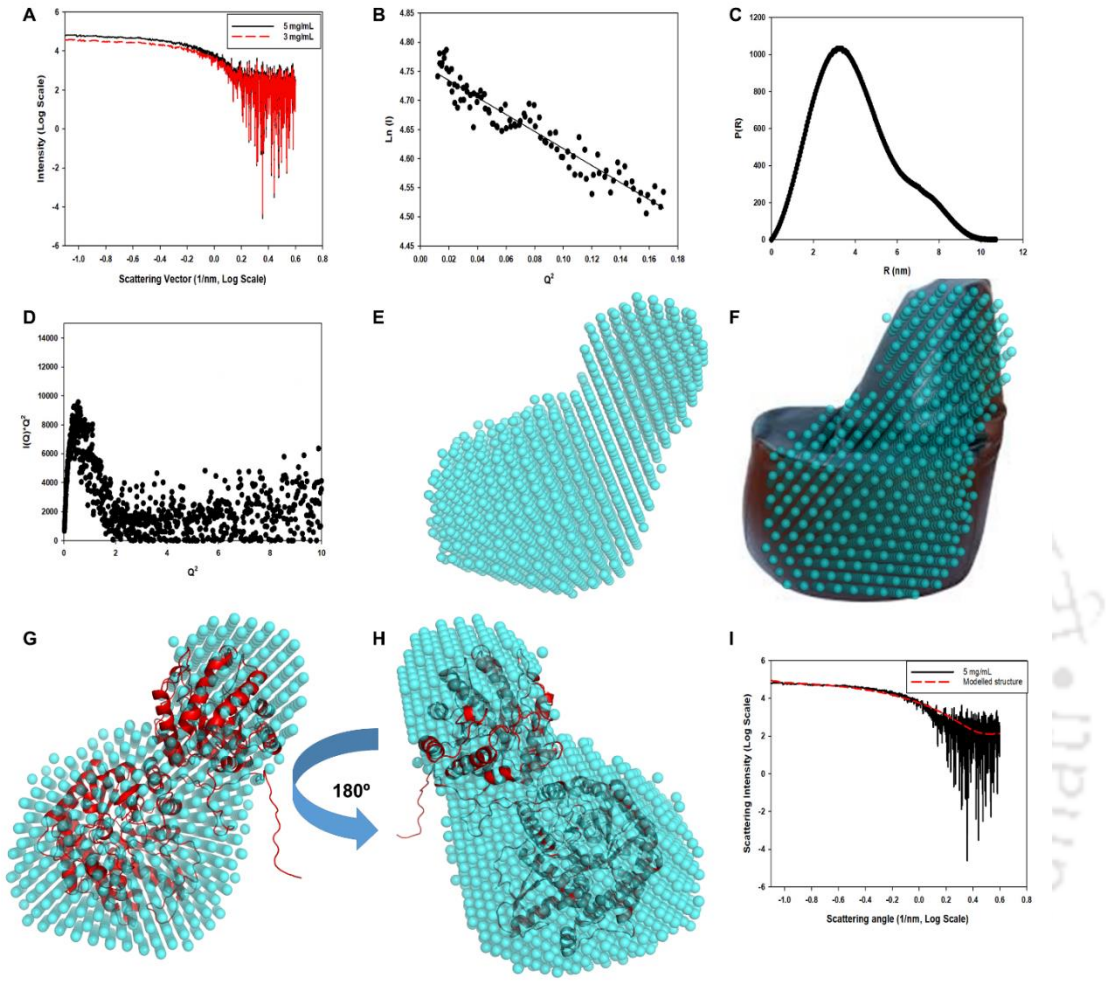


Fig. 4.6. Small angle X-ray scattering analysis of chimera (*Ct*GH1-L1-*Ct*GH5-F194A). A) SAXS intensity profile, (B) Guinier plot of the SAXS intensities, C) P(R) curve of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) as a function of R, (D) Kratky plot, (E) *ab initio* derived shape of chimera (*Ct*GH1-L1-*Ct*GH5-F194A), (F) Gasbor generated model superposed on bean-bag shape, (G & H) Superposition of chimera (*Ct*GH1-L1-*Ct*GH5-F194A) modeled structure with the *ab initio* model Gasbor derived structure and (I) CRYSOL analysis of modeled structure and experimentally collected SAXS data.

4.4 Conclusion

The amino acid sequence of the chimera (*Ct*GH1-L1-*Ct*GH5-F194A) composed of β-glucosidase (*Ct*GH1) and endoglucanase (*Ct*GH5-F194A) was evaluated for secondary structure analysis by using Psi-Pred and RaptorX prediction tools and compared with the CD data. The CD analysis revealed that the secondary structure of chimera consists of 38% of α-helix, 9.32% of β-sheet and 52.68% of the loops corroborating with the predicted secondary structure elements. The 3-dimensional modeled structure of chimera was generated by RaptorX, which showed close homology with β-glucosidase, 3AHX from *Clostridium cellulovorans* and an endoglucanase, 5BYW from *Clostridium thermocellum.* The modeled structure generated was validated by Ramachandran plot, that showed 99.9% amino acid residues in the allowed region. The modeled structure of chimera represented a modular structure consisting of N-terminal β-glucosidase (*Ct*GH1) and C-terminal endoglucanase (*Ct*GH5-F194A) modules connected by the linker. The overall modular structure of chimera represented a classical (α/β)-TIM barrel fold. The molecular dynamics study of the Chimera showed stable RMSD till 70 ns with 0.01 oscillation suggesting the stable conformation of the two modules, *Ct*GH1 (β-glucosidase) and *Ct*GH5-F194A (endoglucanase) in the chimeric form. The average SASA of the chimera remained stable with an average value of 320 nm^2. The minor variation in the SASA also suggested that the accessibility of the substrate in the catalytic sites were not affected in the designed chimera. This indicated that the binding sites were not obstructed by deformation or steric effect between the two modules of chimeric enzyme. The docking studies of the chimera suggested that both the modules

*Ct*GH1(β-glucosidase) and *Ct*GH5-F194A (endoglucanase) are active in chimeric form and involved in the catalysis. The molecular dynamics simulation of the docked complex of chimera showed stable interaction of its β-glucosidase and endoglucanase modules with ligands. MD simulation results of cellohexaose bound and unbound endoglucanase module of chimera displayed favourable flexibilities in the loops resulting in stable binding and efficiently releasing the product from the catalytic cleft after catalysis. Higher short range interaction energy -383 kJ/mol obtained for cellohexaose bound chimera for endoglucanase activity displayed stronger interactions while the individual endoglucanase enzyme showed -254 kJ/mol against cellohexaose thereby resulting efficient binding and hydrolysis of cellohexaose into cellobiose.

MD simulation results showed that the loops near the binding site of β-glucosidase module of chimera and individual β-glucosidase showed no significant changes. However, the outer loops of the β-glucosidase module of chimera showed fluctuations. These fluctuations showed that there might be conformational changes in the β-glucosidase module of chimera that is contributing to improved hydrolysis of cellobiose to glucose. These favourable interactions of both endoglucanase and β-glucosidase modules of chimera against the ligands led to the improved catalytic efficiency of chimera. The SAXS analysis of chimera displayed elongated structure with two modules in fully folded form in the solution and an overall shape similar to a bean-bag contour. The outcome of this study of chimera by computational and SAXS data studies revealed that the peptide linker keeps the two modules of chimera independent for their biological function.

References

Akasako, A., Haruki, M., Oobatake, M., & Kanaya, S. (1997). Conformational stabilities of Escherichia coli RNase HI variants with a series of amino acid substitutions at a cavity within the hydrophobic core. *Journal of Biological Chemistry, 272*, 18686-18693.

Andrade, M. A., Chacon, P., Merelo, J. J., & Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Engineering, Design and Selection, 6*, 383-390.

Bernado, P. (2010). Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *European Biophysics Journal, 39*, 769-780.

Beguin, P., (1990). Molecular biology of cellulose degradation. *Annual Reviews in Microbiology, 44*, 219-248.

Collins, M. N., Dalton, E., Schaller, B., Leahy, J. J., & Birkinshaw, C. (2012). Crystal morphology of strained ultra high molecular weight polyethylenes. *Polymer Testing, 31*, 629-637.

Cao, H., Li, X., Zhang, Y., Shi, P., Bai, Y., & Yao, B. (2018). Glucose-tolerance molecular modification of GH1 β-glucosidase from Alicyclobacillus sp. A4. *Journal of Agricultural Science and Technology (Beijing), 20*(5), 26-33.

Crasto, C. J., & Feng, J. A. (2000). LINKER: a program to generate linker sequences for fusion proteins. *Protein Engineering, 13*, 309-312.

Cota, J., Oliveira, L. C., Damásio, A. R., Citadini, A. P., Hoffmam, Z. B., Alvarez, T. M., & Murakami, M. T. (2013). Assembling a xylanase-lichenase chimera

through all-atom molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1834*, 1492-1500.

Dalton, E., & Collins, M. N. (2014). Lamella alignment ratio: a SAXS analysis technique for macromolecules. *Journal of Applied Crystallography*, *47*, 847-851.

Fan, Z., Werkman, J. R., & Yuan, L. (2009). Engineering of a multifunctional hemicellulase. *Biotechnology Letters*, *31*, 751-757.

Fischer, H. D. O. N., Oliveira Neto, M. D., Napolitano, H. B., Polikarpov, I., & Craievich, A. F. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography*, *43*, 101-109.

Fournet, G., & Guinier, A. (1955). Small angle scattering of X-rays. *Translated by Walker, CB and Yudowitch, KL In: New York: John Wiley & Sons*, 7-78.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M. & Svergun, D.I. (2017). ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, *50*, 1212-1225.

Grabnitz, F., Seiss, M., RucknageL, K. P., & Staudenbauer, W. L. (1991). Structure of the β-glucosidase gene bglA of *Clostridium thermocellum*: Sequence analysis reveals a superfamily of cellulases and β-glycosidases including human lactase/phlorizin hydrolase. *European Journal of Biochemistry*, *200*, 301-309.

Hess, B., Bekker, H., Berendsen, H. J., & Fraaije, J. G. (1997). LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*, 1463-1472.

Jeng, W. Y., Wang, N. C., Lin, M. H., Lin, C. T., Liaw, Y. C., Chang, W. J., & Wang, A. H. J. (2011). Structural and functional analysis of three β-glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *Journal of Structural Biology*, *173*, 46-56.

Onuchic, J. N., & Wolynes, P. G. (2004). Theory of protein folding. *Current Opinion in Structural Biology*, *14*, 70-75.

Kallberg, M., Margaryan, G., Wang, S., Ma, J., & Xu, J. (2014). RaptorX server: a resource for template-based protein structure modeling. *Protein Structure Prediction*, Humana Press, New York, NY, 17-25

Kelly, S. M., Jess, T. J., & Price, N. C. (2005). How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1751*, 119-139.

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

Kozin, M. B., & Svergun, D. I. (2001). Automated matching of high-and low-resolution structural models. *Journal of Applied Crystallography*, *34*, 33-41.

Lu, P., & Feng, M. G. (2008). Bifunctional enhancement of a β-glucanase-xylanase fusion enzyme by optimization of peptide linkers. *Applied Microbiology and Biotechnology*, *79*, 579-587.

Reddy Chichili, V. P., Kumar, V., & Sivaraman, J. (2013). Linkers in the structural biology of protein–protein interactions. *Protein Science*, *22*(2), 153-167.

Jamros, M. A., Oliveria, L. C., Whitford, P. C., Onuchic, J. N., Adams, J. A., Blumenthal, D. K., & Jennings, P. A. (2010). Proteins at work: A combined SAXS and theoretical determination of the multiple structures involved on the protein kinase functional landscape. *Journal of Biological Chemistry*, *285,* 36121-36128.

McKee, L. S., Peña, M. J., Rogowski, A., Jackson, A., Lewis, R. J., York, W. S., & Marles-Wright, J. (2012). Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proceedings of the National Academy of Sciences USA, 109,* 6537-6542.

Meleiro, L. P., Salgado, J. C. S., Maldonado, R. F., Carli, S., Moraes, L. A. B., Ward, R. J.,& Furriel, R. P. M. (2017). Engineering the GH1 β-glucosidase from Humicola insolens: insights on the stimulation of activity by glucose and xylose. *PloS One*, *12*, 0188254.

Nath, P., Dhillon, A., Kumar, K., Sharma, K., Jamaldheen, S. B., Moholkar, V. S., & Goyal, A. (2019). Development of bi-functional chimeric enzyme (*Ct*GH1-L1-CtGH5-F194A) from endoglucanase (*Ct*GH5) mutant F194A and β-1, 4-glucosidase (*Ct*GH1) from *Clostridium thermocellum* with enhanced activity and structural integrity. *Bioresource Technology*, *282*, 494-501.

Peng, J., & Xu, J. (2011). RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins: Structure, Function and Bioinformatics*, *79*, 161-171.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera-a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*, 1605-1612.

Schrodinger, L. L. C. The PyMOL Molecular Graphics System, Version 2.0. 2019.

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

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

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

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

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

Yang, S., Blachowicz, L., Makowski, L., & Roux, B. (2010). Multidomain assembled states of Hck tyrosine kinase in solution. *Proceedings of the National Academy of Sciences USA*, *107*, 15757-15762.

Urbanowicz, B. R., Bennett, A. B., del Campillo, E., Catala, C., Hayashi, T., Henrissat, B., & Teeri, T. T. (2007). Structural organization and a standardized nomenclature for plant endo-1,4-β-glucanases (cellulases) of glycosyl hydrolase family 9. *Plant Physiology*, *144*, 1693-1696.

Woolfson, D. N. (2001). Core-directed protein design. *Current Opinion in Structural Biology*, *11*, 464-471.

Wu, T. H., Huang, C. H., Ko, T. P., Lai, H. L., Ma, Y., Chen, C. C., & Guo, R. T. (2011). Diverse substrate recognition mechanism revealed by *Thermotoga maritima* Cel5A structures in complex with cellotetraose, cellobiose and mannotriose. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1814*, 1832-1840.

Yuan, S. F., Wu, T. H., Lee, H. L., Hsieh, H. Y., Lin, W. L., Yang, B. & Ho, M. C. (2015). Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*. *Journal of Biological Chemistry*, *290*, 5739-5748.

Chapter 5

Saccharification of alkali and organosolv pretreated sugarcane bagasse by cocktail of cellulase- Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and Cellobiohydrolase (*Ct*CBH5A)

5.1 Introduction

Extensive consumption of energy resources have led the focus towards the research and technology development in the field of sustainable energy resources and their production (Perry *et al.,* 1973, Ntaikou *et al.,* 2014). Lignocellulosic biomass is one of the sustainable energy resources for the production of bioethanol. The cell walls of this lignocellulosic biomass contain cellulose, hemicellulose which can be converted into alcohol. Therefore, plant lignocellulosic biomass can be used for the generation of alternative energy source (Chen *et al.,* 2016). The recalcitrance nature of this biomass is one of the barriers against the action of the various chemical and microbial degradation processes (Zhao *et al.,* 2009).

The plant cell wall is made up of three layers which are the lamella, the primery cell wall and the secondary cell wall, the lamella contain pectin it is present in the outermost layer forms the interfaces between adjacent plant cells and primary cell wall is thin, flexible and extensible layer formed during thee plant cell wall growth. The

secondary cell wall is made up of cellulose, hemicellulose and lignin. In the secondary cell wall, the interaction between cellulose and hemicellulose and degree of lignification is responsible for recalcitrance of the lignocellulosic biomass (Zhao *et al.*, 2009). Therefore, effective pretreatment processes are required for disruption of the plant cell wall as well as cellulose crystallinity and lignin association so that hydrolytic enzymes can act more efficiently for releasing the fermentable sugars (Zhao *et al.*, 2009).

The amount of these three components (cellulose, hemicellulose,and lignin) within the cell walls of the plants dependend on plant species, growing conditions, genotype expression as well as the developmental stages of the plants (Chakar *et al.*, 2004)

There are composition difference between the hardwood and softwood species. The hardwoods contain about 40-50% cellulose, 24-40% hemicellulose, and 18-25% lignin, while the composition of the softwoods cell wall ranges between 45-50% cellulose, 25-35% hemicellulose and 25-35% lignin (Zeronian *et al.*, 1995). The composition in grass varies from 25-40% cellulose, 35-50% hemicellulose and 10-30% lignin (Zeronian *et al.*, 1995). The amount of lignin varies in this species. Therefore, optimal feedstock can be selected based on composition for specific purpose but the pretreatment processes may vary from one biomasses to another (Chakar *et al.*, 2004).

Lignocellulosic substances such as agricultural wastes are promising feedstocks for the production of bioethanol. Agricultural wastes are renewable, abundant and cost-effective sources. Sugarcane bagasse is one of the largest cellulosic agro-industrial waste (Pandey *et al.*, 2000). Sugarcane bagasse contains approximately 50% of cellulose, 25% each of hemicellulose and lignin (Pandey *et al.*, 2000), therefore it can be used for the production of bioethanol. The overall process for the bioethanol

production has several limitations and challenges such as transportation of biomass, biomass handling and efficient pretreatment methods for total delignification of lignocellulosics. The pretreatment of lignocellulosic biomass hybrid poplar remove lignin to increase the accessibility of potential carbohydrate such as cellulose and hemicellulose for the production of fermentable sugars after enzymatic saccharification (Zhu *et al.* 2010). Bioethanol production from bagasse requires cellulases for saccharification (Travaini *et al,* 2015). The production of these cellulases is cost-intensive process. The efficiency of the hydrolysis process depends on the usage of enzyme cocktail that has higher catalytic efficiency along with the synergistic actions of enzymes (McKee *et al.*, 2012). To reduce the process cost, economic development of cellulases, with bi-functional or multifunctional activities were initiated (Lee *et al.*, 2011). Application of these engineered enzymes for deconstruction of the pretreated biomass into monosugar such as glucose can be in bioethanol production (Lee *et al.*, 2011). In the present study, sugarcane bagasse was subjected to alkali pretreatment following by organosolv pretreatment process for its delignification. The deliginfied pretreated biomass was further enzymatically hydrolysed using a cocktail of cellulases. The cocktail was of Chimera having bi-functional β-glucosidase and endoglucanase activities and Cellobiohydrolase (*Ct*CBH5A) activity. The best pretreatment and saccharification of biomass giving a maximum yield of glucose was further analysed using different analytic processes for determining the efficiency of the pretreatment on the biomass for delignification and the cellulose acceesibility.

5.2 Materials and Methods

5.2.1 Chemical and Reagents

Phosphoric acid (85%, AR) and *p*NP-β-D-glucopyranoside were purchased from Sisco Research Laboratories Pvt. Ltd. India. Carboxy methylcellulose (CMC), D-glucose, D-cellobiose, D-xylose and L-arabinose were obtained Sigma-Aldrich, India. Sodium carbonate, Sodium potassium tartarate, Sodium bicarbonate, Sodium sulphate, sodium phosphate (monobasic), Sodium phosphate (dibasic) were procured from HiMedia Laboratories Pvt. Ltd., India. Sodium arsenate, ammonium molybdate, sulphuric acid, acetone were purchased from Merck Limited, India. Sugarcane bagasse (scb) was gathered from the local juice center at IIT Guwahati campus. The lignocellulosic feedstock (sugarcane bagasse) was washed trice with water dried in the oven at 70°C. The dried substrate was grinded and sieved through a sieve with 1 mm pore size.

5.2.2 Pretreatment of sugarcane bagasse.

5.2.2.1 Stage 1 Alkali pretreatment

The 1g of sieved sugarcane bagasse was chemically pretreated by adding in 15 ml of 1% (w/v) NaOH for different time intervals, viz. 1h, 2h, 3h and 4h and incubating at 50°C. The alkali pretreated sugarcane bagasse (scb) pulp was washed with distilled water until the neutral pH 7.0 was achieved. The neutralised samples were then filtered through muslin cloth and dried in oven at 60°C for overnight. The hydrolysate was preserved for further analysis. The dried filtrate residues were further pretreated using phosphoric acid-acetone method mentioned in section 5.2.2.2. The raw sugarcane bagasse was also subjected single stage organosolv pretreatment using phosphoric acid-acetone pretreatment as mentioned earlier by Gupta *et al.*, 2017.

5.2.2.2 Stage 2 Organosolv (phosphoric acid-acetone) pretreatment of alkali pretreated sugarcane bagasse

The dried alkali pretreated scb obtained from the stage 1 (Section 5.2.2.1) pretreatment for different time interval of 1h, 2h, 3h and 4h were further treated using phosphoric acid-acetone pretreatment. 1g of each dried pretreated samples from stage 1 was mixed in 85% (v/v) phosphoric acid and kept at 50°C at 120 rpm for 1h in 50 ml conical flask. After 1h, 24 ml of chilled acetone was added and centrifuged at 6000g for 10 min. The supernatant was discarded and another 24 ml of chilled acetone was added to the slurry and centrifuged at 6000g for 10 min. After centrifugation the slurry was separated from the supernatant and washed with 24 ml of distilled water by centrifuging at 6000g for 10 min. This process centrifugation of the slurry with distilled water was repeated twice. In the final washing step, the pH of the slurry was adjusted to 6.0 by 1.5M NaOH. The neutralised slurry was then separated from the supernatant using muslin cloth and dried at room temperature for overnight. The fermenting sugar (hexoses and pentoses) in the pretreated hydrolysates from all the single (alkali and organosolv) and dual stage pretreatments of alkali pretreated followed organosolv preetreatment were analysed by high performance liquid chromatography (HPLC coupled with RI detector, Shimadzu, Japan). The samples were filtered through PVDF membrane (pore size 0.2 µm, MF-Millipore, Merck, Massachusetts) using the sterile syringe filter. The HPLC column Phenomenex Rezex ROA (300 x 7.8 mm) with guard column (50 x 7.8 mm) was used for reducing sugar analysis according to the method as described earlier (Jamaldheen *et al.*, 2018). The column temperature was maintained constant at 37ºC throughout the analysis. 5 mM H_2SO_4 was used as the mobile phase with a flow rate of 0.5 mL/min and retention time 40 min. Glucose, xylose and

cellobiose standards (Sigma Aldrich, USA) were prepared from 0.1 to 1 mg/mL concentrations.

5.2.3 Production of recombinant enzymes Chimera and Cellobiohydrolase (*Ct*CBH5A)

The recombinant enzymes, Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and Cellobiohydrolase (*Ct*CBH5A) were used for saccharification of pretreated sugarcane bagasse. Chimeric enzyme *Ct*GH1-L1-*Ct*GH5-F194A (β-glucosidase and endo β-1,4 glucanase) mentioned in chapter 3 and section 3.3.4**.** Cellobiohydrolase (*Ct*CBH5A) cloned from *Clostridium thermocellum* was a gift from Professor Carlos M.G.A. Fontes, NZYTech Ltd., Lisbon, Portugal. Both Chimera and Cellobiohydrolase were expressed in *Escherichia coli* BL21 (DE3) cells and purified by a method mention in chapter 2 and section 2.3.4. Both, Chimera and *Ct*CBH5A proteins were expressed in *Escherichia coli* BL21 (DE3) cells at temperature 24ºC. The cells were centrifuged at 8000g for 10 min at 4ºC and resuspended in 4 mL of 50 mM sodium phosphate buffer, pH 7. This cell suspension was sonicated at 33% amplitude (pulse 5s on and 10s off, Sonics, Vibra cell) for 10 min. The supernatant containing the extracellular proteins Chimera and *Ct*CBH5A were purified by Ni^{2+} ions charged 5 mL, Sepharose column by using immobilized metal-ion affinity chromatography as described earlier in chapter 2 and section 2.2.11.1.

5.2.4 Determination of enzyme concentration of recombinant enzymes (Chimera and Cellobiohydrolase)

The protein concentrations of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and Cellobiohydrolase (*Ct*CBH5A) were determined by the Bradford method (Bradford *et al.*, 1976) by using BSA standard. The reaction mixture contained 10 µL of protein in 90 µL of 50 mM sodium phosphate buffer, pH 7 and 1 mL Bradford reagent. The reaction mixture was incubated at 25°C for 15 min and absorbance at 595 nm was measured by UV-visible spectrophotometer (Perkin Elmer, Lambda-45).

5.2.5 Determination of enzyme activity of recombinant enzymes (Chimera and Cellobiohydrolase)

The enzyme activity of *Ct*GH1-L1-*Ct*GH5-F194A and *Ct*CBH5A was determined by reducing sugar estimation method as described earlier by Nelson (Nelson *et al.,* 1944) and Somogyi (Somogyi *et al.*, 1945). Chimera showed β-glucosidase activity and endoglucanase 810 U/µmol and 7,000 U/µmol, respectively, as described earlier in section 3.3.4. The specific activity of Cellobiohydrolase *(Ct*CBH5A) was 20 U/mg in 50 mM sodium phosphate buffer, pH 6.4 at 65°C for 2 min (unpublished work).

5.2.6 Determination of enzyme stability of recombinant enzymes (Chimera and Cellobiohydrolase)

The Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was stable in 50 mM citrate phosphate buffer, in pH range 4.0 to 5.5 as mentioned in Chapter3 and section 3.3.2. The *Ct*CBH5A was stable in 50 mM sodium phosphate buffer in pH range 6.2 to 6.6 (Unpublished work). For the enzymatic saccharification of sugarcane bagasse the

cocktail comprising of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and Cellobiohydrolase (*Ct*CBH5A) common pH and the same stable buffer was considered which was 50 mM sodium phosphate buffer, pH 6.4 (Unpublished work).

5.2.7 Saccharification of pretreated sugarcane bagasse by Chimera (*Ct*GH1-L1-*Ct*GH5- F194A) and Cellobiohydrolase (*Ct*CBH5A)

The saccharification of all the pretreated samples from stage 1 and stage2 section 5.2.2.2 (only organosolv, alkali treated for 1h+ organosolv, alkali treated for 2h+ organosolv, alkali treated for 3h + organosolv and alkali treated for 4h+ organosolv) were subjected to enzymatic saccharification by using cocktail of Chimera and CBH5A. The activity of Chimera was 7000 U/μmol for endoglucanase activity and 810 for β-glucosidase activity. The specific activity of *Ct*CBH5A was 20 U/mg. The total enzyme loading used was 600 U/g of pretreated biomass. The Chimera and *Ct*CBH5A were used in a previously optimized ratio of 2:3 ratio in saccharification reaction. The substrate in each case was kept constant by using 2% (w/v) of the pretreated biomass. The saccharification reaction was carried out in 1 mL reaction volume, in 50 mM sodium phosphate buffer, pH 6.4 at 30ºC for the different time intervals of 48h, 72h and 96h in shaking incubator at 150 rpm. At each time period the total reducing sugar and glucose yield were quantified. The enzymatic hydrolysates were analysed for reducing sugar by a method described in chapter 3 and section 3.2.2.1 by Nelson (Nelson *et al.*, 1945) and Somogyi (Somogyi *et al.*, 1945). The glucose present in enzymatic hydrolysates obtained from all the pretreatment methods was performed by HPLC as mentioned in section 5.2.2.2.

5.2.8 Fourier-transform infrared spectroscopy (FTIR) analysis of untreated and pretreated sugarcane bagasse

The FTIR-spectral analysis of oven dried raw and pretreated sugarcane bagasse biomass was performed by grinding the biomass with KBr in the ratio 1:100 (w/w) and pelleting them in hydrolytic press. The pelleted samples were scanned between wavenumber 400-4000 cm^1 on FTIR (Perkin Elmer, USA) and the peaks corresponding to lignin, hemicelluloses and cellulose were analysed.

5.2.9 Field Emission Scanning Electron Microscope (FESEM) analysis of untreated and pretreated sugarcane bagasse

The FESEM analysis of untreated, pretreated and dual pretreated samples of sugarcane bagasse from stage 1 and stage 2 was performed. The untreated and pretreated biomasses were placed over a thin aluminium sheet and the samples was incubated at 50°C for 24h. A double coating of gold film was performed over samples using Polaron Sputter Coater, SC7620 “Mini” (Quorum technologies, England). Each sample was then examined under FESEM (carl Zeiss, SIGMA VP instrument, Germany). The FESEM images were obtained to determine the variation in the structural topology of untreated and pretreated sugarcane bagasse.

5.3 Results and Discussion

5.3.1 Reducing sugar analysis of hydrolysates from pretreatment of sugarcane bagasse

In single stage alkali pretreatment sugarcane bagasse by 1% (w/v) NaOH for 1h, 2h, 3h and 4h gave a total reducing sugar (TRS) loss of 20.3, 13.5, 9.4 and 5.9 mg/g, of raw biomass respectively in the pretreated hydrolysate (Table 5.1). The loss of TRS in the hydrolysate after pretreeatment was because the sodium hydroxide disrupts the scb

cell wall by solubilizing the hemicellulose and lignin. Sodium hydroxide mechanistically cleaves the alpha-aryl ester bonds from its polyphenolic monomers along with the weakening of hydrogen bonds, which in turn promotes the swelling of cellulose was mentioned earlier by Rocha, *et al.,* 2012. The TRS loss from the hydrolysate of alkali pretreated biomass from different time intervals 1h, 2h, 3h and 4h followed by organosolv pretreatment using phosphoric acid-acetone method gave comparatively lesser TRS loss of 0.6, 1.2, 1.04 and 0.8 respectively (Table 5.1). Also, since the TRS loss in the stage 2 pretreated hydrolysate was less the hexose and pentose sugars were not detected in the HPLC. The less TRS loss in the hydrolysate during phosphoric acid/acetone pretreatment method was because the phosphoric acid pretreatment can reduce the crystallinity of cellulose and avoid its further degradation into water-soluble sugar (Zhang *et al.,* 2006) and the acetone further helps in the solubilisation of lignin (Zhang *et al.,* 2006). As a result, high purity of amorphous cellulosic residues is obtained separated from hemicellulose, which significantly facilitates the enzymatic hydrolysis (Zhang *et al.,* 2006).

5.3.2 Saccharification of pretreated sugarcane bagasse by Chimera (*Ct*GH1-L1-*Ct*GH5- F194A) and Cellobiohydrolase (*Ct*CBH5A)

The enzymatic hydrolysates of the single organosolv pretreated and all the 4 dual pretreated sugarcane bagasse obtained after saccharification with 2:3 loading of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and Cellobiohydrolase (*Ct*CBH5A) at 30°C for 46h, 72h and 96h, were analysed for reducing sugar (Table 5.2). The maximum TRS yield obtained was 230 mg/g of pretreated biomass at 96h for dual stage pretreated i.e alkali treated for 2h using 1% (w/v) NaOH followed by organosolv pretreatment (Table

5.2). The maximum TRS yield obtained was179 mg/g of pretreated biomass at 96h for dual stage pretreated i.e alkali treated for 1h using 1% (w/v) NaOH followed by organosolv pretreatment (Table 5.2). The maximum TRS yield were 96 mg/g and 71 mg/g of pretreated biomass at 96h (Table 5.2) for alkali treated for 3h and 4h followed by organosolv pretreatment respectively. Also, the TRS yield obtained from single organosolv pretreatment was 112.5 mg/g of pretreated biomass (Table 5.2). The highest TRS yield was reported for 2h alkaline pretreatment followed by organosolv pretreatment (230 mg/g of pretreated biomass). From this result it can be concluded that the 2h alkali pretreatment followed by organosolv method can give more accessibility to the cellulose for the cellulase enzyme than the other pretreatments. This could be because in 2h alkali pretreatment the cellulose content was higher and delignification could have been occurring more significant than the other pretreatment condition. The similar observation was made by Maryana *et al.*, 2014 on the alkaline pretreated sugarcane bagasse where delignification and cellulose contents were significant when alkaline pretreatment was conducted with 2N NaOH for 40 min.

5.3.3 Fermentable hexose sugar analysis in enzymatic hydrolysates from saccharification of pretreated sugarcane bagasse

The enzymatic hydrolysates of all 5 pretreated sugarcane bagasse obtained after the saccharification with Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and CBH (*Ct*CBH5A) at 30°C for 48h, 72h and 96h, were analysed for hexose sugars by HPLC. The dual pretreated sugarcane bagasse using phosphoric acid-acetone method resulted in maximum yield of 137 mg/g of glucose from the pretreated biomass which was 1.9-fold higher as compared to single organosolv pretreated sugarcane bagasse (71.5 mg/g of

pretreated biomass) (Table 5.2). This result showed that the effect of NaOH efficiently removes lignin while from the organosolv pretreatment the cellulose obtained is nearly pure and amorphous which aids the cellulase enzymes to act efficiently on the degradation of cellulose into hexose sugar (Zhang *et al.*, 2006).

5.1 Reducing sugar analysis of hydrolysates from pretreatment of sugarcane bagasse

Stage 1 pretreatment NaOH soaking	TRS yield [a *] (mg/g raw biomass)	Glucose [b*] (mg/g)	Xylose [b*] (mg/g)	Cellobiose [b*] (mg/g)	TRS yield [b*] (mg/g)	Stage 2 pretreatment Organosolv pretreatment	TRS yield[a*] (mg/g biomass)
1h	20.3± 0.2	3.7±0.01	3.2±0.12	3.1±0.04	10	Simple organosolv	1.5±0.02
2h	13.5±0.2	2.6±0.4	2.5±0.2	2.1±0.06	7.2	1h NaOH soaking+ organosolv	0.6±0.01
3h	9.4±0.1	1.3±0.04	1.3±0.02	1.8±0.02	4.5	2h NaOH soaking+ organosolv	1.2±0.02
4h	5.9±0.3	0.5±0.01	0.6±0.01	1.7±0.03	3.0	3h NaOH soaking+ organosolv	1.04±0.01
						4h NaOH soaking+ organosolv	0.8±0.02

TRS- Total reducing sugar *Mean ± SE (n=2)
TRS yield [a*] estimated by Nelson and Somogyi method; TRS yield [b*] estimated by sum of hexose and pentose sugars (HPLC analysis) *Mean ± SE (n=2)

Table 5.2. Quantitative analysis of reducing sugars released from pretreatment of scb biomass after enzymatic saccharification.

Pretreatment	Time of saccharification	TRS yield[a*] (mg/g of pretreated biomass)	Glucose[b*] (mg/g)
Simple organosolv	48h	40.8±1.1	15.5±0.02
	72h	104.4±0.7	53.5±0.5
	96h	112.5±0.2	71.5±0.3
1h NaOH soaking+ organosolv	48h	138.8±1.6	52.5±1.8
	72h	170.9±0.5	68.8±0.6
	96h	179.6±1.8	100.5±0.6
2h NaOH soaking+ organosolv	48h	75.5±3.5	40.8±0.1
	72h	159.8±1.6	66.4±0.3
	96h	230.1±2.1	137.4±0.6
3h NaOH soaking+ organosolv	48h	95.5±1.2	31.5±0.4
	72h	98.8±0.7	57.3±0.9
	96h	95.8±1.1	48.8±0.4
4h NaOH soaking+ organosolv	48h	39.9±0.5	15.5±0.4
	72h	58.2±1.2	27.8±0.2
	96h	70.7±0.2	37.8±0.5

TRS- Total Reducing Sugar; [a*] TRS yield estimated by Nelson and Somogyi method.
[1]Mean ± SE (n=2); [b*]Glucose estimation by HPLC

5.3.4 FESEM analysis of untreated and pretreated sugarcane bagasse

FESEM analysis of untreated, single and dual stage pretreated sugarcane bagasse showed the morphological changes in the pretreated sugarcane bagasse after the pretreatment (Fig. 5.1). The rigid and non-porous surface of raw biomass was due to the intact structural integrity of cellulose, hemicelluloses and lignin (Fig. 5.1A). Alkali treated using 1% (w/v) NaOH for 1h and 2h pretreated sugarcane bagasse showed visible voids (Fig. 5.1 B and C). These changes in the surface morphology of biomass can be attributed to the delignification of biomass. Sant'Anna *et al.,* 2012 showed that alkali pretreatment of bagasse form pores on the surface of biomass mainly because of the lignin removal due to which cellulase accessibility for the alkali pretreated substrate gets increased. The simple Organosolv pretreated sugarcane bagasse by phosphoric acid-acetone pretreatment method (Fig. 5.1D) and Dual pretreated sugarcane bagasse from Alkali using 1%(w/v) NaOH for 1h and 2h followed by organosolv pretreatment (Fig. 5.1E and F) showed rough surfaces. This indicated that the pretreatment removed external fibers which in turn increased the surface area of pretreated biomass in which the cellulase enzymes are more accessible to cellulose fibers. Similar structural changes were earlier reported for rice straw pretreated with electron beam irradiation (Bak *et al.*, 2009) and for rice straw pretreated with aqueous ammonia pretreatment (Ko *et al.*. 2009). In rice straw pretreated by acetone in presence of 0.6% H_2SO_4 as catalyst became rough and porous with a number of visible fibrous structures (Ko *et al.*, 2009).

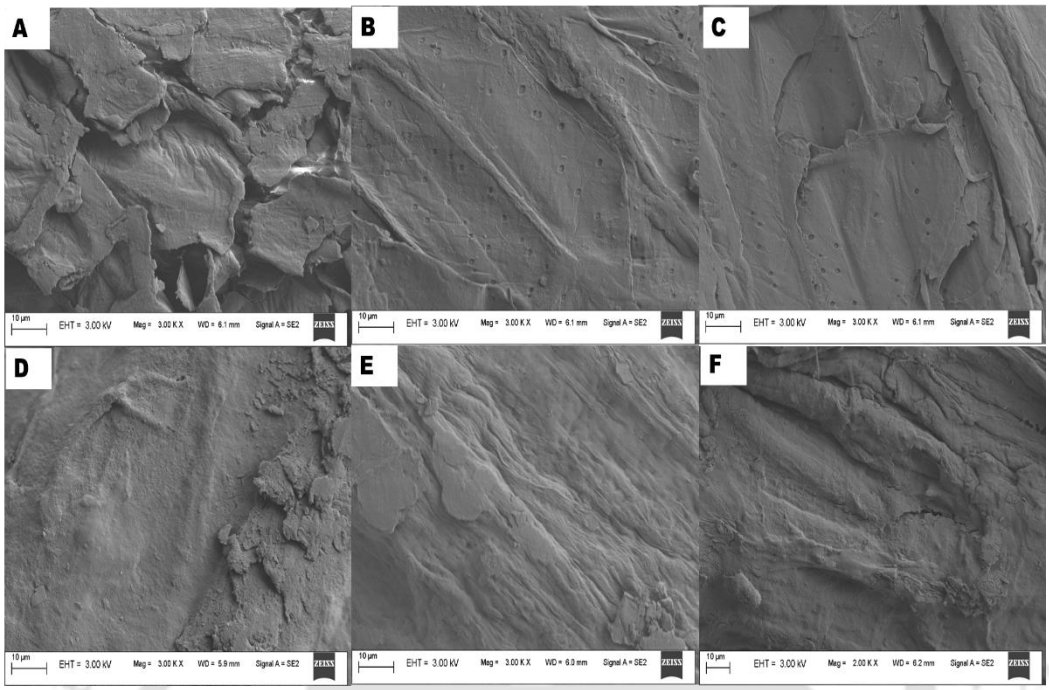


Fig. 5.1 FESEM examination of sugarcane bagasse biomass A) Untreated biomass, B) Alkali pretreated biomass using 1% (w/v) NaOH for 1h at 50°C, C) Alkali pretreated biomass using 1% (w/v) NaOH for 2h at 50°C, D) Organosolv pretreated biomass using 85% (v/v) phosphoric acid-acetone method for 1h at 50°C, E) Dual pretreated using alkali pretreated biomass using 1% (w/v) NaOH for 1h + organosolv pretreatment using 85% (v/v) phosphoric acid-acetone method at 50°C and F) Dual pretreated using alkali pretreated biomass using 1% (w/v) NaOH for 2h + organosolv pretreatment using 85% (v/v) phosphoric acid-acetone method at 50°C.

5.3.5 FT-IR analysis of untreated and pretreated sugarcane bagasse

The chemical shifts differences in the single alkali pretreated and dual pretreated sugarcane bagasse were determined by FT-IR. The FT-IR spectrum showed reduction of peak at wavenumber 1265 cm^{-1} and 1749 cm^{-1} in single alkali pretreated samples (Fig. 5.2). The presence of lignin and aromatic lignin ring in biomass was analysed by the peak at wavenumber 1215 cm^{-1}, 1595 cm^{-1} and 1770 cm^{-1} as also reported earlier (Xu *et al.*, 2013, Shuaizhu *et al.*, 2016). This analysis showed that the single, 1% (w/v) NaOH pretreatment for 2h, efficiently removes lignin from the sugarcane bagasse

biomass. Further, FT-IR analysis of the dual pretreated sugarcane bagasse (alkali pretreated using 1% (w/v) NaOH for 2h at 50°C followed by organosolv pretreatment using phosphoric acid-acetone method) showed the changes in the cellulose structure. It showed increased OH stretching of hydrogen bond between 3405 cm^{-1} and 3420 cm^{-1} (Fig. 5.2) This is predicted to be formed by glycosidic bond in the cellulose and various other groups of lignin. The bands near 1062 cm^{-1} and 1171 cm^{-1} (Fig. 5.2) were associated with the structural feature of cellulose and hemicellulose respectively, as reported earlier (Liu *et al.,* 2009). After the phosphoric acid-acetone pretreatment on 2h alkali pretreated sugarcane bagasse the intensity of peak at1062 cm^{-1} and 1171 cm^{-1} was increased (Fig. 5.2). This suggested that the cellulose content in the solid residue during the dual pretreated sample has increased, since hemicellulose fraction is removed by the phosphoric acid pretreatment. Similar, observation was made in corn stover pretreated biomass by FeCl3 (Liu *et al.,* 2009). The higher stretching of OH bond near 3420 $cm^{-1}$and CH_2 bond near 2913 cm^{-1} (Fig. 5.2) are the distinguished features of cellulose as mentioned earlier by Gupta *et al.,* 2017. This suggested that the intramolecular hydrogen bond is formed when the organosolvent acetone has penetrated the crystalline lattices of the cellulose. A similar observation was reported during phosphoric acid-acetone pretreatment of sugarcane leaf top and Kans grass by Gupta *et al.,* 2017.

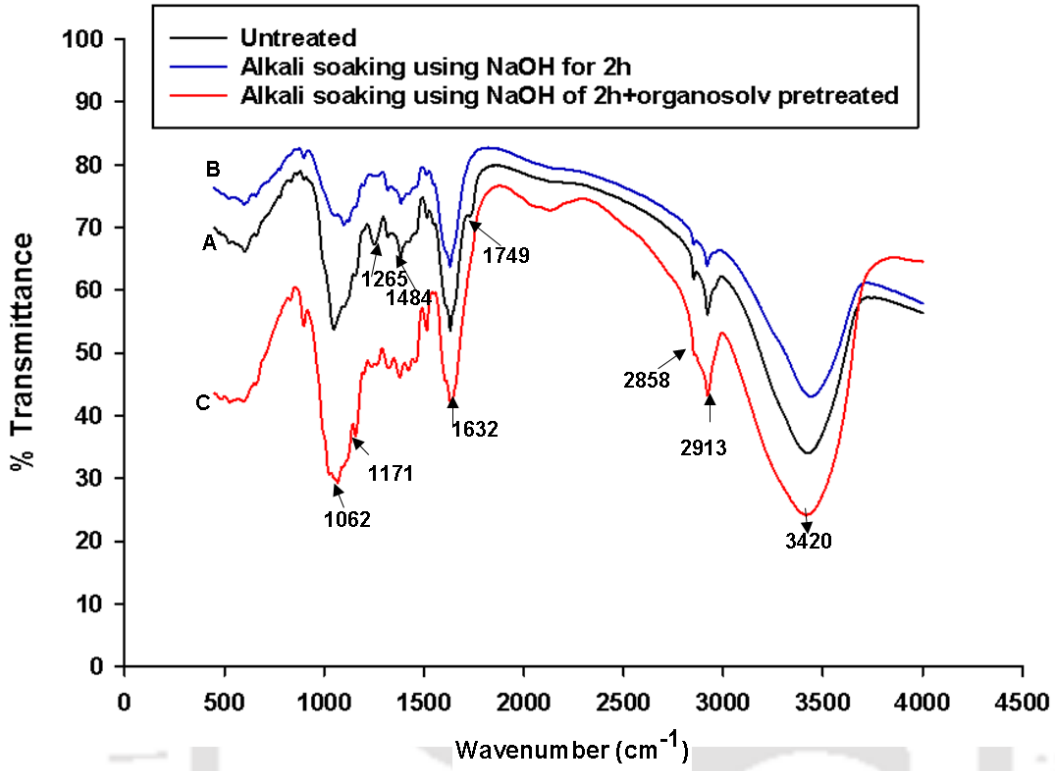


Fig. 5.2 FT-IR spectra of sugarcane bagasse A) Untreated, B) Alkali treated using 1% (w/v) NaOH for 2h at 50°C and C) Alkali treated using 1%(w/v) NaOH for 2h at 50°C + Organosolv pretreatment using 85% (v/v) phosphoric acid-acetone method.

5.4 Conclusion

In the study, the optimization of pretreatment of sugarcane bagasse using alkali followed by organosolv pretreatment method for efficient enzymatic saccharification in context to get maximum TRS yield was carried out. In single stage alkali pretreatment sugarcane bagasse by 1% (w/v) NaOH for 1h, 2h, 3h and 4h gave a TRS loss of 20.3, 13.5, 9.4 and 5.9 mg/g, of raw biomass respectively, in the pretreated hydrolysate The loss of TRS was because the sodium hydroxide disrupts the scb cell wall by solubilizing the hemicellulose and lignin. The dual stage pretreatments of sugarcane bagasse by alkali pretreatment followed by organosolv pretreatment method resulted in a minimum loss in TRS yield in the pretreated hydrolysate. The less TRS loss in the hydrolysate during phosphoric acid/acetone pretreatment method was because the phosphoric acid pretreatment can reduce the crystallinity of cellulose and avoid its further degradation into water-soluble sugar. The enzymatic saccharification using Chimera and Cellobiohydrolase of all 5 pretreated sugarcane bagasse which were single stage organosolv pretreted and dual stage pretreted by alkali pretreatment for different time interval of 1h, 2h, 3h and 4h followed by organosolv pretreatment were analysed for hexose sugars by HPLC. The dual pretreated sugarcane bagasse from alkali pretreated using 1% (w/v) NaOH at 50°C for 2h followed by organosolv pretreatment using 85% (v/v) phosphoric acid-acetone method resulted maximum yield of 137 mg/g of glucose from the pretreated biomass which was 1.9-fold higher as compared to single organosolv pretreated sugarcane bagasse (71.5 mg/g of pretreated biomass). The FESEM analysis showed Alkali using 1% (w/v) NaOH for 1h and 2h pretreated sugarcane bagasse showed visible voids. These changes in the surface morphology of biomass can be attributed because of the delignification of biomass. The single

organosolv pretreated sugarcane bagasse by phosphoric acid-acetone pretreatment method and dual pretreated sugarcane bagasse from Alkali using 1%(w/v) NaOH for 1h and 2h followed by organosolv pretreatment showed rough surfaces. This indicates the pretreatment has removed external fibers which in turn increased the surface area of pretreated biomass in which the cellulase enzymes are more accessible to cellulose fibers. The FT-IR analysis showed that the phosphoric acid-acetone pretreatment on 2h alkali soaked pretreated sugarcane bagasse the intensity of peak in 1062 cm^{-1} and 1171 cm^{-1} has increased. This suggested that the cellulose content in the solid residue during the dual pretreated sample has increased since hemicellulose fraction was removed by the phosphoric acid pretreatment. This result showed that alkaline NaOH soaking followed by organosolv pretreatment using phosphoric acid-acetone method is an efficient pretreatment strategy than the single organosolv pretreatment. This dual stage pretreatment strategy aids the cocktail cellulase enzymes comprising of Chimera and CBH5A to act efficiently on the degradation of cellulose into glucose which can be further utilized for bioethanol production.

References

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

Chakar, F. S., & Ragauskas, A. J. (2004). Review of current and future softwood kraft lignin process chemistry. *Industrial Crops and Products*, *20*, 131-141.

Ennaert, T., de Beeck, B. O., Vanneste, J., Smit, A. T., Huijgen, W. J., Vanhulsel, A., & Sels, B. F. (2016). The importance of pretreatment and feedstock purity in the reductive splitting of (ligno) cellulose by metal supported USY zeolite. *Green Chemistry*, *18*, 2095-2105.

Gupta, A., Rajulapati, V., Das, D., & Goyal, A. (2017). Comparative analysis of bioethanol production involving saccharification by mixed recombinant clostridial enzymes using sugarcane leaves and kans grass as sustainable feed stocks from North-east India. *Indian Journal of Biotechnology*,*16*, 199-210.

Jamaldheen, S.B, Sharma, K., Rani, A., Moholkar, V.S., & Goyal, A. (2018) Comparative analysis of pretreatment methods on sorghum (*Sorghum durra*) stalk agrowaste for holocellulose content. *Preparative Biochemistry* and *Biotechnology*. *48*, 457-464.

Ko, J. K., Bak, J. S., Jung, M. W., Lee, H. J., Choi, I. G., Kim, T. H., & Kim, K. H. (2009). Ethanol production from rice straw using optimized aqueous-ammonia soaking pretreatment and simultaneous saccharification and fermentation processes. *Bioresource Technology*, *100*, 4374-4380.

Lee, H. L., Chang, C. K., Teng, K. H., & Liang, P. H. (2011). Construction and characterization of different fusion proteins between cellulases and β-

glucosidase to improve glucose production and thermostability. *Bioresource Technology*, *102*, 3973-3976.

Liu, L., Sun, J., Li, M., Wang, S., Pei, H., & Zhang, J. (2009). Enhanced enzymatic hydrolysis and structural features of corn stover by FeCl3 pretreatment. *Bioresource Technology*, *100*, 5853-5858.

Maryana, R., Marifatun, D., Wheni, A. I., Satriyo, K. W., & Rizal, W. A. (2014). Alkaline pretreatment on sugarcane bagasse for bioethanol production. *Energy Procedia*, *47*, 250-254.

McKee, L. S., Peña, M. J., Rogowski, A., Jackson, A., Lewis, R. J., York, W. S., & Marles-Wright, J. (2012). Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proceedings of the National Academy of Sciences USA*, *109*, 6537-6542.

Nath, P., Dhillon, A., Kumar, K., Sharma, K., Jamaldheen, S. B., Moholkar, V. S., & Goyal, A. (2019). Development of bi-functional chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A) from endoglucanase (*Ct*GH5) mutant F194A and β-1, 4-glucosidase (*Ct*GH1) from *Clostridium thermocellum* with enhanced activity and structural integrity. *Bioresource Technology*, *282*, 494-501.

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

Nixon, A. E., Ostermeier, M., & Benkovic, S. J. (1998). Hybrid enzymes: manipulating enzyme design. *Trends in Biotechnology*, *16*, 258-264.

Pandey, A., Soccol, C. R., Nigam, P., & Soccol, V. T. (2000). Biotechnological potential of agro-industrial residues. I: sugarcane bagasse. *Bioresource Technology*, *74*, 69-80.

Rocha, G. J., Martín, C., da Silva, V. F., Gómez, E. O., & Gonçalves, A. R. (2012). Mass balance of pilot-scale pretreatment of sugarcane bagasse by steam explosion followed by alkaline delignification. *Bioresource Technology*, *111*, 447-452.

Sant'Anna, C., & Souza, W. (2012) Microscopy as a tool to follow deconstruction of lignocellulosic biomass. Current Microscopy Contributions to Advances in Science and Technology. 639-645.

Shuaizhu, Bi, Lincai, P., Keli, C., & Zhengliang, Z. (2016). Enhanced enzymatic saccharification of sugarcane bagasse pretreated by combining O2 and NaOH. *Bioresource Technology 214*, 692-699.

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

Xu, F., Yu, J., Tesso, T., Dowell, F., & Wang, D. (2013). Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: a mini-review. *Applied Energy*, *104*, 801-809.

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

Zhao, X., Cheng, K., & Liu, D. (2009). Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Applied Microbiology and Biotechnology*, *82*, 815.

Zhu, L., O'Dwyer, J. P., Chang, V. S., Granda, C. B., & Holtzapple, M. T. (2010). Multiple linear regression model for predicting biomass digestibility from structural features. *Bioresource Technology*, *101*, 4971-4979.

FUTURE PROSPECTS

The Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) developed by fusion of β-glucosidase (*Ct*GH1) and endoglucanase *Ct*GH5-F194A from *Clostridium thermocellum* showed improved bi-functional activity. Thus, the Chimera can be used as a single enzyme for hydrolysis of lignocellulosic biomass into glucose. The enzymatic hydrolysis using Chimera can reduce the cost and time for production of enzymes which are required for the complete conversion of cellulosic component of the biomass into glucose. The Chimera was also found thermostable therefore, it can be used in industrial scale for bioethanol, biobutanol or lactic acid production.

Journal publications

From Thesis:

1. **Nath, P.,** Dhillon, A., Kumar, K., Sharma, K., Jamaldheen, S. B., Moholkar, V. S., & Goyal, A. (2019). Development of bi-functional chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A) from endoglucanase (*Ct*GH5) mutant F194A and β-1, 4-glucosidase (*Ct*GH1) from *Clostridium thermocellum* with enhanced activity and structural integrity. *Bioresource Technology*, *282*, 494-501.
2. **Nath, P.,** K., Sharma, & Goyal A. (2019). Combined SAXS and computational approaches for structure determination and binding characteristics of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) generated by assembling β-glucosidase (*Ct*GH1) and a mutant endoglucanase (*Ct*GH5-F194A) from *Clostridium thermocellum*. *International Journal of Biological Macromolecules*, *148*, 364-377.
3. Nath[¥], P., Maibam[¥], P. D., Singh, S., Rajulapati, V., & Goyal, A. (2020). Comparative pretreatment using mild alkali and organosolv method for improving enzymatic digestibility of sugarcane bagasse by recombinant Chimera and Cellobiohydrolase for bioethanol production (submitted).
4. **Other Publications:**

1. Kumar, K., **Nath, P**., and Goyal, A. (2018). Structural characterization of an endo β-1, 3-glucanase of family 81 glycoside hydrolase (CtLam81A) from *Clostridium thermocellum*. *Journal of Proteins & Proteomics*, *9*(3).
2. Mohanapriya.N., Singh, S., Jamaldheen, S. B., **Nath, P.,** Moholkar, V. S., & Goyal, A. (2020)**.** Assessment of combination of pretreatment of *Sorghum durra* stalk and production of chimeric enzyme (β-glucosidase and endo β-1,4 glucanase, *Ct*GH1-L1-*Ct*GH5-F194A) and cellobiohydrolase (*Ct*CBH5A) for saccharification to produce bioethanol. Preparative biochemistry & biotech (in press)

Conferences

1. Presented a poster entitled "Construction and characterization of chimeric enzyme developed by fusing β-glucosidase (*Ct*GH1) and endoglucanase (*Ct*GH5-F194A) both from *Clostridium thermocellum* for enhanced catalytic efficiency and thermostability" in "International Conference on Biotechnological Research and Innovation for Sustainable Development" 15th BRSI convention. CSIR- Indian Institute of Chemical Technology (CSIR-IICT), Nov. 22-25, 2018, Hyderabad, India.
2. Presented a poster entitled "Protein engineering of endo β-1-4 glucanase (*Ct*GH5) from *Clostridium thermocellum* by site-directed mutagenesis for development of mutant with enhanced activity" in "Bioprocessing India, Recent Trends in Bioprocessing for Healthcare, Energy and Environment" Dec 9-11, 2017, IIT Guwahati, Assam India.
3. Presented a poster entitled "Identification of promising functional residues capable of introducing endo-xylanase activity into an exo-acting arabinofuranosidase (*Ct*43Ara*f*) with enhanced activity: An *in silico* approach" in "56th International Annual Conference of Association of Microbiologists of India (AMI)", December 7-10, 2015, Jawaher Lal Nehru University, New Delhi.

Awards

1. Best poster award on the work entitled "Identification of promising functional residues capable of introducing endo-xylanase activity into an exo-acting arabinofuranosidase (*Ct*43Ara*f*) with enhanced activity: An *in silico* approach*"* in "56th International Annual Conference of Association of Microbiologists of India (AMI)" organized by Jawaher Lal Nehru University, New Delhi during December 7-10, 2015.

VITAE

The author was born on December 20, 1989 in Itanagar (Arunachal Pradesh). She passed the Secondary Examination (10th Class) in 2006 from Central board of secondary education and *Higher Secondary Examination (12th Class) in 2008 under the Assam Higher Secondary Education Council. She completed Integrated MSc. (Bioscience and Bioinformatics) from Tezpur Central University in July, 2014.*

Miss Priyanka Nath joined the Ph.D. program in December 2014 at Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781 039, Assam, India. She successfully completed the course work with 7.14/10 CPI. She received DST Inspire Fellowship from Department of Science and technology, India from April 2015. She delivered the open (PhD Synopsis) Seminar on October 24, 2019 and presented her thesis work before the Doctoral Committee and her performance was satisfactory.

Bioresource Technology 282 (2019) 494–501

Contents lists available at ScienceDirect

Bioresource Technology

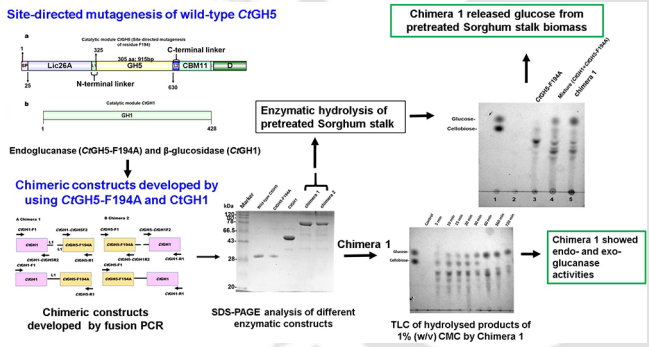
journal homepage: www.elsevier.com/locate/biortech

Development of bi-functional chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A) from endoglucanase (*Ct*GH5) mutant F194A and β-1,4-glucosidase (*Ct*GH1) from *Clostridium thermocellum* with enhanced activity and structural integrity

Priyanka Nath[a,b], Arun Dhillon[a], Krishan Kumar[a], Kedar Sharma[a], Sumitha Banu Jamaldheen[a,c], Vijayanand Suryakant Moholkar[c,d], Arun Goyal[a,b,c,*]

[a] Carbohydrate Enzyme Biotechnology Laboratory, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India
[b] DBT PAN-IIT Center for Bioenergy, Indian Institute of Technology Guwahati, India
[c] Centre for Energy, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India
[d] Department of Chemical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Endoglucanase
β-Glucosidase
Chimera
Site-directed mutagenesis
Bioethanol

ABSTRACT

Site-directed mutagenesis of β-1,4-endoglucanase from family 5 glycoside hydrolase (*Ct*GH5) from *Clostridium thermocellum* was performed to develop a mutant *Ct*GH5-F194A that gave 40 U/mg specific activity against carboxymethyl cellulose, resulting 2-fold higher activity than wild-type *Ct*GH5. *Ct*GH5-F194A was fused with a β-1,4-glucosidase, *Ct*GH1 from *Clostridium thermocellum* to develop a chimeric enzyme. The chimera (*Ct*GH1-L1-*Ct*GH5-F194A) expressed as a soluble protein using *E. coli* BL-21cells displaying 3- to 5-fold higher catalytic efficiency for endoglucanase and β-glucosidase activities. TLC analysis of hydrolysed product of CMC by chimera 1 revealed glucose as final product confirming both β-1,4-endoglucanase and β-1,4-glucosidase activities, while the products of *Ct*GH5-F194A were cellobiose and cello-oligosaccharides. Protein melting studies of *Ct*GH5-F194A showed melting temperature (T_m), 68 °C and of *Ct*GH1, 79 °C, whereas, chimera showed 78 °C. The im-

* Corresponding author at: Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India.
E-mail address: arungoyl@iitg.ac.in (A. Goyal).

https://doi.org/10.1016/j.biortech.2019.03.051
Received 11 January 2019; Received in revised form 8 March 2019; Accepted 9 March 2019
Available online 12 March 2019
0960-8524/
TH-2322_146106042

proved structural integrity, thermostability and enhanced bi-functional enzyme activities of chimera makes it potentially useful for industrial application in converting biomass to glucose and thus bioethanol.

1. Introduction

The risk of depletion of fossil fuels as an energy source has led to the development of alternative strategies to overcome this challenge. Production of bio-ethanol from cheap lignocellulosic biomass is one of the most promising approaches to address this problem. Plant cell walls contain cellulose, hemicellulose and lignin which can be converted into alcohol, thus plant biomass can serve as alternative energy source. Cellulose is the most abundant polymer (polysaccharide) and for its complete break down three enzymes are required i.e. endoglucanase, cellobiohydrolase and β-glucosidase (Beguin, 1990; Yennamalli et al., 2013). Endo-β-1,4-glucanase hydrolyses cellulose backbone randomly and produces cellodextrin as final product (Urbanowicz et al., 2007). Cellobiohydrolase binds the end of the cellulose chain and releases cellobiose as the major product (Barr et al., 1996), while β-glucosidase cleaves cellobiose to form two molecules of glucose (Grabnitz et al., 1991).

For efficient hydrolysis of cellulosic biomass, it is important to engineer the enzymes with improved catalytic efficiency. Moreover, the requirement of multiple enzymes for complete degradation of lignocellulosic biomass is a cost intensive process (McKee et al., 2012). Therefore, to reduce the cost of production of multiple enzymes, the development of multifunctional chimera is beneficial that can catalyze two or more reactions for efficient hydrolysis of biomass (Lu and Feng, 2008, Fan, et al., 2009). Efforts were made to develop chimeric cellulases by attaching a Carbohydrate Binding Module (CBM3) to a family 9 endoglucanase, Cel9A (Telke et al., 2012) and to a family 5 endoglucanase, Cel5H (Shi et al., 2013) resulting in 8- to 12-fold higher catalytic efficiency. Telke et al., 2012 also proved from the computational protein modelling that the active site configuration of native catalytic module does not get affected upon fusing the two modules. Therefore, other catalytic modules with different substrate specificities and mode of actions can be introduced in a single polypeptide chain for development of chimeric cellulases with increased catalytic efficiency. Adlakha et al., 2012 showed the development of new endoglucanase (Endo5A) and β-glucosidase (Gluc1C) from *Paenibacillus* sp. having higher catalytic efficiencies. They observed that Gluc1C improves the efficiency of Endo5A and releases the reducing sugar from carboxymethyl cellulose. Therefore, the fusion of Gluc1C with Endo5A resulted in 3.3- and 2- fold higher activity towards β-glucosidase and endoglucanase, respectively.

The family 5 glycoside hydrolase, GH5 is one of the largest and well characterized families (http://www.cazy.org/GH5.html). One of the members of family 5 GH, Cel5E from *C. thermocellum,* is a part of cellulosomal *celH,* comprising two N-terminal catalytic modules (Lic26A and Cel5E) and a C-terminal family 11 carbohydrate binding module (CBM11) and two Type I dockerins at C-terminal connected by C-terminal linker of CBM11 (Taylor et al., 2005). This family 5 GH (Cel5E) was shown to hydrolyse both soluble and insoluble cellulolosic substrates (Bharali et al., 2005). The enzyme activity of the family 5 GH, (Cel5E) endoglucanase (Gen Bank Acc No. ABN52701.1) was enhanced by attaching the linker at both N- and C-terminal of GH5 and by mutating Phe267 to Ala (Yuan et al., 2015). The TLC analysis of Cel5E hydrolysed carboxymethyl cellulose (CMC) showed the production of cellodextrin as the final product (Yuan et al., 2015). β-Glucosidase from family 1 GH was found to hydrolyse cellodextrin up to a five glucose chain length (Adlakha et al., 2012). The fusion of catalytic domains of family 5 GH and family 1 GH will produce bifunctional enzyme that could alone convert cellulolosic biomass to glucose as a major product. In present study, the catalytic efficiency of endoglucanase (*Ct*GH5) (ABN52701.1) of family 5 glycoside hydrolase without the N- or C-terminal linker from *Clostridium thermocellum* was enhanced by site-directed mutagenesis of amino acid residue Phe194 of *Ct*GH5 following the method as mentioned by Yuan et al.,2015. Further, mutant F194A was fused to β-1,4-glucosidase (*Ct*GH1) of family 1 glucoside hydrolase (CAA42814.1) from the same species *Clostridium thermocellum* for development of bifunctional enzyme.

2. Materials and methods

2.1. Site-directed mutagenesis of wild-type CtGH5

The residue Phe194 of *Ct*GH5 was mutated to alanine as mentioned by Yuan et al., 2015 by site-directed mutagenesis using site-directed mutagenesis kit (Agilent Technologies India Pvt. Ltd.) and the mutant (*Ct*GH5-F194A) was developed. The wild-type plasmid *Ct*GH5-pET21a (+) was used as the template and the mutagenic forward and reverse primers were used for PCR. The details of primers are provided in supplementary material. The generated plasmid was sequenced to confirm the mutation at the desired location. This plasmid was transformed in *E. coli* BL21 cells for expression of *Ct*GH5-F194A following the method as described previously (Bharali et al., 2005, Ahmed et al., 2009).

2.2. Expression and purification of recombinant proteins

The recombinant genes were expressed using *E. coli* BL21 cells as described earlier (Taylor et al., 2005). The cells were cultured in the 100 ml LB medium supplemented with 100 μg/ml of ampicillin for wild-type *Ct*GH5 and mutant F194A enzymes (Bharali et al., 2005). The enzyme, *Ct*GH1 was expressed using *E. coli* BL21 cells in 100 ml LB medium supplemented with 50 μg/ml of kanamycin as mentioned earlier (Jamaldheen et al., 2018). The chimeric gene was cloned in pET-28a(+) vector, therefore, 50 μg/ml of kanamycin was used for growing chimeric construct containing *E. coli* cells in LB medium at 37 °C and 180 rpm for 4 h. At the mid-exponential phase (A_{550} 0.6) the cells were induced with 1 mM final concentration of isopropyl-1-thio-β-D-galactopyranoside (IPTG) for over-expression of recombinant proteins. After addition of IPTG, the cells of wild-type *Ct*GH5 and mutant F194A were kept at 37 °C and at 180 rpm for 6 h whereas, the cells of chimeric protein were kept at 24 °C and 180 rpm for 16 h. The cells were harvested by centrifugation at 7000*g* and at 4 °C for 10 min. The cell pellet was resuspended in 10 ml 50 mM sodium phosphate buffer, pH 7.0, containing 300 mM NaCl. The resuspended cells were sonicated on ice for 10 min using 10 s off; 10 s on; 33% amplitude followed by centrifugation at15000 × *g* and at 4 °C for 50 min and the cell free extract was separated from the cell pellet and filtered through 0.2 μm membrane. Purification of the recombinant proteins were carried out by immobilized metal ion affinity chromatography (Carvalho et al., 2004). The filtered cell free extract was loaded on the 1 ml sepharose columns (GE Healthcare, HiTrap chelating) which was pre-equilibrated with binding buffer comprising 50 mM sodium phosphate buffer, pH 7.0 containing 50 mM imidazole and 300 mM NaCl (Bharali et al.,2005). The column was washed with 50 ml of binding buffer containing50mM imidazole and 300 mM NaCl. The elution of bound protein fractions was carried by 50 mM sodium phosphate buffer, pH 7.0 containing 500 mM imidazole and 300 mM NaCl linear gradient. The fractions were desalted and was analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS–PAGE) using 12% (w/v) gels. The protein concentrations were estimated by Bradford method using BSA

as standard (Bradford, 1976).

2.3. Enzyme assays

The enzyme assays for wild-type *Ct*GH5 and mutant F194A were performed with 20 mM citrate phosphate buffer, pH 4.2 and 5.5, respectively. For enzymatic assay of chimera 1, *Ct*GH1-L1-*Ct*GH5-F194A, 20 mM citrate phosphate buffer (pH 5.0) was used. 100 µl reaction mixture contained 1.0%, w/v carboxymethylcellulose as substrate and 10 µl of enzyme (wild-type *Ct*GH5, 0.1 mg/ml), (mutant F194A, 0.05 mg/ml) and (*Ct*GH1-L-*Ct*GH5-F194A, 0.05 mg/ml) and incubated at 60 °C for wild-type *Ct*GH5 and chimera 1 and at 50 °C for the mutant F194A for 4 min. The enzyme activity was calculated by measuring the reducing sugar concentration by method as described earlier (Nelson, 1944; Somogyi, 1945). The concentration of the reducing sugar was estimated by using glucose as standard. One unit of activity was defined as the amount of enzyme which produced 1 µmole of glucose per min. For the assay of β-glucosidase activity of GH1, the substrate *p*-Nitrophenyl β-D-glucopyranoside (*p*NPG) was used. The enzymatic reaction was performed in a 1 ml reaction mixture in 20 mM citrate phosphate buffer, pH 5.5 containing 1 mM of the substrate and 20 µl *Ct*GH1-L1-*Ct*GH5-F194A or *Ct*GH1 (0.05 mg/ml). The reaction was performed at 70 °C for 5 min for chimera 1 and 65 °C for *Ct*GH1 for 5 min on a UV visible spectrophotometer (Varian, Cary 100 Bio) with a peltier temperature controller. The reactions were performed for 5 min and continuous absorbance at 405 nm was monitored. The quantification was done by the release of *p*-nitrophenol using 24150 M^{-1} cm^{-1} as an extinction coefficient described by Cartmell et al., 2011.

2.4. pH and temperatures effect on different enzyme constructs

The pH profiles of the enzymes were performed under wide range of pH. The buffers 20 mM citrate phosphate, for pH (3–7) and 20 mM sodium phosphate, for pH (6.5–8) were used. The pH profile was generated using substrate 1% (w/v) CMC using temperature 60 °C for wild-type *Ct*GH5 and *Ct*GH1-L1-*Ct*GH5-F194A, 50 °C for mutant F194A. Similarly, the optimum temperatures of enzymes were determined by incubating the enzymes at varying temperatures between 30 and 80 °C, at their respective optimised pH and optimised time period. The pH stability analysis for chimera 1 and its individual enzymes (*Ct*GH5-F194A) for endoglucanase activity was performed by incubating 0.05–0.1 mg/ml of enzyme at different pH, using 50 mM citrate phosphate, pH (3–7) and 50 mM sodium phosphate, ranging from pH (6.5–8.0) for 1 h at 25 °C. An aliquot of 10 µl was taken and the assay for endoglucanase activity was performed at their respective optimum pH and temperature. The thermal stability analysis of chimera 1and its individual enzymes (wild-type *Ct*GH5 and *Ct*GH5-F194A) for endoglucanase activity was performed by incubation of 0.05–0.1 mg/ml of enzyme at varying temperature from 20 to 80 °C, for 60 min and activity was calculated by incubating the wild-type *Ct*GH5 at pH 4.2 and 60 °C, *Ct*GH5-F194A at pH 5.5 and 50 °C and chimera 1 at pH 5.0 and 60 °C for 4 min. For the β-glucosidase activity of chimera 1 the optimum pH was obtained by using *p*NPG (1 mM) as a substrate in 1 ml of reaction mixture under different pH ranges, of 20 mM citrate phosphate, pH (3–7) and sodium phosphate, pH (6.5–8). The temperature optima for β-glucosidase activity of chimera 1 was performed by assaying it by *p*NPG as substrate at different temperature ranging from 30 to 80 °C for 5 min. The pH stability analysis for chimera 1 and its individual enzymes (*Ct*GH1) for β-glucosidase activity was performed by incubating 0.05 mg/ml of enzyme at different pH, using 50 mM citrate phosphate, pH (3–7) and 50 mM sodium phosphate, ranging from pH (6.5–8) for 1 h at 25 °C. An aliquot of 20 µl was taken and the assay was performed in triplicates for chimera 1 at pH 5.5 and 70 °C and for *Ct*GH1 at pH 6.5 and 65 °C for 5 min. The thermal stability analysis of chimera 1 and its individual enzymes (*Ct*GH1) for β-glucosidase activity was performed by incubation of 0.05 mg/ml of enzyme at varying temperature from 20 to 80 °C for 60 min and the assay was performed by incubating the chimera 1 at pH 5.5 and 70 °C for and *Ct*GH1 at pH 6.5 and 65 °C for 5 min.

2.5. Protein melting studies

The protein melting studies for chimera 1 and the individual enzymes, wild-type *Ct*GH5, mutant F194A and *Ct*GH1 were performed at different temperatures (25 °C to 100 °C) by monitoring the change in absorbance at 280 nm using an UV–Visible spectrophotometer (Varian, Cary 100-Bio) coupled with a peltier temperature controller. The concentration of the enzymes used was 0.05 mg/ml for each enzyme. chimera 1 and wild-type *Ct*GH5 was dissolved in 50 mM sodium acetate buffer, pH 5.0 while mutant F194A and *Ct*GH1 were dissolved in 50 mM MES buffer, pH 5.5. The 1 ml protein sample were kept at room temperature for 10 min to reach the equilibrium. A curve was generated by plotting the change in the absorbance at 280 nm due to protein unfolding with respect to the change in temperature.

2.6. Substrate specificities and kinetic parameters for different enzymatic construct

The kinetic parameters of (wild-type *Ct*GH5, mutant F194A and chimera 1) for endoglucanase activity against various substrates i.e. CMC, Lichenan and β-Glucan was determined. The optimum conditions used for wild-type *Ct*GH5 20 mM citrate phosphate buffer, pH 4.2 at 60 °C, for mutant F194A 20 mM citrate phosphate buffer, pH 5.5 at 50 °C and chimera 120 mM citrate phosphate buffer, pH 5.0 at 60 °C. The reaction was performed in 100 µl reaction mixture for 4 min containing varying substrate concentration ranging from 0.01% (w/v) to 2% (w/v) and using enzyme 10 µl (0.005–0.1 mg/ml). The kinetic parameters were determined by putting the initial velocities in Michaelis-Menten equation. The kinetic parameter of (*Ct*GH1and chimera 1) for β-glucosidase activity against substrate *p*NPG was determined by varying its concentration from (0.01 mM to 1 mM). The reactions were performed at their respective optimized conditions for *Ct*GH1, 20 mM citrate phosphate buffer, pH 6.5 at 65 °C and for chimera 1, 20 mM citrate phosphate buffer, pH 5.5 at 70 °C. The reaction was monitored by taking the absorbance at 405 nm (A_{405}) for 5 min. The quantification was done by calculating the concentration of *p*-nitrophenol released using 24150 M^{-1} cm^{-1} as an extinction coefficient as described by Cartmell et al., 2011.

2.7. Enzymatic hydrolysis of pretreated biomass

Enzymatic hydrolysis of 1% (w/v) Sorghum stalk pretreated by 1% NaOH and autoclave (Jamaldheen et al., 2018) was performed in 20 mM citrate phosphate buffer, pH 5.0 for chimera1 and 20 mM citrate phosphate buffer, pH 5.5 for *Ct*GH5-F194A and mixture of *Ct*GH1 + *Ct*GH5-F194A. The 1 ml reaction was carried by maintaining the equimolar concentrations of each enzyme and incubated in a shaker incubator at 50 °C and 180 rpm for 48 h. The reaction was stopped by heating the reaction mixture in a boiling water bath for 5 min. The reaction mixture was then centrifuged at 13,000*g* at 25 °C for 10 min and the supernatant containing the released sugar were separated out in different 1.5 ml microcentrifuge tube. The reducing sugar for each reaction was estimated as described earlier (Nelson, 1944. Somogyi, 1945). The experiments were carried out in triplicate and all data are calculated as average of three independent experiments ± standard error.

2.8. End-product determination by thin layer chromatography

The mode of action of chimera 1 with respect to its individual enzymes was determined by the TLC using wild-type *Ct*GH5, mutant F194A and chimera 1 with substrate 1% (w/v) CMC and 1% (w/v)

cellobiose in 100 µl reaction mixture using 20 mM citrate phosphate buffer under their respective optimum conditions. The enzyme concentration 0.2 µM (was kept constant) for each case. The time dependent hydrolysis of substrate by chimera 1 was performed from 5 min to 12 h using 1% (w/v) CMC under optimum condition of 20 mM citrate phosphate buffer, pH 5.0 at 60 °C using 0.2 µM concentration of enzyme. The TLC analysis of hydrolysed products of pretreated Sorghum stalk biomass as substrate released by chimera 1 and its individual enzymes was also performed.

The enzymatic hydrolysis of 1% (w/v) Sorghum stalk pretreated by 1% NaOH and autoclaving was performed in 20 mM citrate phosphate buffer, pH 5.0 for chimera 1 and 20 mM citrate phosphate buffer, pH 5.5 for *Ct*GH5-F194A and mixture of *Ct*GH1 + *Ct*GH5-F194A. One ml reactions were carried out by maintaining equimolar concentrations of each enzyme and incubating in a shaker incubator at 50 °C and 180 rpm for 48 h. The reactions were stopped by adding two volumes absolute ethanol. The reaction mixtures were centrifuged at 13000*g* for 10 min. The supernatants were separated and dried under oven at 80 °C. The hydrolysed products were analysed by Thin Layer Chromatography (TLC) on silica gel coated aluminium plate (TLC Silica gel 60 F254, Merck India Ltd.). The standard used were glucose and cellobiose at concentrations of 2 mg/ml. The mobile phase used was butanol/acetic acid/MiliQ water in the ratio of 2:1:1, for separation. The visualization of the migrated sugars was achieved by immersing the TLC plate in the solution containing sulphuric acid: methanol 5:95, (v/v); and α-naphthol 0.5% (w/v). The plate was dried under oven at 80 °C for 15 min to visualize the spots of the standard and hydrolysed products.

3. Results and discussion

3.1. Cloning, expression and purification of chimera 1 and its individual enzymes

The wild-type *Ct*GH5 (endoglucanase, mutant F194A, *Ct*GH1 (β-glucosidase) and chimera 1 (*Ct*GH1-L1-*Ct*GH5-F194A were expressed as soluble proteins and purified as shown in Fig. 1. *Ct*GH1 was fused with mutant F194A to create a chimeric enzyme. The mutant F194A was used because it showed enhanced activity as compared with its wild-type enzyme as shown in Table 1. The wild-type *Ct*GH5 belonging to cellulosomal enzyme CelH has N-terminal and C-terminal linkers and the details of the molecular architecture of the *Ct*CelH are reported in supplementary material. Thus these natural linkers were used for fusing the two catalytic modules together using fusion PCR approach. The chimera 1 (2322 bp) and the chimera 2 (2337 bp) were amplified by using fusion PCR approach and the insert in pet28a(+) vector was confirmed by restriction digestion with *Nhe*I (GCTAGC) and *Xho*I (CTCGAG) restriction enzymes and the positive clones were confirmed. The details of the clone confirmation are reported in the supplementary material. The chimera 1expressed as soluble protein and was purified (Fig. 1). The molecular masses of different compared (Fig. 1). The chimera 1 showed a molecular mass of 89 kDa. It showed activity against both CMC and *p*NPG (Table 1), whereas, the chimera 2 did not show any activity against CMC or *p*NPG (data not shown). Therefore, the results showed that by fusing *Ct*GH1 at the N-terminal of mutant F194A was favourable for retaining both β-glucosidase and endoglucanase activities in the chimeric enzyme. Therefore, for further studies only chimera 1 was used.

3.2. Effect of temperature and pH on different enzymatic construct

The chimera 1 (*Ct*GH1-L1-*Ct*GH5-F194A) showed maximum endoglucanase activity using 1% (w/v) CMC at temperature 60 °C (Fig. 2a) and pH 5.0 (Fig. 2b), while the mutant F194A showed maximum activity at 50 °C (Fig. 2a) and pH 5.5 (Fig. 2b). The wild-type *Ct*GH5 enzyme showed maximum activity at 60 °C (Fig. 2a) and pH 4.2 (Fig. 2b). The chimera 1 showed maximum β-glucosidase activity using substrate, *p*NPG at 70 °C (Fig. 3a) and at pH 5.5 (Fig. 3b), while, the individual enzyme *Ct*GH1 showed maximum activity at 65 °C (Fig. 3a) and pH 6.5 (Fig. 3b).

The thermal stability analysis showed that the wild-type *Ct*GH5 is stable up to 60 °C and the individual mutant enzyme *Ct*GH5-F194A is stable up to 55 °C. The thermostability analysis of chimera 1 showed that its endoglucanase activity is stable at 55 °C up to 1 h. This could be because of the side chain flexibility incorporated through mutagenesis of Phe194 by Ala near the substrate binding site of the wild-type *Ct*GH5 which was used for development of the chimeric construct. This has been also reported earlier that the thermostability of an enzyme is related with the conformational flexibility (Anbar et al., 2012). The thermal stability of both, *Ct*GH1 and β-glucosidase domain of chimera 1 was found up to 60 °C for 1 h.

The pH stability analysis of chimera 1 for endoglucanase activity showed that the enzyme is stable between pH 4.0 and 5.0 and retained 80% activity at pH 5.5 (Fig. 4a). The individual mutant F194A was stable between pH 5.0 to 6.0 (Fig. 4a), while its activity declines at higher and lower pHs. For β-glucosidase activity the chimera 1 showed pH stability between pH 5.0 to 5.5 and retained 80% activity at pH 4.0 (Fig. 4b), while the individual enzyme *Ct*GH1 was stable between pH 5.5 to 6.0 (Fig. 4b). The advantageous outcome was that the chimera 1 displayed pH stability between pH 4.0–5.5 for both the enzyme activities. This may prove beneficial in the application as both the activities displayed same pH stability.

3.3. Melting temperature of chimera 1

The melting temperature (T_m) of wild-type *Ct*GH5 was 79 °C and mutant F194A was 68 °C as shown in Fig. 5a. This decrease in T_m of the mutant can be correlated with the results of the decrease in thermostability of the mutant *Ct*GH5-F194A as compared with the wild-type *Ct*GH5 (Section 3.2). The melting temperature for the chimera 1 was 78 °C (Fig. 5b) and of *Ct*GH1 was 79 °C (Fig. 5b). The overall T_m of the chimera 1 was retained at similar to that of GH1. This might be due to the effect of the N-terminal *Ct*GH1 in the chimeric construct which has T_m of 79 °C (Fig. 5b). This positive effect of N-terminal domain exerted on the enzyme has been also described in a previous study (Lee et al., 2011).

3.4. Comparison of kinetic parameters of enzyme constructs

The K_m and k_{cat} of chimera 1 were compared against their individual modules for substrates Lichenan, β-Glucan, CMC and *p*NPG

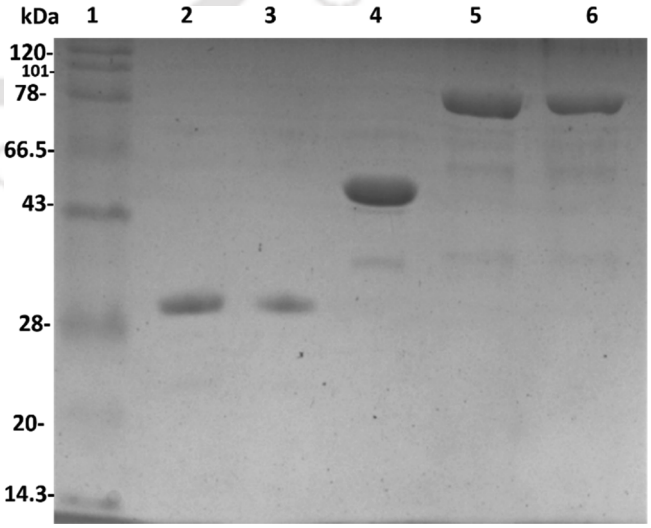


Fig. 1. SDS-PAGE (12%, w/v) analysis of the different purified enzymatic constructs. Lanes: 1, Protein marker; 2, wild-type *Ct*GH5; 3, mutant *Ct*GH5-F194A; 4, *Ct*GH1 5, *Ct*GH1-L1-*Ct*GH5-F194A (chimera 1).

Table 1
Kinetic parameters of individual and chimeric enzyme against different substrates.

Substrate	Wild-type *Ct*GH5			Mutant F194A			Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)			*Ct*GH1		
	K_m (mg.ml^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (ml.min^{-1}mg^{-1})	K_m (mg.ml^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (ml.min^{-1}mg^{-1})	K_m (mg.ml^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (ml.min^{-1}mg^{-1})	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m (mM^{-1}min^{-1})
Lichenan	0.17 ± 0.08	6.0 ± 0.1 × 10^3	3.8 ± 1.3 × 10^4	0.15 ± 0.05	8.7 ± 0.9 × 10^3	5.8 ± 1.3 × 10^4	0.11 ± 0.01	3.2 ± 0.9 × 10^4	2.9. ± 0.8 × 10^5	–	–	–
β-Glucan	0.16 ± 0.05	5.4 ± 0.4 × 10^3	3.3 ± 0.8 × 10^4	0.18 ± 0.04	8.1 ± 0.6 × 10^3	4.5 ± 1.5 × 10^4	0.16 ± 0.08	2.7. ± 3.5 × 10^4	1.6. ± 4.7 × 10^5	–	–	–
CMC[a]	0.35 ± 0.10	9.0 ± 0.1 × 10^2	2.5 ± 0.9 × 10^3	0.38 ± 0.11	1.57 ± 0.13 × 10^3	4.1 ± 1.2 × 10^3	0.54 ± 0.15	6.9 ± 0.7 × 10^3	1.2 ± 0.4 × 10^4	–	–	–
*p*NPG[b]	–	–	–	–	–	–	[c]0.02 ± 0.034	8.1 ± 0.7 × 10^2	[d]4.1 ± 0.2 × 10^4	0.04 ± 0.02	2.8 ± 0.3 × 10^2	7.1 ± 0.2 × 10^3

[a] CMC- Carboxymethylcellulose.
[b] *p*NPG- *p*-Nitrophenyl β-D-glucopyranoside.
[c] Unit of K_m in mM.
[d] Unit of k_{cat}/K_m in mM^{-1} min^{-1}.

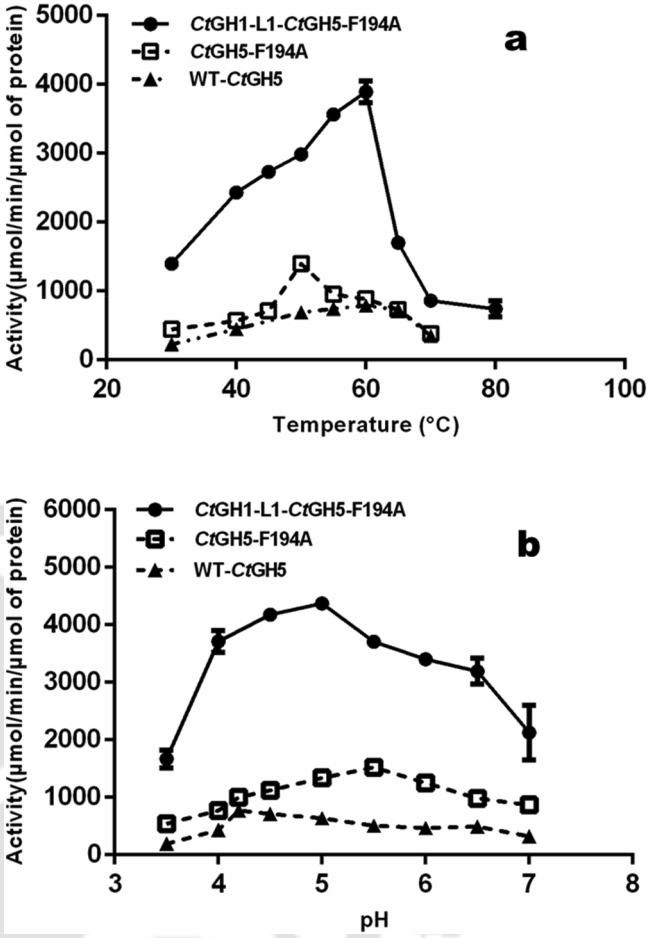


Fig. 2. Endoglucanase activity of chimera 1 and individual enzymes (wild-type *Ct*GH5 and *Ct*GH5-F194A) was measured a) by incubating the enzymes with 1% (w/v) CMC at between temperature range, 30–80 °C and b) by incubating the enzymes with 1% (w/v) CMC at different pH. The experimental details are mentioned in the methods. The experiment was performed in triplicate. Each data point represents the mean ± SE (n = 3).

(Table 1) under optimal conditions. For endoglucanase activity chimera 1 showed higher catalytic efficiency (k_{cat}/K_m) against Lichenan, β-Glucan, and CMC i.e. 2.9 ± 0.8 × 10^5 ml min^{-1} mg^{-1}, 1.6 ± 4.7 × 10^5 ml min^{-1} mg^{-1}, 1.2 ± 0.4 × 10^4 ml min^{-1} mg^{-1} which was 3–5 fold higher than the individual mutant F194A. For β-glucosidase activity chimera 1 showed 4.1 ± 0.2 × 10^4 mM^{-1} min^{-1} catalytic efficiency (k_{cat}/K_m) against *p*NPG which was 5.5 folds higher than the individual module *Ct*GH1. Previous report on hybrid enzyme between exoglucanase (cellobiohydrolase and β-glucosidase) *Ct*CD-*Cc*BG showed 2-fold higher activity with phosphoric acid swollen cellulose (PASC) while retaining similar activity for β-glucosidase as (Lee et al., 2011). Another fusion enzyme between endoglucanase (Endo5A) and β-glucosidase (Gluc1C) from *Paenibacillus* sp. with synthetic Glycine-serine linkers (G_4S_3) in between, showed 3.2- and 2- fold higher activity against both β-glucosidase and endoglucanase activities, respectively (Adlakha et al., 2012). It was also previously reported that inter-domain flexibility between two modules in a hybrid enzyme results in significant increase in activity (Lu and Feng, 2008). Thus, the results from present study revealed that the chimeric enzyme constructed by fusing β-glucosidase (*Ct*GH1) at N-terminal and endoglucanase (*Ct*GH5-F194A) at C-terminal using natural linker gave much higher catalytic efficiency for β-glucosidase and endoglucanase activities than the individual enzymes.

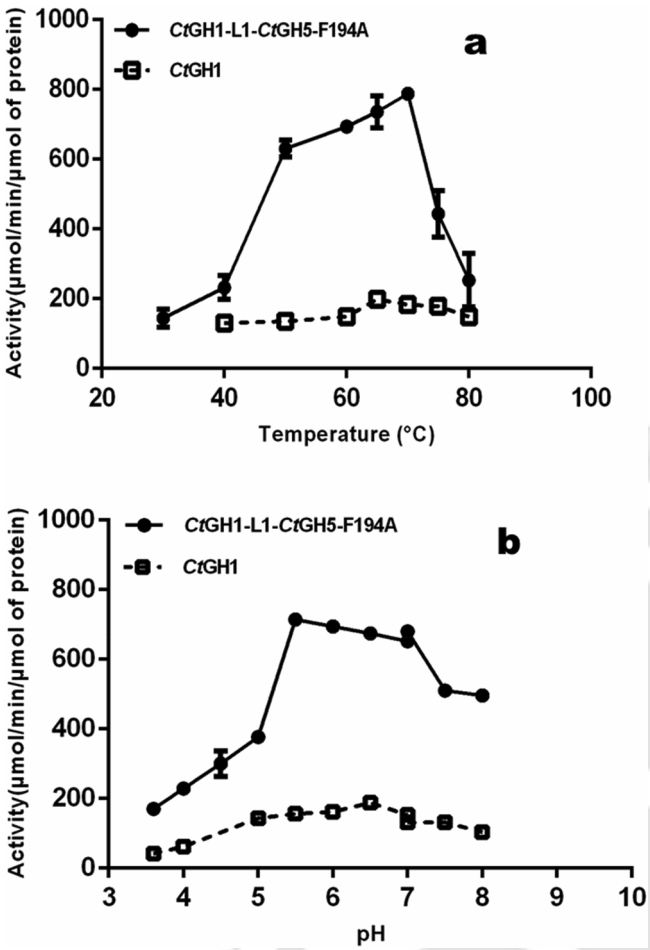


Fig. 3. β-Glucosidase activity of chimera and individual enzyme (*Ct*GH1) was measured, a) by incubating the enzymes with 1 mM *p*NPG at different temperatures ranging between 30 and 80 °C and b) by incubating the enzymes with 1 mM *p*NPG at different pH. The experimental details are mentioned in the methods. The experiment was performed in triplicate. Each data point represents the mean ± SE (n = 3).

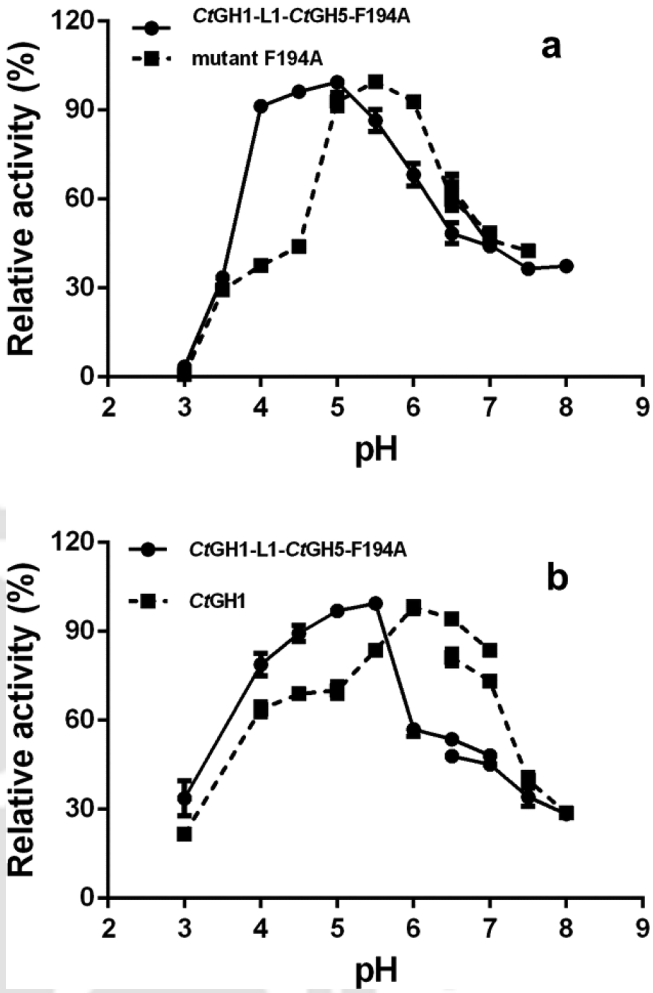


Fig. 4. pH stability analysis of different enzymatic constructs. a) endoglucanase activity of chimera 1 and *Ct*GH5-F194A and b) β-glucosidase activity of chimera 1 and *Ct*GH1. The pH stability was assayed for endoglucanase and β-glucosidase activities by incubating the enzymes at different pH for 60 min. In panel a) for endoglucanase activity, the reactions were performed in 100 µl reaction mixture containing 1% (w/v) CMC under optimized conditions. In panel b) for β-glucosidase activity, the assays were performed in 1 ml reaction containing 1 mM *p*NPG under optimized conditions. The assays were performed in triplicates and each data point represents the mean ± SE (n = 3).

3.5. Enzymatic hydrolysis of pretreated substrate Sorghum stalk with chimera 1 and individual enzymes

The enzymatic hydrolysis of pretreated biomass Sorghum stalk (1%, w/v NaOH + autoclaving for 20 min) was performed with chimera 1 and its individual enzymes (mix *Ct*GH1 + *Ct*GH5-F194A) for 48 h. The chimera showed 47.8 ± 0.7 mg/g of reducing sugar from pretreated biomass and the mix enzyme *Ct*GH1 + *Ct*GH5-F194A showed 29.4 ± 0.4 mg/g of reducing sugar. Therefore, the chimera 1 showed 1.6- fold higher yield of glucose/g from pretreated biomass when compared to the mix enzyme *Ct*GH1 and *Ct*GH5-F194A. The increased yield of reducing sugar by chimera 1 was might be because of its higher catalytic efficiency and stability then the individual enzymes. It was also reported earlier that fusion enzyme showed 2-fold increased yield of glucose from pretreated biomass than the mix enzymes added separately (Lee et al., 2011). In a previous study by Jamaldheen et al. (2018) it was reported that when endoglucanase (*Ct*Cel8A) and β-glucosidase (*Ct*Bgl1A) added separately for enzymatic hydrolysis of pretreated Sorghum stalk (1%, w/v NaOH + autoclaving for 20 min) it showed the yield of 34.2 mg/g of reducing sugar from pretreated biomass.

3.6. Analysis of hydrolysed products by purified chimera 1 using TLC

The hydrolysed products of CMC by chimera 1 at different time interval was analysed by TLC (Fig. 6a). In the initial 5 min major cello-oligosaccharides were formed and after that the major-product was only glucose till 12 h. This showed that both endo- and *exo*-activities are incorporated in the chimera 1. The hydrolysed product of individual enzymes i.e. the wild-type *Ct*GH5, the mutant F194A and the mixed enzymes at equimolar concentrations were compared with the chimera 1 (Fig. 6b). It was confirmed that the individual enzymes showed cellobiose and higher oligosaccharide as the major product but the chimera1 alone forms glucose along with cello-oligosacharides as a final product in 1 h of incubation time. Therefore, this showed that both β-glucosidase and endoglucanase activities were incorporated into the engineered chimeric construct and is more efficient than its individual enzymes. The action of chimera 1 on the pretreated biomass was also analysed by TLC. The individual enzyme *Ct*GH5-F194A (endoglucanase) hydrolysed the pretreated biomass and produced cellobiose and the higher cello-oligosaccharides (Fig. 6c). In the presence of mixture of *Ct*GH1 (β-glucosidase) and *Ct*GH5-F194A (endoglucanase)

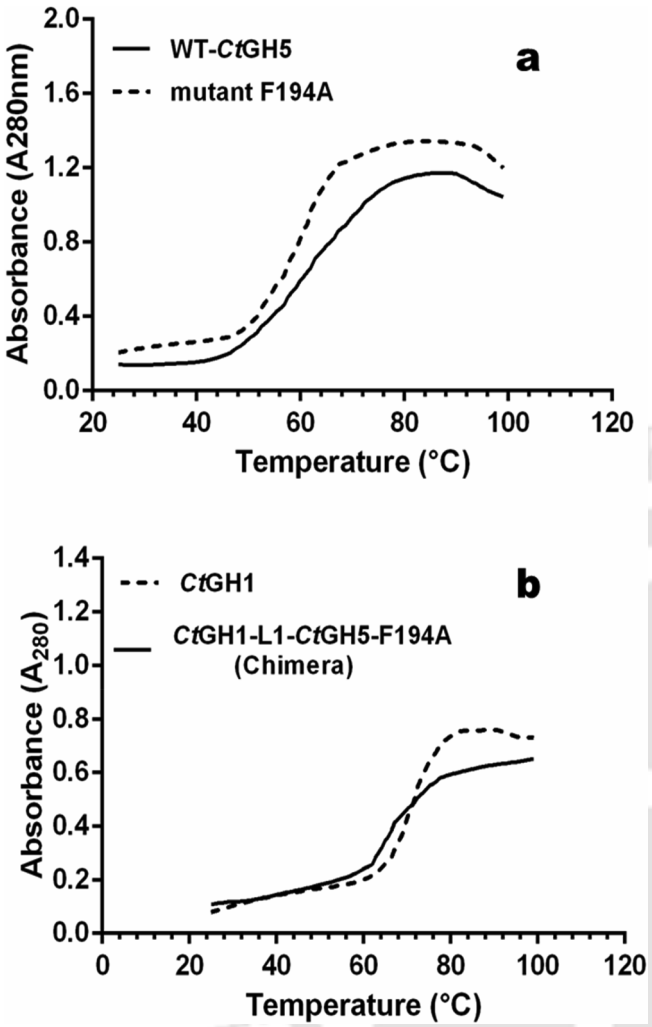


Fig. 5. Melting temperature analysis of chimera 1 and different enzymatic constructs. a) wild-type *Ct*GH5 and mutant *Ct*GH5-F194A and b) *Ct*GH1 and chimera 1. 50 µg/ml of each, chimera 1 or wild-type *Ct*GH5 was dissolved in 1 ml 50 mM sodium acetate buffer (pH 5.0), while 50 µg/ml of each, mutant *Ct*GH5-F194A and *Ct*GH1 was dissolved in 1 ml 50 mM MES buffer (pH 5.5) and the absorbance was measured a UV–Visible spectrophotometer at different temperatures ranging between 30 and 100 °C.

the glucose was formed as the final product from the pretreated biomass (Fig. 6c). However, the chimera 1 displayed higher accumulation of glucose after 48 h than the mixture of *Ct*GH1 and *Ct*GH5-F194A. Therefore, chimera 1 acts on the pretreated Sorghum stalk biomass and efficiently releases the glucose as final product.

4. Conclusion

The site-directed mutagenesis of Phenylalanine by flexible residue, Alanine near the active site of wild-type *Ct*GH5 resulted in two-fold increase in its endoglucanase activity. This mutant, *Ct*GH5-F194A was fused with β-glucosidase (*Ct*GH1) using the natural linker to develop a chimeric enzyme. The chimeric enzyme displayed enhanced endoglucanase and β-glucosidase activities. Higher melting temperature of chimeric enzyme was due to the positive effect conferred by N-terminal thermostable *Ct*GH1. Chimeric enzyme resulted in higher accumulation of glucose from pretreated Sorghum stalk than the individual enzymes. Thus, all these attributes make the chimeric enzyme potentially beneficial for bioethanol production at large scale.

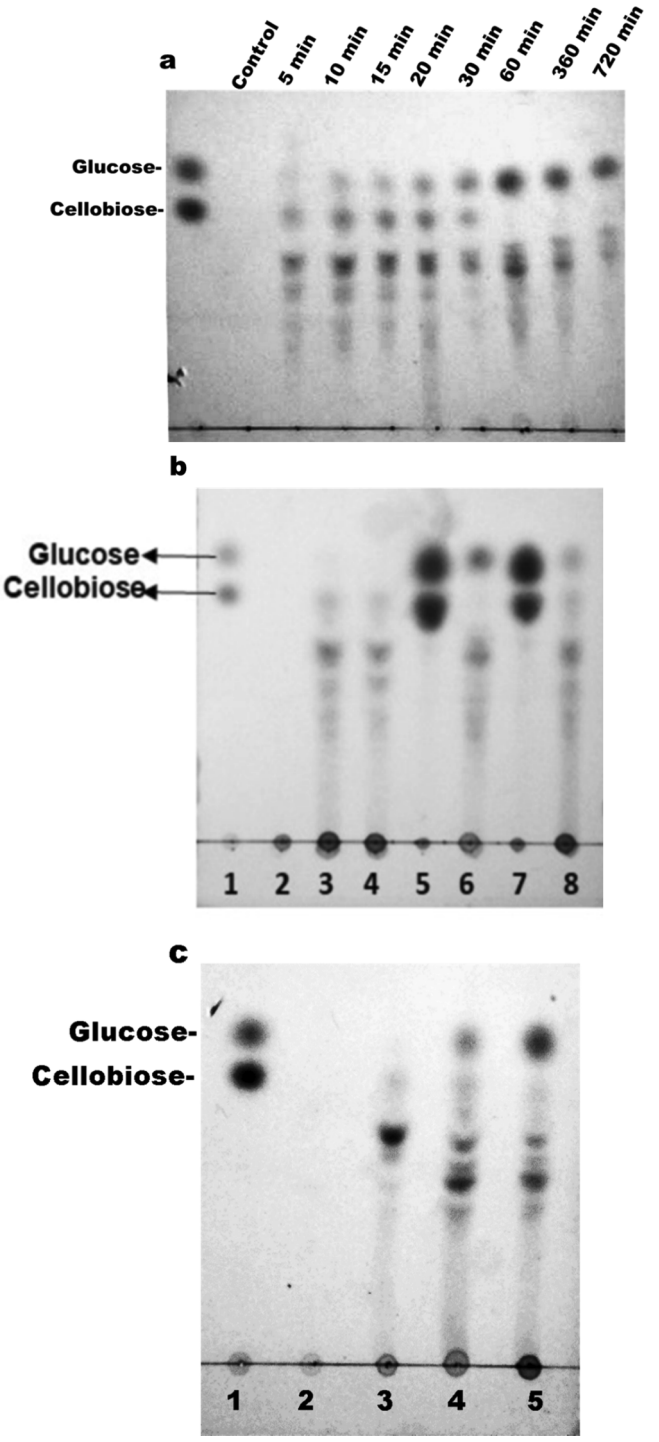


Fig. 6. a) Time dependent analysis of hydrolysed product of 1% (w/v) CMC by chimera 1. The reaction was performed at 60 °C using 20 mM citrate phosphate buffer, pH 5.0 using 0.2 µM enzyme, b) TLC analysis of hydrolysed products of 1% (w/v) CMC or 1% (w/v) cellobiose by different enzymatic constructs. Lanes: 1, Standard (Glucose + cellobiose); 2, CMC as control; 3, CMC + wild-type *Ct*GH5; 4, CMC + mutant *Ct*GH5-F194A; 5, Cellobiose + *Ct*GH1; 6, CMC + chimera 1; 7, Cellobiose + chimera 1; 8, CMC + mixture *Ct*GH1 + mutant *Ct*GH5-F194A. Each reaction was performed at optimal conditions using 0.2 µM of each enzyme for 1 h. c) Enzyme hydrolysed products of 1% (w/v) pretreated Sorghum stalk. Lanes: 1, Standard (Glucose + cellobiose); 2, pretreated biomass as control; 3, mutant *Ct*GH5-F194A; 4, mix (*Ct*GH1 + *Ct*GH5-F194A); 5, chimera 1. The reactions were performed at 50 °C for 48 h using 1.1 µM of each enzyme in 20 mM citrate phosphate buffer, pH 5.0 for chimera 1 and 20 mM citrate phosphate buffer, pH 5.5 for individual enzymes (*Ct*GH5-F194A and mixture of *Ct*GH1 + *Ct*GH5-F194A).

Acknowledgements

Authors acknowledge the financial support provided by DBT-Pan-IIT Grant (BT/EB/PAN IIT 2012), Centre for Bioenergy from Department of Biotechnology, Ministry of Science and Technology, New Delhi, India to Prof. Arun Goyal. The fellowship to Priyanka Nath was supported by funding from DST Inspire fellowship from Department of Science and Technology, New Delhi, India. The authors are grateful to Prof. Carlos M.G.A. Fontes for fruitful discussion and valuable suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biortech.2019.03.051.

References

Adlakha, N., Sawant, S., Anil, A., Lali, A., Yazdani, S.S., 2012. Specific fusion of β-1,4-endoglucanase and β-1,4-glucosidase enhances cellulolytic activity and helps in channeling of intermediates. Appl. Environ. Microbiol. 78 (20), 7447–7454.

Ahmed, S., Deka, D., Jawed, M., Goyal, D., Fontes, C.M., Goyal, A., 2009. Biochemical characterization of a recombinant derivative (*Ct*Lic26A-Cel5) of a cellulosomal cellulase from *Clostridium thermocellum*. Curr. Trends Biotechnol. Pharm. 3 (1), 56–63.

Anbar, M., Gul, O., Lamed, R., Sezerman, U.O., Bayer, E.A., 2012. Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A using consensus-guided mutagenesis. Appl. Environ. Microbiol. AEM-07985.

Barr, B.K., Hsieh, Y.L., Ganem, B., Wilson, D.B., 1996. Identification of two functionally different classes of exocellulases. Biochemistry 35 (2), 586–592.

Beguin, P., 1990. Molecular biology of cellulose degradation. Ann. Rev. Microbiol. 44 (1), 219–248.

Bharali, S., Purama, R.K., Majumder, A., Fontes, C.M., Goyal, A., 2005. Molecular cloning and biochemical properties of family 5 glycoside hydrolase of bi-functional cellulose from *Clostridium thermocellum*. Indian J. Microbiol. 45 (4), 317–321.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72 (1–2), 248–254.

Cartmell, A., McKee, L., Pena, M.J., Larsbrink, J., Brumer, H., Kaneko, S., Marles-Wright, J., 2011. The structure and function of an arabinan-specific alpha-1, 2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. J. Biol. Chem. 286 (17), 15483–15495.

Carvalho, A.L., Goyal, A., Prates, J.A., Bolam, D.N., Gilbert, H.J., Pires, V.M., Fontes, C.M., 2004. The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates β-1,4 and β-1,3–1,4-mixed linked glucans at a single binding site. J. Biol. Chem. 279 (33), 34785–35347.

Fan, Z., Werkman, J.R., Yuan, L., 2009. Engineering of a multifunctional hemicellulase. Biotechnol. Lett. 31 (5), 751–757.

Grabnitz, F., Seiss, M., RucknageL, K.P., Staudenbauer, W.L., 1991. Structure of the β-glucosidase gene bglA of Clostridium thermocellum: sequence analysis reveals a superfamily of cellulases and β-glycosidases including human lactase/phlorizin hydrolase. Eur. J. Biochem. 200 (2), 301–309.

Jamaldheen, S.B., Sharma, K., Rani, A., Moholkar, V.S., Goyal, A., 2018. Comparative analysis of pretreatment methods on Sorghum (*Sorghum durra*) stalk agrowaste for holocellulose content. Prep. Biochem. Biotech. 48, 1–27.

McKee, L.S., Peña, M.J., Rogowski, A., Jackson, A., Lewis, R.J., York, W.S., Marles-Wright, J., 2012. Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. Proc. Natl. Acad. Sci. 109 (17), 6537–6542.

Nelson, N., 1944. A photometric adaptation of the Somogyi method for the determination of glucose. J. biol. Chem 153 (2), 375–380.

Lee, H.L., Chang, C.K., Teng, K.H., Liang, P.H., 2011. Construction and characterization of different fusion proteins between cellulases and β-glucosidase to improve glucose production and thermostability. Bioresour. Technol. 102 (4), 3973–3976.

Lu, P., Feng, M.G., 2008. Bifunctional enhancement of a β-glucanase-xylanase fusion enzyme by optimization of peptide linkers. Appl. Microbiol. Biotechnol. 79 (4), 579–587.

Shi, R., Li, Z., Ye, Q., Xu, J., Liu, Y., 2013. Heterologous expression and characterization of a novel thermo-halotolerant endoglucanase Cel5H from *Dictyoglomus thermophilum*. Bioresour. Technol. 142, 338–344.

Somogyi, M., 1945. Determination of blood sugar. J. Biol. Chem. 160, 69–73.

Taylor, E.J., Goyal, A., Guerreiro, C.I., Prates, J.A., Money, V., Ferry, N., Gilbert, H.J., 2005. How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides. Structure and activity of a *Clostridium thermocellum* lichenase, *Ct*Lic26A. J. Biol. Chem. 280 (38), 32761–32767.

Telke, A.A., Ghatge, S.S., Kang, S.H., Thangapandian, S., Lee, K.W., Shin, H.D., Kim, S.W., 2012. Construction and characterization of chimeric cellulases with enhanced catalytic activity towards insoluble cellulosic substrates. Bioresour. Technol. 112, 10–17.

Urbanowicz, B.R., Bennett, A.B., del Campillo, E., Catala, C., Hayashi, T., Henrissat, B., Teeri, T.T., 2007. Structural organization and a standardized nomenclature for plant endo-1,4-β-glucanases (cellulases) of glycosyl hydrolase family 9. Plant Physiol. 144 (4), 1693–1696.

Yennamalli, R.M., Rader, A.J., Kenny, A.J., Wolt, J.D., Sen, T.Z., 2013. Endoglucanases: insights into thermostability for biofuel applications. Biotechnol. Biofuels 6 (1), 136.

Yuan, S.F., Wu, T.H., Lee, H.L., Hsieh, H.Y., Lin, W.L., Yang, B., Ho, M.C., 2015. Biochemical characterization and structural analysis of a bifunctionalcellulase/xylanase from *Clostridium thermocellum*. J. Biol. Chem. 290 (9), 5739–5748.

International Journal of Biological Macromolecules 148 (2020) 364–377

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: http://www.elsevier.com/locate/ijbiomac

Combined SAXS and computational approaches for structure determination and binding characteristics of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) generated by assembling β-glucosidase (*Ct*GH1) and a mutant endoglucanase (*Ct*GH5-F194A) from *Clostridium thermocellum*

Priyanka Nath [a,b], Kedar Sharma [a], Krishan Kumar [a], Arun Goyal [a,b,*]

[a] Carbohydrate Enzyme Biotechnology Laboratory, Department of Biosciences and Bioengineering, India
[b] DBT PAN-IIT Center for Bioenergy, Indian Institute of Technology Guwahati, Guwahati, Assam, India

ARTICLE INFO

Article history:
Received 1 October 2019
Received in revised form 10 January 2020
Accepted 11 January 2020
Available online 13 January 2020

Keywords:
Chimera
MD simulation
SAXS

ABSTRACT

Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) developed by fusing β-glucosidase (*Ct*GH1) at N-terminal and endoglucanase (*Ct*GH5-F194A) at C-terminal was structurally characterized. Its secondary structure analysis by CD showed 38% α-helix, 9.3% β-sheets and 52.7% random coils corroborating with prediction. *In-silico* modeled structure of Chimera comprised two modules, *Ct*GH1 and *Ct*GH5-F194A displaying $(\alpha/\beta)_8$ fold. Ramachandran plot of Chimera showed 99.9% residues in allowed region. Binding interaction of Chimera with cello-oligosaccharides suggested active forms of *Ct*GH1 and *Ct*GH5-F194A and their involvement in catalysis. MD simulation of cellohexaose bound endoglucanase module of Chimera showed favourable flexibility in loops, LA with H-bond formation with Asn510 and in loop LC relocation of Tyr687 away from active site efficiently releasing the product after catalysis. Higher short range interaction energy of Chimera, −383 kJ/mol than the individual endoglucanase, 254 kJ/mol against cellohexaose suggested higher efficient catalysis by Chimera. β-Glucosidase module of Chimera showed fluctuations in outer loops suggesting conformational changes that might be contributing to improved hydrolysis. SAXS analysis of Chimera displayed monodispersed state. Guinier analysis of Chimera showed globular shape ($R_g = 3.15 \pm 0.10$ nm). Kratky plot confirmed fully folded and flexible behaviour in solution. Gasbor modeled structure of Chimera displayed an elongated structure with two modules having shape similar to bean-bag contour.

1. Introduction

The lignocellulosic plant biomass contains cellulose, hemicellulose and lignin. Cellulose can be hydrolysed to glucose while hemicellulose can be hydrolysed to various C5 sugars such as xylose, arabinose and C6 sugars such as galactose, mannose and glucose. The monomeric sugars glucose, xylose and arabinose can be utilized by the yeast to produce alcohol and the lignocellulosic plant biomass can thus serve as an alternative source of energy. The lignocellulosic plant biomass is rich in cellulose. The cellulose can be entirely hydrolyzed to glucose by treating with three enzymes *viz*. endoglucanase, cellobiohydrolase and β-glucosidase [1]. Endo-β-1,4-glucanase acts randomly on the cellulose chain and produces cellodextrin which are larger cellooligosaccharides as hydrolyzed products [2]. Cellobiohydrolase acts at the end of the cellodextrin and releases cellobiose as the product [3], while β-glucosidase hydrolyzes the cellobiose to form two molecules of glucose [4]. The production of all three enzymes for cellulose hydrolysis is a cost-intensive method [5]. The production cost can be reduced by developing enzymes with improved catalytic efficiency and multifunctional activity [6,7]. The multifunctional enzymes with two or more fused enzymes in a single polypeptide chain can act synergistically to hydrolyze the polymer into simple sugar [7]. An essential aspect for the construction of active multifunctional Chimera with synergistic catalytic activity and improved catalytic efficiency is to determine the structural organization of the modules in the chimeric construct and selection of the appropriate linker sequence [8]. The structural organization of individual modules in the Chimera can be predicted by generating a homology model using computational approaches [9]. The molecular dynamics can provide successful validation of the protein model to explore the protein conformation [10]. Efforts were made to create chimeric enzyme using a computational approach where bi-functional *endo-*

Abbreviations: GH, glycoside hydrolase; MD, molecular dynamics; SAXS, small-angle X-ray scattering; RMSD, root-mean-square deviation; RMSF, root-mean-square-fluctuation.

* Corresponding author at: Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India.
E-mail address: arungoyl@iitg.ac.in (A. Goyal).

https://doi.org/10.1016/j.ijbiomac.2020.01.116
0141-8130/
TH-2322_146106042

xylanase (xynA) and lichenase (BglS) modules from *Bacillus subtilis* were fused [11]. For example, substrate accessibility to its active site and its dynamic behaviour in the solution were also determined [11]. The resulting Chimera of *endo*-xylanase (xynA) and lichenase (BglS) preserved the biochemical characteristics of the parental enzyme with slight variation in the catalytic efficiency [11]. The computational development for the prediction of stable conformation and the arrangement of the modules in a chimeric enzyme in the solution resulted in the development of multifunctional enzyme and predicted their extended conformation and validation of their experimental data [12].

The family, GH5 is one of the largest and well-characterized families of glycoside hydrolases (http://www.cazy.org/GH5.html). The catalytic activity of family 5 GH (Cel5E) (ABN52701.1) from *Clostridium thermocellum* was enhanced by attaching the linker at both N- and C-terminals of GH5 and by mutating Phe267 to Ala [13]. In our previous study the catalytic efficiency of endoglucanase (*Ct*GH5) (ABN52701.1) of family 5 glycoside hydrolase from *Clostridium thermocellum* without the N- or C-terminal linker was cloned and its mutant F194A was generated by site-directed mutagenesis [14], by following the method reported earlier [13]. The mutant *Ct*GH5-F194A resulted in enhanced enzyme activity and was fused with β-glucosidase (*Ct*GH1) to develop a Chimera, *Ct*GH1-L1-*Ct*GH5-F194A [14]. *Ct*GH1 (β-glucosidase) used in the chimeric construct was earlier biochemically and structurally characterized [15]. The fusion of the two constructs *Ct*GH1 and *Ct*GH5-F194A to develop a Chimera was performed using a natural N-terminal linker from the cellulosomal gene *cel*H from *C. thermocellum* [16]. The chimeric enzyme (*Ct*GH1-L1-*Ct*GH5-F194A) developed by using the construct β-glucosidase (*Ct*GH1) at the N-terminal and endoglucanase (*Ct*GH5-F194A) at C-terminal showed significant enhancement in both the activities [14]. The Chimera showed 3- to 5-fold increase in the catalytic efficiency of β-glucosidase and endoglucanase activities [14].

In the present study, the structural analysis of the developed Chimera, *Ct*GH1-L1-*Ct*GH5-F194A was undertaken to find out its organization, stability and the folding behaviour. The secondary structure composition of the Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), previously constructed by fusing β-glucosidase (*Ct*GH1) at N-terminal and mutant endoglucanase (*Ct*GH5-F194A) at the C-terminal was determined [14]. The computer based prediction softwares were used to study its three-dimensional structure and the structural parameters. The binding interaction of the two modules of Chimera was performed by molecular docking against different ligands. The Chimera and its individual modules in the presence and absence of the ligands were MD simulated to elucidate the role of loops and the key residues involved in the catalysis or binding. The small angle X-ray scattering (SAXS) has been used to analyse the average particle size, molecular shape and other properties of various type of macromolecules such as biomacromolecules [15] and orthopaedics [17,18]. SAXS analysis of the chimeric enzyme was performed that displayed the arrangement of the two modules of Chimera in the solution form.

2. Materials and methods

2.1. Experimental and in silico prediction analysis of secondary structure of Chimera

The Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was cloned, expressed and biochemically characterized earlier was used in the present study [14]. The secondary structure of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) *i.e.* the content α helices, β strands and random coils were predicted by different available webservers for prediction of secondary structure such as Psi-Pred v 3.3 web server (*http://bioinf.cs.ucl.ac.uk/psipred/*) and RaptorX (http://raptorx.uchicago.edu/). The secondary structure composition of Chimera was also determined by circular dichroism (CD) analysis. The purified chimeric enzyme at a concentration of 5.3 μM in 50 mM sodium phosphate buffer, pH 7.0 was used for CD analysis. The CD spectrum was observed on a spectropolarimeter (JascoJ-815, Japan) at 25 °C using 1 nm bandwidth over far UV region between 190 and 250 nm at a scanning rate of 50 nm/min with an average of three scans. The CD experimental data was analyzed by the difference in molar extinction coefficient (Δε, deciliter mol^{-1} cm^{-1}) described as a function of wavelength [19]. The percentage of α-helix and β-sheet were obtained from web-based K_2D_2 software (http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d2/) [20].

2.2. Computer-based structural prediction of Chimera and its validation

The 3D-structure prediction of the Chimera was performed by using RaptorX web server (http://raptorx.uchicago.edu/). RaptorX is also an online server for protein structure and function prediction [21]. This program uses a statistical method for template-based modeling of protein. This methodology improves alignment accuracy by utilizing the structural data in single or multiple templates [22]. RaptorX also calculates *p*-value for the relative global quality, GDT (global distance test) and uGDT (un-normalized GDT) for the absolute global quality, and RMSD for the absolute local quality of each residue in the model. The prediction of the threading method is reliable when the p-value is $<10^{-4}$ [22]. For generating the modeled structure of Chimera, RaptorX predicted two modules for the input sequence of Chimera, The Chimera is made up of two modules, *viz.* N-terminal module 1 (*Ct*GH1) comprising 445 amino acids, the C-terminal module 2 (*Ct*GH1-F194A) comprising 309 amino acids and the two modules are connected by a natural linker made up of 19 amino acids. The difference in the number of amino acid residues between the endoglucanase module of Chimera and the individual endoglucanase is of 487 (pET28a(+)-His-tag region, 23 + *Ct*GH1, 445 + Natural Linker L1, 19 = 487). For module 1, amino acid residues (1-474) *p*-value predicted was 7.14×10^{-12} and for module 2, amino acid residue (475-796) a p-value of 1.41×10^{-8} was obtained. The webserver RaptorX showed closest homology for the input sequence with 3AHX an β-glucosidase from *Clostridium cellulovorans* and 5BYW an endoglucanase from *Clostridium thermocellum*. The energy minimization of the modeled structure was performed using YASSARA energy minimization tool (http://www.yasara.org/minimizationserver.php). The modeled structure generated was analyzed by the software UCSF Chimera [23]. The energy minimized structure was validated by using the saves web server (http://servicesn.mbi.ucla.edu/SAVES/).

2.3. Molecular dynamics simulation of modeled Chimera structure

Molecular dynamics simulation (MD) for the Chimera modeled structure was accomplished by using Gromacs v 5.14 [24] in a supercomputer facility Param-Ishan available in the Indian Institute of Technology Guwahati. GROMOS96 53a6 was used as a force field to calculate the protein forces. The modeled Chimera was placed inside a cubic box with dimension 6.38 × 7.73 × 10.12 and volume 1949.8 nm^3 with single point charge (SPC) along with the water molecules. The MD simulation contained 60,088 water molecule and 22 Na^+ counter ions were added for the neutralization of the charges on the Chimera. The chimeric structure was minimized using steepest descent method using 50,000 iteration steps with cut-off till 1000 kJ mol^{-1}. Then the equilibration of the whole system was carried out for 500 ps using NVT ensemble in the constant number of particles, volume, and temperature. The next calibration was performed for 500 ps by the NPT ensemble in a constant number of particles, volume, temperature and pressure. The production run for 70 ns was performed with NPT ensemble acquiring a 2 fs integration time. The linear constraint solver (LINC) algorithm [25] was utilized to constrain the bonds linked with hydrogen atoms and radius of gyration. Throughout the simulation, the modeled Chimera structure was analyzed as a time-dependent function to verify its stability in the solvent system. The root-mean-square deviation (RMSD) of the simulated system was analyzed by gmx rms command, and gmx rmsf was used for determining the

fluctuation (RMSF). The radius of gyration (Rg), solvent accessible surface area (SASA) was determined by using the program gmx gyrate and gmx sasa, respectively.

2.4. Binding interaction analysis of Chimera

Molecular docking of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) with the cello-oligosaccharides was performed by using SwisDock, the web based server (http://www.swissdock.ch/docking). Cello-oligosaccharides were downloaded using the GLYCAM server [26]. The modeled Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was saved in PDB format and the ligands were saved in Mol2 file format for docking analysis in the Swiss Dock web-based tool. Swiss Dock produced large number of ligand binding results. The enzyme-ligand docked complex showing the strongest binding with highest negative binding free energy were selected. This ligand bound structure was extracted and visualized in PyMol 2.0 [27]. The representation of ligand interaction with the amino acid residues of the protein was generated using the PDBsum Generate tool (*https://www.ebi.ac.uk/thorntonsrv/databases/pdbsum/Generate.html*).

2.5. Molecular dynamics of ligand bound Chimera

Molecular dynamics (MD) simulation for determining the interaction of Chimera with the ligands was performed in the presence of cellobiose and cellohexaose. The cellobiose was placed in the catalytic cleft of β-glucosidase module in the Chimera, while the cellohexaose was kept in the catalytic cleft of endoglucanase module in the Chimera. The docked complexes of Chimera with cellobiose and cellohexaose from Section 2.4 were selected and were simulated by using GROMACS v 5.14. Similarly, the cellohexaose bound individual endoglucanase (*Ct*GH5-F194A) and cellobiose bound individual β-glucosidase (*Ct*GH1) were also simulated for the comparative analysis. Protein forces were calculated through GROMOS96 43A1 force field and ligand topologies for cellobiose and cellohexaose were generated by using PRODRG server. The docked complexes of the Chimera or ligand bound Chimera were placed in a cubic box with volume 1949.8 nm^3 and filled with water molecules. The MD simulation system containing 60,088 water molecules were added with 23 Na^+ counter ions for the neutralization of the charges on ligand bound Chimera for cellobiose in β-glucosidase module and 28 Na^+ counter ions for cellohexaose in endoglucanase module of Chimera. Similarly, the docked complex of individual endoglucanase (*Ct*GH5-F194A) with cellohexaose and the only individual endoglucanase (*Ct*GH5-F194A) were placed inside a cubic box with a volume 489 nm^3 with single point charge (SPC) and 14,509 water molecules were neutralized by 24 and 19 Na^+ counter ions. The docked complex of individual β-glucosidase (*Ct*GH1) with cellobiose and the only individual β-glucosidase (*Ct*GH1) were placed inside a cubic box with a volume 674 nm^3 with single point charge (SPC) and 25,581 water molecules were neutralized by 11 and 10 Na^+ counter ions. Initially, the equilibration of entire system was accomplished with restraints in NVT ensemble (constant number of particles, volume and temperature) for 500 ps. The entire system was equilibrated twice, first of all with restraints and secondly without restraints for 500 ps in NPT ensemble (constant number of particles, pressure and temperature). The 70 ns production run was setup along with 2 fs integration time of NPT ensemble. The docked complexes after 70 ns production were analyzed with the help of PyMOL 2.0 software [27]. The root-mean-square deviation (RMSD) of the simulated ligand bound form of Chimera and its individual endoglucanase (*Ct*GH5-F194A) was analyzed by gmx rms command. The gmx rmsf was used for determining the fluctuation (RMSF) and the gmx energy command was used to obtain the interaction energy between proteins and ligands. Average energies for short range interactions between protein and ligand was given in the terms of Lennard Jones and coulombic interaction energy. The representation of ligand interaction with the amino acid residues of the protein after molecular dynamics simulation was generated by using the PDBsum Generate tool (*https://www.ebi.ac.uk/thorntonsrv/databases/pdbsum/Generate.html*).

2.6. Small angle X-ray scattering analysis of Chimera

The scattering profiles of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) at two different concentrations (3 mg/mL and 5 mg/mL) along with its matched buffer (50 mm Sodium phosphate, pH 7.0) were collected as per the method reported previously [28] by a home source small-angle X-ray scattering system (SAXSpace, Anton Paar GmbH, Graz, Austria) established at CSIR-Institute of Microbial Technology, Chandigarh, India. Incident X-rays were originated by using line collimation system and passed through the sample, filled in a thermostatic quartz capillary of a 1 mm diameter and two independent exposures of 30 min each was collected at 10 °C. The SAXS analysis software was used for processing and subtraction of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and its matched buffer SAXS profiles. The integrity and non-aggregation effect on the Chimera scattering profiles were confirmed by visual inspection in the SAXS data analysis program of ATSAS 2.84 suite [29]. The radius of gyration (Rg) of globular and rod shape (Rc) of the Chimera was calculated by applying the Guinier approximation equation in the primusqt [30,31]. The length (L) of the Chimera was calculated by implementing [$\sqrt{12}\{(Rg^2)-(Rc^2)\}$] formula. The maximum dimension (D_{max}) of the particle and the distance distribution function plot (P(R)) were evaluated by employing GNOM software integrated with ATSAS v2.84 [32]. The molecular mass of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was calculated with the assistance of SAXSMoW program [33]. Gasbor program was used for constructing 50 *ab initio* models by taking dummy aspartic acid residues equivalent to the molecular mass of *Ct*GH1-L1-*Ct*GH5-F194A. These 50 low-resolution shape *ab initio* models were averaged by applying DAMAVER program [34]. The final averaged *ab initio* model was superposed with *Ct*GH1-L1-*Ct*GH5-F194A modeled structure using the SUPCOMB program [35]. The CRYSOL program was used to generate theoretical scattering for *Ct*GH1-L1-*Ct*GH5-F194A modeled structure, and it was fitted with the experimental data generated from SAXS [36].

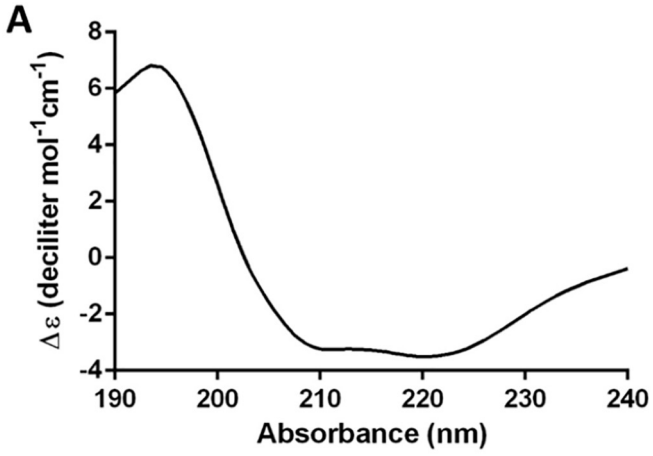
3. Results and discussion

3.1. Secondary structure analysis by circular dichroism and web server's prediction

The circular dichroism (CD) spectrum of the Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was analyzed from the K_2D_2 web server, where the data points are measured with the already existing secondary structures of the known proteins [37]. The CD analysis displayed 38% of α-helix, 9.32% of β-sheets and 52.68% of random coils (Fig. 1A and Table 1). The secondary structure prediction analysis of the Chimera by Psi-Pred web server showed 33% of α-helix, 10% of β-sheets and 54% of random coils (Fig. 1B and Table 1). The RaptorX prediction web server showed 34% of α-helix, 11% of β-sheets and 53% of random coils (Table 1). Thus, the predicted secondary structure result of the Chimera corroborated with that obtained by CD analysis.

3.2. 3-Dimensional structure prediction of Chimera and its validation

The homology based modeled 3D structure of Chimera was generated by using RaptorX webserver. The 3D-model structure generated by the RaptorX server predicted a two-module protein structure with *p*-value of 7.14×10^{-12} for module 1, *Ct*GH1 (β-glucosidase), amino acid residues (1-468) and 1.41×10^{-8} for module 2, *Ct*GH5-F194A (endoglucanase), amino acid residue (488-796). The 3D structure prediction of the Chimera by RaptorX was considered for further analysis because the p-value was $<10^{-4}$. The modeled structure of Chimera



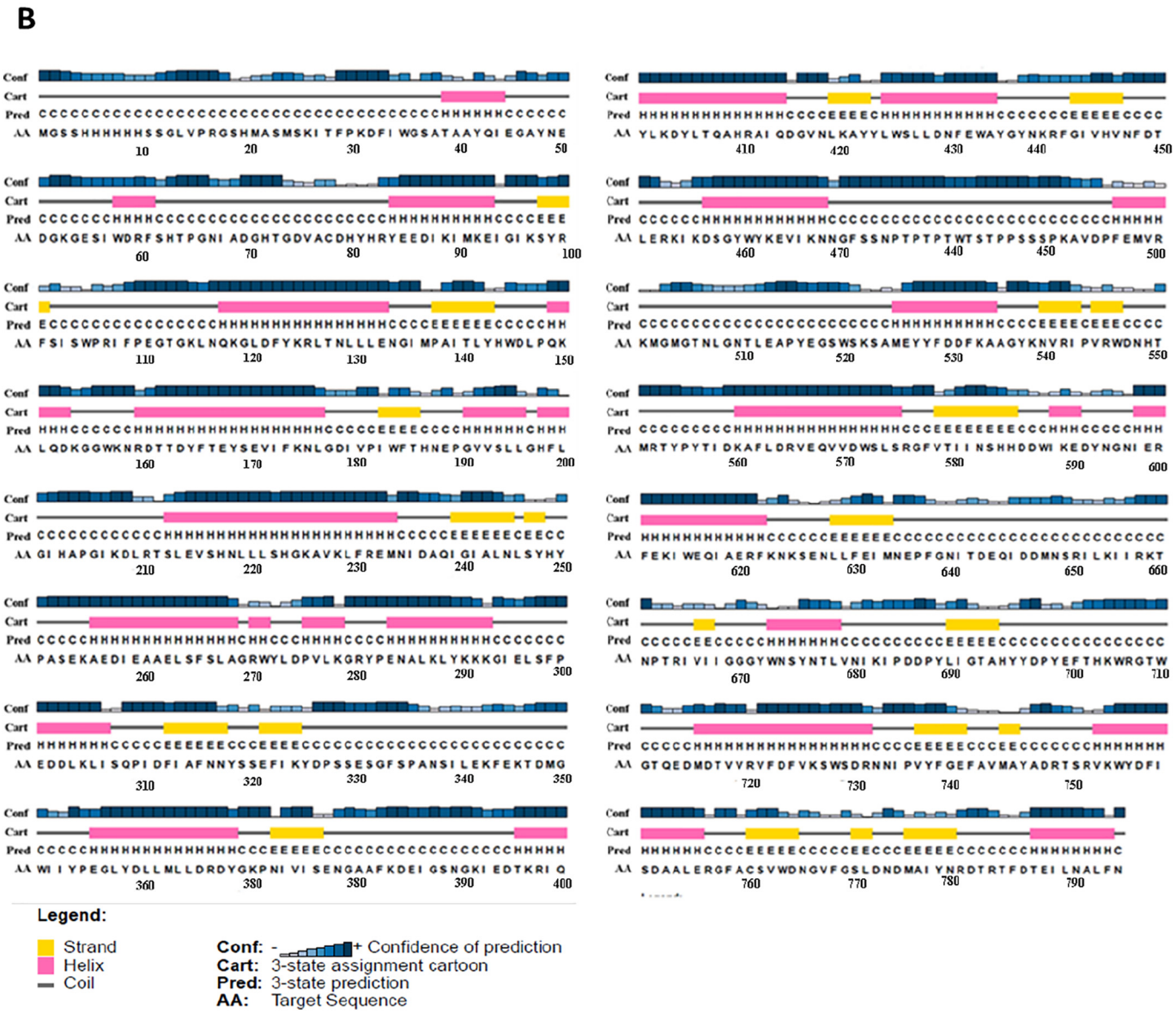


Fig. 1. A) Circular dichroism (CD) spectrum of Chimera for estimating the percentage of secondary structure components. The CD data were explained by the difference in molar extinction coefficients (decilitre mol^{-1} cm^{-1}) as a function of wavelength. The purified Chimera purified with a concentration of 5.3 μM in 50 mM sodium phosphate buffer, pH 7.0 was utilized for CD analysis. The CD spectrum was observed at 25 °C using 1 nm bandwidth extending over far UV region between 190 and 250 nm at a scanning rate of 50 nm/min. B) Secondary structure prediction of Chimera by Psi-Pred server tool demonstrating the amino acid residues involved in formation of α -helix (cylinders), β -sheet (arrow) and random coil (continuous line).

Table 1
Secondary structure analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

Prediction	α-helix (%)	β-sheet (%)	Random coils (%)
CD analysis (K_2D_2)	38	9.3	53
RaptorX	34	11	54
PSI-PRED	33	10	51

was energy minimized by using YASSARA energy minimization tool. The starting energy of the modeled structure was 1.1×10^5 kJ mol^{-1} while after minimization it showed energy of -4.36×10^5 kJ mol^{-1} for the end structure. The end structure of the Chimera after minimization was extracted from the server and was used for further structural characterization. The modeled structure of the Chimera showed N-terminal β-glucosidase (*Ct*GH1) module (grey colour) and C-terminal endoglucanase (*Ct*GH5-F194A) (green colour) connected by a linker (red colour) (Fig. 2A and B). The overall modeled structure of the Chimera showed classical (α/β)-TIM barrel fold (Fig. 2A and B). The surface view of Chimera shows the active site for both β-glucosidase (Glu189 and Glu378) (Fig. 2C) and endoglucanase (Glu626 and Glu731) (Fig. 2D).

This minimized structure was validated by using different validation tools available in SAVES v5.0 (http://servicesn.mbi.ucla.edu/SAVES/). The verify3D showed 91.4% of the residues gave average 3D-1D score ≥ 0.2 (Fig. 2E). Quality assessment of the modeled structure of Chimera by Ramachandran plot was carried out, before taking it further for MD simulation and other computational studies. It showed 90.9% non-glycine and non-proline amino acid residues in the most favourable region, 8.4% within the additional allowed region, 0.6% in the generously allowed region and only 0.1% amino acid residues in the disallowed region which were Ala77 and Trp433 (Fig. 2F). This result demonstrated that the amino acid residues in the modeled chimeric structure are involved in favourable phi (Φ) and psi (Ψ) backbone dihedral angles. The results were fairly in agreement with the individual modules even after a linker was introduced between them and the orientation of the amino acids residues were not disturbed. Quality assessment of the model from ProSA revealed that the Chimera model matched X-ray region of the plot with *Z*-score (−12.54) which is reliable to the Z-scores of the templates; (−10.79) for 3AHX an β-glucosidase from *Clostridium cellulovorans*, and (−10.21) for 5BYW an endoglucanase from *Clostridium thermocellum* (Fig. 2G). Similarly, local model quality (Fig. 2H) showed that absence of problematic or erroneous regions in the predicted structure of Chimera as there were absence of positive energy values (thick green line indicates residue wise energy plot calculated based on smoothened average energy score over each-40 residues). From all these structure validation parameters it can be considered that the predicted protein structure is reliable and satisfactory.

3.3. Molecular dynamics simulation of Chimera

The modeled structure of Chimera was further subjected to molecular dynamics simulation for 70 ns. The dynamics study was carried out in order to determine the stability as well as the compactness of the modeled structure over the defined time scale. Molecular dynamics simulations along with the trajectory leads to the energetically more acceptable conformation for the protein in aqueous solution and can be represented by the RMSD between the starting structure and those obtained during the simulation. The RMSD profile shown in Fig. 3A demonstrated that the RMSD reached a plateau phase value of 0.7 nm at 40 ns and remained stable till 70 ns with ~0.01 nm oscillations in the RMSD profile after 40 ns (Fig. 3A). The stable RMSD with less oscillations represented the stable conformation of the two modules *Ct*GH1 (β-glucosidase) and *Ct*GH5-F194A endoglucanase in the chimeric form. The global compactness of Chimera during the MD simulation was determined by using gmx gyrate program. The fluctuation in the R_g values (3.2–3.35 nm) for the Chimera till the time scale of 20 ns was due to the flexibility, while after 20 ns the Chimera achieved stability and compactness at 3.2 nm and remained stable till 70 ns (Fig. 3B). These results displayed that the chimeric modeled structure has a stable conformation. RMSF is measured by using gmx rmsf program. The RMSF calculates the displacement of a particular atom, or group of atoms, relative to the reference structure. The flexible residues were denoted as the residues showing the highest α-carbon flexibility among all the residues in a polypeptide chain, thus showing higher RMSF values. Rigid residues were denoted as the residues showing the lowest flexibility or lower RMSF values. In the Chimera starting amino acid residues of N-terminal and end residues of C-terminal showed highest flexibility. The fluctuation in RMSF was also observed between residues between 470 and 480 this region comprising the linker connecting both N-terminal and C-terminal modules of the Chimera signifying the flexibility of the linker in the chimeric structure. The amino acid residues within the protein active site core regions of the Chimera for module 1 *Ct*GH1 (β-glucosidase) from amino acid residues 189–358 and module 2 endoglucanase (*Ct*GH5-F194A) 600–800 amino acid residues, respectively, showed comparatively lesser fluctuation than the starting N-terminal and end C-terminal residues (Fig. 3C). Previous studies also suggested that the protein core is the main determinant of protein stability and activity [38,39]. Therefore, the RMSF described that both catalytic cores of the Chimera maintained their stability and activity in the chimeric form. The solvent accessible surface area was calculated using gmx sas programme. The average SASA of the Chimera remained fairly stable with an average value of 320 nm^2 till 70 ns (Fig. 3D). This suggested that the overall exposure of the Chimera towards the solvent area remains fairly unchanged. The minor variation in the SASA also suggested that the accessibility of the substrate in the catalytic sites were not affected in the designed Chimera. This also suggested that the active sites of both the modules are not obstructed by steric effects between the modules due to the presence of the linker. The absence of inter-modular interaction between the two modules was also confirmed in a Chimera of xylanase and lichenase from *Bacillus subtilis* that resulted in the active form of chimera as reported earlier [11]. The superposition of initial structure of the Chimera and structure obtained by MD simulation showed variation in the loop regions and showed an RMSD of 1.244 Å (Fig. 3E). The MD simulated structure of Chimera was stable and can be used for further analysis.

3.4. Molecular docking analysis of Chimera

The molecular docking study of Chimera was performed in order to determine the interaction of different cello-oligosaccharides along with the *para*-nitrophenyl derivative with the two catalytic clefts of the Chimera which involved module 1, *Ct*GH1 (β-glucosidase) and module 2, *Ct*GH5-F194A (endoglucanase). The molecular docking in the SwissDock showed >20 possible docking solutions for both modules of the Chimera. The binding interaction were observed in both catalytic sites of module 1, *Ct*GH1 and module 2, *Ct*GH5-F194A. This demonstrated the active involvement of both catalytic sites in the chimeric form. The module 1 showed binding with *p*NPG, cellobiose, cellotriose, cellotetraose, cellopentaose, cellohexaose and celloheptaose with free energy of binding (ΔG) of −6.92, −7.37, −7.62, −10.05, −10.86, −11.59 and −11.13 kcal/mol, respectively, as shown in Table 2. The binding studies demonstrated that the binding site of module 1, *Ct*GH1 (β-glucosidase) of the Chimera can accommodate cello-oligosaccharides up to 7 degree of polymerization, DP (Table 2). Similar binding interactions have been reported for β-glucosidase (*Ht*Bgl) from *Hungateiclostridium thermocellum* which, showed that the active site of *Ht*Bgl displays strong binding affinities against cellobiose, cellotriose, cellotetraose up to cello-heptaose [15] and also for other glucosidases of family 1 glycoside hydrolases from different organisms such as *Humicola insolens* [40] and *Alicyclobacillus* sp. [41]. Similarly, module 2, *Ct*GH5-F194A (endoglucanase) in the Chimera showed binding interaction with cellobiose, cellotriose, cellotetraose, cellopentaose,

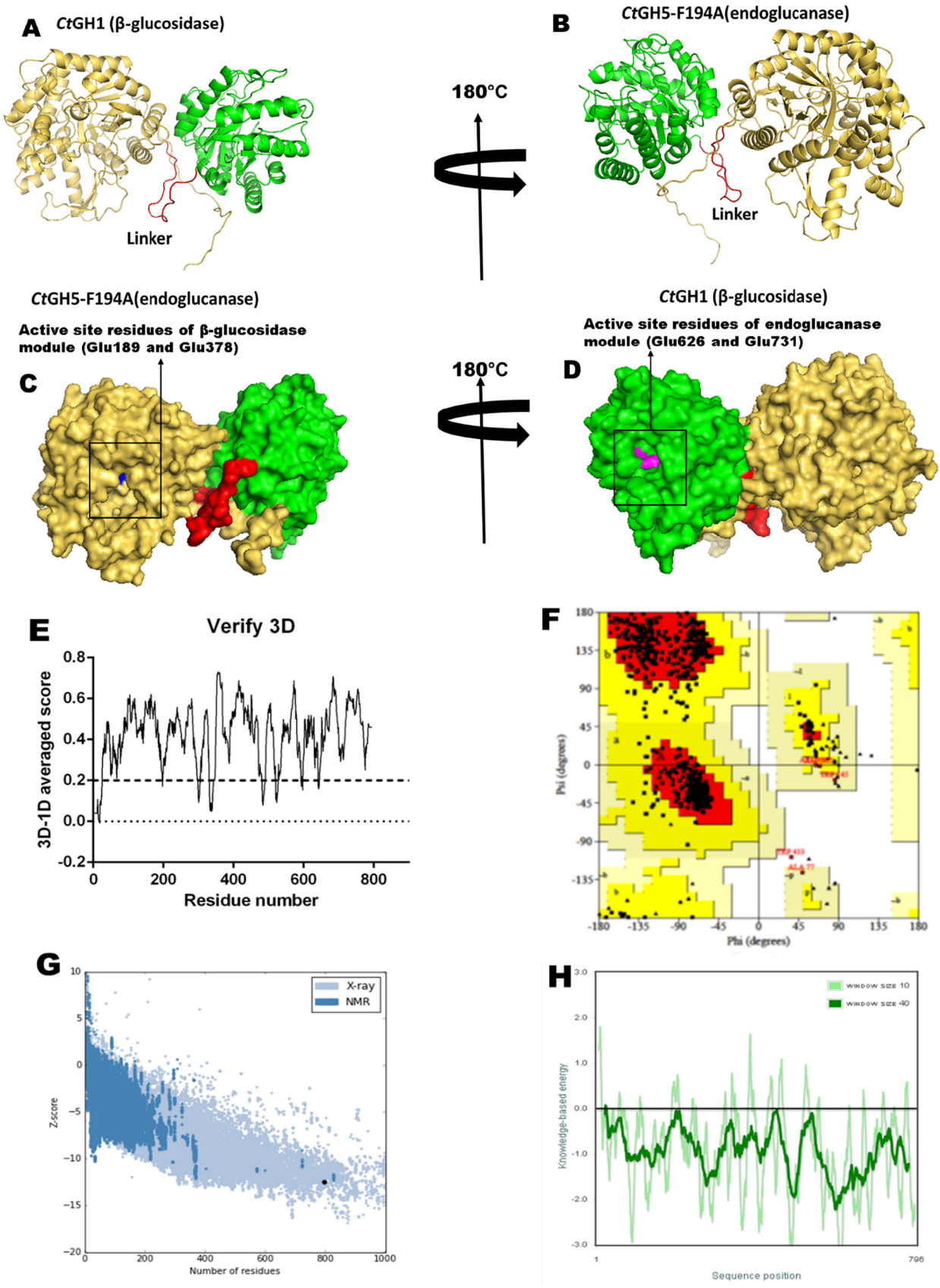


Fig. 2. A & B) Cartoon representation of 3D- modeled structure of Chimera showing two modules with N- terminal containing *Ct*GH1 (β-glucosidase) module I (Yellow) and C-terminal containing *Ct*GH5-F194A (endoglucanase) module II (green) connected by a linker (red). Molecular surface view of Chimera with active site highlighted, C) blue (β-glucosidase) and D) magenta (endoglucanase). Quality assessment of Chimera by E) Verify-3D F) Ramachandran plot of modeled Chimera structure, G) ProSA analysis and H) ProSA amino acid residue scores.

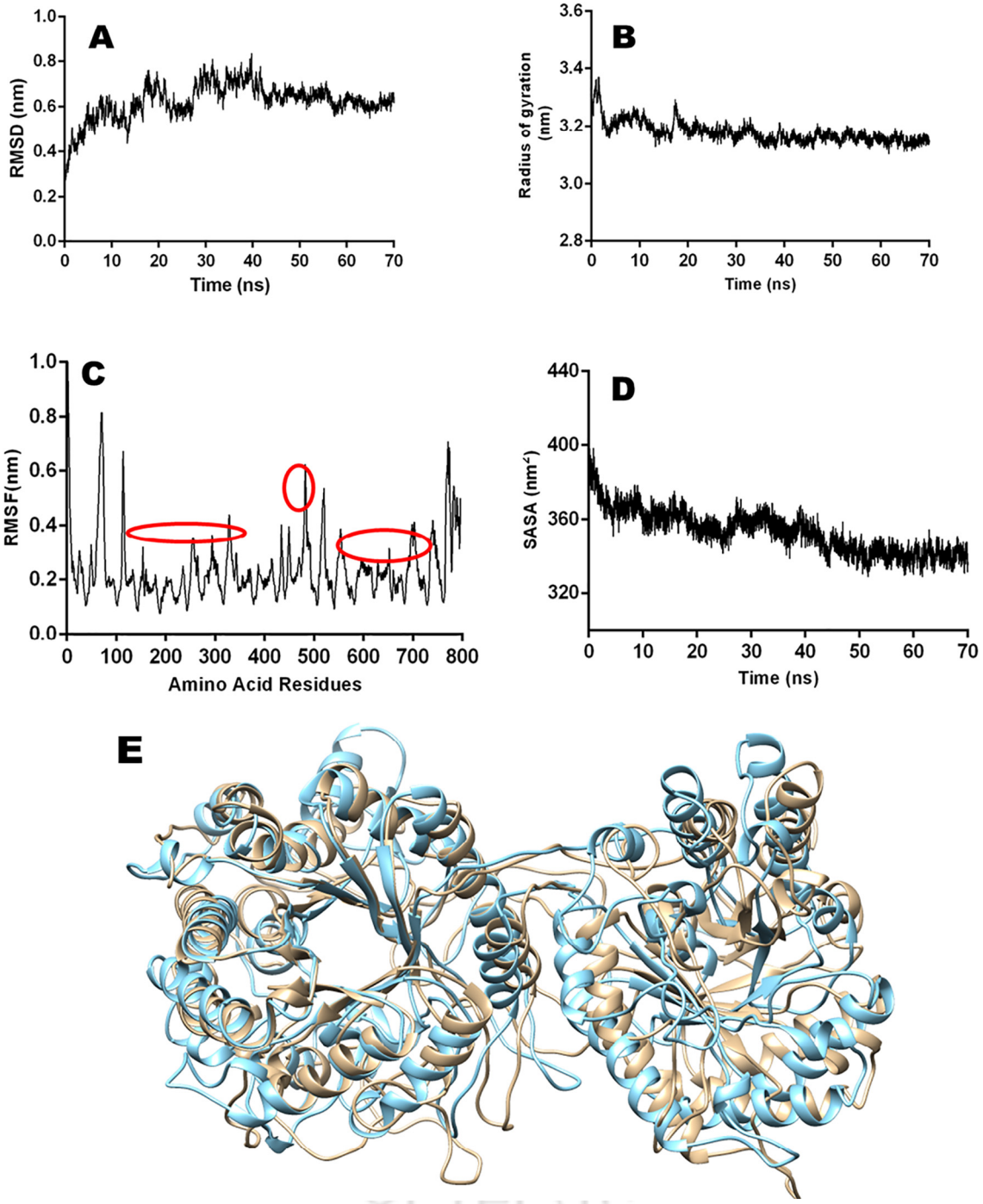


Fig. 3. Molecular dynamics simulation of Chimera A) RMSD plot, B) Radius of gyration plot, C) RMSF plot, D) SASA plot and E) Superposition of simulated structure of Chimera (cyan) with modeled Chimera structure (light brown).

cellohexaose and celloheptaose with binding free energy (ΔG) of −7.62, −7.91, −10.05, −10.86, −11.59 and −11.54 kcal/mol, respectively as shown in Table 2. This also indicated that the binding site of module 2, *Ct*GH5-F194A of Chimera can accommodate cello-oligosaccharides with 7 DP (Table 2). The docked structure of moduled Chimera and its non-covalent interactions were analyzed. The binding interaction of the *p*NPG derivative and in the catalytic cleft of the module 1 *Ct*GH1 (β-glucosidase) of Chimera showed involvement of amino acid residues Trp145, Glu189, Val192, Tyr319, Trp351 and Trp433 in hydrophobic interaction and Asn245, Leu246, Ser247, and Ser320 involved in hydrogen bond interaction (Table 2, Fig. 4A and B). Similarly, for the binding interaction of cellobiose with the catalytic cleft of the module 1 *Ct*GH1 (β-glucosidase) of Chimera showed involvement of amino acid residues Trp145, Val192, Leu200, His203, Trp351, Trp433 and Trp435 in hydrophobic interaction and Glu189 and Asn245 in hydrogen bond interaction (Table 2, Fig. 4C and Fig. 4D). The binding interaction of the hydrolysed products such as short celloliosaccharides and cellobiose in the catalytic cleft of module 2 *Ct*GH5-F194A (endoglucanase) of Chimera showed that the amino acid residues His586, His585, Trp764, Met776, Met735, His694 are involved in hydrophobic interaction, while Glu626, Glu731, Asn774 are involved in hydrogen bond formation (Table 2, Fig. 4E and Fig. 4F). The celloheptaose showed binding interaction in the catalytic cleft of the module 2, *Ct*GH5-F194A (endoglucanase) of Chimera and amino acid residues Pro689, Trp663,

Table 2
Molecular docking analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) with cello-oligosaccharides at catalytic sites of *Ct*GH1 and *Ct*GH5-F194A modules.

Cello-oligo-saccharides	β-Glucosidase (*Ct*GH1)			Endoglucanase (*Ct*GH5-F194A)		
	Binding free energy (kcal/mol)	Amino acids involved in hydrogen bond formation	Amino acids involved in hydrophobic bond formation	Binding free energy (kcal/mol)	Amino acids involved in hydrogen bond formation	Amino acids involved in hydrophobic bond formation
pNPG	−6.92	Ser247, Ser320, Leu246, Asn 245	Glu189, Val192, Trp351, Trp145, Trp433, Tyr 319	nd	nd	Nd
Cellobiose	−7.37	Glu 189, Asn 245	Leu200, Trp351, Val192, Tyr435, Trp145, Trp433, His203	−7.62	Glu626, Glu731, Asn774	His586, His585, Trp764, Met776, Met735, His694
Cellotriose	−7.62	Glu189, Glu378, Tyr319, Trp433	Glu432, Hi 203, Met349, Trp351, Leu193, Glu322, His249, Ser247, Asn 245	−7.91	His685, Asp668	Glu626, His694, Ser665, Tyr687, Trp663, Tyr690
Cellotetraose	−10.05	Glu331, Asn 245, Ser332, Glu432	Met349 Trp351, Glu189, Tyr319, Leu195, Ser335, Phe199, Leu200, Leu196, Trp433, Trp435	−10.05	His586, Glu731, His694, Asn766, Asn510	Glu626, Trp663, Tyr690, Tyr687, His694, Met776, Asn774, Trp520
Cellopentaose	−10.86	Trp433, Glu331, Ser335, Ser332, His 249	Glu189, Glu322, Ser247, Asn245, Tyr319, Val192, Glu432, Met349, Leu196, Trp351, Ala337, Leu268	−10.86	Asn766, Trp764, Asn688, Tyr690, Asn510	Asn774, Tyr687, Trp663, Tyr662, Glu626, Ser665, Trp520
Cellohexaose	−11.59	Glu331, Glu322, Glu350, Asn348	Tyr435, Ser265, Pro336, Leu200, Ser247, Ala269, Ile324, Leu268, His249, Trp351, Met349	−11.59	Glu691, Asn688, Glu626	Pro689, Trp663, Tyr690, His586, Trp520, Asn774, Val768, Asn510, Met776, Tyr687, His694
Celloheptaose	−11.13	Glu 342, Ser339, Glu32, Ala337, Asn 245	Glu189, Trp145, Tyr319, His203, Ser320, His249, Ser247, Ala267, Leu195, Leu268, Leu200, Glu331, Pro336	−11.54	Tyr690, Asp668, Glu691, Glu626, Trp764, Ser519	Asn766,Asn510, Asn774, Trp520, His586, Typ687, Trp663

nd - not determined.

Tyr690, His586, Trp520, Asn774, Val768, Asn510, Met776, Tyr687, His694 involved in hydrophobic interaction and Pro689, Trp663, Tyr690, His586, Trp520, Asn774, Val768, Asn510, Met776, Tyr687, His694 were involved in hydrogen bond interactions (Table 2, Fig. 4G and H). The docking studies of the Chimera with the ligands giving favourable binding energies suggested that both the modules *Ct*GH1 (β-glucosidase) and *Ct*GH5-F194A (endoglucanase) are active in the chimeric form and involved in the catalysis.

3.5. Molecular dynamics of ligand-bound Chimera (CtGH1-L1-CtGH5-F194A) and corresponding individual enzymes

The molecular dynamics of Chimera in the presence of ligands cellobiose or cellohexaose was performed up to 70 ns. The structure stability of Chimera in presence of the ligands was determined by analysing the time evolution of the C_α root-mean-square deviation (RMSD) and compared with the RMSD of only Chimera. The chimeric enzyme containing cellobiose in the β-glucosidase module and cellohexaose in the endoglucanase module attained the stability after 15 ns and remained stable till 70 ns with a mean RMSD value of 0.5 nm (Fig. 5A). The MD simulation of the individual endoglucanase (*Ct*GH5-F194A) in presence of cellohexaose was also undertaken for 70 ns for comparative analysis with the cellohexaose bound endoglucanase module of Chimera. The ligand bound form of individual endoglucanase (*Ct*GH5-F194A) showed stable RMSD till 70 ns with mean RMSD value of 0.3 nm (Fig. 5B). The cellohexaose in the catalytic cleft of endoglucanase module of Chimera interacts with the active site residues Glu626 and Glu731 (Fig. 5C). To determine the dynamics at the level of each residue the root mean square fluctuation (RMSF) was calculated. The active site residues Glu 626 and Glu731 showed stability with RMSF ≤ 0.2 nm for both ligand bound and unbound Chimera (Fig. 5D). The loop LA present in the endoglucanase module of the Chimera, is composed of amino acid residues ranging from 509 to 525 remains flexible during the MD simulation in both ligand bound and unbound forms of Chimera (Fig. 5D). This loop (LA) is present near the flexible linker comprising amino acid residues (487-506), which is between β-glucosidase and endoglucanase modules (Fig. 5D). The loop LA is located near the −3 binding site of the catalytic cleft of endoglucanase module of the Chimera (Fig. 5C).

The structure superposition of ligand bound Chimera from 0 ns (Blue) and after 70 ns MD simulation (Magenta), showed that the loop LA of MD simulated structure moves near the active site cleft of the endoglucanase module of Chimera and the distance of residue Asn510 between 0 ns and 70 ns structures is 4.7 Å (Fig. 5F). This relocation of the loop LA resulted in the stable hydrogen bond formation with a bond distance, ≤0.35 nm between the cellohexaose and the amino acid Asn510 in the catalytic site of endoglucanase module of Chimera as shown in 2D representation (Fig. 5H). The similar corresponding amino acid residue, Asn93 of a bifunctional cellulase/xylanase (*Ct*Cel5E) from *Clostridium thermocellum* was involved in the formation of hydrogen bond with the substrate for catalysis as earlier reported [13]. Therefore, it can be concluded that Asn510 residue present in the binding site of the endoglucanase module in Chimera is important for binding and catalysis of cellohexaose. Another loop LC present near the binding site of the endoglucanase module of chimeric enzyme comprising amino acid residues (685-689) displayed the flexibility upon ligand binding (Fig. 5D). In loop LC, the residue Tyr687 after 70 ns (Magenta) is displaced from the active site of the endoglucanase module of Chimera and the distance of the residue Tyr687 between 0 ns and 70 ns structures is 5.5 Å (Fig. 5F). The relocation of Tyr687 (corresponding residue Tyr270 for *Ct*Cel5E) was important for the catalysis, as it can facilitate the release of the product after the enzymatic cleavage by the acid/base catalysis carried out by Glu626/Glu731 (corresponding residues Glu209/Glu314 for *Ct*Cel5E) as also reported earlier [13], while the corresponding residues in individual endoglucanase are Glu139/Glu244. The RMSF for the loop LA for individual endoglucanase (*Ct*GH5-F194A) comprising amino acid residues (22-31) showed minimal flexibility in the ligand bound and unbound forms (Fig. 5E). The structure superposition of the ligand bound individual endoglucanase (*Ct*GH5-F194A) (green) with its 70 ns MD simulated ligand bound structure (Red) showed that due to the lack of flexibility in the loop LA of individual endoglucanase (*Ct*GH5-F194A), the residue Asn23 (corresponding residue Asn510 in Chimera) cannot relocate towards the active site for hydrogen bond formation (Fig. 5G). Therefore, stable hydrogen bond formation was not initiated in the catalytic cleft of individual endoglucanase (*Ct*GH5-F194A) (Fig. 5I). The 2D representation of the ligand bound cellohexaose in the individual endoglucanase (*Ct*GH5-F194A) after 70 ns of simulation also confirmed that Asn23 (corresponding Asn510 in Chimera) is not involved in hydrogen bond

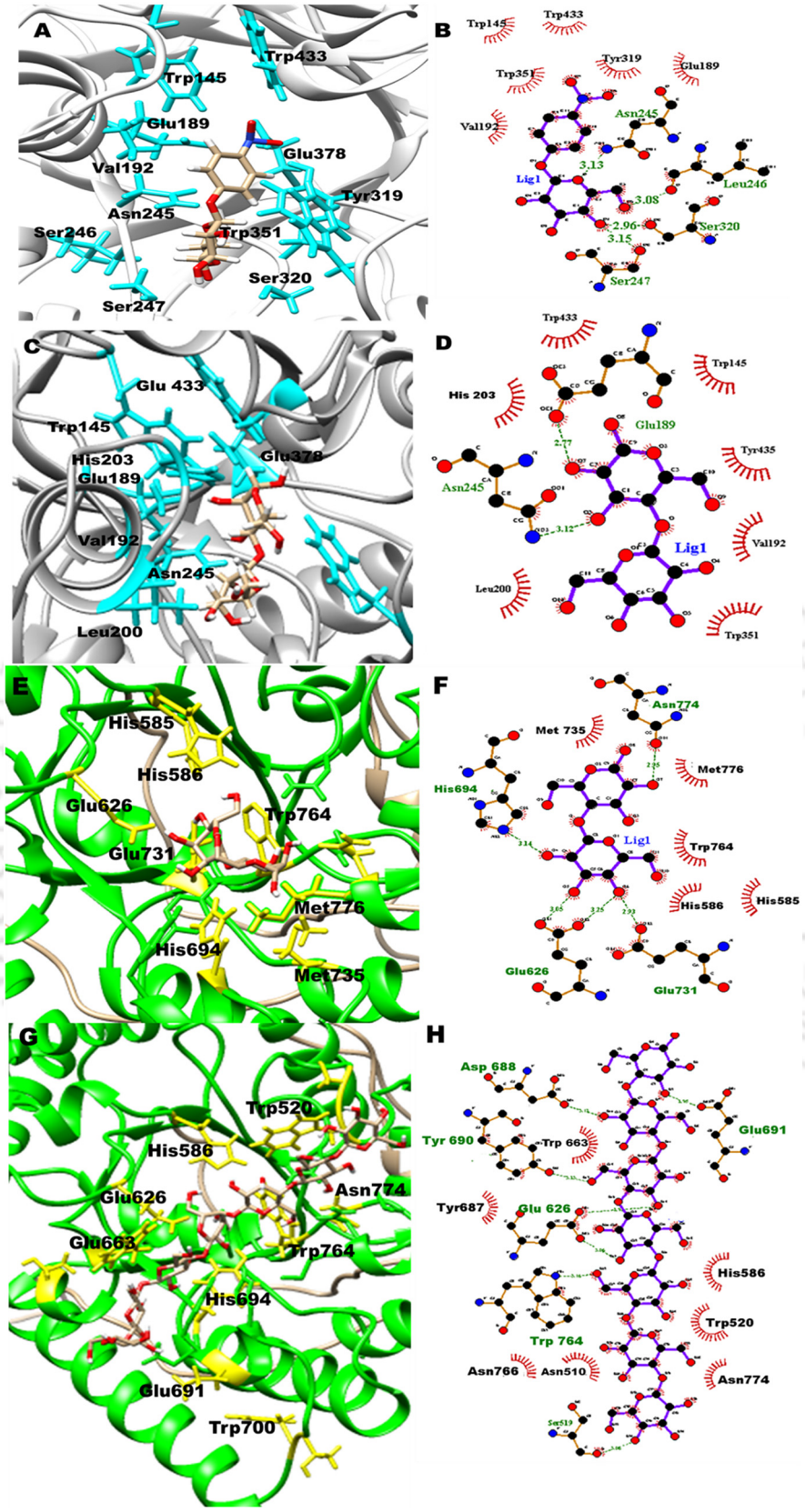


Fig. 4. Molecular docking analysis of Chimera composed of module 1, *Ct*GH1(β-glucosidase) and module 2, *Ct*GH5-F194A (endoglucanase) with ligands. A) the active-site of *Ct*GH1 (β-glucosidase) showing the interaction of *p*NPG in the catalytic cleft, B) two-dimensional schematic presentation of *p*NPG with the active site residues of *Ct*GH1(β-glucosidase), C) the active-site of *Ct*GH1 (β-glucosidase) showing interaction of cellobiose in the catalytic cleft, D) two-dimensional schematic presentation of cellobiose with active-site residues of *Ct*GH1 (β-glucosidase), E) the active-site of *Ct*GH5-F194A (endoglucanase) showing interaction of cellobiose in the catalytic cleft, F) two-dimensional schematic presentation of cellobiose with the active-site residues of *Ct*GH5-F194A (endoglucanase), G) the active- site of *Ct*GH5-F194A (endoglucanase) showing the interaction of celloheptaose in the catalytic cleft and H) two-dimensional schematic presentation of celloheptaose with the active-site residues of *Ct*GH5-F194A (endoglucanase) of the Chimera.

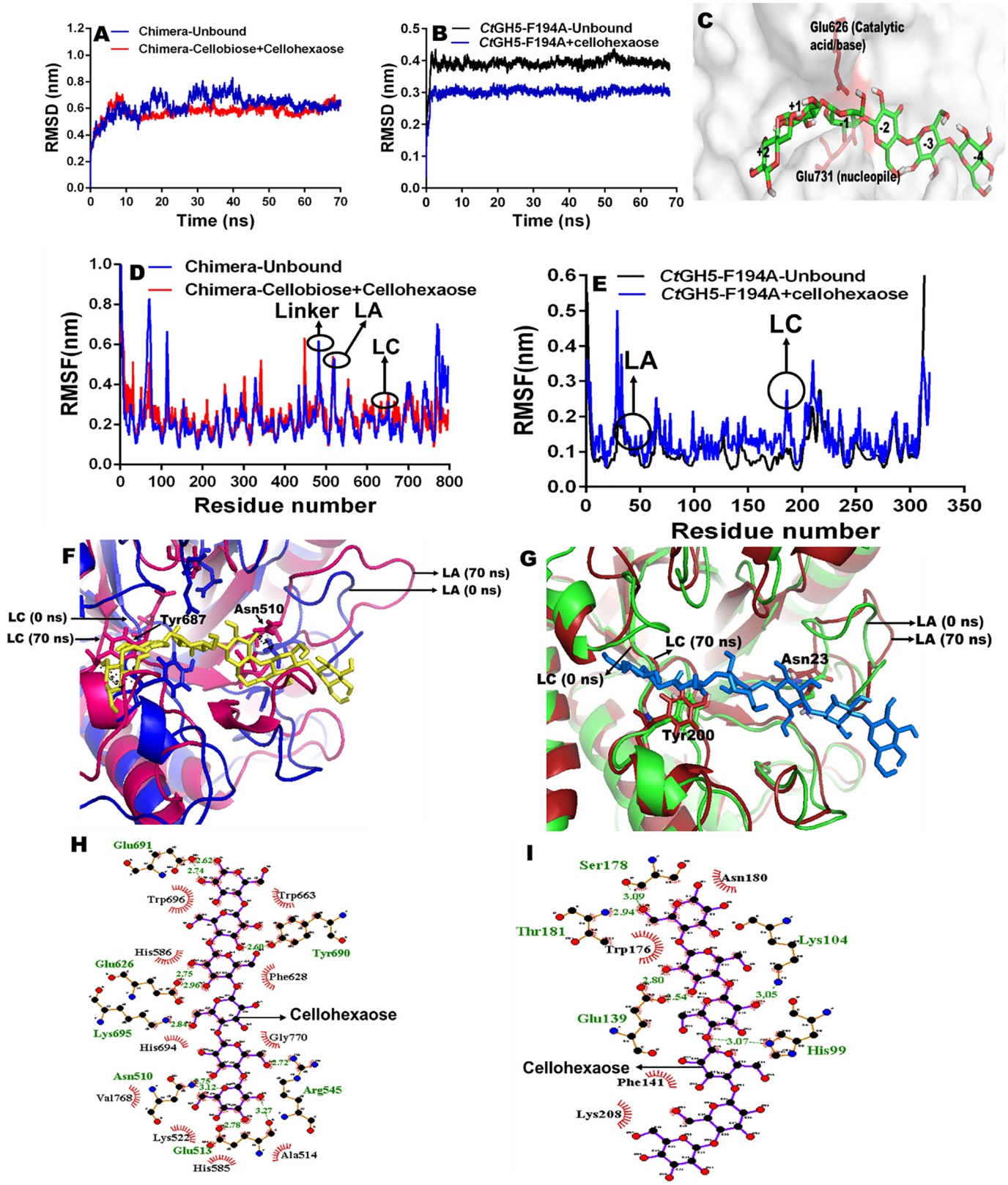


Fig. 5. Conformational dynamics for endoglucanase activity of substrate-bound Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) and individual endoglucanase (*Ct*GH5-F194A) (A) RMSD of ligand bound and unbound forms of Chimera (B) RMSD of ligand bound and unbound forms of individual endoglucanase (*Ct*GH5-F194A) (C) position of cellohexaose (green) at the active site of endoglucanase module of Chimera (D) RMSF of ligand bound and unbound forms of Chimera (E) RMSF of ligand bound and unbound forms of individual endoglucanase (*Ct*GH5-F194A) (F) Active site residues of endoglucanase module of Chimera with cellohexaose (dark yellow) after 70 ns simulation (G) The interaction of active site residues of individual endoglucanase (*Ct*GH5-F194A) with cellohexaose (blue) after MD simulation of 70 ns (H) two-dimensional schematic presentation of cellohexaose with the active-site residues of endoglucanase module (*Ct*GH5-F194A) of Chimera after MD simulation for 70 ns (I) two-dimensional schematic presentation of cellohexaose with the active-site residues of individual endoglucanase (*Ct*GH5-F194A) after MD simulation for 70 ns.

Table 3
SAXS data collection parameter and derived parameters of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A).

Data-collection parameter	Chimera (*Ct*GH1-L1-*Ct*GH5-F194A)
Instrument	SAXSpace Anton-Paar
Wavelength (Å)	1.54 Å
Q range (nm^{-1})	0.135–5.95
Exposure time (min)	30 × 2
Temperature (°C)	10
Protein concentration (mg/mL)	5
Structural parameter	
Q range (nm^{-1}) used for Rg analysis	0.10–0.41
I(0) au from Guinier	57,810 ± 693
Rg nm from Guinier	3.13 ± 0.1
I(0) au from P(r)	57,810
Rg nm from P(r)	3.12 ± 0.07
D_{max} (nm)	10.0
Porod volume estimate (nm^3)	120.82
Molecular Mass Determination	
Theoretical molecular mass (kDa)	89
Molecular Mass from SAXSMOW (kDa)	87
Modeling parameters	
Q range (nm^{-1}) used for structure modeling	0.12–2.53
Resolution (nm)	33 ± 0.5
NSD	0.93
Software employed	
Data processing	Primus
P(r) function calculation	GNOM
Ab initio modeling	GASBOR
Validation and averaging	DAMAVER
Structure superposition	SUPCOMB
3-D graphical representation	PyMOL

formation with cellohexaose (Fig. 5I). The loop LC comprising the amino acid residues (197-201) in the individual endoglucanase (*Ct*GH5-F194A) showed the flexibility, when the cellohexaose is present in the structure (Fig. 5E). The superposition of cellohexaose bound individual endoglucanase (*Ct*GH5-F194A) at 0 ns (green) with 70 ns MD simulated structure (red) demonstrated the restriction in the movement of Tyr200 residue (corresponding residue Tyr 687 in Chimera) from the active site (Fig. 5G). This restricted movement of residue Tyr200 in individual endoglucanase (*Ct*GH5-F194A) might be causing delayed product release from its active site. Therefore, its catalytic efficiency is lower than the endoglucanase module of the Chimera. The structural comparison after MD simulation concluded that the flexibility of the loops LA and LC and their relocation efficiency in the endoglucanase module of the Chimera (*Ct*GH5-L1-*Ct*GH5-F194A) is higher than the individual endoglucanase (*Ct*GH5-F194A). This favourable flexibility in the loops of endoglucanase module of Chimera resulted in stable ligand binding and efficient release of product after catalysis from the catalytic cleft.

For further analysis, the average short range interaction energy was calculated for ligand bound Chimera and its individual enzymes. The interaction energy of endoglucanase module of Chimera with cellohexaose was −383 kJ/mol, while the individual endoglucanase (CtGH5-F194A) showed interaction energy of −254 kJ/mol with cellohexaose after the MD simulation for 70 ns. Therefore, taking into consideration the MD simulation and the short range interaction energy calculation, a conclusion can be drawn that the endoglucanase module of Chimera is more stable for interaction with the ligands in its catalytic cleft. Therefore, this may lead to efficient catalysis of the ligand. Moreover, the β-glucosidase module of Chimera also showed sufficiently strong interaction with cellobiose giving an average short range interaction energy of −64 kJ/mol. Similarly, the individual β-glucosidase (*Ct*GH1) showed no substantial change in the interaction energy. Furthermore, no significant changes in the loops, near the binding site of both β-glucosidase module of Chimera and its individual β-glucosidase (*Ct*GH1) after the MD simulation for 70 ns were found (data not shown). Although, the MD simulated ligand bound form of Chimera showed minor fluctuations in the outer loops of β-glucosidase module after 70 ns (data not shown). This fluctuation in the outer loops might be responsible for the conformational change that might be causing improved catalysis in β-glucosidase module of the Chimera. Therefore, from the MD simulation study of the interaction of Chimera and its individual enzymes with ligands, it was found that endoglucanase module of Chimera shows stable binding with cellohexaose and efficient release of the hydrolysed product (cellobiose) followed by the action of fused β-glucosidase module of Chimera with changed conformation leading to its improved catalytic efficiency. Therefore, all these favourable interactions made this Chimera a potent enzyme for hydrolysis of cellulosic substrates into glucose.

The overall MD simulation studies amply demonstrated that, the two modules in the Chimera, are present without any unfavourable inter-modular interactions and are in active form and independently functioning. The property of a natural linker for preventing unfavourable inter-modular interactions was also reported earlier [42]. Therefore, the independent functioning of the two modules might be because of the natural linker, that has been used for connecting the two modules. The MD simulation results of protein-ligand complex displayed that the relocation and orientation of the ligand binding site loops LA and LC in endoglucanase module in Chimera plays a key role, that might be responsible for enhanced catalytic efficiency.

3.6. Small angle X-ray scattering analysis of Chimera

The small angle X-ray scattering analysis of Chimera showed the conformational behaviour and molecular arrangement of both the linked modules in the solution form. The SAXS data of Chimera at both concentrations (3 mg/mL and 5 mg/mL) were processed, and the results were given in Table 3. The initial processing and through visual examination of the scattering pattern generated by the Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) for different concentrations revealed monodisperse nature for the Chimera in solution (Fig. 6A). The SAXS profiles at both concentrations were similar and lacked the aggregation or inter-particle interaction effect. The SAXS data collected at 5 mg/mL was used for further processing and analysis. Guinier estimation analysis demonstrated that the radius of gyration (Rg) for Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) was 0.57 ± 0.05 nm for rod-shape and 3.15 ± 0.10 nm for globular shape. The length of the Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) molecule was 10.7 nm. The linearity of Guinier plot at low q region unveiled that the Chimera exists in a monodispersed state in the solution form (Fig. 6B). The evaluation of P(R) function acquired by Fourier transformation of the Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) scattering profile displayed the symmetric profile (Fig. 6C). The P(R) plot in Fig. 6(C) displayed a shoulder signifying that there is an inter-modular separation between the two modules. The maximum dimension (D_{max}) and the radius of gyration (Rg) of Chimera estimated from P(R) plot were found to be 10 nm and 3.17 nm, respectively (Fig. 6C). The global compactness and flexibility in solution for the Chimera were determined by Kratky plot analysis (Fig. 6D). The analysis using Kratky plot for the Chimera displayed bell-shape peaks at low Q region, confirming a compact and folded structure of Chimera. The molecular mass of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) procured from SAXSMow server employing SAXS scattering profile was 87 kDa, which is similar to the theoretical and experimental molecular mass (89 kDa) as described earlier [14]. The averaged Gasbor generated *ab initio* models for Chimera displayed two-module structure (Fig. 6E). The Gasbor modeled structure of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) displayed an elongated structure with two modules existing as independent units

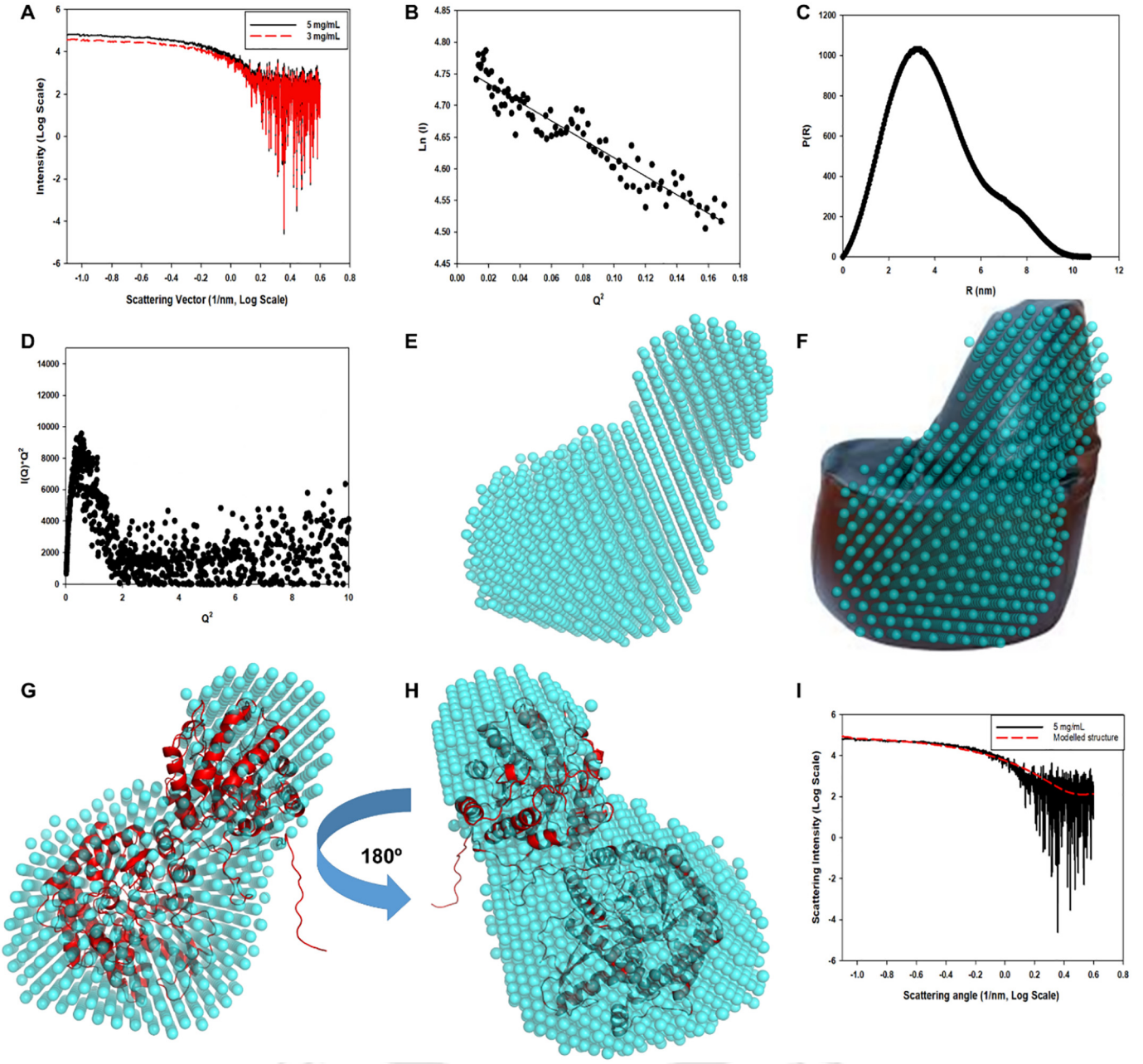


Fig. 6. Small angle X-ray scattering analysis of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A). A) SAXS intensity profile, (B) Guinier plot of the SAXS intensities, C) P(R) curve of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) as a function of R, (D) Kratky plot, (E) *ab initio* derived shape of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A), (F) Gasbor generated model superposed on bean-bag shape, (G & H) Superposition of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) modeled structure with the *ab initio* model Gasbor derived structure and (I) CRYSOL analysis of modeled structure and experimentally collected SAXS data.

with no inter-modular contact, and having an overall shape similar to a bean-bag contour (Fig. 6F). The superposition of modeled structure of Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) with Gasbor derived *ab initio* model from the collected SAXS data, displayed an excellent fitting with a minor deviation in the hexa-histidine tag region at the N-terminal (Fig. 6G & H). The phenomenon of independent existence of different modules in a multi-modular ubiquitin protein with no inter-modular contact by *ab initio* shape reconstruction using SAXS studies was also reported earlier [43]. The CRYSOL fitting of the Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) modeled structure with the SAXS experimental data (Fig. 6I) also corroborated with the results obtained by SAXS data processing. Therefore, the SAXS data supported the results of MD simulation as the two modules of Chimera displayed no interaction between them. Therefore, the results clearly showed that the two modules of Chimera are functionally independent and are in active form as also elucidated earlier by biochemical studies [14].

4. Conclusion

The amino acid sequence of the Chimera (*Ct*GH1-L1-*Ct*GH5-F194A) composed of β-glucosidase (*Ct*GH1) and endoglucanase (*Ct*GH5-F194A) was evaluated for secondary structure analysis by using Psi-Pred and RaptorX prediction tools and compared with the CD data. The CD analysis revealed that the secondary structure of Chimera consists of 38% of α-helix, 9.32% of β-sheet and 52.68% of the loops corroborating with the predicted secondary structure elements. The 3-dimensional modeled structure of Chimera was generated by RaptorX, which showed close homology with β-glucosidase, 3AHX from *Clostridium cellulovorans* and an

endoglucanase, 5BYW from *Clostridium thermocellum*. The modeled structure generated was validated by Ramachandran plot, that showed 99.9% amino acid residues in the allowed region. The modeled structure of Chimera represented a modular structure consisting of N-terminal β-glucosidase (*Ct*GH1) and C-terminal endoglucanase (*Ct*GH5-F194A) modules connected by the linker. The overall modular structure of Chimera represented a classical (α/β)-TIM barrel fold. The molecular dynamics study of the Chimera showed stable RMSD till 70 ns with 0.01 oscillation suggesting the stable conformation of the two modules, *Ct*GH1 (β-glucosidase) and *Ct*GH5-F194A (endoglucanase) in the chimeric form. The average SASA of the Chimera remained stable with an average value of 320 nm^2. The minor variation in the SASA also suggested that the accessibility of the substrate in the catalytic sites were not affected in the designed Chimera. This indicated that the binding sites were not obstructed by deformation or steric effect between the two modules of chimeric enzyme. The docking studies of the Chimera suggested that both the modules *Ct*GH1(β-glucosidase) and *Ct*GH5-F194A (endoglucanase) are active in chimeric form and involved in the catalysis. The molecular dynamics simulation of the docked complex of Chimera showed stable interaction of its β-glucosidase and endoglucanase modules with ligands. MD simulation results of cellohexaose bound and unbound endoglucanase module of Chimera displayed favourable flexibilities in the loops resulting in stable binding and efficiently releasing the product from the catalytic cleft after catalysis. Higher short range interaction energy −383 kJ/mol obtained for cellohexaose bound Chimera for endoglucanase activity displayed stronger interactions while the individual endoglucanase enzyme showed −254 kJ/mol against cellohexaose thereby resulting efficient binding and hydrolysis of cellohexaose into cellobiose.

MD simulation results showed that the loops near the binding site of β-glucosidase module of Chimera and individual β-glucosidase showed no significant changes. However, the outer loops of the β-glucosidase module of Chimera showed fluctuations. These fluctuations showed that there might be conformational changes in the β-glucosidase module of Chimera that is contributing to improved hydrolysis of cellobiose to glucose. These favourable interactions of both endoglucanase and β-glucosidase modules of Chimera against the ligands led to the improved catalytic efficiency of Chimera. The SAXS analysis of Chimera displayed elongated structure with two modules in fully folded form in the solution and an overall shape similar to a bean-bag contour. The outcome of this study of chimera by computational and SAXS data studies revealed that the peptide linker keeps the two modules of chimera independent for their biological function.

CRediT authorship contribution statement

Priyanka Nath: Methodology, Formal analysis, Writing - original draft, Writing - review & editing.**Kedar Sharma:**Formal analysis, Writing - original draft, Writing - review & editing.**Krishan Kumar:**Investigation.**Arun Goyal:**Conceptualization, Methodology, Writing - original draft, Writing - review & editing.

Acknowledgements

Authors acknowledge the financial support provided by DBT-PAN-IIT Grant (BT/EB/PAN IIT 2012), Centre for Bioenergy from Department of Biotechnology, Ministry of Science and Technology, New Delhi, India to Prof. Arun Goyal. The fellowship to Priyanka Nath was supported by funding from DST Inspire fellowship from Department of Science and Technology, New Delhi, India. The authors are also thankful to Dr. Ashish from the Institute of Microbial Technology, Chandigarh, India, for providing SAXS Facility. The authors are thankful to Mr. Prasad Gosavi and Ms. Navjot Kaur Saini, Anton Paar India for collecting the SAXS data at Institute of Microbial Technology, Chandigarh, India.

Declaration of competing interest

The authors declare that they have no conflicts of interest with the content of this article.

References

[1] P. Beguin, Molecular biology of cellulose degradation, Annu. Rev. Microbiol. 44 (1990) 219–248.
[2] B.R. Urbanowicz, A.B. Bennett, E. del Campillo, C. Catala, T. Hayashi, B. Henrissat, T.T. Teeri, Structural organization and a standardized nomenclature for plant endo-1,4-β-glucanases (cellulases) of glycosyl hydrolase family 9, Plant Physiol. 144 (2007) 1693–1696.
[3] B.K. Barr, Y.L. Hsieh, B. Ganem, D.B. Wilson, Identification of two functionally different classes of exocellulases, Biochemistry 35 (1996) 586–592.
[4] F. Grabnitz, M. Seiss, K.P. RucknageL, W.L. Staudenbauer, Structure of the β-glucosidase gene bglA of *Clostridium thermocellum*: sequence analysis reveals a superfamily of cellulases and β-glycosidases including human lactase/phlorizin hydrolase, Eur. J. Biochem. 200 (1991) 301–309.
[5] L.S. McKee, M.J. Pena, A. Rogowski, A. Jackson, R.J. Lewis, W.S. York, J. Marles-Wright, Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 6537–6542.
[6] P. Lu, M.G. Feng, Bifunctional enhancement of a β-glucanase-xylanase fusion enzyme by optimization of peptide linkers, Appl. Microbiol. Biotechnol. 79 (4) (2008) 579–587.
[7] Z. Fan, J.R. Werkman, L. Yuan, Engineering of a multifunctional hemicellulose, Biotechnol. Lett. 31 (2009) 751–757.
[8] C.J. Crasto, J.A. Feng, LINKER: a program to generate linker sequences for fusion proteins, Protein Eng. 13 (2000) 309–312.
[9] J.N. Onuchic, P.G. Wolynes, Theory of protein folding, Curr. Opin. Struc. Biol. 14 (2004) 70–75.
[10] S. Yang, L. Blachowicz, L. Makowski, B. Roux, Multidomain assembled states of Hck tyrosine kinase in solution, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 15757–15762.
[11] J. Cota, L.C. Oliveira, A.R. Damásio, A.P. Citadini, Z.B. Hoffmam, T.M. Alvarez, M.T. Murakami, Assembling a xylanase–lichenase Chimera through all-atom molecular dynamics simulations, Biochim. Biophys. Acta 1834 (2013) 1492–1500.
[12] M.A. Jamros, L.C. Oliveria, P.C. Whitford, J.N. Onuchic, J.A. Adams, D.K. Blumenthal, P.A. Jennings, Proteins at work: a combined SAXS and theoretical determination of the multiple structures involved on the protein kinase functional landscape, J. Biol. Chem. 285 (2010) 36121–36128.
[13] S.F. Yuan, T.H. Wu, H.L. Lee, H.Y. Hsieh, W.L. Lin, B. Yang, M.C. Ho, Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*, J. Biol. Chem. 290 (2015) 5739–5748.
[14] P. Nath, A. Dhillon, K. Kumar, K. Sharma, S.B. Jamaldheen, V.S. Moholkar, A. Goyal, Development of bi-functional chimeric enzyme (*Ct*GH1-L1-CtGH5-F194A) from endoglucanase (*Ct*GH5) mutant F194A and β-1,4-glucosidase (*Ct*GH1) from *Clostridium thermocellum* with enhanced activity and structural integrity, Bioresour. Technol. 282 (2019) 494–501.
[15] K. Sharma, A. Thakur, R. Kumar, A. Goyal, Structure and biochemical characterization of glucose tolerant β-1,4 glucosidase (*Ht*Bgl) of family 1 glycoside hydrolase from *Hungateiclostridium thermocellum*, Carbohydr. Res. 483 (2019), 107750.
[16] E.J. Taylor, A. Goyal, C.I. Guerreiro, J.A. Prates, V. Money, N. Ferry, H.J. Gilbert, How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides. Structure and activity of a *Clostridium thermocellum* lichenase, *Ct*Lic26A, J. Biol. Chem. 280 (2005) 32761–32767.
[17] M.N. Collins, E. Dalton, B. Leahy, J.J. Leahy, C. Birkinshaw, Crystal morphology of strained ultra-high molecular weight polyethylenes, Polym. Test. 31 (2012) 629–637.
[18] E. Dalton, M.N. Collins, Lamella alignment ratio: a SAXS analysis technique for macromolecules, J. Appl. Crystallogr. 47 (2014) 847–851.
[19] S.M. Kelly, T.J. Jess, N.C. Price, How to study proteins by circular dichroism? Biochim. Biophys. Acta 1751 (2005) 119–139.
[20] M.A. Andrade, P. Chacon, J.J. Merelo, F. Moran, Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network, Protein Eng., Des. Sel. 6 (1993) 383–390.
[21] J. Peng, J. Xu, RaptorX: exploiting structure information for protein alignment by statistical inference, Proteins 79 (2011) 161–171.
[22] M. Kallberg, G. Margaryan, S. Wang, J. Ma, J. Xu, RaptorX server: a resource for template-based protein structure modelling, Protein Structure Prediction, Humana Press, New York, NY 2014, pp. 17–25.
[23] 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. 2 (2004) 1605–1612.
[24] H.J. Berendsen, D. van der Spoel, van R. Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56.
[25] B. Hess, H. Bekker, H.J. Berendsen, J.G. Fraaije, LINCS: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472.
[26] K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. González Outeiriño, C.R. Daniels, B.L. Foley, R.J. Woods, GLYCAM06: a generalizable biomolecular force field. Carbohydrates, J. Comput. Chem. 29 (2008) 622–655.
[27] L.L.C. Schrodinger, The PyMOL Molecular Graphics System, Version 2.0, 2019.

[28] K. Sharma, I.L. Antunes, V. Rajulapati, A. Goyal, Low-resolution SAXS and comparative modeling based structure analysis of endo-β-1, 4-xylanase a family 10 glycoside hydrolase from *Pseudopedobacter saltans* comb. nov, Int. J. Boil. Macromol. 112 (2018) 1104–1114.
[29] 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, J. Appl. Crystallogr. 50 (2017) 1212–1225.
[30] G. Fournet, A. Guinier, in: C.B. Walker, K.L. Yudowitch (Eds.), Small Angle Scattering of X-Rays, John Wiley & Sons, New York 1955, pp. 7–78.
[31] P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H. Koch, D.I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis, J. Appl. Crystallogr. 36 (2003) 1277–1282.
[32] D.I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495–503.
[33] H.D.O.N. Fischer, M.D. 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.
[34] V.V. Volkov, D.I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering, J. Appl. Crystallogr. 36 (2003) 860–864.
[35] M.B. Kozin, D.I. Svergun, Automated matching of high-and low-resolution structural models, J. Appl. Crystallogr. 34 (2001) 33–41.
[36] D.I.B.C. Svergun, C. Barberato, M.H. Koch, CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, J. Appl. Crystallogr. 28 (1995) 768–773.
[37] C. Louis Jeune, M.A. Andrade Navarro, C. Perez Iratxeta, Prediction of protein secondary structure from circular dichroism using theoretically derived spectra, Proteins 80 (2012) 374–381.
[38] A. Akasako, M. Haruki, M. Oobatake, S. Kanaya, Conformational stabilities of *Escherichia coli* RNase HI variants with a series of amino acid substitutions at a cavity within the hydrophobic core, J. Biol. Chem. 272 (1997) 18686–18693.
[39] D.N. Woolfson, Core-directed protein design, Curr. Opin. Struc. Biol. 11 (2001) 464–471.
[40] L.P. Meleiro, J.C.S. Salgado, R.F. Maldonado, S. Carli, L.A.B. Moraes, R.J. Ward, R.P.M. Furriel, Engineering the GH1 β-glucosidase from *Humicola insolens*: insights on the stimulation of activity by glucose and xylose, PLoS One 12 (2017), 0188254.
[41] H. Cao, X. Li, Y. Zhang, P. Shi, Y. Bai, B. Yao, Glucose-tolerance molecular modification of GH1 β-glucosidase from *Alicyclobacillus* sp. A4, J. Agric. Food Chem. (Beijing) 20 (2018) 26–33.
[42] V.P. Reddy Chichili, V. Kumar, J. Sivaraman, Linkers in the structural biology of protein–protein interactions, Protein Sci. (2013) 153–167.
[43] P. Bernado, Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering, Eur. Biophys. J. 39 (2010) 769–780.