STUDIES ON β-GLUCOSIDASE AND ACID PHOSPHATASE ENZYMES OF HYDROLASE FAMILY

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

BIBEKANANDA KAR

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE - 247 667 (INDIA)

MAY, 2015

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled, **"STUDIES ON β-GLUCOSIDASE AND ACID PHOSPHATASE ENZYMES OF HYDROLASE FAMILY"** in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during the period from December, 2009 to May, 2015 under the supervision of Dr. Ashwani Kumar Sharma, Associate professor, Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree to this or any other institute.

Bibekananda Kar
(BIBEKANANDA KAR)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dated: 08-05-2015

(Ashwani Kumar Shrama)
(Supervisor)

The PhD Viva-Voce Examination of **Mr. Bibekananda Kar**, Research Scholar has been held on.....27.8.2015

Supervisor

Chairman SRC

External Examiner

Head of the Department & Chairman ODC

ABSTRACT

Proteins are highly complex most abundant biological macromolecules common to all life present on earth today and they are responsible for most of the complex functions that make life possible. All the forms of life i.e. plant, animal, fungi, bacteria, virus contain thousands of proteins for proper functioning. In modern classification proteins are classified in three major classes; storage proteins, structural proteins and biological active proteins and some of these proteins also play combined roles. The metabolic proteins mainly comprised of enzymes. Enzymes are awesome machines with a suitable level of complexity. They are central to every biochemical process and have extraordinary catalytic power. Among six different classes of enzymes, 'Hydrolase' is one of the most important group of enzymes which catalyze the hydrolysis of a chemical bond like C-C, C-O, C-N, ether, ester and halide bonds etc. Hydrolases are classified into several subclasses, based upon the bonds they act upon e.g. glycosyl hydrolase, esterase, phosphatase, lipase, DNA glycosylases etc.

In the hydrolase family, glycosyl hydrolases and phosphatases are two important groups of enzymes as they play very crucial role in cell metabolism. We have studied two enzymes from each of these two groups. First enzyme is a β-glucosidase from family 1 glycosyl hydrolase. Glycosyl hydrolases or Glycoside hydrolases or Glycosidases (GHs; EC 3.2.1.x) catalyze the hydrolysis of O-, N- or S-linked glycosidic bonds between a carbohydrate and non-carbohydrate moiety or between two carbohydrates. The cleavage of these glycosidic bonds is crucial for the processes like hydrolysis of structural polysaccharides during penetration of pathogens, recycling of cell surface carbohydrates, defense against pathogens expansion of cell wall, energy uptake, starch metabolism, symbiosis and recycling of signaling molecules etc. On the basis of amino acid sequence and structural analogy, at present there are 133 Glysoside hydrolase families available on the CAZY web server (URL- http://www.cazy.org/). Among them, family 1 glycosyl hydrolase is most studied. Glycoside hydrolases family 1 contains different enzymes with some well-known functions, such as β-glucosidase (EC 3.2.1.21), β-fucosidase (EC 3.2.1.38), β-galactosidase (EC 3.2.1.23), β-mannosidase (EC 3.2.1.25), exo-β-1, 4-glucanase (EC 3.2.1.74) etc. β-glucosidase characterized till date fall primarily in glycoside hydrolase families 1, 3 and 5 with family 1 β-glucosidases being more abundant in plants. β-glucosidase of this family may have high specificity for glucosides or in addition to this may hydrolyse fucosides and/or galactosides.

Carbohydrates are the most abundant biomolecules on earth and also in plant cells. The controlled regulation of synthesis, breakdown and modification of these macromolecules in nature is one of the most fundamentally important processes. This carbohydrate metabolism is only possible due to some enzymes i.e. glycosyl hydrolases (hydrolyzes and/or rearrange the glycosidic bonds), glycosyl transferases (build glycosidic bonds), polysaccharide lyases (non-hydrolytic cleavage of glycosidic bonds) and polysaccharide lyases (hydrolyzes carbohydrate esters). The importance of these processes can be assessed by the fact that 1-2% of an organism's total gene dedicates for the synthesis of glycosyl hydrolases and glycosyl transferases alone.

The second enzyme, we have studied is a phosphatase from HAD superfamily. Like glycosyl hydrolase HAD superfamily (HADSF) hydrolases are also very important in their functional point of view. The HADSF is the major family found with almost 48,000 sequences and is present ubiquitously in the living cells. HADSF constitute different enzymes like dehalogenase, phosphatase, phosphonatase, β-phosphoglucomutase and ATPases. Phosphatase forming the majority in HADSF can vary in figure from 30 in prokaryotes to 200-300 in the eukaryotes. They help in phosphorous transfer reactions by transferring the phosphate group to an Asp as an active site residue.

In the present work, a β-glucosidase enzyme from *Putranjiva roxburghii* plant (PRGH1) was characterized, cloned and expressed in both prokaryotic and eukaryotic system. A comparison study between the native enzyme and bacterial expressed recombinant enzyme has been carried out to understand the possible role of post translational modifications on the enzyme. Several mutation studies were done to locate the possible active site. Meticulous bioinformatics work was carried out to understand the substrate preference of this enzyme. The possible role of N-linked glycosylation in the stability and the activities of the enzyme were examined. We have used this enzyme in bioethanol production by expressing this enzyme in eukaryotic system. Along with this enzyme we have characterized another hydrolase enzyme from *Staphylococcus lugdunensis*. This enzyme belongs to HADSF and showed phosphatase activity.

This thesis is divided into five chapters and covers the studies carried out on two hydrolases. The first one is a glycosyl hydraolase sourced from the *Putranjiva roxburghii* plant and the second one is a phosphatase from HAD family sourced from *Staphylococcus lugdunensis*, which is a human pathogen.

Chapter **1** reviews the literature; describing the recent studies on family 1 Glycosyl hydrolase (GH1), role of N-linked glycosylation, application of β-glucosidase in bioethanol production and studies on HADSF.

Chapter **2** describes the cloning, overexpression, purification and characterization of a glycosidase 1 enzyme. A 66 kDa, thermostable enzyme with β-fucosidase, β-glucosidase and β-galactosidase activities was purified from the seeds of *Putranjiva roxburghii* by employing concanavalin-A affinity chromatography. The deduced amino acid sequence showed considerable similarity with plant β-glucosidases. The enzyme hydrolyzes *p*-nitrophenyl-β-D-glucopyranoside (*p*NP-Fuc) with higher efficiency ($K_{cat}/K_m = 3.79 \times 10^4\ M^{-1}s^{-1}$) as compared to *p*NP-Glc ($K_{cat}/K_m = 2.27 \times 10^4\ M^{-1}s^{-1}$) and *p*NP-Gal ($K_{cat}/K_m = 1.15 \times 10^4\ M^{-1}s^{-1}$). Both the native and recombinant protein has pH optima of 4.8. The thermostability of the recombinant enzyme is much lower than native enzyme. Mutational study showed that disruption of active site residues affect the activity of the enzyme. Oligomerization study showed that at higher concentration the enzyme form various species of oligomers having molecular mass around a decamer. Mixed substrate analysis showed that all the three activities (glucosidase, galactosidase and fucosidase) of the enzyme were performed by a single active site. This preference for the substrate has also been studied and proved using meticulous bioinformatics work.

Chapter **3** describes the role of glycosylation for the stability and consequence effect on the activity of a very efficient thermostable β-glucosidase from *Putranjiva roxburghii* plant (PRGH1). We successfully produced deglycosylated PRGH1 by using PNGase F. We compared the activities of both forms of this enzyme under various conditions like different temperature, pH and organic solvents. At higher pH the deglycosylated PRGH1 showed a sharp decrease in activity. The temperature profile of both the glycosylated and deglycosylated enzyme clearly reflect that glycosylated form of this enzyme have greater stability at higher temperature. CD and intrinsic fluorescence studies of both glycosylated and deglycosylated enzyme showed that the conformation of the native protein changed to a certain extent after removal of the N-linked sugars. Proteolysis study along with the spectral studies suggests that the structure of native glycosylated PRGH1 is quite compact and rigid than the deglycosylated counterpart. Mutagenesis studied shows that out of seven potential glycosylation site three sites were glycosylated.

Chapter **4** describes the expression of a plant β-glucosidase gene (PRGH1) in the *S. cerevisiae* Y294 and studies on the properties of the enzyme. The optimal pH and temperature of the enzyme activity was found to be 5.0 and 65 °C respectively indicating this enzyme is a

thermostable enzyme with preference towards moderate acidic condition. The enzyme showed broad substrate specificity and able to hydrolyze cellobiose significantly. The enzyme showed resistance towards alcohols, suggesting the enzyme can be used in fermentation industry more efficiently. The recombinant *S. cerevisiae* Y294 harbouring *prgh1* gene showed better growth profile, cellobiose consumption and ethanol production. In addition to this, complementing with commercial cellulose enzyme the recombinant *S. cerevisiae* Y294 was used in SSF experiment using CMC, rice straw, sugarcane bagesse as sole carbon source. The study demonstrated the feasibility of using the β-glucosidase gene to enhance the second generation cellulosic ethanol production.

Chapter **5** describes the cloning, overexpression and characterization of a acid phosphatase from Haloacid dehalogenase superfamily. SHFD gene with ~840 bp has been cloned and overexpressed in *E.coli*. The enzyme with a molecular mass of ~32 kDa has been purified using Ni^{2+}-NTA affinity chromatography. SHFD showed phosphatase activity with an optimum temperature of 25 °C. SHFD is an acid phosphatase with an optimum pH of 5.0. The kinetic parameters (K_m= 0.32 mM, V_{max} = 0.36 U/mg, k_{cat} = 21.43 ± 0.85 s^{-1} and k_{cat}/K_m = 66.96 $mM^{-1}s^{-1}$) indicate that it is a very efficient enzyme. SHFD is a mixed α/β protein as predicted by the ESPript and CD spectrum. Multiple sequence alignment shows the conservation of nucleophilic Asp10, acid/base catalyst Asp12, phosphate binding Ser43, and many other catalytic residues like Arg45, Lys210, Asp233, Asn236 and Asp237. SHFD is a two domain protein, with a larger core domain comprising four conserved loops surrounding the active site. The core domain also has a modified Rossmann fold with six stranded β-sheets surrounded by six α-helices. It is responsible for binding, reorienting of the phosphate group along with the co-factor Mg^{2+} and also in preparing the Asp10 for the nucleophilic attack.

ACKNOWLEDGEMENTS

This thesis is the result of five years of work whereby I have been accompanied and supported by many people. It is a pleasant aspect that I have now the opportunity to express my gratitude to all those people who contributed directly or indirectly to this endeavour.

Foremost, I would like to express my sincere gratitude to my revered and esteemed supervisor Dr. Ashwani Kumar Sharma for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study. I am very much gratified to him for his inspiration and great efforts to explain things clearly and simply. I owe him more than words can express.

I feel overwhelmed in thanking to Prof. Partha Roy (Present) and Prof. Ramasare Prasad (Former) Head, Department of Biotechnology for providing necessary facilities, support and cooperation in the Department. I gratefully acknowledge the help rendered from time to time by the members of my Student Research Committee (SRC) Prof. R. P. Singh, Prof. Partha Roy (Internal member) and Dr. Kaushik Ghosh (External member) for their scholarly suggestions; prudent admonitions, immense interest and hard questions that have made this task a success. I am courteous to the honorable faculty members Prof. Ritu Barthwal, Prof. Vikas Pruthi, Prof. G. S. Randhawa, Dr. Sanjay Ghosh, Dr. Prabhat Kumar Mandal, Dr. Ranjana Pathania, Dr. Naveen Navani, Dr. M. Nayar, Dr. Krishna Mohan Poluri, Dr. Debabrata Sircar, for their support and encouragement.

I would like to express my sincere thanks towards Dr. Pravindra Kumar and Dr. Shailly Tomar, Department of Biotechnology for their unflinching help and support in crystallographic study and helpful discussion. I am much indebted to both of them for their valuable advice and precious times to read my work. I feel overwhelmed in thanking Dr. Bijan Choudhury for helpful discussion especially during strain development work. It is my privilege to place on record my gratitude to by Prof. Willem Heber van Zyl, Stellenbosch University, South Africa for providing vector and yeast strain for my strain development work. I would like to express my sincere thanks to Dr. Savita Yadav for the DLS study at AIIMS New Delhi.

I gratefully acknowledge Prof. S. C. Sabat, ILS Bhubaneswar for his support and motivation during my master degree project work. I must place on record my thankfulness to Dr. Abhay Kumar Dalai, Department of Biotechnology, Ravenshaw University for providing me the opportunity to start my first steps in research. I am extremely thankful for his advice, supervision and crusial contribution. His involvement with his originality has triggered and nourished my intellectual maturing that I will benefit from, for a long time to come.

The road to my Ph.D started with training at biochemistry laboratory. My thanks go in particular to Dr. Girijash Kumar Patel with whom I started this project and many rounds of discussions with him helped me a lot. My vocabulary falls short when I try switching my thoughts into words to express the deep sense of gratitude towards Ms. Preeti for assistance, help and support in the successful completion of this thesis. This thesis is the result of her sincere effort and her motivation when I really needed. I owe gratitude to Dr. P.selva kumar who willingly devoted so much time in giving guidance to me.

I am obliged to my labmates for their help and support. I am thankful to Dr. P.selva kumar and Mr. Prabhat for their amiable attitude and cooperation while working in lab and making a fun filled

environment outside the lab. I have enjoyed a lot of memorable time with them. Special thanks are due to my batchmates Ms. Nidhi and Ms. Anamika for their support and inspiration. I would like to thank Dr. Deepankar galoth, Ms. Gunjan, Mr. Pranav, and Mr. Vishvajeet for their assistance, help and support in the successful completion of this thesis. They all make my lab life easy. I convey special acknowledgement to Rajat, for his persistent support, affection and care throughout this journey. I am also thankful to all M. Sc. students Mr. Ashish, Mr. Hrishikesh, Mr.Krishnakant and Mr. Sujeet for their sincere cooperation.

Also to be acknowledged with love and appreciation in the memorable concern, affection and care from my dear friends Dr. Shivendra, Mr. Umesh, Dr. Satya, Dr. Jitendra, Dr.Preeti, Ms. Manju, Mr. Paramesh, Mr. Pradeep Dr. Megha, Dr. Dipak, Dr. Aditya, Dr. Pramod, , Dr. Sonali, Dr. Shilpi, Ms. Pooja, Mr. Tapas, Mr. Narendar, Mr. Rajesh, Mr. Madhu, Mr. Harvijay, Mr. Vijay, Mr. Atin, Mr. Tamogan, Mr. Siddharth, Mr. Shailendra, Mr. Madan, Mr. Abhijeet, Mr. Sitakant, Mr., Shaktikant, Ms. Monu, Ms. Ramanjeet, Ms. Benazir, Ms. Anjali, Ms. Anchal, Ms. Neha and many others who made my stay in IITR memorable and wonderful. I take the opportunity to extend my hearty gratitude to Mrs. Manjula Sharma, wife of Dr. A. K, Sharma, who had taken every care for me. I also wish to convey my feelings and love to Abhinav and Swati to always making us feel at home.

Deep heartfelt admiration is owed to my friends from my native place Late Mr. Sameer, Mr. Padmanabh, Mr. Sujit, Mr. Paramananda, Mr. Biswajeet, Mr. Biswajit, Mr., Rakesh, Mr.Sashibhusan, Mr. Sakrajit, Mr. Arjun, Mr. Deepak, Mr. Satya, Mr. Simanta, Mr. Swadhin, Mr. Sunil, and Mr. Pradipta. I am thanking all for the good time, days filled with pleasure, fond, memories and the feelings that I will always treasure.

Special thanks are due to all the technical and office staff of the Biotechnology Department in particular Mr. Ved Pal Singh Saini, Mr. Subhash Jain, Mrs. Shashi Prabha, Mr. Lokesh, Ms. Surita, Mr.Yogendar Valthare, Mr.Rajesh Pal, Mr.Pradeep and Mr. Anuj. I would like to special thank to Mr.Ajit and Mr.Vikas for supplying me chemical requirements on time. I would like to thank Mr. Anil for providing the Putranjiva roxburghii seeds and the locally available cellulosic feedstocks. I would like to special thank to Mr.Ajit and Mr.Vikas for supplying me chemical requirements on time.

It would not have been possible to undertake the journey of my career, and to reach where I am today without the support of my beloved Mother, my brother Jibanananda, sister Chinmayee and my brother-in-law Satyapriya. I thank them for the unconditional love and care they showed on me. Their faith on me has always been a great strength which helped me throughout the way. I would like to dedicate this thesis to my late Father. His blessings are always with me during my good and bad times. Thanks for giving such a wonderful life to me. I would like to thanks my beloved Kakei for his sincere encouragement and inspiration throughout my career. I would like to thank my native village 'Khanati', for teaching me the importance to lead a life as simple as possible.

I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

I owe my gratitude to the Council of Scientific and Industrial Research, New Delhi for providing the financial support to conduct the research work.

Finally, I thanks to God for my life through all tests in the past years. You have made my life more beautiful. May your name be exalted honoured and glorified!

Bibekananda Kar

(Bibekananda Kar)

LIST OF PUBLICATIONS

1. Patel G.K., **Kar, B**. and Sharma A.K. (2011). Characterization of a thermostable family 1 glycosyl hydrolase enzyme from *Putranjiva roxburghii* seeds. ***Appl. Biochem. Biotechnol.*** DOI: 10.1007/s12010-011-9445-2
2. **Kar, B**., Verma, P., and Sharma A.K. (2015). Cloning, expression, characterization and bioinformatic studies of of family 1 glycosyl hydrolase enzyme from *Putranjiva roxburghii*. (Communicated)
3. **Kar, B**., Verma, P., and Sharma A.K. (2015). Studies on the application of *Putranjiva roxburghii* β-glucosidase for bioethanol production. (Communicated)

GENES SUBMITTED TO NCBI DATABASE

1. Gene sequence of Glycosyl hydrolase family 1 from *Putranjiva roxburghii* with accession no- **KF006311**.
2. Gene sequence of Trypsin inhibitor from *Putranjiva roxburghii* with accession no- **HQ332518**.
3. Gene sequence of PNP_UDP_1 Superfamily protein from *Acalypha wilkesiana* with accession no- **KF018085**.
4. Gene sequence of PNP_UDP_1 Superfamily protein from *Euphorbia cotinifolia* with accession no- **KF018086**.
5. Gene sequence of PNP_UDP_1 Superfamily protein from *Euphorbia tithymaloides* with accession no- **KF018087**.

WORKSHOPS/CONFERENCE ATTENDED AND POSTER PRESENTATION

1. **Kar B**., Patel G.K., Chaudhary N.S., and Sharma A.K. 'Characterization, cloning, sequence analysis and modeling of a trypsin inhibitor from *Putranjiva roxburghii'*. ***Proceedings of 7th ABA&IBS,*** 2011, P-17, poster presented in 7th Asian Biophysics Association (ABA) Symposium & Annual Meeting of the Indian Biophysical Society (IBS) at Habitat Center, New Delhi, India, from 30th January to 2nd February, 2011.
2. Patel G.K., **Kar, B**., and Sharma A.K. 'Biochemical and biophysical characterization of a glycosyl hydrolase family 1 protein from *Putranjiva roxburghii'*. ***Proceedings of 7th ABA&IBS,*** 2011, P-202, poster presented in 7th Asian Biophysics Association

(ABA) Symposium & Annual Meeting of the Indian Biophysical Society (IBS) at Habitat Center, New Delhi, India, from 30th January to 2nd February, 2011.

3. Attended '**42nd National Seminar on Crystallography and International workshop on application of X- ray diffraction for drug discovery**' at New Delhi.

4. **Kar B.**, Verma P., Patel G.K., and Sharma A.K. 'Biochemical and Bioinformatic Characterization of a Thermostable Family 1 Glycosyl Hydrolase Enzyme with Fucosidase, Glucosidase and Galactosidase activities'. Poster presented in '**INDO-US international conference/workshop on recent advances in structural biology & drug discovery'** held in IIT Roorkee, Roorkee from October 9-11, 2014.

5. Verma P., **Kar B.**, and Sharma A.K. 'Interaction and oligomerization analysis of BceR/S, a Two-Component Signal Transduction System Sensing Bacitracin in *Staphylococcus lugdunensis*. . Poster presented in '**International Conference on Molecular Signaling: Recent trends in Biomedical and Translational Research**' held in IIT Roorkee, Roorkee from December 17-19, 2014.

CONTENTS

INTRODUCTION

INTRODUCTION

Proteins are fundamental components in all life forms and their functional versatility is due to their structural range. Proteins carry out their activities owing to their ability to bind to several inorganic and organic molecules and to other macromolecules like proteins or DNA [73]. In modern classification proteins are classified in three major classes; storage proteins, structural proteins and biological active proteins. The major biologically active proteins include enzymes. Enzymes are awesome machines with a suitable level of complexity. They are central to every biochemical process and have extraordinary catalytic power. Enzymes act in a highly coordinated manner to sustain life. Enzymes are essential to each and every biochemical reactions. They catalyze thousands of reactions and the mechanism by which they speed up these reactions is one of the most interesting fields in biological research. Thus this mechanism has attracted researcher from various fields i.e. biochemistry, physical chemistry, genetics, microbiology, agriculture, toxicology, medicine, chemical engineering, pharmacology etc. The study of enzymes has immense practical importance. The cumulative effect of a numbers of factors such as solvent, salt concentration, temperature, pH and other effectors may determine the unique structure and function of a protein molecule [173, 249] and may adopt different structure by changing the medium conditions [199] which have important implications for protein design and engineering [192]. Enzymes are grouped in to six major classes depending upon the reactions they catalyzed i.e. oxidoreductase, transferases, hydrolases, lyases, isomerases and ligases. Among these different classes of enzymes, 'Hydrolase' is one of the most important group of enzymes which catalyze the hydrolysis of a chemical bond like C-C, C-O, C-N, ether, ester and halide bonds etc. Hydrolases are classified into several subclasses, based upon the bonds they act upon e.g. glycosyl hydrolase, esterase, phosphatase, lipase, DNA glycosylases etc.

Carbohydrate metabolism is only possible due to some enzymes i.e. glycosyl hydrolases (hydrolyzes and/or rearrange the glycosidic bonds), glycosyl transferases (build glycosidic bonds), polysaccharide lyases (non-hydrolytic cleavage of glycosidic bonds) and polysaccharide lyases (hydrolyzes carbohydrate esters). Glycosyl hydrolases or Glycoside hydrolases or Glycosidases (GHs; EC 3.2.1.x) catalyze the hydrolysis of O-, N- or S-linked glycosidic bonds between a carbohydrate and non-carbohydrate moiety or between two carbohydrates. The cleavage of these glycosidic bonds is crucial for the processes like hydrolysis of structural polysaccharides during penetration of pathogens, recycling of cell surface carbohydrates, defense against pathogens expansion of cell wall, energy uptake, starch

metabolism, symbiosis and recycling of signaling molecules etc. On the basis of amino acid sequence and structural analogy, at present there are 133 Glysoside hydrolase families available on the CAZY web server (URL- http://www.cazy.org/) [27]. Among them, family 1 glycosyl hydrolase is most studied. Glycoside hydrolases family 1 contains different enzymes with some well-known functions, such as β-glucosidase (EC 3.2.1.21), β-fucosidase (EC 3.2.1.38), β-galactosidase (EC 3.2.1.23), β-mannosidase (EC 3.2.1.25), exo-β-1, 4-glucanase (EC 3.2.1.74) etc. β-glucosidase characterized till date fall primarily in glycoside hydrolase families 1, 3 and 5 with family 1 β-glucosidases being more abundant in plants. β-glucosidase of this family may have high specificity for glucosides or in addition to this may hydrolyse fucosides and/or galactosides.

Like glycosyl hydrolase HAD superfamily (HADSF) hydrolases are also very important in their functional point of view. The HADSF is the major family found with almost 48,000 sequences and is present ubiquitously in the living cells. HADSF constitute different enzymes like dehalogenase, phosphatase, phosphonatase, β-phosphoglucomutase and ATPases [79, 97]. Phosphatase forming the majority in HADSF can vary in figure from 30 in prokaryotes to 200-300 in the eukaryotes. They help in phosphorous transfer reactions by transferring the phosphate group to an Asp as an active site residue. phosphatase plays a crucial role to sustain life. This group of enzyme transfers phosphorous group. Transfer of phosphorous group is generally used for cell signalling in all the living kingdoms from prokaryotes to eukaryotes. It is used as an approach to counter to different internal and external stimuli [56]. Phosphatases help in phosphorous transfer reactions by removing the phosphorous group from biological molecules [122]. Almost 35-40% of the bacterial genome is composed of molecules carrying a phosphate group [131]. Although there are many enzymes responsible for phosphoryl transfer but they belong to different super families which are evolutionary dissimilar and have different structural folds.

In this work we have studied two enzymes from hydrolase family: First, a ~66 kDa, thermostable and highly efficient glycosidase 1 or family 1 glycosyl hydrolase (GH1) enzyme sourced from *Putranjiva roxburghii*, a medicinal plant belonging to *Euphrbiaceae* family and Second, ~32 kDa novel HAD superfamily phosphatase sourced from *Staphylococcus lugdunensis*, a human pathogen belonging to *Staphylococcaceae* family.

CHAPTER 1
LITERATURE REVIEW

1. LITERATURE REVIEW

1.1. Introduction

Proteins are most important and abundant functional biomacromolecules present in all types of cell and they are constructed from 20 different amino acids. The structural and functional diversity of this macromolecule is very high. Each protein exhibits a specific biological activity which depends on the three dimensional structure of the protein determined by the amino acid sequence and the folding [175]. In some cases post-translational modifications and cofactors are required for executing the biological functions. The diverse combinations of twenty different amino acids bring about varied functional products such as enzymes, antibodies, hormones, transporters, lens protein of the eye, muscle fibres, milk proteins, antibiotics, horns, spider web, poisons, antibiotics etc. having specific biological activities. Proteins carry out their activities owing to their ability to bind to several inorganic and organic molecules and to other macromolecules like proteins or DNA [74]. The interactions involve in binding are van der Waals, hydrophobic, electrostatic forces and hydrogen bonding. Hydrogen bonding is very important owing to their number and location. Mutations bring the disruption of these interactions which leads to complete loss or reduction of biological activity.

A complicated organization of disciplined biochemical reaction catalyzed by distinct enzymes is very critical to sustain life. Enzymes are essential to each and every biochemical reaction. They catalyze thousands of reactions and the mechanism by which they speed up these reactions is one of the most interesting fields in biological research. Thus this catalytic machinery has attracted researcher from various fields i.e. biochemistry, physical chemistry, genetics, microbiology, agriculture, toxicology, medicine, chemical engineering, pharmacology etc. This catalytic power also has received attention from industries like biofuel, food, leather, paper, detergent etc. Owing to the complex structure and biological instability of these enzymes, our understanding to this mechanism is still limited. The structural and functional aspect of these enzymes can be verified both *in vitro* and *in vivo* environments. Proteomics study is an interesting field to know the basics of different biophysical and biochemical functions and to understand their various mode of action [62, 91, 107, 120, 130, 132, 186, 258]. Proteins are also the major source of dietary supplements [111].

Plant genome encodes thousands of enzymes which show differential expression in different parts of the plant. Plant seeds are very rich source of vital proteins and enzymes

which has been used for hundreds of years in industries and medicines. Seed proteins are mainly storage proteins which play crucial roles to provide nutrition during seedling stage, maintain viability, and protect from microbes and insects. By providing the dominant share of dietary components it plays an essential role in animal and human nutrition. The proteins in plants may be categorized as biological active proteins, structural and storage. Besides these proteins, seeds are rich sources of alkaloids, oils, resins, tannins, pigments etc. Seed proteins are used in different structure function studies [129].

'Hydrolase' group of enzymes is one of the important group of enzymes among six different classes of enzymes. It catalyzes the cleavage of C-C, C-N, and C-O bonds along with some other bonds like phosphoric anhydride bonds. On earth and in plant cells, carbohydrates are the most copious biomolecule where as in animal cells proteins have the largest share among macromolecules. Recycling of these macromolecules is critical to sustain life and it is possible due to controlled action of enzymes and their respective inhibitors [252]. Several hydrolases are drug targets [113].

In this chapter we have reviewed the recent development of our understanding regarding the properties of Glycoside Hydrolase family 1 enzymes in section **1.2**, the role of glycosylation on the structural and functional aspects of the enzymes in section **1.3**, yeast strain development and its application in SSF (simultaneous saccharification and fermentation) in section **1.4**, and understanding regarding the Haloacid Dehalogenase (HAD) family hydrolase enzymes in section **1.5** with a short introduction.

1.2. Overview-Glycoside Hydrolases

Carbohydrates are the most copious and diverse group of biological molecules on the earth. They are present in the form of monosaccharides (e.g. glucose, ribose, arabinose, galactose etc.), disaccharides (cellobiose, lactose, maltose, sucrose etc.), oligosaccharides (cellulose, starch, chitin, glycogen etc.) and glycoconjugates (peptidoglycans, glycoproteins, glycolipids, lipopolysaccharides etc.). These molecules exist in a large number of stereochemical variations. Polysaccharides are formed by the monosaccharides units which are linked by glycosidic bonds. This bond is among the most stable bonds, with a half-life of more than five million years [256]. Glycosyl hydrolases or Glycoside hydrolases or Glycosidases (GHs; EC 3.2.1.x) catalyze the hydrolysis of O-, N- or S-linked glycosidic bonds between a non-carbohydrate and carbohydrate moiety or between two or more carbohydrates (Figure 1.1). The cleavage of these glycosidic bonds is crucial for operations

like hydrolysis of structural polysaccharides during penetration of pathogens, recycling of cell surface carbohydrates, defense against pathogens [96], expansion of cell wall, energy uptake, and recycling of signaling molecules etc. [42]. Glycosidases are present everywhere in nature, reported even in chlorella virus [220].

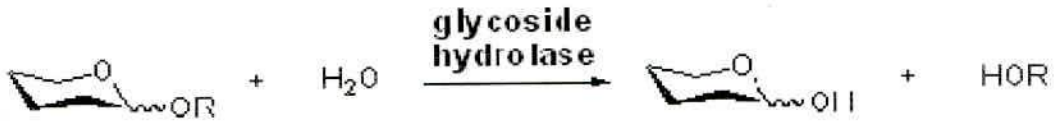


Figure 1.1. Hydrolysis of glycoside by glycosyl hydrolase enzyme, figure adopted from CAZypedia.

Along with glycosyltransferases, they form the major catalytic machinery for the synthesis and cleavage of glycosidic bonds. Despite of the various protein folds seen in glycosidase family, their active site architecture fall into three generalized morphological classes: tunnel, cleft or groove and pocket crater.

1.2.1. Glycoside Hydrolase classifications

The Enzyme nomenclature system (IUBMB) of Glycosidases is primarily based on their substrate specificity and the reaction they catalyze. However, this classification can't segregate the Glycosidases family owing to their multi substrate specificities e.g. endoglucanase and myrosinase [95, 23]. This classification based on substrate specificities fail to reflect the structural characteristics and evolutionary events of glycosidases. Hence, glycoside hydrolases are classified in families based on their amino acid sequence similarities was proposed by Henrissat (1991) [93, 94].

Recently there are 133 glycosidases families existing on the CAZy web server (URL-http://www.cazy.org/Glycoside-Hydrolases.html) on MAY 2nd, 2015, basing on the amino acid sequence similarities [27]. This method of classification is very convenient as sequence has direct relation with the folding and function. After the sequence specific grouping, systematic comparison of the enzymes showed that sometime enzymes with different substrate specificities fall in the same family explaining divergence in evolution e.g. cellulase enzyme is found in eleven different families [42]. In general, folds of proteins are more conserved than the sequences. So, some families were grouped in **'Clans'** on the basis of their overall structural similarities. Currently, all the 133 glycosidases families were divided in 14 different clans. Among these, GH-A is the largest clan consisting of 19 different families.

These families of glycosidases show different substrate specificities but in their 3-D structures they possesses common $(\beta/\alpha)_8$ TIM barrel fold [108].

1.2.2. Overview of Family 1 glycoside hydrolase (GH1)

GH1 is the most studied family of glycoside hydrolase enzymes. The β-D-fucosidases (EC 3.2.1.38), β-D-glucosidases (EC 3.2.1.21), and β-D-galactosidases (EC 3.2.1.23) enzymes mainly fall in this family and are responsible for the cleavage of terminal non reducing β-D-fucose, β-D-glucose and β-D-galactose respectively. Retaining mechanism is followed by this group of enzymes. Structural and functional characters of all glucosidases are similar in all families. These enzymes shows common active site signature sequence [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN]. The β-glucosidases are divided into a narrow specific and broad specific group based on their specificity towards substrates [60. 211].

1.2.3. Catalytic machinery and mechanism

In β-glucosidases, two glutamate residues are responsible for the catalysis and were first observed in the *Agrobacterium*. The Glu170 is a general acid/base catalytic residue identified by detailed mutation study and Glu358 which is a nucleophile is conserved in the conserved sequence YITENG in the same enzyme [246, 255]. In some cases in GH-A clan possesses an asparagine residue before the general acid/base catalytic residue in a conserved NEP sequence believed to be involved in hydrogen bonding interactions at 2-hydroxyl position with the substrate. A single mutation can bring drastic changes to the characters of a protein [62, 121, 160].

The glycosidases show two types of catalytic mechanisms, inverting (single displacement) and retaining (double displacement). The enzymes from Family 1 glycoside hydrolase family function through retaining mechanism only. The catalytic hydrolysis with inversion mechanism is carried out by involving an oxocarbenium-ion-like transition state (Figure 1.2 A). The general base detaches a proton from water molecule and then attacks the glycosidic bond's C1 carbon atom [265]. As a result of the addition of water between the anomeric C1 atom and the general base, the two catalytic carboxyl groups are located on the reverse side of the glycosidic bond; usually the distance is about 10.5 Å [160, 246]. The retaining mechanism (double displacement) involved the construction of a glycoside enzyme intermediate (covalently bound) which is hydrolyzed in each step by oxocarbenium-ion-like

transition states (Figure 1.2 B). In this case nucleophile directly attacks the C1 atom not the water molecule. The distance between the two reactive groups is around 5 to 5.5 Å [160].

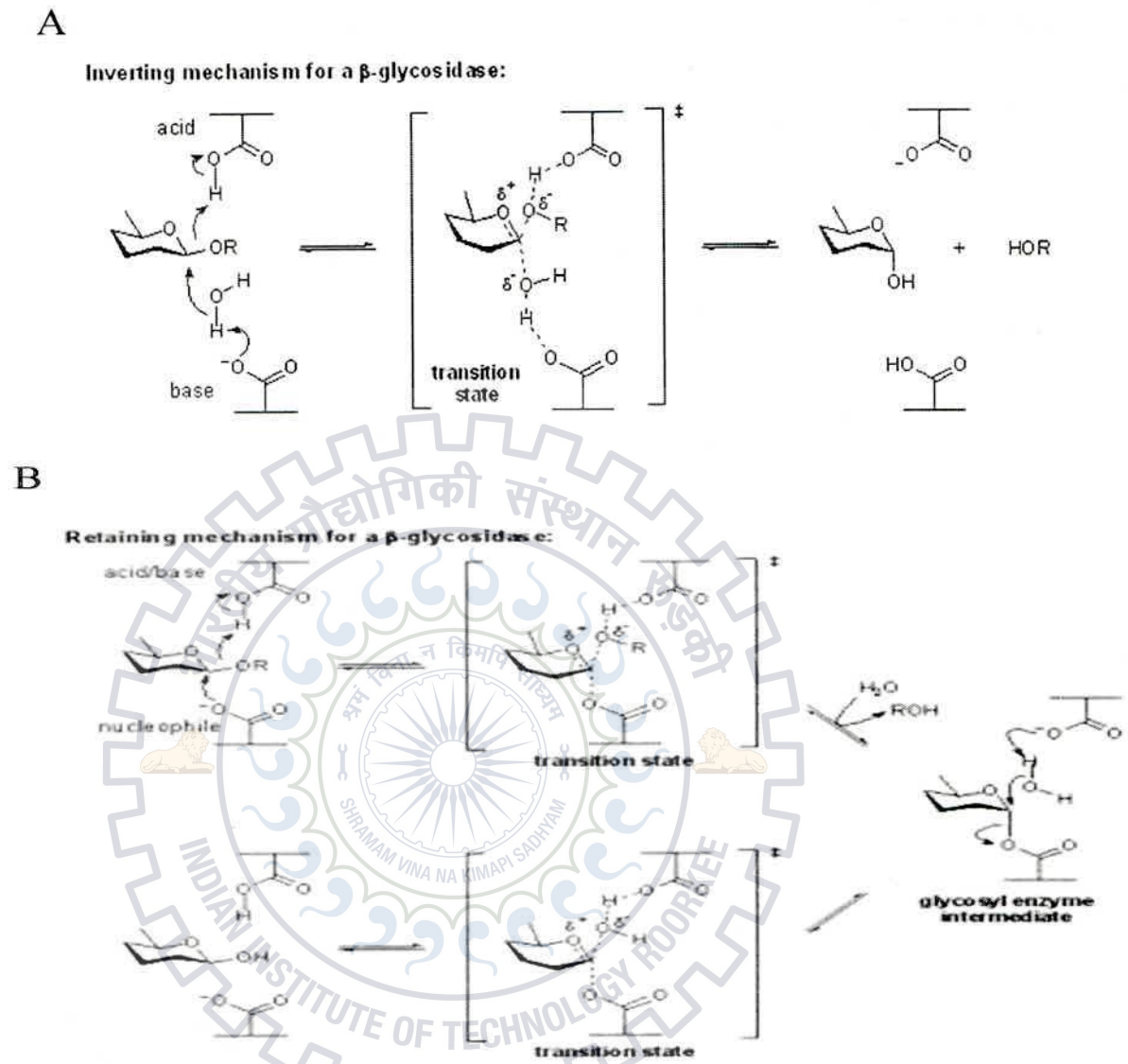


Figure 1.2. General mechanisms of Glycosyl hydrolases adopted from CAZypedia (A) Inverting mechanism (single displacement mechanism) and (B) Retaining mechanism (double displacement mechanism)

1.2.4. Plant Glycoside Hydrolase enzymes

Several plant parts including seeds, stem, root, flowers, leaves, and bark are sources of Glycoside hydrolase enzymes. These enzymes mainly fall under the family 1 of Glycoside hydrolase. They play very crucial role in different plant metabolisms. They execute several functions like lignification, vitamin metabolism, plant defence, growth and development, activation of phytohormones and secondary metabolites hydrolysis. Plant defence is an important physiology and studied by many workers [39]. Strictosidine β-D-glucosidase act on

its substrate 'strictosidine' to produce a reactive aglycone. This aglycone enters various biochemical pathways to form quinoline and indole (vindoline, ajmaciline, strychnine, reserpine, camptothecin), and these compounds help in the damaged induced defence system of plants. This type of β-glucosidase mediated alkaloid metabolism is observed in Loganiaceae, Apocynaceae and Rubiaceae plant families [10]. β-glucosidases of *Pinus contorta* has performed a crucial role in lignification process of pine stem by discharging the 'monolignol coniferol' from the conifers which is the natural substrate of *Pinus* [50]. BGLU45 and BGLU46 are the two β-glucosidases, sourced from *Arabidopsis thaliana* showing similarity with the *Pinus* β-glucosidase and also able to hydrolyze monolignol indicating their participation in biosynthesis of lignin. Various metabolites like extract of neem leaf inhibit mammary carcinogenesis [7].

1.2.5. Substrate specificity of plant β–glucosidases

With respect to the aglycone moiety, β-glucosidases have a very wide range of substrate specificities with varied preference for aglycone moiety. ZmGlu1 and ZmGlu2, the β-glucosidase isoenzymes of maize hydrolyze a wide range of both natural and synthetic substrates. The β–glucosidase of sorghum (Dhurrinase, SbDhr1) has 72% sequence similarity with ZmGlu1, but it can't act on the substrates of ZmGlu1, instead it hydrolyzes the natural substrate dhurrin. Interestingly, dhurrin act as an inhibitor to ZmGlu1. This difference in the substrate specificities is due to the variation in the amino acid residues at the aglycone binding site. Dhurrinase interact with its natural substrate (dhurrin) by using Ser62, Phe261 and Asn259 as interacting residues and they are different from the maize β-glucosidase. Mutation studies were carried out in ZmGlu1 at the corresponding residues of sorghum dhurrinase [239]. Mutational study showed that Phe198Val mutant had a major effect on enzyme activity. Likewise, a Glu191Asp mutant of ZmGlu1 showed a major change in active site architecture. The active site appeared like a flattened crater with two separate regions.

1.2.6. Industrial applications of Glycosidases

β-galactosidases hydrolyse the milk protein lactose to galactose and glucose. The enzyme can be used in dairy industry for the production of milk with low lactose level. Undesirable microbial growth during processing of milk can be minimized by doing this hydrolysis by β-galactosidases at the high temperature. So the identification and characterization of thermostable β-galactosidases are prerequisite for dairy industry [20, 196]. Yang et al. (2009) reported that by converting isoflavone glycosides to their respective

aglycones improves the nutritional values of several food materials and this conversion is usually done by the β-glucosidases [261, 262]. Decrease in lysosomal β-glucosidase activity leads to Gaucher disease. In this case the glucosylceramide accumulate in the macrophages. The symptoms are hepato-splenomegaly, skeletal lesions, anaemia etc. Sawkar et al. (2002) suggested that the addition of subinhibitory concentrations of N-(n-nonyl) deoxynojirimycin to a fibroblast culture medium leads to two fold increase in the β-glucosidase activity [206]. This technology is likely to replace the costly enzyme replacement therapy in near future. Immobilized β-glucosidase increases the clarity of wine to 100% [215]. β-glucosidase enzymes also convert the cellobiose to glucose, so it is used in the fermentation industry. Random drift mutagenesis (RNDM) is a direct evolution technique used to increase the activity of heat resistance β-glucosidase of *Caldicellulosiruptor saccharolyticus* [88]. This enzyme is used for bioethanol production. Immobilization of glycosyl hydrolases also has important applications [55, 231].

1.2.7. Cloning of β–glucosidases genes

Large number of bacterial, yeast, plant and animal β-glucosidase genes have been cloned and expressed in both prokaryotic as well as eukaryotic system. Generally *Escherichia coli* and *Saccharomyces cerevisiae* used as prokaryotic and eukaryotic hosts respectively. Cloning has been performed by two methods, (1) starting with a cDNA library or a genomic library, nucleotide probes are designed basing on the prior knowledge of the polypeptide (2) formation of a genomic DNA library followed by selection of the recombinant clones by screening for β-glucosidase production [13]. Plants and fungi are better producer of β-glucosidase enzymes; however studies on cloning of β-glucosidases from these organisms are relatively low owing to its complexities associated with the presence of introns in their genes. Yet, several workers have cloned and expressed the β-glucosidases gene successfully from plant e.g. *Brassica napus* [61], rice [259] and from fungal sources e.g. *Aspergilli* [104, 41], *Humicola grisea* and *Trichoderma reesei* [224]. Studies also showed the existence of multiple gene and gene products that are expressed differentially. Fungal β-glucosidases genes are mostly expressed in heterologous fungal systems like *Saccharomyces cerevisiae* [41], *Aspergillus* sp. [224], *Pichia pastoris* [41], and *Trichoderma reesei* [11]. Different other glycoproteins are also cloned in *E. coli* and successfully studied [106].

1.3. Overview-Glycosylation

The significant rise in complexity from the genome to the proteome in nature is enabled by different mechanisms which produce different mRNA transcripts from a single gene like genomic recombination, differential transcription termination, alternative splicing etc. Protein post-translational modifications (PTMs) further enhances the protein diversity and functions by covalent attachment of functional groups, changing the nature of amino acid, cleavage of regulatory subunits, degradation of entire proteins etc. These chemical modifications play significant role in proteomics by regulating activity, cellular localization of protein and complex formation with other cellular molecules such as proteins, nucleic acids, lipids, and cofactors. The most common modifications include proteolysis, acetylation, formylation, lipidation, methylation, phosphorylation, ubiquitination, and glycosylation. Therefore, PTM identification and understanding is crucial for cell biology studies as well as for the disease treatment and prevention.

The addition of sugar residue to a protein is the most complex post-translational modification of a protein as seen in case of most secretory and plasma-membrane proteins which have one or more carbohydrate chains [214]. These glycans play diverse structural and functional roles in secreted and membrane proteins. Glycosylation is a modification in which a glycosyl donor i.e. a carbohydrate is bonded to a hydroxy or other functional moiety of a glycosyl acceptor. It is an enzymatic process as opposed to the non-enzymatic reaction called aglycation. In biological system, it refers to the attachment of glycans to other macromolecules like proteins, lipids or other organic macromolecules through a series of post-translational enzymatic steps. These post translational modifications produce diverse protein-bound oligosaccharides which are also among one of the essential biopolymers found in cells amongst DNA, RNA, and proteins. Glycosylation occurs in endoplasmic reticulum lumen or another places like the lumen of Golgi *trans*-, *medial*-, or *cis*-cisternae, in cytoplasm and nucleus. However, large numbers of proteins synthesized in endoplasmic reticulum (ER) go through glycosylation process. Thus the carbohydrate residues also act as indicators for marking their movement from ER and Golgi bodies. Engineered antibodies use aglycosylation to bypass glycosylation [112]. ER also helps in sub-cellular trafficking [182].

1.3.1. Occurrence of Glycosylation

Glycosylation reaction occurs in almost every organism from archaea, eubacteria and eukaryotes [143, 163]. However, eukaryotes have the large array of organisms ranging from single-celled to complex multi-cellular organisms that express glycosylated proteins. The

specificity of the reaction lies in the fact that it is well-ordered and has features that vary amongst the different cells and the species [214].

1.3.2. Glycoprotein diversity

Glycosylation has a lot to contribute in the proteomic diversity. Further each feature of glycosylation may also be reformed like:

- **Composition of glycan** – sugar types attached to a specific protein
- **Glycan arrangement** – branching and unbranching of chain
- **Length of glycan** – length of oligosaccharides; short-chain or long-chain
- **Glycoside linking** – site where oligosaccharide binds

The process of glycosylation involves a large number of enzymatic events like linking monosaccharides together, transporting sugar moiety from an entity to the other and sugar trimming. All these steps are not necessary in every glycosylation event. Unlike transcription or translation, glycosylation is a non-template process and many enzymes are involved to produce the enormous diversity of glycoproteins in the cell like either addition or removal of sugars from one substrate to other. In fact enzyme activity also varies from one cell to the other and also within the intracellular compartment which leads to different glycoproteins suitable for the need of that particular cell. Mechanism of glycosylation is highly-ordered and sequential in which the activity of each enzyme is reliant on the preceding enzymatic reaction [245].

There are various enzymes involved in glycosylation process like Glycosyltransferases (Gtfs), Glycosidases etc. Glycosyltransferases (Gtfs) are the enzymes responsible for transferring mono- or oligosaccharides from donor molecules to acceptor molecules like the growing oligosaccharide chains or proteins. Each Gtf is very specific in its function of associating a precise sugar from a dolichol or a sugar nucleotide which acts as a donor and works individually from other glycosyltransferases. These Gtfs are very wide in their nature as the glycoside linkages are identified on nearly each functional group of protein, and the most commonly used monosaccharides are included in the glycosylation reactions [214]. On the other side, glycosidases remove sugars from the proteins by catalyzing the hydrolysis of glycosidic bonds. These enzymes are very specific for eliminating a specific sugar (e.g. mannosidase) and are also very crucial in the ER and Golgi for processing of glycans.

1.3.3. Types of Glycosylation

Glycosylation incorporates the addition of a variety of sugar moieties from a monosaccharide to highly complex branched oligosaccharides to the proteins. The protein-sugar bond is classified into precise groups depending on the bonding nature and the attachment of oligosaccharide. They can be categorized as:

1. **N-linked glycosylation:** This is the most common type of glycosylation occurring in the lumen of the ER. The carbohydrate moiety binds to the amino group of asparagine and is often required for the proper folding of some eukaryotic proteins in addition to the cell-cell and cell-extracellular attachment function. This type of glycosylation can be seen in eukaryotes, archaea but is very rare in bacteria.

2. O-linked glycosylation: This type of glycosylation can be found in all the three groups; eukaryotes, archaea and bacteria. It includes the binding of monosaccharide to the hydroxyl group of serine or threonine in the ER, Golgi bodies, cytoplasm and nucleus.

3. Glypiation: This type of glycosylation is used for the localization of the proteins to cell membranes and can be seen on surface glycoproteins in eukaryotes and some archaea. The protein is attached via its C terminus to the GPI anchor i.e. a phosphoethanolamine linker through a glycan chain. Glycan structure consists of a phospholipid tail for anchoring to the membrane. The vast differences in the lipid and the sugar moiety used in the tail and the glycan core enables unique functions like signal transduction, immune recognition and cell adhesion. These GPI anchors can be cleaved with enzymes like phospholipase C for the localization of proteins to the plasma membrane.

4. C-linked glycosylation: This can also be called as C-mannosylation as a mannose sugar is added to the indole ring of the first tryptophan residue in this W-X-X-W sequence, where W is a tryptophan residue and X is any amino acid. It is an unusual form of glycosylation as the sugar is not linked to any reactive atom such as nitrogen or oxygen but to a carbon atom. Thrombospondin are the commonest example of C-mannosylated proteins.

5. Phosphoglycosylation: In this type of glycosylation, glycan binds to the amino acid serine via phosphodiester bond. Xylose, mannose, fucose and N-acetylglucosamine (GlcNAc) phosphoserine glycans are the most common examples seen like xylose in *Trypanosoma cruzi*, mannose in alpha dystroglycan i.e. a cell-surface laminin receptor in *Mus musculus* (mouse).

Among these, N- and O-glycosylation are the most frequently seen glycosylations as the aspargine-linked i.e. N-linked or serine/threonine-linked i.e. O-linked oligosaccharides are the key structural components in many secreted and cell surface proteins. Although the O-linked and N-linked oligosaccharides forms majority but both are considerably different from each other in their structures, incorporation of different sugar residues. Generally N-linked glycans have a number of branches with a negatively charged residue called sialic acid present at the end, while O-linked glycans have one to four sugar moieties and are thus small. Many nuclear and cytosolic proteins are not glycosylated with exclusions of few nuclear-pore complex proteins and transcription factors, where a single GlcNAc is attached to the threonine or serine hydroxyl group. A pictorial view of these glycosylations types can be seen below (Figure 1.3):

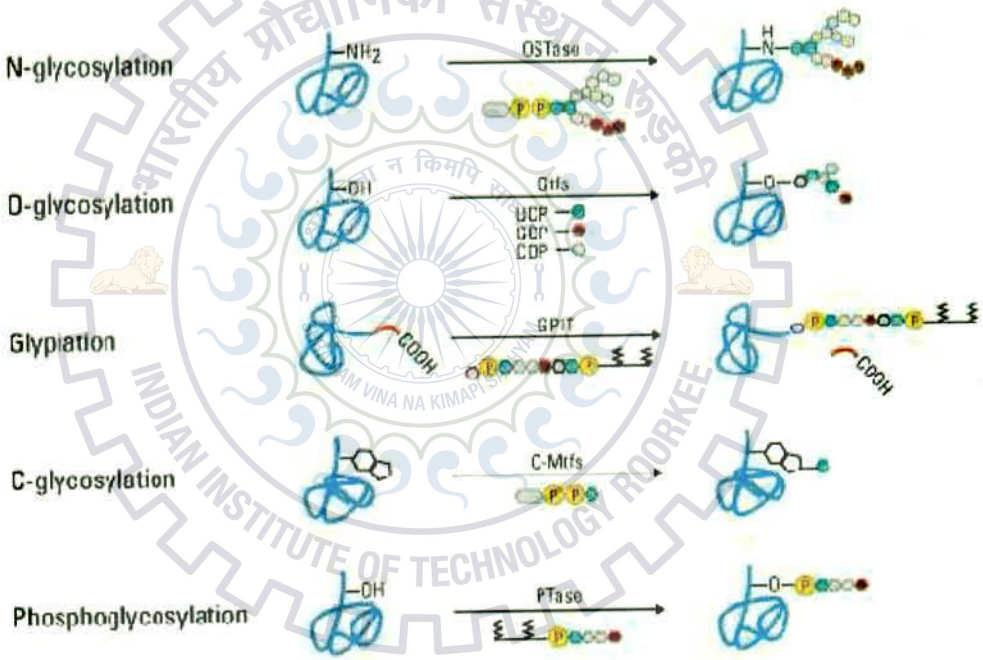


Figure 1.3. Different types of glycosidic bonds

1.3.4. Mechanism of glycosylation

The difference in the configuration of N- and O-linked glycans lies in the difference in their synthesis. O-linked oligosaccharides are linked sequentially one by one in contrast to the N-linked oligosaccharides where a big oligosaccharide which contains fourteen sugar residues is added. Moreover, different steps in the formation of these two different oligosaccharides occur in specific organelles. The mechanisms used for the glycosylation are:

1.3.4.1. N-Linked glycosylation

The N-linked glycosylation although characterized as a post-translational modification but takes place along with the translation i.e. the sugar moiety is being added to the proteins undergoing translation and ER transportation. Most membrane bound and secreted proteins are N-linked glycoproteins as the location of their processing and translation is ER. The β-N-glycosidic bond between GlcNAc and amide nitrogen of Asn is the most commonly seen sugar-protein bond which also forms the spot for addition of diverse range of sugar moiety in proteins which play important functions [168, 238]. This β-glycosylamine linkage was first defined in an ovalbumin which contains 3% carbohydrate [111, 207].

N-glycoproteins have very huge and far branched oligosaccharides which underwent numerous processing rounds when they are added to the proteins like subsequent addition and removal of certain sugar residues, one by one, in a precise manner with an unrelated enzyme catalyzing each step [22, 47]. Overall, N-linked glycosylation constitutes several steps like preformed oligosaccharide assembly, attachment, trimming and further their maturation and a different enzyme is used in each step. It is identical for all the proteins in the beginning and the diversity is seen after the trimming and maturation of glycans [232].

The biological synthesis of the N-linked oligosaccharides initiates in the ER where the preformed oligosaccharide is attached to dolichol (large chain having number of isoprenoid lipids) present in the ER by a pyrophosphoryl residue [22, 43]. This 14-residue oligosaccharide precursor molecule consists of glucose (Glc), GlcNAc and mannose (Man). The pyrophosphoryldolichol glycan is made on the membrane of ER in a series of steps catalysed initially with the help of enzymes present on the rough ER cytosolic site and later on the luminal face. The configuration of the preformed precursor is conserved in single-cell eukaryotes, plants and animals. This conserved glycan configuration i.e. $Glc_3Man_9(GlcNAc)_2$ is then transported from the carrier (dolichol) to the translating polypeptide asparagine residue with the help of oligosaccharide-protein transferase (OST). The consensus sequence for this attachment is Asn-X-Thr or Asn-X-Ser, where X may be any of the amino acid. Generally the N-terminal consensus sequences are being glycosylated than the C-terminal due to the folding of the polypeptide as it enters the ER which might hinder access for OSTs. Trimming of oligosaccharides occur both in the Golgi bodies and ER with the help of glycosidases. The purpose of glycosylation is different in both the ER and the Golgi bodies. In ER, it is used to cross-check protein folding and degradation of proteins. Glycosidases I and II along with ER-resident Mannosidase (ERManI) catalyse the trimming via hydrolyses [57, 58]. The

glycoproteins that are entered to Golgi bodies after proper folding are having $Man_9GlcNAc_2$ glycan structure in higher eukaryotes. The maturation pathway in Golgi is well organized in a manner with each phase relying on the earlier one. This includes both trimming and addition of sugars in order to generate diversity among the oligosaccharides. Golgi contains various enzymes in different cisternae in order to facilitate this step-wise process. The final glycan structures are complex oligosaccharides or high-mannose oligosaccharides or the hybrid of them [200].

1.3.4.2. O-linked glycosylation

In cellular biology, O-linked glycoproteins have an important function [74] like in the biosynthesis of mucins (high-molecular weight proteins which form mucus), formation of proteoglycan core proteins used in the synthesis of extracellular matrix components. Moreover, the antibodies playing key role in our immune system are also profoundly O-linked glycosylated. O-linked glycosylation follows after the translation on glycoproteins that are N-glycosylated in the ER. In contrast to N-glycosylation, there is no precursor glycan synthesis instead the oligosaccharides are added consecutively once in a step to the threonine or serine residue hydroxyl group and O-linked glycoproteins have very simple glycan structure than the N-linked glycoproteins. It can also occur on the oxidized forms of lysine and proline i.e. hydroxylysine and hydroxyproline found in collagen [74]. O-glycosylation occurs with the transport of a solo GalNAc residue to the β-hydroxyl group of threonine or serine in the Golgi bodies by N-acetylgalactosamine (GalNAc) transferase. GlcNAc, fucose, xylose, galactose or mannose are the sugar moieties used for O-glycosylation depending on the cell and species [46, 214]. Following the addition of first sugar, a variable number of sugars are added to the growing chain with the help of monosaccharide donors (sugar nucleotides). These O-linked glycans can also be seen in cytosol and nucleus for regulation of gene expression or signal transduction through other glycosyltransferases [139]

1.3.5. N-linked glycan precursors from prokaryotes to eukaryotes

These glycan precursors are conserved in eukaryotes while diversity can be seen in case of prokaryotes. The glycan substrate for N-glycosylation is gathered at the ER membrane in eukaryotes while at the plasma membrane in prokaryotes. The glycan moiety is gathered on a lipid anchor with the stepwise addition of monosaccharides by various glycosyltransferases. In prokaryotes and eukaryotes, the lipid anchor is conserved while there is very less similarity in the glycan moiety between the two kingdoms. In archaea, even larger diversity can be seen as the glycans can be modified via methylation, sulfation etc. [114]. In fact as in majority of

eukaryotes, the structure of the preformed oligosaccharide is $Glc_3Man_9(GlcNAc)_2$ [22] while the archaeal and bacterial glycan precursors are heterogeneous [264].

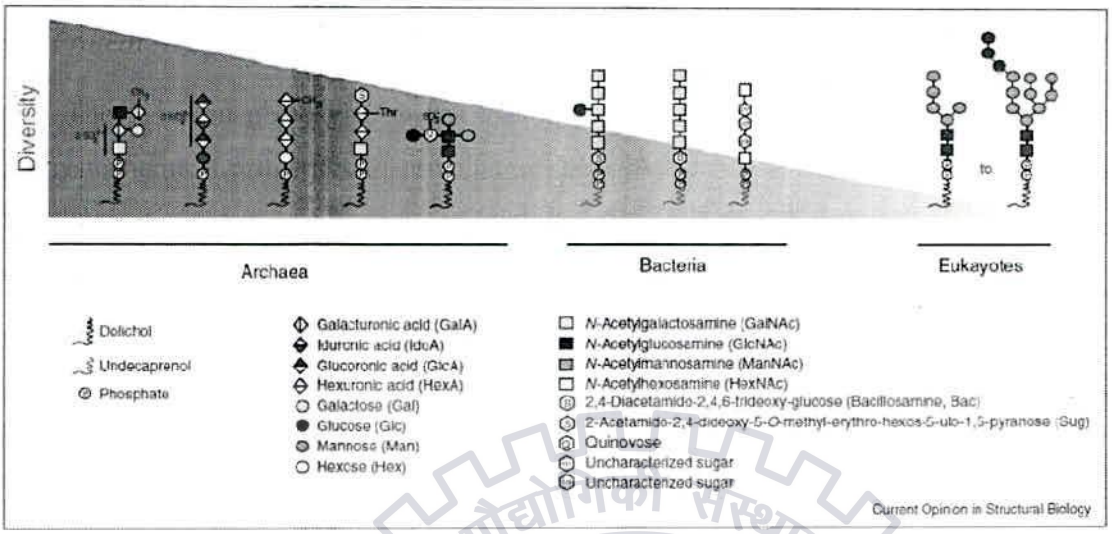


Figure 1.4. Diversity of N-linked glycosylation in different living kingdoms. Prokaryotes produce lipid-linked glycans while eukaryotes are shown to synthesize preserved structure. Different glycan structures are transferred and these glycans are represented by the colored symbols.

1.3.6. Protein N-glycosylation in plants

The *N*-linked glycosylation has a pronounced influence on the biological functions and physiochemical properties of proteins in plants. Most of the endomembrane system and extracellular compartment proteins are N-glycosylated [191]. Recently it has been reported that the enzymes required in cellulose biosynthesis and normal embryogenesis belong to the N-glycosylation pathway like the α-glucosidase I of ER which trims glucose moiety from the recently produced *N*-glycoproteins and the mannose-1-phosphate guanylyltransferase which synthesize GDP-mannose [17, 156]. N-glycans in plants can be categorized in two clusters as the high mannose-type N-linked glycans ($Man_{5-9}GlcNAc_2$) and complex-type N-linked glycans which have been derived by the high-mannose oligosaccharides after the trimming of their mannose residues and addition of sugar moieties to them [191]. The great characteristic plant assemblies contains β1,2-link xylose and α1,3-linked fucose linked with the nearest β-attached mannose and the GlcNAc moiety on the consensus $Man_3GlcNAc_2$ core rather than the oligomannose structures ($Man_{5-9}GlcNAc_2$) present in mammals. These fucose and xylose glycans are not seen in mammals [70].

Nowadays, plants are gaining importance as they can be a platform for the production of recombinant proteins. The reason behind this importance lies in their capability of carrying

out post-translational protein modifications in a manner similar to the mammalian cells. Largely homogeneous, human-type oligosaccharides can be further produced by manipulating the N-glycosylation pathway in plant. Reports are there which shows the ability of plants to carry out human-like complex glycosylation [18]. Plants can also be used as another organism for therapeutic glycoproteins production by knocking out the N-glycan processing genes specific to plant or by inserting the machinery required for the addition, synthesis of human sugars and their transport [81].

1.3.7. Roles of N-linked plant glycans

The N-glycosylation play important functions in eukaryotes such as proper folding of the protein and hence its function, inhibition of the degradation carried out by the proteases. The can also be used as signals for targeting to particular organelles, for cell-cell adhesion and protein identification process. Currently in order to access the importance of N-glycosylation in plants, many studies have been done using N-linked glycosylation and N-linked glycosylation processing inhibitors, mutation of Asn involved in N-linked glycosylation and mutants which affect the N-glycan maturation. The above methods help to completely suppress the N-linked glycosylation or its processing and in turn study the effects of these alterations on plants.

1.3.8. Functional stability of plant proteins

N-glycans help in protecting the protein from proteolytic degradation in addition to conferring thermal stability, solubility and function to the protein. The importance of N-glycosylation can be understood from a simple example of concanavalin A, also called as Con A. Jack bean seed lectin is produced as an inactive glycosylated pro-Con A protein [33]. The processing of inactive pro-Con A to its active form i.e. Con A happens in the proteinaceious body with the involvement of proteolytic tasks to hydrolyze the glycopeptide bond of pro-Con A molecule [19]. Faye and Chrispeels (1986) found that N-glycosylation inhibition reduces pro-Con A transportation from ER towards the vacuole. In addition, it has been reported that the inactive form does not have any lectin activity but can be converted into its active form with the help of deglycosylation enzymes EndoH [64, 190]. Therefore, the results suggest that N-glycosylation is significant for the transportation of Con A to the storage vacuole as well as for its regulation of lectin activity. Con A is a stable protein that binds to the glycosylated proteins [1, 228] and lectins are used in detection purposes [48].

The oligosaccharide side-chains are also thought to be involved in the glycoproteins secretion into the extracellular part as seen in the case of cell wall β-fructosidase in carrot. Faye and Chrispeels (1989) reported that the unglycosylated protein degrades in the secretory path or soon when it enters in the cell wall. It was also described that most of the N-linked glycosylated proteins are not secreted in company of an inhibitor of N-glycosylation (e.g. tunicamycin) [53].

Glycosylation of proteins has been gaining tremendous importance due to the key functions played by the carbohydrate moiety attached to the protein. In case of stem bromelain from *Ananas comosus*, glycosylation favours the functional stability of the enzyme than the deglycosylation [115]. Deglycosylated bromelain show decreased enzyme activity and less stability in the organic solvents as compared to the glycosylated forms. Deglycosylated forms also show disturbed fluorescence and circular spectra with changes in temperature and pH. Function of glycosylation in providing functional stability to the protein is of paramount importance and can be used as a strategy by protein pharmaceuticals to engineer long term stable proteins. FTIR and CD studies are important techniques in studies related to protein [105].

1.4. β-glucosidase in bioethanol production

Saccharomyces cerevisiae and *Trichoderma reesei* are mostly used fungus in the industries for bioethanol production. The cellulolytic enzymes secreted by these strains are used in industrial bioethanol production. The hydrolysis step which converts cellulose to glucose is acknowledged as the main rate limiting step in the conversion of cellulosic material to biofuels. These lignocellulosic raw materials are used because of the low cost and huge availability of these materials. The enzymatic hydrolysis of these materials is a complex and multi-enzyme used process. The last step is the involvement of β-glucosidase enzyme in the conversion of oligosaccharides to monosaccharides, which is further used by different fermenting enzymes to bioethanol [157]. Cellobiose, a disaccharide which is formed during the hydrolysis, is a very strong inhibitor of all the three enzymes involved in the conversion of cellulosic material to glucose. In addition to this, glucose, which is the end product of hydrolysis, also inhibits this hydrolysis of cellulosic material [44]. The foremost cause of this glucose inhibition is that, at high concentration of glucose, it binds to the active site of the enzymes and also prevents the active site to release the product [126, 127]. *Saccharomyces cerevisiae* produces very low amount of secreted β-glucosidase enzyme, which is the main limitation for using the native *Saccharomyces cerevisiae* as a bioethanol producing strain

from cellulosic biomass preparation. This limitation can be overcome by adding external β-glucosidase from other sources or by developing recombinant overexpressing β-glucosidase *Saccharomyces cerevisiae* strain [128].

β-glucosidase with glucose tolerance property which is helpful to overcome feedback inhibition and along with the available commercial enzyme cocktail can improve the hydrolysis of cellulosic wastage [219]. Some fungal β-glucosidase from Aspergillus family have shown this glucose tolerance, and cloned to utilise them [198]. Researchers are doing work in this regard and trying to produce some better strain for bioethanol production by changing the genome complex of the existing microorganisms.

1.4.1. Raw materials for bioethanol production

1.4.1.1. Lignocellulosic materials

Lignocellulosic materials are mainly industrial and agricultural wastage, which is prospective cheap feedstock and can be used for sustainable bioethanol production. These materials are available in large quantities in many countries irrespective of the geographical differences. Annually, 200×109 tons of plant biomass is produced globally and about 9-20×109 tons is principal biomass and remain available for bioethanol production [147]. Out of these plant biomass 90% is of lignocellulosic content. Over the last few years, extensive research has been carried out for the conversion of these materials to bioethanol [188]. Lignocellulosic materials are compound mixtures of polymers of carbohydrate such as lignin, cellulose, hemicelluloses and small amount of extractives.

Lignin is a three dimensional aromatic polymer of phenylpropane and it is complex and hydrophobic in nature. Owing to its heterogeneity and complex structure, it is the most complicated natural biopolymer and provides the necessary resistance towards biological and chemical degradation [144]. The integration of this lignin in to the plant cell walls provides the essential mechanical strength to the large plants, which helps these plants to remain straight. The building blocks and primary precursors of this lignin are (i) P-coumaryl alcohol (ii) coniferyl alcohol (iii) sinapyl alcohol (Figure 1.5).

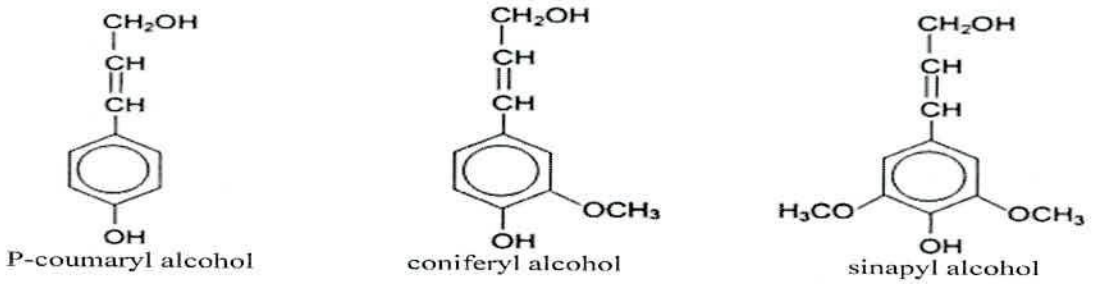


Figure 1.5. The building units of lignin

Hemicellulose is a comparatively less complex polymer consisting of pentoses, hexoses, deoxy-hexoses and hexuronic acids. It has an unsystematic, amorphous structure with very little mechanical strength. Owing to this nature, it is effortlessly hydrolyzed by dilute base or acid and by hemicellulosic enzymes [66]. It differs from celluloses by much smaller and branched molecular chain and by its composition of different sugar units. In contrast to cellulose, it has lower degree of polymerization (DP) and is highly branched. In the side chains, some of the hydroxyl groups have been replaced by acetyl groups, so that during hydrolysis, acetic acid is also liberated along with the monomeric sugars. Hemicellulose is linked to lignin by covalent bonds and to cellulose by hydrogen bonds [109].

Cellulose is a linear chain homopolysaccharide consisting of several hundreds to many thousands of β-D-glucose units linked by β (1→4) glycosidic bonds. It is a very important structural part of the primary plant cell wall. It is also present in many species of algae, fungi, bacteria and even in some animals [181]. It is the most available organic polymer on this planet Earth. The property of the cellulose varies according to the degree of polymerization. In nature, the degree of polymerization of cellulose chains in cotton and wood celluloses are 15,000 and 10,000 glucopyranose units respectively [212]. Even though cellulose is a polar molecule having several hydroxyl groups, it is insoluble in water [148].

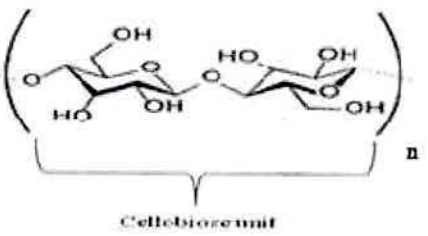


Figure 1.6. Chemical structure of cellulose

The percentage of cellulose, hemicellulose, and lignin (compositional structure) in the general agricultural residues is shown in Table1.1.

Table1.1. Percentage of cellulose, hemicellulose, and lignin in the agricultural residues

Agricultural residue	Cellulose	Hemicellulose	Lignin
Grasses	25–40	35–50	10–30
Rice straw	40	18	5.5
Wheat straw	33–40	20–25	15–20
News paper	40–55	25–40	18–30
Solid cattle manure	1.6–4.7	1.4–3.3	2.7–5.7
Softwood stem	45–50	25–35	25–35
Hardwood stem	40–50	24–40	18–25
Corn cobs	45	35	15
Primary wastewater solids	8–15	NA	24–29
Switch grass	30–50	10–40	5–20

1.4.1.2. Orange peel

Citrus waste is a low cost and renewable agricultural residue, which can be used for bioethanol production. Conversion of this citrus waste to bioethanol can avoid the disposal related problems and is helpful for oxygenating the fuel additives. Orange is the primary citrus fruit in India and since 1980s; the production rate has been improved tremendously. After juice extraction, the processed orange consists of peels, seeds and segment membranes [83]. This wastage contains various polymers of carbohydrates, which make it an obvious choice for bioethanol production by using suitable microorganisms. Individual or combined efforts of mechanical, chemical and biological pre-treatments are required to breakdown the complex polymer of cellulose and pectin and convert these cell wall polymers to monosaccharide units [83, 84]. In enzymatic hydrolysis, the synergic activity of β-glucosidases and cellulases is vital for the hydrolysis. Pentose and hexose sugars including xylose, arabinose, glucose, galactose and galacturonic acid are produced during enzymatic hydrolysis. Along with these sugars peel oil also liberated, which is consisted of 95% D-limonene. It is highly toxic for the fermenting microorganisms, hence reducing the efficiency of bioethanol production. By using processes like aeration and filtration this limonene can be separated from this feedstock which is vital prior to fermentation [84, 253].

1.4.2. Enzymes for biomass to bioethanol conversion

Most of the bio-refinery industries is now directed towards carbohydrate based products and it is obvious that most of the industrial research concentrating on this project.

For sustainable production of renewable biofuels, ethanol production from lignocelluloses is rising as one of the most significant technology. Bioethanol is widely accepted as an alternative and/or stabiliser to gasoline owing to its fewer emissions and higher octane rating than gasoline [257]. Increase in the interest of the commercialization of this technology has been observed due to this highly demanded technology, which utilises the inexpensive biomass. Currently majority of the world's bioethanol production depends upon the sucrose or starchy biomass such as molasses, beet starch, corn starch, cane juice etc. But these feedstocks are also used as food, which creates a direct competition between being used as food and feed. Owing to this major disadvantage, economical switch of non-food lignocellulosic biomass to ethanol is actively sought.

The production of bioethanol from these raw materials includes various steps of pre-treatment, saccharification and finally the ethanol recovery from the fermentation product. The saccharification (hydrolysis) is a very important as it produces fermentable sugars. Two types of hydrolysis mainly employed for biomass conversion i.e. enzymatic hydrolysis and acid hydrolysis. Acid hydrolysis is a conventional method, which has several disadvantages. However, enzyme hydrolysis is more advanced and efficient method, function without generating any toxic substance. This method is in a phase of rapid development and carries immense potential for establishing a cost effective technology [165]. This technology can be widespread if we able to produce cellulose enzyme more efficiently. Enzyme production can be enhanced by using the raw material wisely or by strain development.

Cellulases are multi enzyme complexes whose secretion and action is very carefully regulated by the regulatory machinery of the organism. Cellulose involves three type of hydrolysis activities i.e. endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91), β-glucosidase (EC 3.2.1.21) [157]. Endoglucanases randomly cleave the internal β (1→4) glycosidic bonds of cellulose; cellobiohydrolases produces cellobiose by attacking the ends of cellulose chain and β-glucosidase cleave the cellobiose (a dimer of two glucose molecules linked by a glycosidic bond) to monomeric glucose. The detail mechanism of action of cellulose is shown in Figure 1.7.

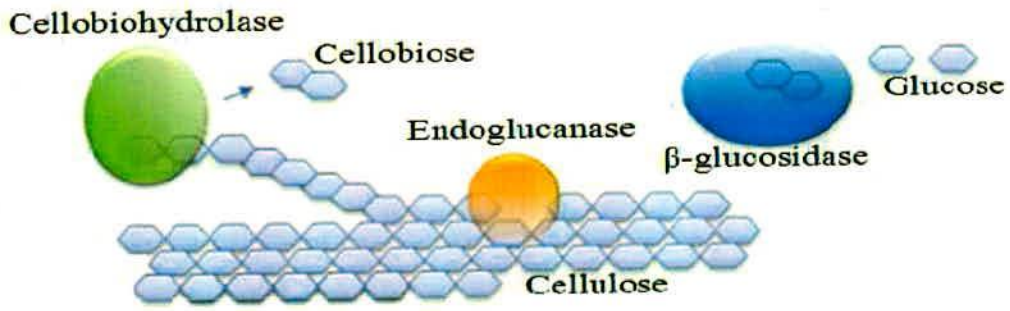


Figure 1.7. Schematic diagram of cellulose hydrolysis by cellulase complex

Saccharomyces cerevisiae produces insufficient β-glucosidases (BGL) resulting in cellobiose accumulation. Cellobiose inhibits both endoglucanases and cellobiohydrolases activity [211]. Overexpression of heterologous β-glucosidases can work at high glucose concentration to improve the saccharification efficiency of natural *Saccharomyces cerevisiae*.

1.4.3. Biomass conversion

The conversion of cellulosic waste material to ethanol involves three major steps i.e. pretreatment, hydrolysis, fermentation and distillation [72] (Figure 1.7). The plant biomass is pretreated to open up the complex architect plant structure and to reduce the size of the feedstocks. This step is considered to be most important step owing to its large impact on the production of toxic by-product that can affect the fermentation process, digestibility of cellulose, and demands of down streaming waste treatment [72]. Biomass can be pretreated by three different methods i.e. physical, chemical, or biological and can be used individually or combination of these three methods [172]. An efficient pretreatment method should execute the following characteristics: should produce cellulose fibres with much more reduced crystallinity and a lesser degree of polymerization, should contain the fraction of hemicellulose, and should decrease energy stress [172]. Further, the amount of released inhibitors should be in limit so that it can't hinder the fermentation step.

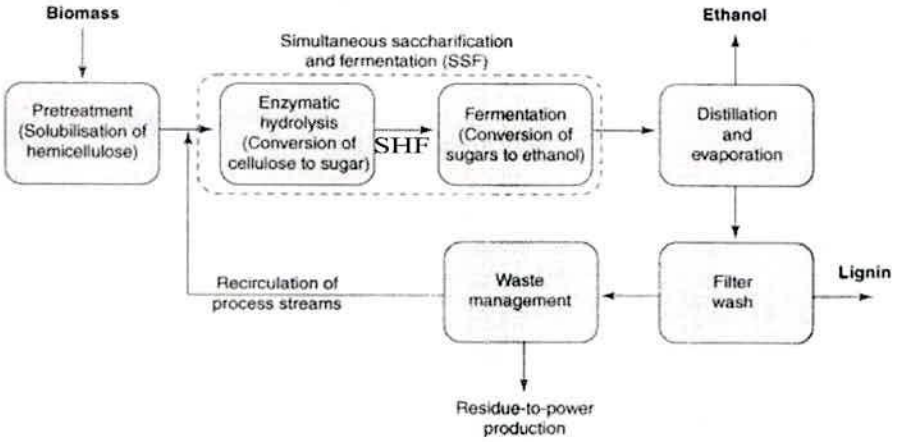


Figure 1.7. Schematic flowsheet of the process of ethanol production from lignocellulosic biomass (SHF: separate hydrolysis and fermentation). Figure adopted from reference no. 72.

Pretreatment is followed by hydrolysis step which include the breakdown of carbohydrate complex polymers to monomeric free sugars. For each glycosidic bond broken it requires the addition of a water molecule. Usually two methods are used to carry out hydrolysis step i.e. enzymatic hydrolysis and acid hydrolysis. Enzymatic hydrolysis is preferred over acid hydrolysis as it produces minimum by-product, require low energy for functioning, need gentle operating condition, produces low amount of wastage for disposal and comparatively gives higher yield. The cellulase enzyme complex comprised of endoglucanases, cellobiohydrolases, and β-glucosidase is used to depolymerise the cellulose to glucose. This cellulolytic enzyme complex may be secreted into the substrate medium called as noncomplexed system or may be attached to the microorganism's cell wall called as complexed system. Organisms having noncomplexed cellulase system are mostly used in industry which produces commercial cellulolytic enzymes as secreted enzyme can be harvested quite easily [157].

The conversion of released sugar into ethanol by ethalogenic microorganisms is known as fermentation and it is the next step in the process. Recently, in industry fermentation is usually done by using *Saccharomyces cerevisiae* owing to its toughness and high ethanol yield. But, it can only ferment monomeric hexose sugars and not the pentoses. The pentose sugars can be fermented by using another pentose utilising organism simultaneously or *Saccharomyces cerevisiae* can be engineered genetically to utilise pentose sugars as well [237]. The high concentration of fermentation product i.e. ethanol can be detrimental to the fermenting organism, so continuous removal of ethanol is required for the smooth functioning of the system.

1.4.4. Strain development to enhance bioethanol production

Generally, large quantities of cellulolytic enzymes are required for the complete hydrolysis of cellulose to ethanol which cost much higher than expected. This problem induces genetic engineers to develop much efficient yeast strains. These strains can show both cellulolytic and hemicellulolytic activity on their cell surface and can directly convert the lignocellulosic material to ethanol [92]. Various technologies have been employed for *Saccharomyces cerevisiae* strain development to better utilize the monomeric sugars. These are (I) disruption of target genes for reconfiguring *Saccharomyces cerevisiae* metabolism (II) optimization of cellobiose consumption pathways (III) concurrent co-fermentation of cellobiose and xylose [116]. Currently, many workers shifted their focus on joint approach of evolutionary adaptation, genetic engineering, and mutagenesis by using ethyl methane sulfonate [151], evolutionary engineering, inverse metabolic engineering, logical metabolic engineering, transposon mutagenesis etc. Nevoigt (2008) used to get recombinant *Saccharomyces cerevisiae* strain having high-capability of xylose fermentation [176]. Recombinant *Saccharomyces cerevisiae* strains with inhibitor tolerance capacity as well as the ability to utilize glucose, arabinose and xylose simultaneously were obtained from evolutionary engineering experiments. Strains with novel fermentation ability have been developed by utilising a combination of intracellular metabolic engineering and cell surface engineering [92]. P6H9 is a *Saccharomyces cerevisiae* mutant strain which has high tolerance toward high concentration of HMF and reported to be used in industries for second generation bioethanol production. This strain is developed from an industrial strain, P6 by using evolutionary engineering [209]. By co-expressing TAL1 and ADH genes in *Saccharomyces cerevisiae*, the production of bioethanol can be enhanced to 127% in the presence of 70 mM furfural, which is a strong inhibitor [92].

1.4.5. Genetic alteration for Yeast improvement

1.4.5.1 Improved ethanol production

Studies showed that, disruption of one or a small group of genes increases the ability of native strains to produce more ethanol. Kong and co-workers (2007) demonstrated that by disrupting GPD2 (glycerol-3-phosphate dehydrogenase) gene and GLT1 (glutamate synthase) gene overexpression in *Saccharomyces cerevisiae*, enhances the ethanol production [124]. This enhancement was achieved by decreasing production of glycerol and by increasing the transformation rate of NADH to NAD+.

1.4.5.2. Resistance to ethanol

Strains from natural sources were used to develop ethanol tolerance strains by utilising different strategies i.e. adaptive evolution. Several workers were able to create mutant strains that are resistance to many natural stresses, such as high temperatures [250], high acetic acid concentration, and high salt concentration [158]. The mutant with deletion of URA7 and GAL6 showed increased resistance to ethanol. A homozygous diploid *Saccharomyces cerevisiae* knockout showed better growth rate than the wild strain in 8% ethanol. Along with this resistance these mutant strains showed higher glucose utilization rate at low temperatures [263]. Global transcription machinery engineering (gTME) technology can use the mutant library TATA-binding protein of *Saccharomyces cerevisiae* to create ethanol resistance strain [261].

1.4.5.3. Improvement of heat and ethanol resistance

Transposon- mediated mutants were currently screened for heat and ethanol resistance. Recently, five strains were developed with high alcohol resistance (15% of ethanol in growth medium) and out of these, two mutants also show resistance towards heat (42 °C). Transposon insertion in the promoter regions of PPG1 and SSK2 genes helps in simultaneous down regulation of PPG1 and SSK2, which produces a heat and ethanol resistance strain. Like this, disruption in the PAM1 gene of *Saccharomyces cerevisiae* produces a resistance strain [117, 118]. The deletion of MSN2 was also linked to ethanol resistance. Msn2p in association with Msn4p controls a global yeast response against stress that includes almost 200 genes [167]. No clear phenotype was observed in single deletion mutants of MSN2 and MSN4, but deletion mutants of *msn*2 and *msn*4 are over sensitive to carbon limitation, oxidative stress, heat shock and osmotic pressure stress.

1.5. Overview- HAD-superfamily Hydrolases

Environment is comprised of numerous halogenated organic compounds and many of them are considered as pollutants. More than 2,400 are found to occur naturally and these halogenated compounds are extensively used as raw material for various industry goods like herbicides, insecticides, soil fumigants, plastics, disinfectants and solvents in the dry-cleaning process [79, 97]. They are the source of environment pollution and cause many human health problems because of their noxiousness, perseverance, and further metabolization into harmful products. Dehalogenases are the enzymes which are of great importance for environmental hygiene as they help in degrading these halogen compounds. They help in recycling and

detoxification of these compounds using a substitution mechanism to cleave the carbon-halogen bond [67, 235].

The haloalkanoic acid dehalogenase enzyme (HAD) superfamily is the major family found with almost 48,000 sequences and is present ubiquitously in the living cells [125]. This family has been named after its first archaeal enzyme; 2-haloacid dehalogenase which is characterized with its structure [5, 99, 193]. HAD superfamily i.e. HADSF constitutes a large variety of HAD-like hydrolases like phosphatases which form 79% of the members and 20% are formed by ATPases [3, 24]. This family is conquered mainly by phosphotransferases and are responsible for the transfer of phosphorous group to the active site aspartate residue. Phosphatases in HAD family are responsible for numerous roles like nutrient uptake, signal transduction, metabolic processes, nucleic acid repair and conservation of metabolic pools [3]. This group of enzymes also have industrial application [222].

Transfer of phosphorous group is generally used for cell signalling in all the living kingdoms from prokaryotes to eukaryotes. It is used as an approach to counter to different internal and external stimuli [56]. Phosphatases help in phosphorous transfer reactions by removing the phosphorous group from biological molecules [184, 122]. Almost 35-40% of the bacterial genome is composed of molecules carrying a phosphate group [131]. Although there are many enzymes responsible for phosphoryl transfer but they belong to different super families which are evolutionary dissimilar and have different structural folds. HADSF are widely distributed from prokaryotes to eukaryotes and comprise more than 30,000 members with 6,805 proteins [4, 12]. The number of HAD genes can vary like from 10 to 20 in bacteria to 115 in *Arabidopsis thaliana* and 100 in humans [166]. Their key catalytic activities are processed through phosphoryl transfer and constitutes a large number of uncharacterized enzymes, some are known to possess dehalogenase (carbon-halogen bond hydrolysis), phosphatase (carbanoyl-phosphate bond hydrolysis), phosphonatase (carbon-phosphate bond hydrolysis), β-phosphoglucomutase activities (carbanoyl-phosphate bond hydrolysis and intramolecular phosphate group transfer) [3, 125]. Mainly they are hydrolases (phosphatases and ATPases) while hexose phosphate mutases form the minority. HADSF being a large superfamily with the HAD fold being seen in large number of diverse phosphohydrolases within a given organism creates a necessary for exploring new catalytic activities and substrate specificities. Divergence of the HADSF catalytic framework is quite significant in terms of embracing both the phosphomutase i.e. the transfer of phosphoryl groups between

hexose hydroxyl substituents and phosphonoacetaldehyde hydrolase i.e. hydrolysis of phosphonate (phosphate-carbon bond) [136,170].

HADSF members have little sequence similarity between them, therefore the reactions catalyzed and their substrate specificities have to be determined biochemically rather than on the basis of their sequence analysis. Kuznetsova et al, (2006) used a set of 80 phosphorylated substrates to illustrate the HAD family substrate specificities in *Escherichia coli* genome. The results showed β-phosphoglucomutase in only one protein out of the 23 soluble HAD members investigated and 21 members showed small molecule phosphatase activities. *E.coli* HADSF members are not specific for a single substrate, instead they show affinity towards a large number of phosphorylated molecules which are transition-intermediates of many biological reactions and thus have wide overlying substrate spectra. Interestingly, it was also seen that they hydrolyse various small phosphate donor molecules like carbamoyl phosphate, acetyl phosphate, phosphoramidate etc. which mimics the autophosphorylation of the response regulator CheY of the two component systems. Thus HADSF have a large number of secondary functions and constitute a pool for evolution of the novel phosphatases. A chemotaxis phosphatase has been characteerized by Zhao et al. (2002) [268].

1.5.1. Haloacid dehalogenases and their classification

Haloacid dehalogenases yield the corresponding 2-hydroxyalkanoic acids after the dehalogenation reaction i.e.

$$RCHXCOO^- + OH^- \longrightarrow RCHOHCOO^- + X^-$$

The halogenases are divided into four types based on their substrate choice and the conformation of the product. The first class catalyse the dehalogenation of the L isomers of 2-haloacids like chloropropionate (2-CPA) and yield the inverting configuration i.e. D-2-hydroxy acids like D-lactate here. The second class i.e. the D-isomer specific is only active on D isomers of the 2-haloacids like D-2-chloropropionate and yields the equivalent L-hydroxy acids. The other 2 classes are formed by the D, L-enantiomer specific, which catalyze the dehalogenation of the L- and D- isomer of haloacids like 2-CPA. One group catalyzes the reaction with inversion of the conformation while the other with retention [80, 149]. Thus their specificity towards the stereochemistry of the substrate is useful to dehalogenate a particular isomer from its racemic mixture for the creation of chiral 2-hydroxyalkanoic acids, which serves as important material for the production of various pharmaceuticals in addition to the cleaning of the hazardous halogenated compounds in the environment.

1.5.2. Unique Rossmann fold

HAD is a large superfamily with its ubiquitous presence in all the kingdoms and its ability to act on the diverse substrates is quite remarkable. HAD superfamily members possess a conserved Rossmann fold consisting of an α/β domain which surrounds the active site. The Rossmann fold is typical to the HADSF in having the acidic catalytic residues at regular positions. They are also unique from other Rossmann fold by the presence of two motifs called as: a **flap** i.e. a beta-hairpin turn and a **squiggle** i.e. a distinct helical turn [3, 24, 153]. These two structural motifs are thought to help the enzyme in having different conformations required for the specificity and catalysis. Also, Rossmann fold is seen with caps or inserts at these two positions.

1.5.3. Overview of the HADSF structure

The catalytic core domain i.e. Rossmann fold has a typical topology with a three layer sandwich with periodically occurring α/β units. Central sheet which forms the active site is composed of five strands which follow 54123 order and are named as S1 to S5 [24]. Although there is little sequence similarity between the members, but they can be recognized by four conserved motifs located around the binding cavity at the C-terminus of the central sheet strands. The active fold is comprised by four loops containing these four motifs which acts as signatures for the family [125]. The motif I,DxD sequence is present at the end of strand 1, where X is any amino acid between the two Aspartate residues. The first aspartate residue corresponds to the nucleophilic residue [36]. These two aspartate residues help in the coordination of Mg^{2+} ions required for the activity of phosphatases. Motif II is the strand S2 with the conserved serine/threonine which helps in binding the phosphorous moiety of the substrate. Motif II and III are thought to help in the stability of the transition intermediate of the catalytic reaction. Motif III is present on the helix located next to strand S4 and contains a conserved Lysine which is expected to stabilize the negative charges on the nucleophilic Asp and the phosphorylated intermediates as seen in case of arginine fingers [2]. This lysine can also be seen on the loop just present succeeding to the helix. Motif IV contains two or three Asp residues at the end of strand S4. These Asp residues follow a particular signature like DD, GDXXXD, GDXXXXD in different subfamilies of HAD [24]. This motif also helps in coordinating the Magnesium ion [99, 185]. This conserved core form the basic catalytic machinery in all the HADSF members facilitating acceptor molecule for its nucleophilic attack and substrate molecule i.e. phosphate moiety in positioning, binding and activation for

its transfer. There are vast diversity of microbes found in soil that are pathogenic to both plants and animals and their proteins used as drug targets [48].

1.5.4. Catalytic mechanism

Mechanism of transfer of phosphorous group and the catalytic machinery involved is conserved throughout the HADSF. They have a common catalytic core Rossmann fold domain comprising of four motifs which binds the Mg^{2+} ion as its cofactor and also reorients the phosphate moiety for nucleophilic attack [3, 136]. The motif I comprises two Asp molecules, of which the first turns for the nucleophilic attack forming an aspartyl-phosphate intermediate [9, 189, 208] while the other i.e. the second Asp acts as a general acid/base catalyst in the phosphorous group transfer in phosphatase and phosphomutase members of the HADSF [40]. This Asp first protonates the substrate leaving group and then deprotonates the nucleophile [136]. However a threonine replaces this Asp in the ATPases members, leading to slower hydrolysis reaction. Also, alanine can be seen at its place in phosphonatases [24]. The aspartyl-intermediate formed then transfers its phosphorous group to a water molecule resulting in the formation of an alcohol in case of phosphatases. In case of mutases, it is transferred to a hydroxyl group of sugar [153]. These catalytic core residues are conserved in all the HADSF phosphotransferases and are thought to form a mold which helps in stabilizing the trigonal bipyramidal transition states intermediates by electrostatic interactions. The beginning of the interaction i.e. the formation of enzyme substrate complex to the aspartyl-intermediate and then to the end product complex is stabilized by the electrostatic interactions with the equatorial and axial atoms of the moving phosphate moiety in the mold [3]. This has been reported in Hexose phosphate phosphatase BT4131 from *Bacteroides thetaiotaomicron* VPI-5482 [154]. The crystal structures of the wild type and the mutant type of Hexose phosphate phosphatase (HPP) in complex with Mg^{2+} and vanadate or tungstate have been solved. The tungstate and the vandate complex forms a trigonal bipyramidal symmetry with the Asp8 (nucleophilic residue) and an oxygen in the apical location. The mutation of the second Asp in motif I i.e. Asp10 which acts as a general acid/base in the catalysis reaction leads to the shattering of the electrostatic interactions and the bipyramidal complex. Thus the chemical mechanism and the catalytic machinery involved is preserved in all HADSF members.

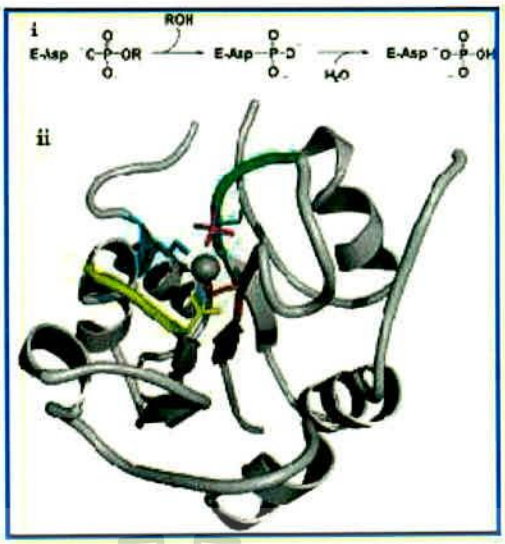


Figure 1.7 (i) Catalytic reaction of HADSF along with the **(ii)** catalytic fold of phosphatases.

1.5.5. Substrate specificities

The catalytic core domain cleft residues is used for the transfer of phosphate group while the residues needed for substrate specificity are present outside this cleft [135, 137, 267]. The reason for the substrate diversity in the HAD subfamilies lies in the different substrate residues present outside the core domain in an additional region called the cap domain. This cap domain shows widespread evolutionary divergence in its fold and location. The HADSF can be divided into three subfamilies based on this cap domain responsible for specific substrate recognition [155]. They are:

1.5.5.1. Subfamily I

The Cap C1 subfamily consists of an α-helical domain located between the catalytic motif I and II. This cap is introduced right into the β-hairpin of the flap motif and is responsible for checking the entry of substrates to the active site. Single substrate specificity loop has been found and its residues interact with the leaving group. The size of the cap is variable from two to eight or more helices [3]. This cap can be found in the cN-I nucleotidase families and acid phosphatases. Some examples are *α*-phosphoglucomutase (*α*-PGM) [138], phosphonoacetaldehyde hydrolase (phosphonatase) [166, 135] and phosphoserine phosphatase (PSP) [247,248].

1.5.5.2. Subfamily II

The capping domain Cap C2 consists of mixed α/β fold and is located between the motif II and III. This sequence is found in the linker region that occurs after strand S3. This cap is found to form a roof over the active site and the displacement of flap helps in regulating

its activities [24]. There can be two substrate specificity loops depending on the size of the cap domain. This C2 cap can be further divided into two subfamilies i.e. C2a (NagD phosphatases) and C2b (phosphoglycolate phosphatase) depending on the α/β topology amongst their members [153, 229].

1.5.5.3. Subfamily III

In this C0 subfamily, no domain is present between any of the three motifs i.e. the first and second or the second and third motif. This subfamily acts on the phosphorylated proteins and can be seen in histidinol-phosphatase, magnesium dependent phosphatases [185]. Example is deoxy-D-mannose-octulosonate 8-phosphate (KDO 8-P) phosphatise [155].

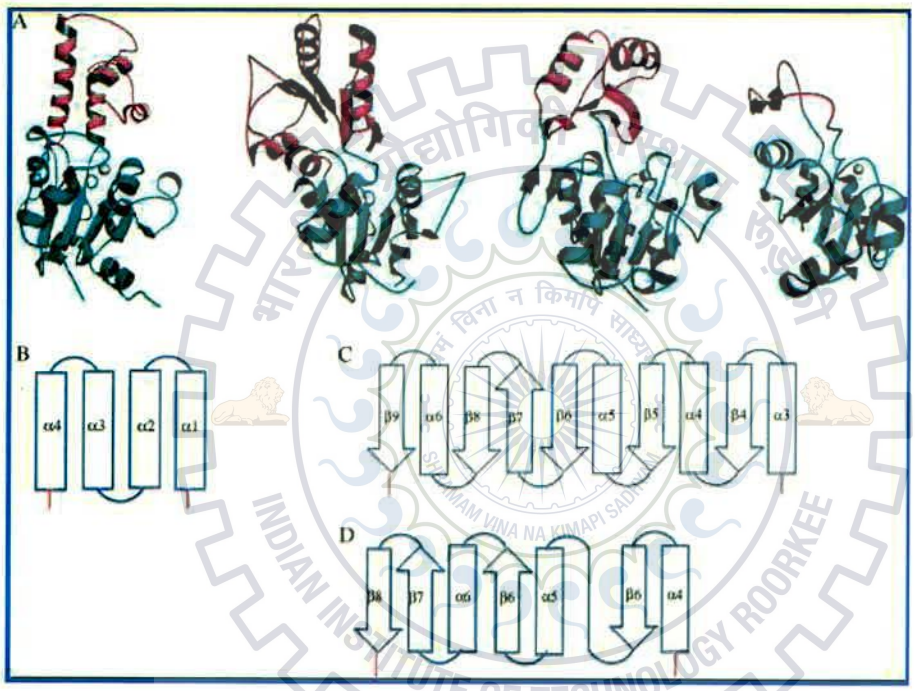


Figure 1.8. Ribbon representations of (A) β-PGM (Subfamily I), NagD (Subfamily IIA), phosphoglycolate phosphatase (Subfamily IIB) and Magnesium-dependent phosphatase (Subfamily III) shown from left to right. Topology of cap domains of type I subfamily (B), type IIA subfamily (C), type IIB subfamily (D).

1.5.6. Functioning of the cap domain

Most of the HADSF members have Cap 1 or Cap 2 or rarely both of them in some members. Rearrangement of the cap domain in respect to the core domain can further result in two configurations i.e. the open and the closed cap conformation [266]. In the closed-cap domain, residues from the cap domain that form the substrate specificity loop protrude into the active site to facilitate binding of the reactant and its catalysis [135, 153]. However in the open-cap domain, the active site faces the solvent directly for reactant binding and release of leaving group [169]. The cap-closing movement differs between the Cap1 and Cap2 members

although both of them act on small substrates only. Cap 1 movement is just like a clam i.e. it moves on a mechanical hinge as compared to the screw motion of the cap domain over the catalytic core region in Cap2 [3]. Thus the cap domain is responsible for the distinct substrate specificity and its catalysis.

However, the idea that the cap domain location and its occurrence can be predicted from the sequence alone is not worthy as absence of the cap domain still doesn't allow an open active site [3]. Oligomerization of the core domain helps in closing the active site as a boundary and also provides specific residues from adjacent region for the catalysis. This can be seen in case of *Bacteriodes thetaiotaomicron* 2-keto-3-deoxy-D-glycero-D-galacto-9-phosphonononic acid phosphate (KDN9P) phosphatase [155].

CHAPTER 2

CLONING, EXPRESSION, CHARACTERIZATION AND BIOINFORMATIC STUDIES OF PRGH1

2. CLONING, EXPRESSION, CHARACTERIZATION AND BIOINFORMATIC STUDIES OF *PUTRANJIVA ROXBURGHII* FAMILY 1 GLYCOSYL HYDROLASE (PRGH1)

2.1. Introduction

The structural diverseness of the carbohydrates provides an advantage for the multiplicity of biological functions, from storage to highly precise signalling roles in the living organisms. This diverseness is the consequence of the action of large spectrum of enzymes involved in the synthesis, modification and breakdown of carbohydrates. Glycoside hydrolases (GHs; EC 3.2.1.x) are a wide spread collection of enzymes. They specifically cleave the glycosidic linkage between a non carbohydrate moiety and a carbohydrate or between two or more carbohydrates. Almost in all living organisms, the selective cleavage of glycosidic bonds is vital in a variety of essential biological processes. These enzymes constitute 133 protein families (CAZy: http://www.cazy.org/) basing on the amino acid sequence similarities [38, 95]. Plant GH enzymes locate in different compartments of cells and play an important roles in various physiological functions like cell wall polysaccharide metabolism, hydrolysis of starch, lignification, seed germination, activation of phytohormones and defence compounds, mobilization of energy, signalling, vitamin B6 metabolism, participation in biosynthesis and remodulation of glycans, symbiosis [21, 141, 164, 171, 213].

Glycoside hydrolases family 1 contains different enzymes with well-known functions i.e. β-fucosidase (EC 3.2.1.38), β-glucosidase (EC 3.2.1.21), β-galactosidase (EC 3.2.1.23), β-mannosidase (EC 3.2.1.25), exo-β-1, 4-glucanase (EC 3.2.1.74) etc. β-glucosidase characterized till date fall primarily in glycoside hydrolase families 1, 3 and 5 with family 1 β-glucosidases being more abundant in plants (93). β-glucosidase of this family may have high specificity for glucosides or in addition to this may hydrolyse fucosides and/or galactosides [25, 77, 90, 142, 244]. In the case of almond enzyme a common catalytic site showed all the above three activities [243] where as enzyme from sheep liver showed two active sites for these three activities [32]. Clearly, β-glucosidases show a range of substrate specificities. Since in plants, β-glucosidases play some vital roles, it is necessary that these enzymes particularly act upon their own substrates and this is achieved by the enzyme localization in the cells and the nature of aglycone specificity of the enzymes. β-glucosidase, β-galactosidase activity are comparatively more studied than the β-fucosidase activity. Most of the extensively studied fucosidases are α-fucosidases. β-fucosidases are less common and

their functions are not understood completely, still from biochemical and biophysical point of view this enzyme is very interesting [77].

The amino acid sequences of the plant GH1 enzymes share high percentage of identity among each other and the members of this family possess a $(\beta/\alpha)_8$-barrel structure. The two catalytic glutamates included in the Y(I/V)TENG and TFNEP conserved peptide motifs of this family act as the catalytic nucleophile and catalytic acid/base respectively [34]. Despite of all these similarities, all the plant GH1 enzymes show great diversity in substrate specificity. For example the β-glucosidases from the two species of *Dalbergia* are sharing over 80% amino acid sequence identity, despite of this they weakly hydrolyzes each other's natural substrate [205]. Similarly, dhurrin is hydrolyzed by dhurrinase (Dhr1) of sorghum and not by β-glucosidase of maize (ZmGlu1) regardless of the high sequence similarities between them. This substrate specificity is attributed by the small variation in the amino acid sequence adjacent to the active site of enzyme.

It is important to compare between native and recombinant form of a same protein. A number of comparison studies demonstrated that, there was considerable distinction between the two forms of protein from structural and functional point of view [34, 177, 269]. Specifically it is very critical when we compare a native form of a eukaryotic protein with its recombinant form heterologously expressed in bacterial system. It helps to depict the role of post translational modifications towards the behaviour of the proteins. The most common and best studied post translational modification is N-linked glycosylation, where oligosaccharides are attached uniquely to the asparagines residues found in Asn-X-Ser/Thr recognition sequence in protein. Several studies showed that the N-linked glycosylation is important for both structure and function of some eukaryotic proteins [29, 76, 174,].

Our earlier study targeted on the detection and partial characterization of a heat resistance family 1 glycoside hydrolase enzyme (PRGH1) having both β- glucosidase and β-galactosidase activities [183]. This native form of this enzyme was sourced from the seeds of *Putranjiva roxburghii* plant. It is a medicinal plant belongs to *Euphorbiaceae* family. The main purpose of this study is to make a comparison between the native (nPRGH1) and recombinant form (rPRGH1) of this enzyme. For this, the *prgh1* gene was cloned and expressed in bacterial system and the meticulous characterization of both native and recombinant form of this enzyme has been carried out. In this work, we also identified the critical amino acids supposed to take part in the activity by making several mutants. A meticulous bioinformatics work has been carried out to support the whole study.

2.2. Materials and methods

2.2.1. Materials

All the substrates and inhibitors, methyl α-D-mannopyranoside, Ni-NTA, imidazole, IPTG, PMSF, inhibitor cocktail were obtained from Sigma-Aldrich Pvt. Ltd. Concanavalin A-agarose was obtained from Merck and restriction enzymes, taq polymerase, phusion polymerase, ligase and PNGase F were obtained from NEB.

2.2.2. Purification of native enzyme

One-step purification strategy for native PRGH1 enzyme (nPRGH1) was employed as compared to the lengthy four step method reported earlier [183]. We employed affinity chromatography using concanavalin A-agarose resin. Mature green seeds of *P. roxburghii* collected locally were used for purification. After removal of the hard seed coat, seeds were crushed and soaked in binding buffer (20 mM potassium phosphate; pH 7.4, 500 mM NaCl). A clear extract was prepared following centrifugation and mixed to the concanavalin A-agarose column pre-equilibrated with binding buffer. Then the column along with the extract was incubated at 30 °C for 1 hr with intermittent mixing at regular time interval. The column was washed extensively with the binding buffer followed by a stepwise elution using 20 mM and 500 mM of α-D-mannopyranoside dissolved in binding buffer. An incubation period of 15 min at 30 °C was given prior to each elution. Eluted protein fractions were evaluated on a 12% SDS-PAGE and pure nPRGH1 protein fractions were dialyzed against the buffers according to the next experiments.

2.2.3. Isolation of mRNA, and synthesis of cDNA

TRIzol-based method by Meng *et al.* was used for total RNA isolation from 7-week-old *P. roxburghii* seeds with slight modifications [161]. 25 mg of polyvinylpyrrolidone (PVPP) per gram of seed was added in the extraction buffer for removal of pigments. The obtained RNA pellet was dissolved in nuclease free water and the RNA purity ($A_{260/280}$) and RNA yield (A_{260}) were calculated spectroscopically. First strand of cDNA was reverse transcribed from total RNA using oligo $(dT)_{17}$ and gene specific primers as per experimental need. Initially total RNA was denatured and to the denatured RNA following components were added to synthesize cDNA.

5X Reaction buffer	10 µl
20 mM Olgo(dT)$_{17}$ primer/GSP	5.0 µl
10 mM mixed dNTPs	5.0 µl
20 unit/µl placental RNAse inhibitor	0.5 µl
100 unit/µl Reverse transcriptase	1.0 µl
H_2O up to	50.0 µl

The above reaction mixture was incubated for 90 minutes at 40 °C in PCR machine. After 90 min incubation, reaction was stopped by heating at 95 °C for 5 minutes. The synthesized cDNA was cleaned and concentrated using kit (Zymo Research, USA) and stored at -20 °C for further use.

2.2.4. Cloning of PRGH1 gene

On the basis of N-terminal sequence of nPRGH1, several combinations of forward degenerate primers were designed [183]. By using these forward degenerate primers and reverse oligo (dT)$_{15}$ adaptor primer (listed in Table 1.1), we tried to perform 3' RACE-PCR to get the partial gene sequence but unfortunately it didn't work. From the sequence alignment of the glycoside hydrolases family 1 proteins of *Euphorbiaceae* family, we found some highly conserved regions. Basing on two conserved regions (FGDRVK and YITENG) degenerate primers were designed. FGDRVK and YITENG were used to design degenerate forward (DP 1) and degenerate reverse primer (DP 2) respectively. The segment flanking by this two amino acid stretches was amplified and sequenced. To get the downstream and upstream gene sequence, on the basis of partial determined nucleotide sequence, gene specific primers were designed. For 3' RACE-PCR, a gene specific forward primer (GSP 1) and reverse oligo (dT)$_{15}$ adaptor primer were used [202]. To get the upstream gene sequence, by using specific reverse primer (GSP 2) specific cDNA was prepared. Then, a nested PCR was carried out by using an N-terminal sequence based forward primer (DP 3) and a specific reverse primer (GSP 3). PCR product was purified, cloned and sequenced. Based on this sequence, two specific reverse primers (GSP 4) and nested primers (GSP 5) were constructed to perform 5'RACE to get the upstream signal sequence of PRGH1 gene [203]. Reverse transcription was carried out with the GSP 4 primer for 5' RACE-PCR, and then following nested PCR was carried out by using GSP 5 as reverse primer and forward oligo (dT)$_{15}$ adaptor primer. The 5' RACE-PCR product was cloned and sequenced. The complete ORF was acquired by using GSP 6 and GSP 7. In this amplification, cDNA was made by using oligo (dT)$_{17}$ primer. The PCR product

was run on 0.8% agarose gel copolymerized with ethidium bromide. The amplification product was sequenced after ligating into pGEM-T cloning vector (pGEM-T-PRGH1).

Table 2.1 List of primers used for the cloning and expression of PRGH1.

Primer specification	Sequence 5'- 3'
GH1nFW1	TCCTGATGATTTCATTTTTGG
GH1nFW2	AAYWSNTTYATHMGNWSNGCNTT
GH1nFW3	GCNTTYCCNGAYGAYTTYATHTT
GHnFW4	AAYWSNTTYATHMGNWSNGCNTTYCC
DP1	CAGCCATATGTTTGGKGAYCGAGTSAAR TAY TAY
DP2	CGCACTCGAGTCAYCCATTYTCAGTAATRTA
GHSPFW1	CAAGAGCTTATGCAACTCGAGGATATGAC
GHSPFW2	GGCATTTGGATCCTATAACCTATGGTCACTATCC
GHSPFW3	CCAAAGATGCGTACAGAAATCCAACAATG
GHSPFW4	GGHAGAGGASCYAGTRTHTGGGAYACATTTAC
GHSPRV1	CATTGTTGGATTTCTGTACGCATCTTTGG
GHSPRV2	GTGACCATAGGTTATAGGATCCAAATGCC
GHSPRV3	GTCATATCCTCGAGTTGCATAAGCTCTTG
5'race1	CATCTCCATTGCTGCGGTCTG
5'race2	CCTTCGTACTGGTAAGCTGC
RACE Adaptor primer	CTGATCTAGAGGTACCGGATCCTTTTTTTTTTTTTT
GHSPFW	ATGATCACCATGGCTCCTCATCAG
GHSPRV1	TTAAGCGGCTGCTGATCTAATAGAATAACC
GHSPRV2	TTAAGCGGCTGCTGATC
Oligo$(dT)_{17}$ Primer	TTTTTTTTTTTTTTTTT
PRGH1EF	CAGCGCTAGCAATTCCTTCAACAGAAGTG
PRGH1ER	CTTGCGGCCGTCATGCTGATCTAATAGAATAAC

2.2.5. Construction of the expression plasmid

The ORF of prgh1 gene, excluding the native signal sequence was amplified from the recombinant plasmid pGEM-T-PRGH1 by using the primers PRGH1EF and PRGH1ER having overhanging regions encoding NheI and EagI restriction sites respectively. The purified amplification product was double digested with NheI and EagI and was inserted in to

pET-28c(+) vector. Correct construction was checked by restriction enzyme digestion and subsequent DNA sequencing.

2.2.6. Overepression and purification of recombinant PRGH1

The recombinant pET-28c(+)-PRGH1 was transformed into freshly prepared *E. coli* BL21 (DE3) cells for expression of recombinant PRGH1 (rPRGH1) enzyme. The rPRGH1 was overexpressed at 16 °C for 12 hrs with 200 μM IPTG. Cells were harvested by using centrifuge and kept at -80 °C until further actions. Expression and solubility of the rPRGH1 was checked by examining the supernatant and pellet of lysed cells on 12% SDS-PAGE.

For the purification of histidine-tagged rPRGH1, metal ion affinity chromatography was employed. The frozen cell pellets were resuspended in buffer A [50 mM Tris-Hcl; pH 8.0, 15 mM imidazole, 500 mM NaCl, 5% glycerol (v/v), 5 mM β-ME with lysozyme (0.5 mg/ml) and 0.3 mM Phenylmethylsulphonyl fluoride and protease inhibitor cocktail included]. Cells were sonicated and subsequently the lysate was clarified by centrifugation. The obtained supernatant was mixed with Ni-NTA column pre-washed with buffer A and extensive washing with buffer A was carried out to remove unbound proteins. The bound proteins were collected by stepwise elution with buffer B [50 mM Tris-Hcl; pH 8.0, 50 mM imidazole, 5 mM β-ME, 500 mM NaCl, 5% glycerol (v/v)] and buffer C [50 mM Tris-Hcl; pH 8.0, 5 mM β-ME, 300 mM imidazole, 500 mM NaCl, 5% glycerol (v/v)]. After examining the eluted fraction on 12% SDS-PAGE, the pure protein fractions were collected together. Further, the pure rPRGH1 was administered to TEV protease cleavage for removal of the N-terminal His-tag at 4 °C for 12 hours and simultaneously dialyzed against dialysis buffer containing 50 mM Tris-Hcl; pH 8.0, 5% glycerol(v/v), 100 mM NaCl. A reverse Ni-NTA column was run to remove the cleaved His-tag, uncleaved protein, and the His-tagged TEV protease. The flow through having His-tag cleaved rPRGH1 was saved and concentrated by using an Amicon Ultra concentrator with a cut off value of 30 kDa (Milipore, Bedford, Massachusettes, USA).

2.2.7. Circular dichroism spectroscopy

CD study of rPRGH1 was carried out on a Chirascan Circular Dichroism Spectrometer. Far-UV CD spectra (190-260 nm) were collected using 1 mm path length quartz cell at 25 °C with band width of 1 nm and time per point was 0.5 s. The CD spectra were analyzed using online DICHROWEB programme. The CD results were expressed in terms of mean residue ellipticity.

2.2.8. Oligomerization study of PRGH1

Previously, the gel filtration chromatography study showed that PRGH1 is a monomeric enzyme with an approximate molecular weight of 66 kDa. Here, we analyzed the effect of concentration on the oligomeric state of nPRGH1 by using NATIVE-PAGE, dynamic light scattering and analytical size exclusion chromatography. Both the nPRGH1 and rPRGH1 were used at two concentrations (15 and 45 μM).

(a) NATIVE-PAGE

NATIVE-PAGE was performed for various concentrations of both nPRGH1 and rPRGH1. Electrophoresis was conducted by using Tris-glycine buffer (pH 8.3) as electrode buffer and at a constant current of 20 mA at 4 °C for 4 h. Subsequently, the gels were stained by Coomassie Brilliant Blue R-250.

(b) Analytical size exclusion chromatography

Analytical size exclusion chromatography was used to determine the oligomeric states of both nPRGH1 and rPRGH1. The chromatography of protein samples were performed using Agilent Bio SEC-5 (5 μm, 300 Å, 7.8 mm × 300 mm) HPLC column. Experiments were carried out at a flow rate of 0.4 ml/min using 50 mM Tris-HCl; pH 8.0, 100 mM NaCl at room temperature. Aliquotes of 20 μl protein samples were injected at two protein concentrations i.e. 15 and 45 μM. A standard curve was created using bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken conalbumin (75 kDa), bovine serum albumin (66 kDa), and chicken ovalbumin (44 kDa). Prior to injection all samples were filtered through 0.22 μm filters.

(c) Dynamic light scattering

The effect of concentration on the oligomeric state of nPRGH1 and rPRGH1 were further investigated using dynamic light scattering techniques. Samples of protein preparation (15 and 45 μM) were filtered through 0.22 μm filter prior to the collection of scattering data and transferred into a quartz cuvette and placed in a thermostated (22 °C) cell holder. To determine the hydrodynamic radii of various protein samples, Dynamic light scattering (DLS) experiments were executed in laser-spectroscatter 201 by RiNA GmbH (Berlin, Germany). Data was analyzed by using PMgr v3.01p17 software provided with the instrument. All the DLS experiments were repeated ten times and an average value were calculated with standard errors.

2.2.9. Enzyme activity assays and kinetics parameters

a) Assay using pNP-Glycosides

In our earlier study we have characterized nPRGH1 using *p*-nitrophenyl β-D-glucopyranoside (*p*NP-Glc) and *p*-nitrophenyl β-D-galactopyranoside (*p*NP-Gal) as substrate. In the present study we have used *p*-nitrophenyl β-D-fucopyranoside (*p*NP-Fuc) as a new substrate to characterized nPRGH1. The kinetics parameters like Km, Vmax, Kcat, and Kcat/Km were determined from Lineweaver-Burk plot by using *p*NP-Fuc. Recombinant PRGH1 was assayed using all the three (*p*NP-Glc/*p*NP-Gal/*p*NP-Fuc) substarate. Kinetic parameters were determined for substrate of interest.

b) Assay using glucooligosaccharides and natural β-D-Glycosides

The activity of PRGH1 on various glucooligosaccharides and natural substrates was analyzed by quantifying the released glucose using glucose assay kit (Sigma). Activity was evaluated in triplicate in 50 mM citrate phosphate buffer; pH 4.8, at 65 °C after 1 hr of incubation. Reactions having no enzymes served as controls.

2.2.10. Mixed substrate analysis

By using equimolar combinations of pNP-glycosides as substrates (pNP-Fuc/pNP-Glc, pNP-Fuc/pNP-Gal, pNP-Glc/pNP-Gal), mutual competition studies were carried out. The theoretical maximal velocities for competing substrates were calculated using the equation used in the previous study [26]. In each competition study the molar ratio between the two substrates was kept one.

2.2.11. Characterization of nPRGH1 and rPRGH1

The optimum pH values for the nPRGH1 activity towards pNP-Glc and pNP-Gal were determined previously [183]. The pH profiles were constructed by determining activity of nPRGH1 towards pNP-Fuc and activities of rPRGH1 towards all three pNP substrates. The reactions were carried out as described in our previous study [183]. The pH stabilities of nPRGH1 and rPRGH1 were calculated by incubating the reaction mixtures without the substrates at 37 °C for 24 hrs in a range of different pH buffers. The residual enzymatic activity was measured with pNP-Glc as a substrate.

In our previous study, the optimum temperature values for the nPRGH1 were determined by using pNP-Glc and pNP-Gal as substrates. The effect of temperature on

rPRGH1 activities were measured by keeping the enzymes at temperature starting from 25 °C to 90 °C for 10 min and pNP-Fuc, pNP-Glc, and pNP-Gal used as substrates. The temperature optimum for the fucosidase activity of nPRGH1 was also studied by using the standard procedure. The thermostability of nPRGH1and rPRGH1 was studied by incubating the reaction mixture for 30 min, with temperatures starting from 25 °C to 90 °C and the residual activity was measured by using pNP-Glc as substrate. All the experiments presented were executed in triplicate.

2.2.12. Construction of PRGH1 mutants

Using the QuickChange site directed mutagenesis kit (Stratagene) mutations were performed by following the manufacturer's instructions. The pET-28c(+)-PRGH1 plasmid was used for mutations. Mutations and synthetic mutagenic oliginucleotide primers are indexed in table 2.2. PCR was carried out with Phusion polymerase, a pair of primers designed for a particular mutation, and pET-28c(+)-PRGH1 plasmid as template. Resulted PCR product was subjected to *DpnI* restriction enzyme digestion to digest the methylated template. XL-1 blue competent cells were transformed with the digested product. The successful introductions of the desired mutations were verified by DNA sequencing of the mutated constructs using pET-28c(+) specific universal primers. The rPRGH1 mutant proteins were expressed, purified and assayed by using *p*NP-Glc as described above for the wild type rPRGH1.

Table 2.2. Mutations and their respective designed primers.

Mutation	Forward Primer (5'-3')	Reverse primer (5'-3')
E173N	AACAATTCCTTCCAGAGAAGT GCCTTTCCCGATGAC	GTCATCGGGAAAGGCACTTCTC TGGAAGGAATTGTT
E389N	GGACTACTTTCAATAACCCAAG AGCTTATGCAACTCGAGG	GCTCTTGGGTTATTGAAAGTAG TCCAATACTTCACTCGGTC
N172Q	CATTACCAACAATGGATACAA TCAAGATGATAATGGCACCGT ACC	CATCTTGATTGTATCCATTGTTG GTAATGTACATTGTTGGATTTC

2.2.13. Bioinformatics analysis

The primary sequence of GH1 enzyme of *Putranjiva roxblurgii* was retrieved from NCBI database using accession no. AHN8564. BLAST was performed to identify the sequences showing significant similarities to GH1 and these homologous sequences were

retrieved from NCBI database. Multiple sequence alignment was done using CLUSTALW taking default parameters [227]. MEGA 6 was used to construct the phylogenetic tree with 1000 bootstrap replicates using Maximum likelihood method based on JTT matrix-based model [225]. Secondary structure predictions were done using ESPript server [195]. Since there is no crystal structure available for GH1 enzyme in PDB, homology modeling was done to predict its three dimensional structure. BLAST search was done against PDB to identify the suitable template for GH1/ homology modeling. Crystal structure of Rice with PDB ID: 3PTK was used as a template based on the 55% sequence identity between query and template protein sequences. Models were generated/Modelling was performed using Modeller 9.10. Out of 10 models generated, models having lowest DOPE score were subjected for stereochemical evaluation using WHAT_CHECK Verify 3D and PROCHECK. Validating the Ramachandran plots, the best model was subjected to energy minimization using Swiss-PDB viewer. Refined model was further verified using ERRAT of SAVES server. Prosa energy plot is also used to evaluate the model for potential errors. Molecular docking studies were done using Autodock 4.2.5.1. The structure of ligand molecules; 4-Nitrophenyl beta-D-fucopyranoside, 4-Nitrophenyl beta-D-galactopyranoside, 4-Nitrophenyl beta-D-glucopyranoside, D-(+)-Cellobiose, D-(+)-Cellotriose and D-(+)-Cellotetraose were retrieved from PubChem database. The energy minimized model of GH1 and the ligands were prepared for docking studies using MGL tools 1.5.6. Grid maps were computed using Autogrid4 having grid dimensions 66 x 72 x 62 Å with grid spacing 0.375 Å centered around the active site. Lamarckian genetic algorithm was used for docking with a population size of 150 and maximum no. of 250000 energy evaluations with maximum no. of 27000 generations. No. of GA runs were set to be 50. After successful completion of docking, results were sorted on the basis of their energy profiles and the conformation with the lowest energy was analyzed for hydrogen bonding interactions with the protein using PyMOL.

2.3. Results and Discussion

2.3.1 Purification of nPRGH1

The nPRGH1 was purified to homogeneity by concanavalin A-agarose affinity chromatography. Most of the protein bound to concanavalin A-agarose was eluted with 500 mM of α-D-mannopyranoside. The purified protein showed a single band on SDS-PAGE with apparent molecular mass of approximately 66 kDa (Figure 2.1). The overall yield using this purification protocol was ~53%. The purity of PRGH1 was estimated to be ~98 fold greater than that of crude seed extract. This result also suggests that it is a glycosylated protein as it is bound to the concanavalin A. By using this method we save the time and got high protein yield as compared to our previous method [183].

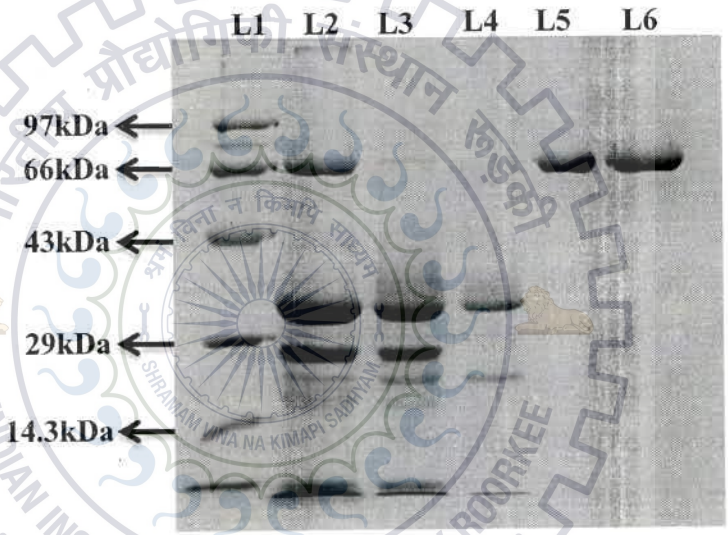


Figure 2.1. SDS-PAGE (12% *w/v*) analysis of native PRGH1 purification. L1, Molecular mass standards; L2, crude extract; L3, flowthrough; L4, first step elution; L5 and L6, second step elution.

2.3.2 Cloning of the PRGH1 gene

Degenerate primers were designed based on the conserved sequences of glycoside hydrolases family 1 proteins of *Euphorbiaceae* family to amplify a partial sequence of PRGH1 gene. Using DP 1 and DP 2 a 693 bp amplicon was obtained. Specific forward primer and reverse primers were designed (Table 1.1) based on the determined 693 bp fragment to amplify the full length gene. Approximately a 600 bp amplicon was acquired after 3′ RACE-PCR which include the 3′ untranslated region along with poly-A tail. Using the forward degenerate primer (DP 3) and reverse GSP 3, a 681 bp fragment was obtained. Using 5′ RACE-PCR a 310 bp amplicon containing the upstream signal sequence was obtained.

Finally, the complete ORF of 1617 bp was obtained by using GSP 6 and GSP 7 primers (Fig. 2.2 A). All the PCR products were cloned into pGEM-T vector and sequenced. Sequence analysis showed that the sequence of PRGH1 contained an ORF of 1617 bp. The deduced protein sequence consists of 538 amino acids which contained an N-terminal signal sequence of 28 amino acids as predicted by the SignalIP server. The prgh1 gene without signal sequence was cloned in pET-28c(+) vector (Figure 2.2 B). The gene sequence was submitted to NCBI database with accession no- KF006311. Nucleotide and deduced amino acid sequence of PRGH1 is shown in figure 2.3.

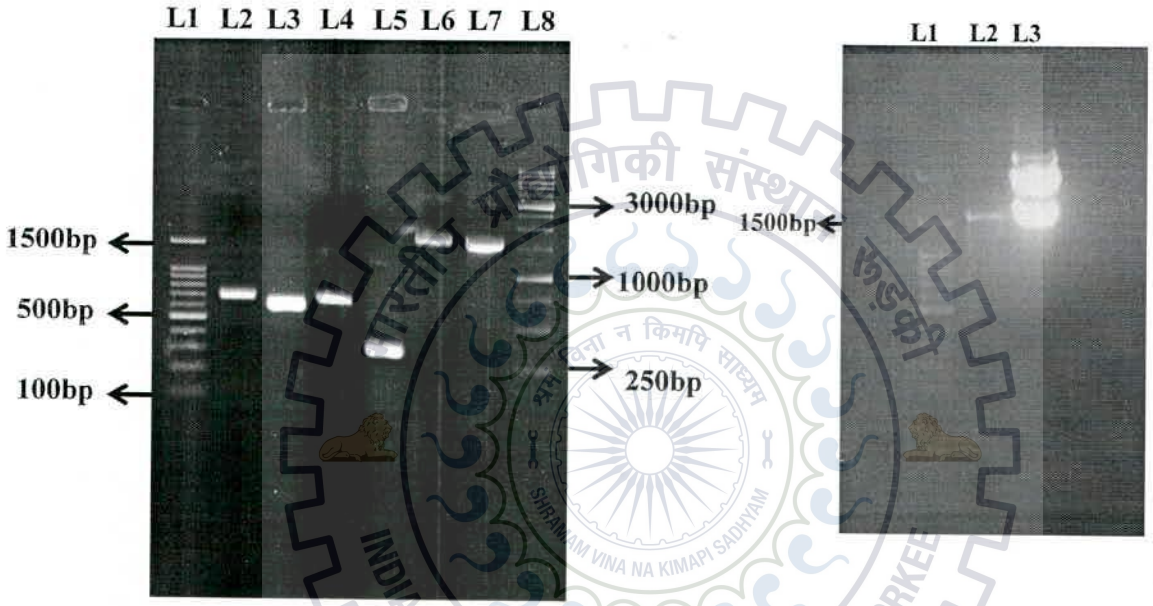


Figure 2.2 (A) Amplification of *prgh1* gene. L1, 100 bp molecular weight marker; L2, L3 and L4, different 3' RACE-PCR products; L5, 5' RACE-PCR product L6, PRGH1 gene with signal sequence; L7, PRGH1 gene without signal sequence. **(B)** Cloning of PRGH1 in expression vector. L1, 100 bp molecular weight marker; L2, PCR amplification of the one of the positive clone; L3, NheI and EagI digestion of the same positive clone.

atgatcaccatggctcctcatcagaaccagtttcttctgtttattggagtatctttggtc
M I T M A P H Q N Q F L L F I G V S L V
ctcttaagttcctatgcaacagcaaacaattccttcaacagaagtgcctttcccgatgac
L L S S Y A T A N N S F N R S A F P D D
ttcatttttggggcatccgcggcagcttaccagtacgaaggtgaagcaaacaaaagtggc
F I F G A S A A A Y Q Y E G E A N K S G
agaggcccaagcgtatgggatatcttcactcacgagtacccagagaaaataacagaccgc
R G P S V W D I F T H E Y P E K I T D R
agcaatggagatgaagcaattgatttctatcatcgatacaaggaagatattcaaaggatg
S N G D E A I D F Y H R Y K E D I Q R M
aagaatatgaatttggatgcttttagattctccatctcgtggacaagaataatacctaat
K N M N L D A F R F S I S W T R I I P N
ggccagataagtgctggagtgaatcaagaaggaatcgacttctataatgatctcattgat
G Q I S A G V N Q E G I D F Y N D L I D
gagctcatatcgaatggccttgagccttttgtgactatatttcattgggatagtccacaa
E L I S N G L E P F V T I F H W D S P Q
ggattagaagacaaatatactggtttttaagccgtagcattgtgaaagattttcaagac
G L E D K Y T G F L S R S I V K D F Q D
tttgcagagctttgctataaaacttttggagaccgagtgaagtattggactactttcaat
F A E L C Y K T F G D R V K Y W T T F N
gagccaagagcttatgcaactcgaggatatgactcgggacttggtgcaccaggccgatgt
E P R A Y A T R G Y D S G L G A P G R C
tcagaatgggtggatcgttcatgcgaagctggaaattcagcgactgaaccatacatagtt
S E W V D R S C E A G N S A T E P Y I V
tctcatcatataattctcgctcatgcagcagctgtacaagtatacaggcaaaaatatcag
S H H I I L A H A A A V Q V Y R Q K Y Q
gcttctcaaaatggtaagattggaataacacttaacgcttactggtacgtgccttactcc
A S Q N G K I G I T L N A Y W Y V P Y S
aacaatacggttgatgaagaagctgcccaagtagcttttgatttctttacgggttggcat
N N T V D E E A A Q V A F D F F T G W H
ttggatcctataacctatggtcactatccgaggaccatgcaggctttagtcggagatcga
L D P I T Y G H Y P R T M Q A L V G D R
ctcccaaaattcaccgaagaagaatttatggttatcaaaggatcttatgacttccttgga
L P K F T E E E F M V I K G S Y D F L G
ttgaattattatggtgcatattatgcatattttaatgatcatccagatccaaacccgctc
L N Y Y G A Y Y A Y F N D H P D P N P L
cataaaagatacaccacagattcacatgtcaatactactgggaagcgagatggaaaacct
H K R Y T T D S H V N T T G K R D G K P
atgggtccgcagggtaccacatcaatgtttaacatttatcctgaggggattcgatatctt
M G P Q G T T S M F N I Y P E G I R Y L
ttgaattacaccaaagatgcgtacagaaatccaacaatgtacattaccgaaaatggatac
L N Y T K D A Y R N P T M Y I T E N G Y
aatcaagatgataatggcaccgtaccgatgagcatactccttaatgatactcgtcggata
N Q D D N G T V P M S I L L N D T R R I
atatactatgaaactcatcttgagaatgtcttaagatcgatcaaggagtacaatgtggat
I Y Y E T H L E N V L R S I K E Y N V D
gtgaaaggttttattgcatggtcatttgaggataattttgagtggtcatctggttacact
V K G F I A W S F E D N F E W S S G Y T
caaagatttggtctctattacatagactacaaaaatcatttggaaagacatgctaaaaat
Q R F G L Y Y I D Y K N H L E R H A K N
tcaacagagtggttcacaaattttctgcagaaaaatcagtccagtactatctctgaagga
S T E W F T N F L Q K N Q S S T I S E G
tctggatcacgatggattcgtccctttggttattctattagatcagcagccgcttaa
S G S R W I R P F G Y S I R S A A A stop

Figure 2.3. Nucleotide and deduced amino acid sequence of PRGH1 including signal sequence which is underlined. PRGH1 ORF consists of 1617 bp and the mature protein codes for 510 amino acids residues.

2.3.3. Overexpression and purification of His_6-rPRGH1

The mature PRGH1 without signal sequence was overproduced as an N-terminally His_6-tagged protein ($rPRGH1_{His}$) as shown in figure 2.4 A. $rPRGH1_{His}$ was purified to homogeneity by Ni^{2+}-NTA affinity chromatography. The N-terminal His-tag was removed using TEV protease and, reverse Ni^{2+}-NTA was performed to get pure untagged rPRGH1. SDS-PAGE analysis showed that rPRGH1 migrated with an apparent molecular mass of approximately 59 kDa which is similar to its calculated molecular mass (Figure 2.4 B). Different β-glucosidases have been heterologously expressed in *E. coli*, purified and characterized (201, 233, 251).

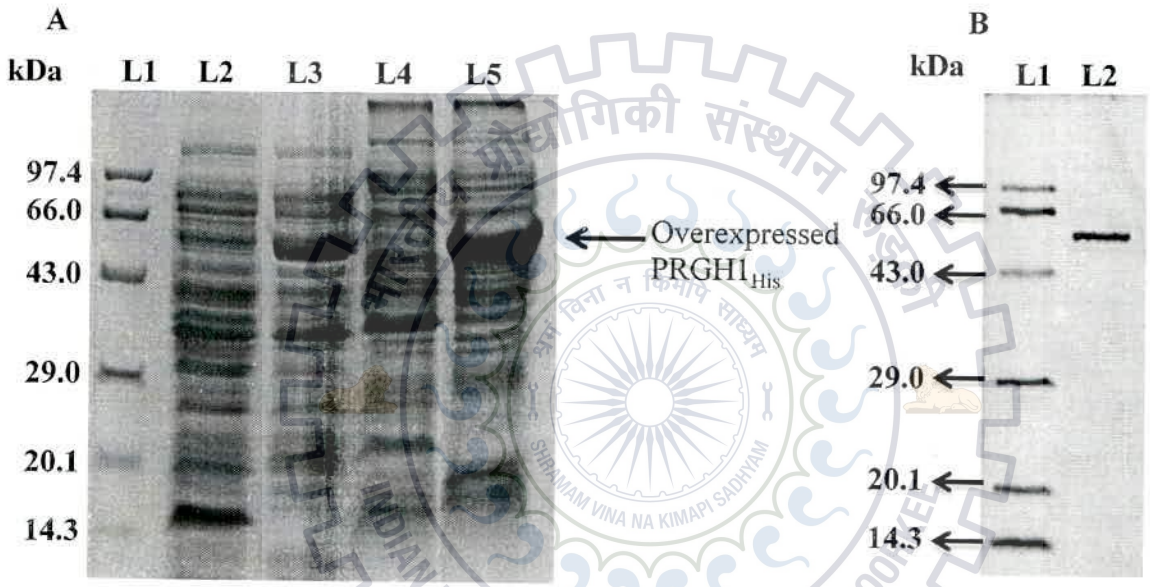


Figure 2.4 (A) *Overexpression of* $rPRGH1_{His}$. L1, molecular weight marker; L2, supernatant of un-induced *E. coli* BL21 (DE3) cell lysate; L3, supernatant of induced *E. coli* BL21 (DE3) cell lysate; L4, pellet of un-induced *E. coli* BL21 (DE3) cell lysate; L5, pellet of induced *E. coli* BL21 (DE3) cell lysate. **(B)** SDS-PAGE (12% *w/v*) analysis of $rPRGH1_{His}$ affinity purification. L1, molecular weight marker; L2, purified $rPRGH1_{His}$ (~59 kDa).

2.3.4. CD measurements

An attempt was made to explicate the difference in the secondary structure of nPRGH1 and rPRGH1 by using far-UV CD. Quantitative study of the CD spectra indicates that the secondary structure of rPRGH1 slightly differ from the nPRGH1 (Figure 2.5). The CD spectrum of both nPRGH1 and rPRGH1 showed them as an α/β protein. The deconvolution of CD data with Dichroweb program indicates the percentage of random coil structures was more in case of rPRGH1. The percentage of the secondary structure of nPRGH1 elements go well with our predicted values from sequence alignments and also support our model. The CD result also suggested that the rPRGH1 maintained its integrity

during the purification process. This difference between the two forms of the protein might be due to post translational modification. Native PRGH1 might be stabilized by the glycosylation process. The detail study was discussed in the next chapter.

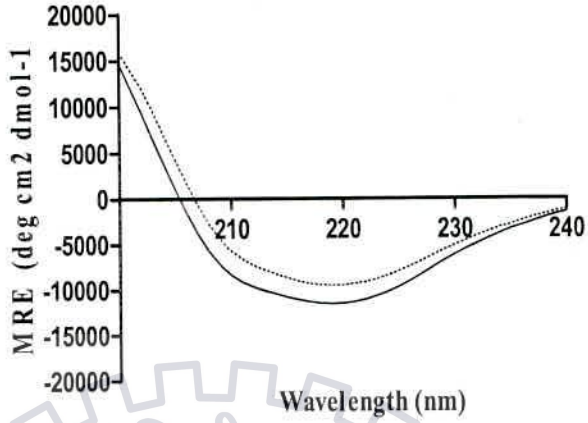


Figure 2.5. Far-UV CD spectrum (200-240 nm) of purified native and recombinant PRGH1 in 20 mM sodium phosphate buffer (pH 8.0).

2.3.5. Concentration-dependent oligomerization of the nPRGH1

We analyzed the profile of two different concentrations of purified nPRGH1 by 8% native-PAGE. The sample with 15 μM of protein showed a single band corresponded to molecular mass of ~66 kDa but the sample with 45 μM of protein showed several closely stacked bands with slow mobility (Figure 2.6). Two nPRGH1 protein samples having concentration of 15 and 45 μM were subjected to gel filtration chromatography by using analytical HPLC. nPRGH1 with 15 μM concentration was eluted essentially as a single peak at 15. 084 min when applied to gel filteration column (Figure 2.7 A). This peak corresponded to 66 kDa mass (according to molecular mass standards) which is similar to monomeric PRGH1. In contrast, when nPRGH1 with 45 μM concentration was applied to the column, it was eluted as different peaks around 11.417 min corresponded to masses around 660 kDa (Figure 2.7 B). This value corresponds similar to the decameric mass of nPRGH1. DLS study also supports this data. At low concentration the hydrodynamic diameter of the enzyme was 18.5 ± 1.75 which corresponds to the monomeric form but at higher concentration the hydrodynamic diameter changed to 67.93 ± 2.59 corresponds to a decameric form (Table 2.3). These results suggest that at low concentration nPRGH1 exists as a monomer and when the concentration increases the monomeric form changes to higher oligomeric forms. However, this concentration dependent oligomerization didn't affect the activities of enzymes. Several

proteins having this type of concentration dependent oligomerization were reported earlier (14, 119, 162).

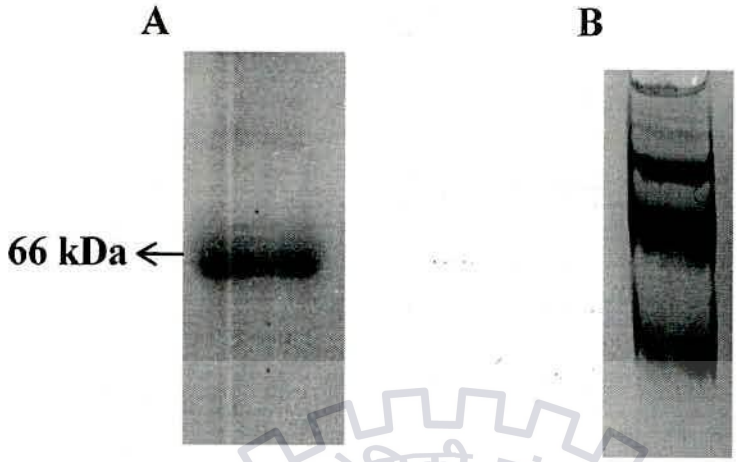
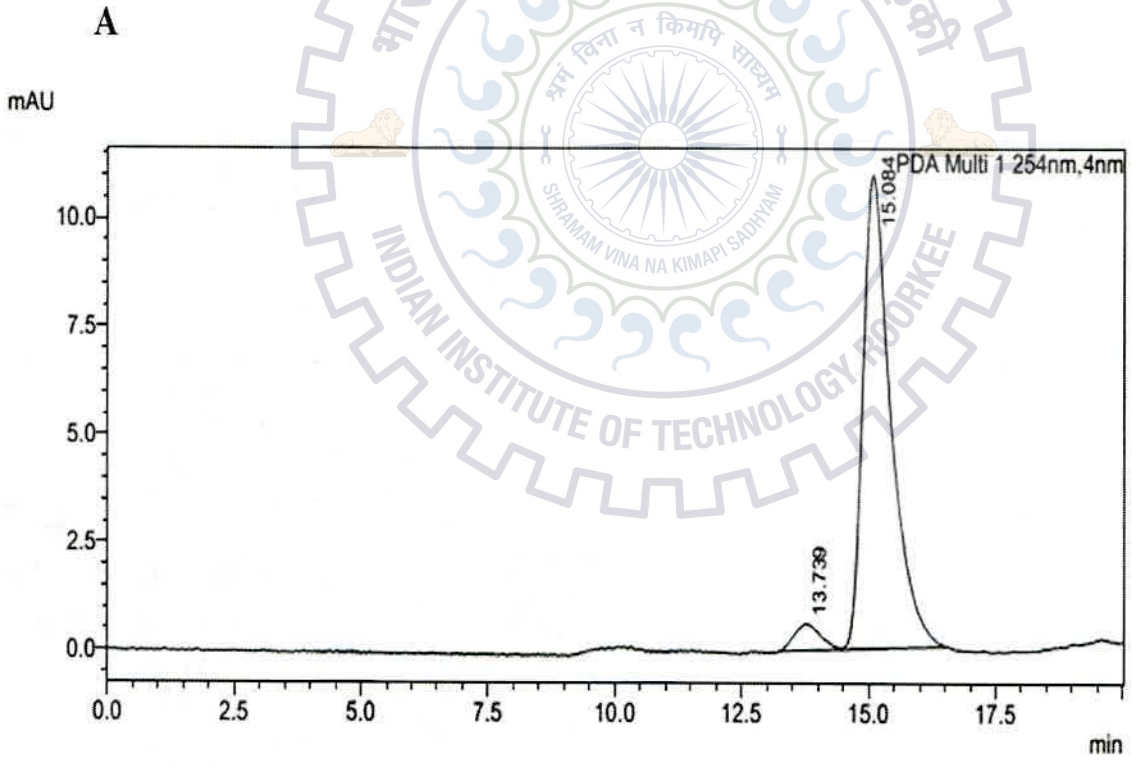


Figure 2.6 (A) NATIVE-PAGE (8% *w/v*) analysis of nPRGH1 at 15 μM **(B)** NATIVE-PAGE (8% *w/v*) analysis of nPRGH1 at 45 μM.



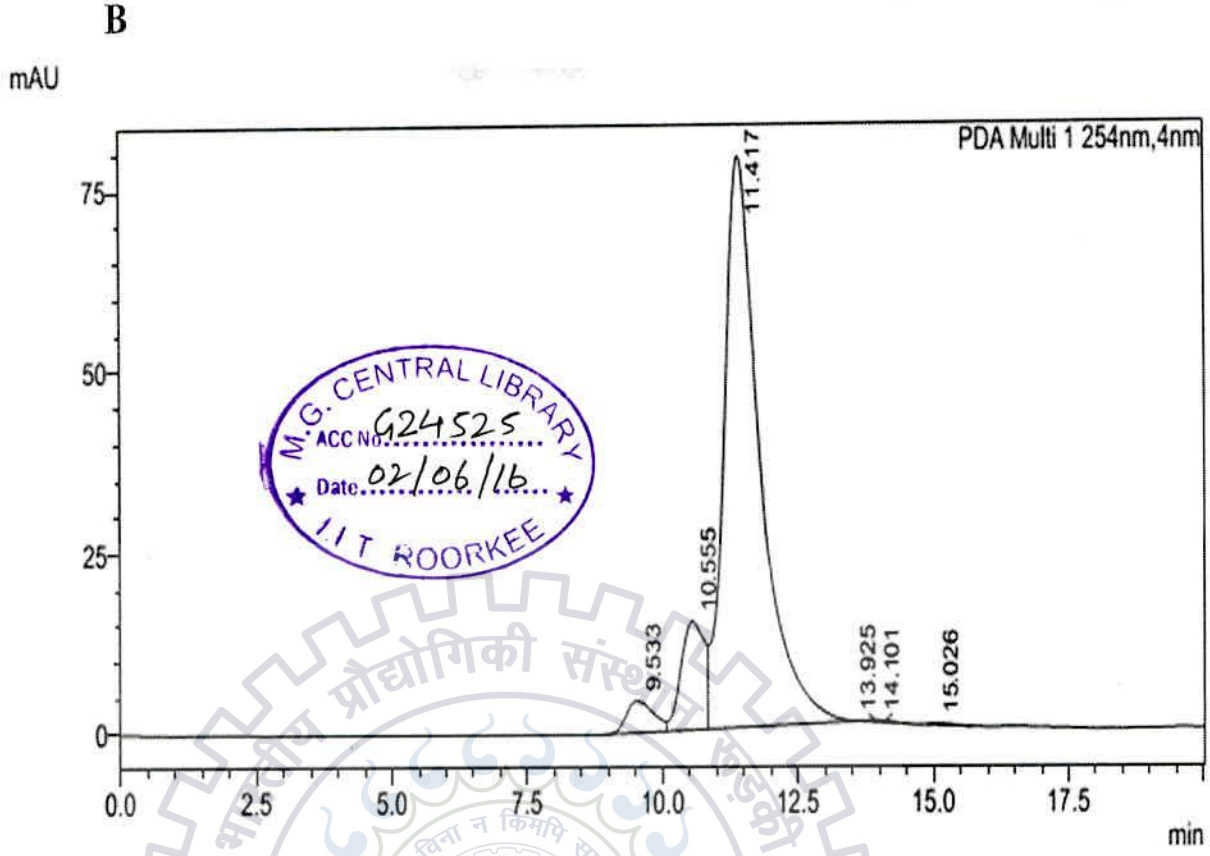


Figure 2.7. Effect of concentration on oligomeric property of nPRGH1. Gel filtration chromatogram showing the elution profiles **(A)** nPRGH1 at 15 µM concentration, and **(B)** nPRGH1 at 45 µM concentration. The column was calibrated with molecular weight standards as described in methods section.

Table 2.3. Hydrodynamic diameters (d_H) obtained from DLS studies.

Concentration (µM)	Size (hydrodynamic diameter) d_H, (nm)	approximate molecular mass
15	18.5±1.75	~66kDa
45	67.93±2.59	~660 kDa

2.3.6. Substrate specificity and kinetic parameters

Generally glycosyl hydrolase family 1 enzymes show a wide range of enzyme activity, so we have used various substrates to characterize this enzyme. Previously nPRGH1 has been characterized by using *p*-nitrophenyl β-D-glucopyranoside (*p*NP-Glc) and *p*-nitrophenyl β-D-galactopyranoside (*p*NP-Gal) [183]. The specificities of recombinant PRGH1 against different substrates are shown in table 2.4. The result showed that the enzyme was active against *p*NP substrates as well as natural substrates. This enzyme was best active against *p*-nitrophenyl β-D-fucopyranoside (*p*NP-Fuc), with relative activity of 122% and also very active against *p*-nitrophenyl β-D-glucopyranoside (*p*NP-Glc) and *p*-nitrophenyl β-D-galactopyranoside (*p*NP-Gal) with relative activity of 100 and 83% respectively. The kinetic parameters of nPRGH1 and rPRGH1 against chromogenic substrates are presented in table 2.5. The results showed that both nPRGH1 and rPRGH1 were maximally active against *p*NP-Fuc, with K_{cat}/K_m values of 3.21×10^4 and 2.49×10^4 respectively. Our study showed that PRGH1 is a broad specific β-glucosidase enzyme. Like PRGH1 several plant β-glucosidases show weak β-galactosidase activity [59, 242].

Table 2.4. Relative hydrolytic activity of purified recombinant PRGH1 enzyme with various substrates.

Substrate	Relative activity (%)
Recombinant PRGH1 expressed in *E. coli*	
***p*-nitrophenyl substrates**	
p-nitrophenyl β-D-glucopyranoside	100
p-nitrophenyl β-D-galactopyranoside	73
p-nitrophenyl β-D-fucopyranoside	122
Glucooligosaccharides and natural β-D-Glycosides	
Cellobiose	62
Cellotriose	41
Salicin	23
Lichenan	13
Amygdalin	31
Arbutin	28
Laminarin	21

Table 2.5. Kinetic parameters of nPRGH1 and rPRGH1 enzyme.

Substrate	K_m (mM)	V_{max}(μkat/mg)	K_{cat} (s^{-1})	K_{cat}/K_m ($M^{-1}s^{-1}$)
Native PRGH1				
*p*NP-Glc	0.53	0.181	12	2.27×10^4 [183]
*p*NP-Gal	0.61	0.112	7.39	1.15×10^4 [183]
*p*NP-Fuc	0.46	0.222	14.77	3.21×10^4 (In this study)
Recombinant PRGH1 expressed in *E. coli*				
*p*NP-Glc	0.58	0.157	10.39	1.79×10^4 (In this study)
*p*NP-Gal	0.68	0.088	5.82	8.57×10^3 (In this study)
*p*NP-Fuc	0.51	0.192	12.71	2.49×10^4 (In this study)

2.3.7. Competition analysis with mixed substrates

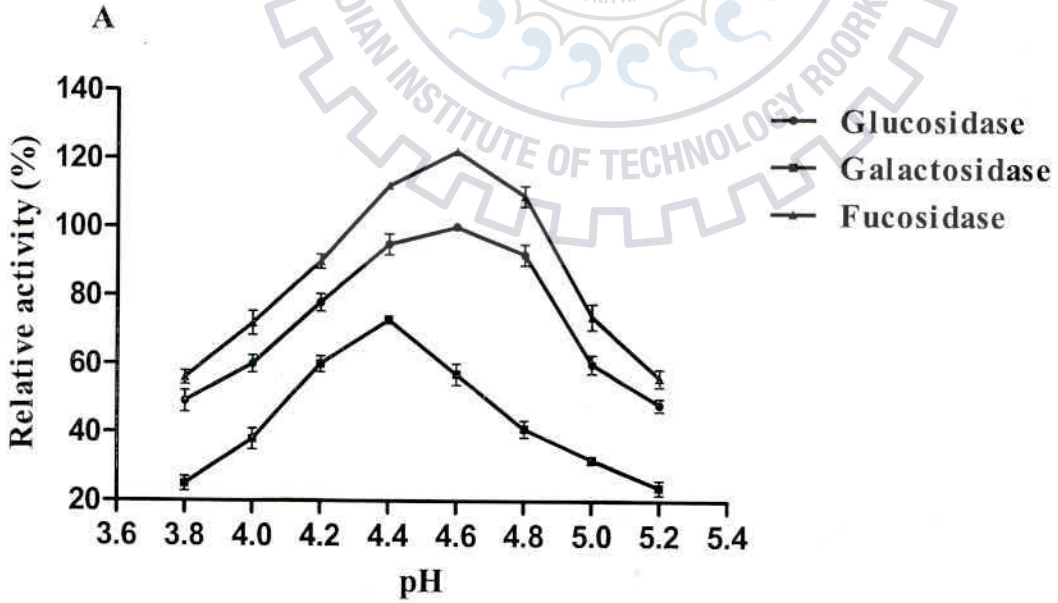
To determine whether nPRGH1 has one or more active sites for the hydrolysis of *p*NP-Glc, *p*NP-Gal, and *p*NP-Fuc, the competition kinetics with mixed substrates were studied. The theoretical maximum velocity (V_{exp}) with the mixture of both substrates was obtained from Lineweaver-Burk plot and the experimental maximum velocity for each reaction determined from the equation applicable for competing substrates. When two substrates are catalyzed in a single active site the experimental maximal velocity decided with the combination of both the substrates must coincide with the theoretical maximum velocity calculated from the equation relevant for competing substrates, and the total rate of the reaction must be less than the sum of the rate of the reactions measured with separated substrates [26]. The result shows that all the three substrates compete for the same active site in PRGH1 enzyme (Table 2.6).

Table 2.6. Competition analysis with mixed substrate.

Substrate mixture	V_{exp} μkat/mg	$V_{maxA} + V_{maxB}$ μkat/mg	V_{theo} μkat/mg
*p*NP-Fuc (A) + *p*NP-Glc (B)	0.198	0.403	0.202
*p*NP-Fuc (A) + *p*NP-Gal (B)	0.180	0.334	0.175
*p*NP-Glc (A) + *p*NP-Gal (B)	0.156	0.293	0.149

2.3.8. Effect of pH and temperature on enzyme activity

Previously, the optimum pH for glucosidase and galactosidase activity was determined as 4.6 and 4.4 respectively. The optimum pH for the fucosidase activity for nPRGH1 was found to be 4.6 (Figure 2.8 A). rPRGH1 follows the same pattern as nPRGH1. Both the enzyme was quite stable over a broad pH range of 4.0 to 8.0 with 70% retention of the original catalytic activity against *p*NP-Glc, but below pH 4.0 the activity declined quickly (Figure 2.8 B). The optimal temperature for all the three activities was found to be 65 °C and both the enzyme followed the same pattern. Thermostability experiments showed that both the enzyme was fairly stable over a broad temperature range of 25-70 °C, and the enzyme activity reduced above 70 °C (Figure 2.9). The result clearly demonstrated that recombinant enzyme is more sensitive towards the high temperature. Both nPRGH1 and rPRGH1 are thermostable enzymes retaining more than 90% and 60% activity at 60 °C respectively. Thus both the enzyme could be used in bioethanol production and food processing industries. Post translational modification might play a role towards the thermostability behaviour of this enzyme, which is characterized in the next chapter. Thermozymes are utilized in various industries [20]. Thermostability study with fungal β-glucanase [180] and human aquaporin 10 protein [178] showed that glycosylation affects the thermostability of the proteins. Compare to this several cold active glycosidases are also reported [45].



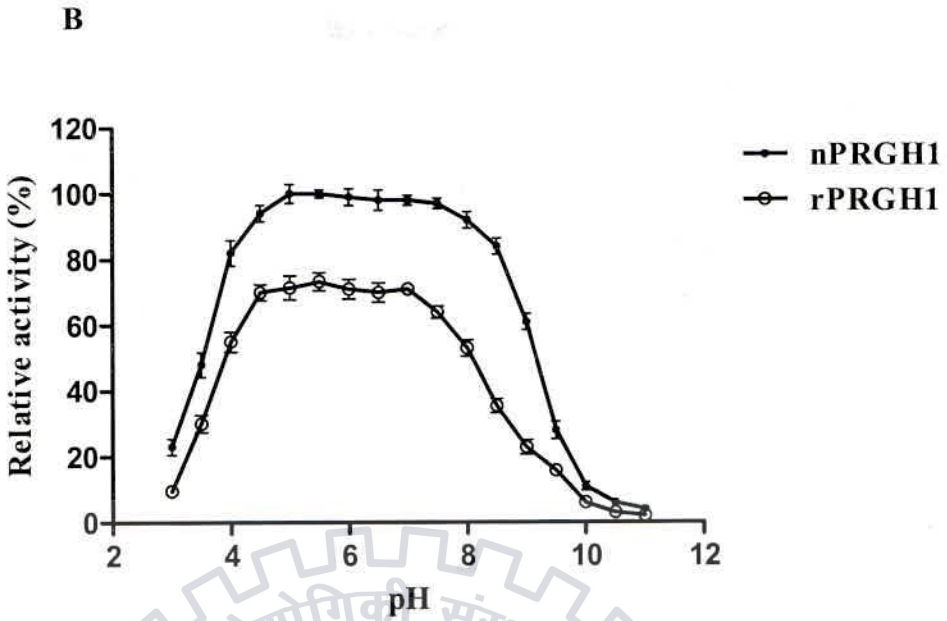


Figure 2.8. The effect of pH on the nPRGH1 activity **(A)**, nPRGH1 and rPRGH1 stability **(B)**. For stability assay, the purified enzymes were incubated at different pH at 37 °C for 1 h and residual activity was measured. Results are the average of triplicate and the bars indicate the standard error.

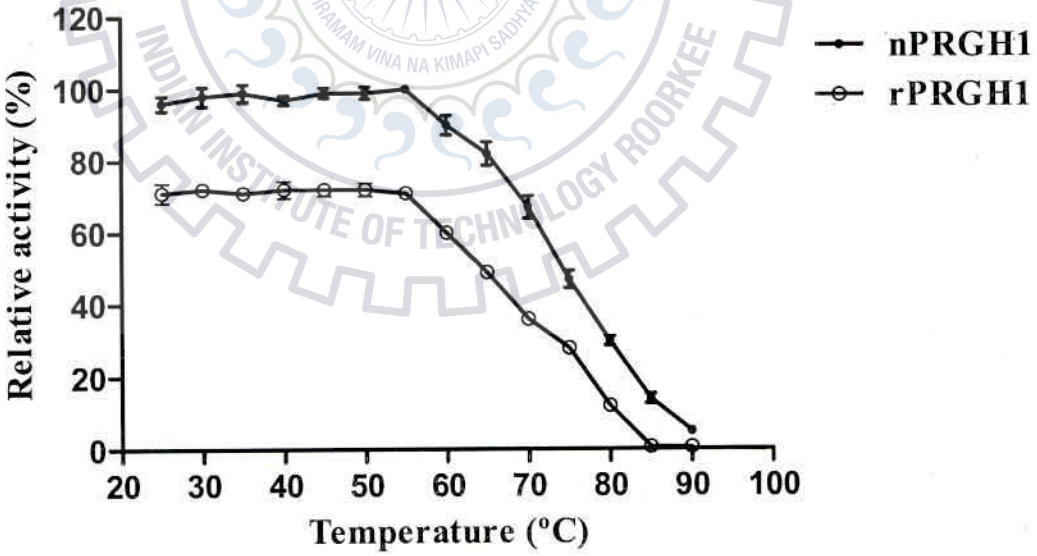


Figure 2.9. The effect of temperature on the stability of nPRGH1 and rPRGH1. For stability assay, the purified enzymes were incubated at different temperatures at pH 4.6 for 30 min and residual activity was measured. Results are the average of triplicate and the bars indicate the standard error.

2.3.9. Characterization of mutant proteins

Three PRGH1 mutants were constructed and verified as containing single codon mutations and overexpressed and purified as described in materials and methods section. The activity of each mutant protein was measured by using *p*NP-Glc as substrate. Mutation of catalytic nucleophile (E389) and catalytic acid/base (E173) decreases the activity (Table 2.3). These two residues are known to play the most important role during hydrolysis. Mutation of E173 to Asn yielded enzyme with 21 fold reductions in their relative activity and a larger reduction was observed in case of mutants at E389. Mutation of another active site residue (N172) decreases the activity significantly (Table 2.3). This mutation studies confirmed that these residues are very important in the catalysis process. Several such studies were done to find out the active site residues of β-glucosidases [254, 241, 234]. Confirmation of these active site residues also helped in the docking studies.

Table 2.7. Mutations and their relative activity.

Mutation	Relative activity (%)
Wild type	100 ± 1.67
E173N	4.7 ± 1.35
E389N	0.32 ± 0.04
N172Q	21.65 ± 2.58

2.3.10. Homology search and ProtParam analysis

The mature amino acid sequence of PRGH1 belonging to Glycosyl hydrolase family 1 (Figure 2.3) consists of 510 amino acids with a predicted molecular weight of 59 kDa and an isoelectric point of 5.45. BLAST sequence similarity results displayed significant homology with beta-Glucosidases from various sources and also with myrosinases (Figure 2.10). PRGH1 showed 43 to 55% identity with different plant β-glucosidases, 46% with myrosinase from *Sinapsis alba*, 35% with humans, 39-41% with fungi and less than 38% with bacteria.

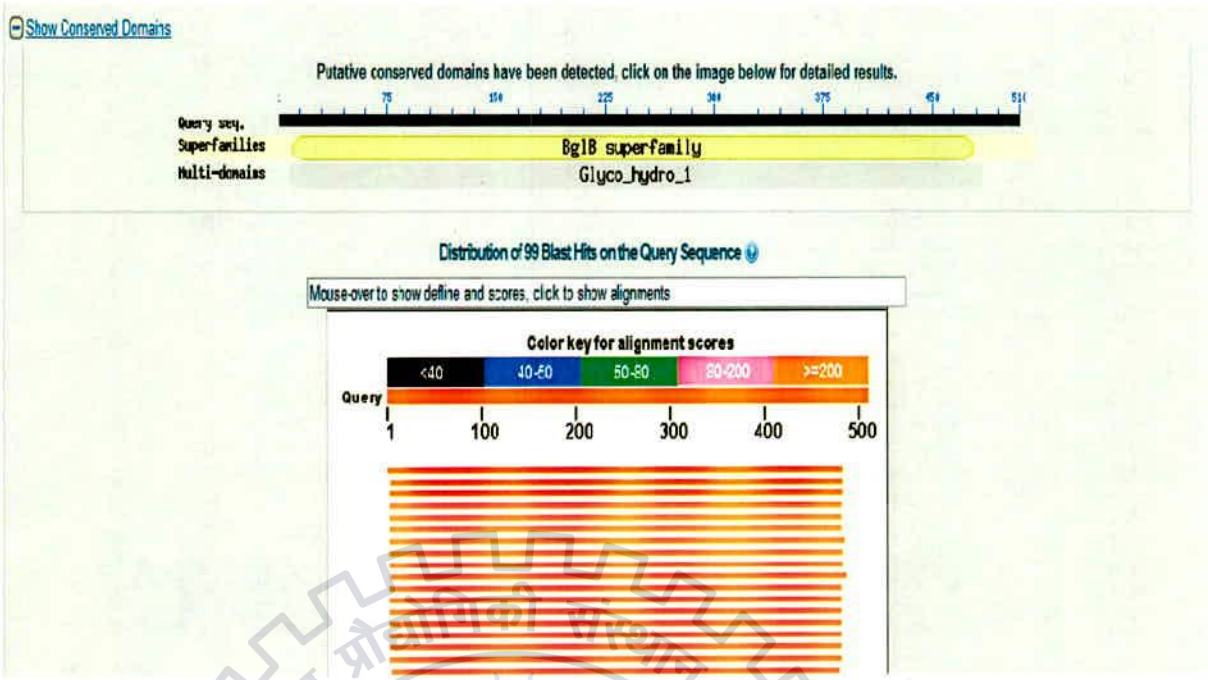


Figure 2.10. Homology search done against Protein Data Bank using NCBI BLASTp.

2.3.11. Multiple sequence alignment analysis

The amino acid sequence alignment of PRGH1 with homologous glucosidases revealed conservation of the catalytic acid/base residue Glu173 and nucleophile Glu389 as shown in Figure 2.11. The residues making the glycone binding site in PRGH1 i.e. Gln23, His127, Trp128, Asn172, Gln446 and Trp447 are also conserved. Some of the aglycone binding residues like Asn314, Tyr316 and Trp439 are also conserved (Figure 2.11), while there are distinct changes at the other aglycone binding regions which help in determination of the specificity of these enzymes. Moreover, the recognition of the sulphate group of substrate by myrosinase is mediated by two basic residues i.e. Arg194 and Arg259, highly conserved among the myrosinases and are absent from the O-glycosidases. In addition, Trp142 is thought to be involved in substrate recognition in myrosinase as it forms van der Waals contacts with the sulphur molecule of the glycosidic bond [23]. However, PRGH1 is an O-glycosidase possess one of the conserved Arg i.e. Arg180 along with the Trp128 which might help them in hydrolyzing S-glycosides. Earlier also Os4Bglu12, an O-glucosidase has been biochemically proved to show thioglucosidase activity also on the basis of the homology of the covalent α-2-fluoroglucoside (G2F) complex of Rice Os4Bglu12 to that of S-glucosidase *S. alba* myrosinase [204].

```
CLUSTAL 2.1 multiple sequence alignment

Rice            AMADVVPKPNWLG--------------GLSRAAFPKRFVFGTVTSAYQVEGMAASGGRGP 46
Os3bglu6        ---SFTMAQQSGG--------------GLTRGSFPEGFVFGTASAAYQYEGAVKEDGRGQ 43
PRGH1           ---------NN----------------SFNRSAFPDDFIFGASAAAYQYEGEANKSGRGP 35
Os4bglu12       AMADITSLYKKAGSAAAPFAYNSAGEPPVSRRSFPKGFIFGTASSSYQYEGGAAEGGRGP 60
1MYR            -DEEITCQENNPFTCG--------NTDGLNSSSFEADFIFGVASSAYQIEGTIG---RGL 48
                          :                  ..  :*   *:**. :::** **      **

Rice            SIWDAFAHT-PGNVAG-NQNGDVATDQYHRYKEDVNLMKSLNFDAYRFSISWSRIFPDG- 103
Os3bglu6        TIWDTFAHT-FGKITD-FSNADVAVDQYHRFEEDIQLMADMGMDAYRFSIAWSRIYPNG- 100
PRGH1           SVWDIFTHEYPEKITD-RSNGDEAIDFYHRYKEDIQRMKNMNLDAFRFSISWTRIIPNGQ 94
Os4bglu12       SIWDTFTHQHPEKIAD-RSNGDVASDSYHLYKEDVRLMKDMGMDAYRFSISWTRILPNGS 119
1MYR            NIWDGFTHRYPDKSGPDHGNGDTTCDSFSYWQKDIDVLDELNATGYRFSIAWSRIIPRGK 108
                .:** *:*    :    '   *.* : * :  :::*:  : .:.  .:****:*:** * *

Rice            -EGRVNQEGVAYYNNLINYLLQKGITPYVNLYHYDLPLALEKKYGGWLNAKMADLFTEYA 162
Os3bglu6        -VGQVNQAGIDHYNKLIDALLAKGIQPYVTLYHWDLPQALEDKYKGWLDRQIVDDFAAYA 159
PRGH1           ISAGVNQEGIDFYNDLIDELISNGLEPFVTIFHWDSPQGLEDKYTGFLSRSIVKDFQDFA 154
Os4bglu12       LRGGVNKEGIKYYNNLINELLSKGVQPFITLFHWDSPQALEDKYNGFLSPNIINDFKDYA 179
1MYR            RSRGVNQKGIDYYHGLIDGLIKKGITPFVTLFHWDLPQTLQDEYEGFLDPQIIDDFKDYA 168
                    **: *: .*: **: *: :*: *:.:::*:* *  *:.:* *:*. .: . *  :*

Rice            DFCFKTFGNRVKHWFTFNEPRVVAVLGYDQGTFPPKRCT----KCAAGGNSATEPYIVAH 218
Os3bglu6        ETCFREFGDRVKHWITLNEPHVVAVQGYDAGLLAPGRCSVLLHLYCKAGNSGTEPYVVAH 219
PRGH1           ELCYKTFGDRVKYWTTFNEPRAYATRGYDSGLGAPGRCSEWVDRSCEAGNSATEPYIVSH 214
Os4bglu12       EICFKEFGDRVKNWITFNEPWVFCVKGYATGLLAPGRCSPWEKGNCSVGDSGREPYTACH 239
1MYR            DLCFEEFGDSVKYWLTINQLYVVPHRGYGSALLAPGRCSPTVDPSCYAGNSSTEPYIVAH 228
                : *:. **: ** * *:*:     **   .  .* **:          .*:*. *** ..*

Rice            NFLLSHAAAVARYRTKYQAAQQGKVGIVLDFNWYEALSNST-EDQAAAQRARDFHIGWYL 277
Os3bglu6        HFILAHAAAASIYRTKYKATQNGQLGIAFDVMWFEPMSNTT-IDIEAAKRAQEFQLGWFA 278
PRGH1           HIILAHAAAVQVYRQKYQASQNGKIGITLNAYWYVPYSNNT-VDEEAAQVAFDFFTGWHL 273
Os4bglu12       HQLLAHAETVRLYKAKYQALQKGKIGITLVSHWFVPFSRSK-SNNDAAKRAIDFMFGWFM 298
1MYR            HQLLAHAKVVDLYRKNYTH-QGGKIGPTMITRWFLPYNDTDRHSIAATERMKQFFLGWFM 287
                : :*:**  .  *: :*   * *::* .:   *:  .  .     *::   :*  **.

Rice            DPLINGHYPQIMQDLVKDRLPKFTPEQARLVKGSADYIGINQYTASYMKGQ--QLMQQTP 335
Os3bglu6        DPFFFGDYPATMRARVGERLPRFTADEAAVVKGALDFVGINHYTTYYTRHNNTNIIGTLL 338
PRGH1           DPITYGHYPRTMQALVGDRLPKFTEEEFMVIKGSYDFLGLNYYHGAYYAYFNDHPDPNPLH 333
Os4bglu12       DPLIRGDYPLSMRGLVGNRLPQFTKEQSKLVKGAFDFIGLNYYTANYADNLP--PSNGLN 356
1MYR            GPLTNGTYPQIMIDTVGARLPTFSPEETNLVKGSYDFLGLNYYFTQYAQPSP-NPVNATN 346
                .*:  * **  *   *  *** *: .:  .:**: *::*:* * : *

Rice            TSYSADWQVTYVFA-KNGKPIGPQA------NSNWLYIVPWGMYGCVNYIKQKYGNPTVV 388
Os3bglu6        NNTLADTGTVSLPF-KNGKPIGDRA------NSIWLYIVPRGMRSLMNYVKERYNSPPVY 391
PRGH1           KRYTTDSHVNTTGK-RDGKPMGPQG------TTSMFNIYPEGIRYLLNYTKDAYRNPTMY 386
Os4bglu12       NSYTTDSRANLTGV-RNGIPIGPQA------ASPWLYVYPQGFRDLLLYVKENYGNPTVY 409
1MYR            HTAMMDAGAKLTYINASGHYIGPLFESDGGDGSSNIYYYPKGIYSVMDYFKNKYYNPLIY 406
                     *  .     .*  :*          :  :   * *:    : * *: * . * :

Rice            ITENGMDQPAN--LSRDQYLRDTTRVHFYRSYLTQLKKAID-EGANVAGYFAWSLLDNFE 445
Os3bglu6        ITENGMDDSNNPFISIKDALKDSKRIKYHNDYLTNLAASIKEDGCDVRGYFAWSLLDNWE 451
PRGH1           ITENGYNQDDNGTVPMSILLNDTRRIIYYETHLENVLRSIKEYNVDVKGFIAWSFEDNFE 446
Os4bglu12       ITENGVDEFNNKTLPLQEALKDDARIEYYHKHLLSLLSAIR-DGANVKGYFAWSLLDNFE 468
1MYR            VTENGISTPGS--ENRKESMLDYTRIDYLCSHLCFLNKVIKEKDVNVKGYLAWALGDNYE 464
                :**** .   .     .  : *  *: :   :*  :   *   . :* *::**:: **:*

Rice            WLSGYTSKFGIVYVDFN-TLERHPKASAYWFRDMLKH---------------------- 481
Os3bglu6        WAAGYSSRFGLYFVDYKDNLKRYPKNSVQWFKALLKT---------------------- 488
PRGH1           WSSGYTQRFGLYYIDYKNHLERHAKNSTEWFTNFLQKNQSSTISEGSGSRWIRPFGYSIR 506
Os4bglu12       WSNGYTVRFGINFVDYNDGRKRYPKNSAHWFKKFLLK---------------------- 505
1MYR            FNNGFTVRFGLSYINWNNVTDRDLKKSGQWYQKFISP---------------------- 501
                :  *:: :**: :::::   .*  * *  *:  ::

Rice            ----
Os3bglu6        ----
PRGH1           SAAA 510
Os4bglu12       ----
1MYR            ----
```

Figure 2.11. Multiple sequence alignment of PRGH1 with other glycosyl hydrolases shows the conserved catalytic acid/base, nucleophilic and glycone residues highlighted in yellow colour. Aglycone residues are highlighted in green and the conserved *S. alba* myrosinase residues responsible for sulphate recognition are shown in red colour.

2.3.12. Phylogenetic tree

The dataset used for phylogenetic tree include 33 protein sequences from plant, bacterial, fungal, human glucosidases and also the myrosinases from plant and insect. PRGH1 glycosyl hydrolase shows maximum homology to plant glycosyl hydrolases and myrosinases and phylogenetic tree further supports the BLAST search. The glycosyl hydrolases from plants, fungi, bacteria make different clusters and are grouped into 4 major groups. Group 1 consists of PRGH1 along with other plant glycosyl hydrolases including both O- and S-glycosyl hydrolases. This phylogenetic tree further confirms that myrosinase (S-glycosidase) and defence related cyanogenic β-glucosidase (O-glycosidase) are derived from a common ancestor as reported in literature [23]. Interestingly PRGH1 showing maximum homology to Os4Bglu12 (55%) is present very close to *S. alba* myrosinase (46%) clustered within Group 1 in the phylogenetic tree. Thus PRGH1 sharing a common ancestor with myrosinase might also have specialized in plant defence mechanisms. Group 2 consists of fungal, insect as well as human β-glucosidases, while bacterial and archaeal β-glucosidases constitute Group 3 and Group 4 (Figure 2.12).

2.3.13. Secondary structure analysis

Secondary structure analysis was done using ESPript which showed that the overall structures are conserved with the presence of 1 disulphide linkage. It reveals that PRGH1 is an α/β protein as shown in Figure 2.13.

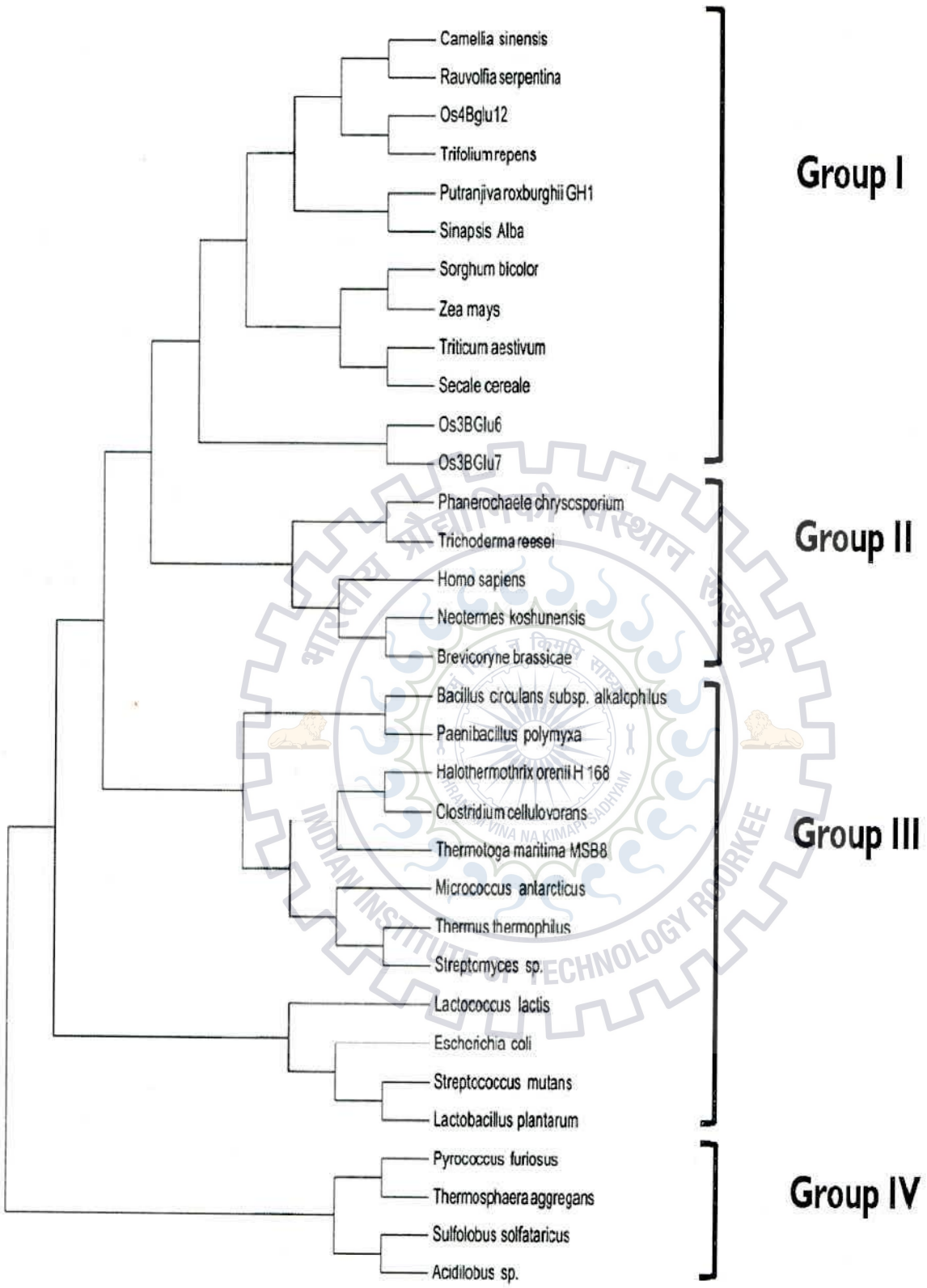


Figure 2.12. Phylogenetic tree of PRGH1 and other related sequences was constructed by the maximum likelihood method using MEGA6.

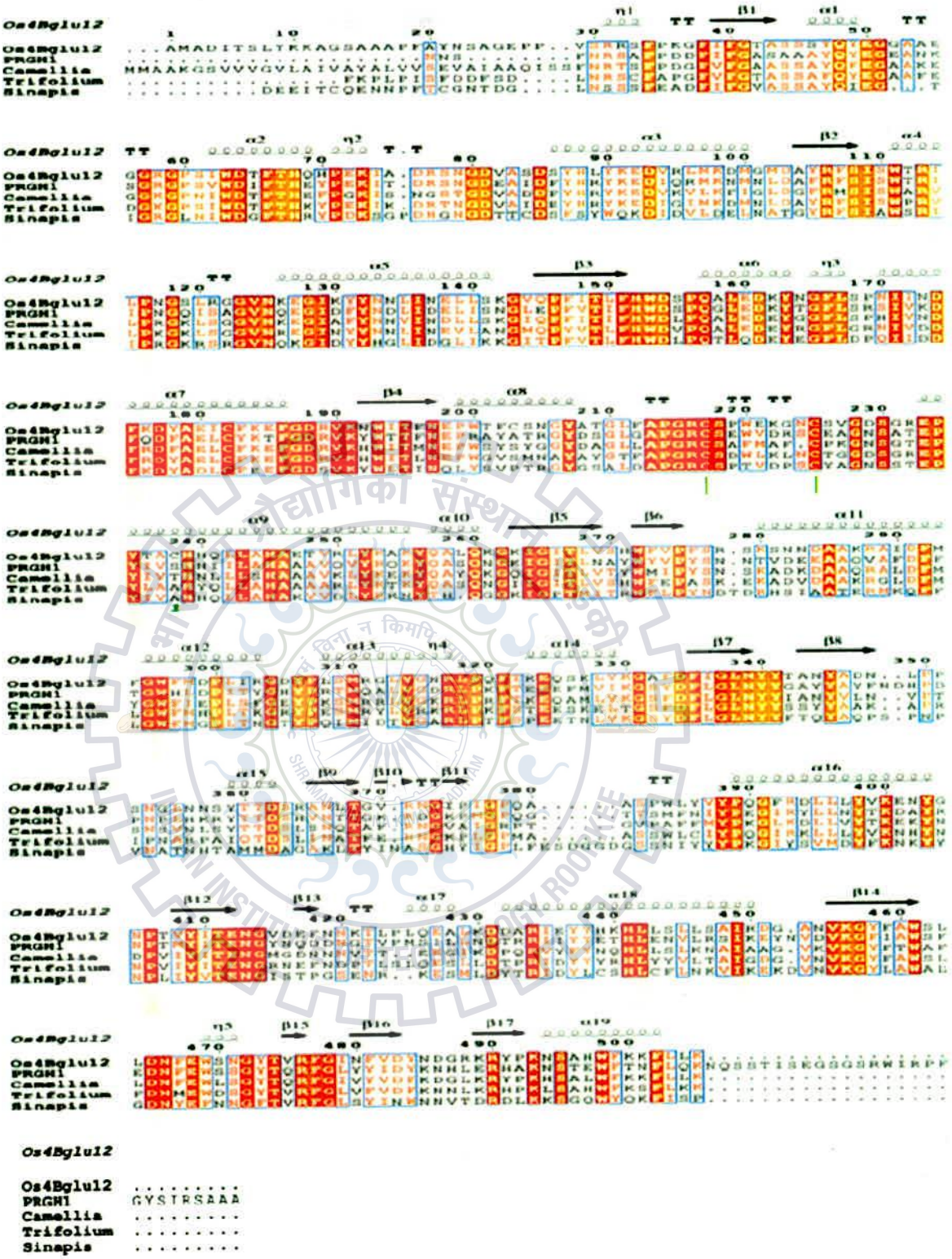


Figure 2.13. Secondary structure analysis by ESPript. α-helices are represented as squiggles while β-sheet are represented as arrows, TT represents β-turn, disulphide bond as green line.

2.3.14. Structural analysis

In order to investigate the structural differences that bring about changes in substrate specificity, homology modeling of PRGH1 was done using Modeller 9v10. The model with the lowest dope score was subjected to energy minimization where PROCHECK, Verify-3D and WHATCHECK were used for the validation of the model. Procheck analysis of PRGH1 model showed 99% residues in the favourable regions. The overall structure of PRGH1 is $(\beta/\alpha)_8$ barrel similar to other glycosyl hydrolases [204]. The catalytic acid/base residue Glu173 and nucleophilic residue Glu389 are placed at the bottom of a cleft at the carboxy end of 4th and 7th β-strands. The active site pocket of PRGH1 is shaped by 4 extended loop regions which join the α-helices and the β-strands of the $(\beta/\alpha)_8$ fold: loop A (Ser18-Asp70 flanked by β1, α1), loop B (Glu173-Thr207 surrounded by β4, α4), loop C (Tyr315- Pro366 flanked by β6, α6) and loop D (Asn390-Asp408 amongst β7, α7) as shown in Figure 2.14. Loop B contains the conserved disulphide bridge Cys192- Cys200 and also contributes to the aglycone binding pocket. Loop C also forms the part of the aglycone binding site and surrounds the active site. Loop C contains 2 extra residues in PRGH1 as compared to the Rice Os4Bglu12 and os3ZBglu7 [35, 204].

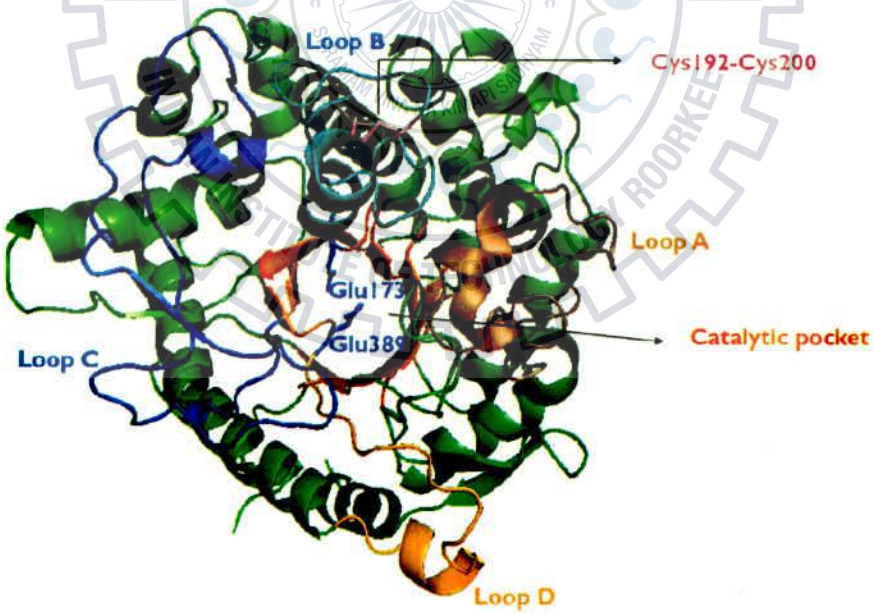


Figure 2.14. The ribbon demonstration of the PRGH1 structure. The catalytic acid/base Glu173 and the nucleophilic Glu389 present on the opposite sides of the catalytic pocket are shown as blue sticks. Four loops regions are also highlighted along with the disulphide bridge (Cys192-Cys200) present in loop B.

This extra residue insertion in loop C is however similar to the *Zea mays* ZmGlu1 and *Sorghum bicolor* SbDhr1 and *S. alba* myrosinase [23, 239]. The two amino acids longer C loop contributes to its narrow and restricted catalytic pocket of PRGH1 as compared to the more open and extended active sites in RiceOs4Bglu12 and OsBglu7.

The retention of the enzyme activity up to 70 °C owing to the thermostable nature of the enzyme is also contributed by the presence of numerous salt links and hydrogen linkages in addition to the single disulphide bridge. The disulphide bridge between Cys192-Cys200 is preserved in the plants secretory pathway and plastids glycosyl hydrolase enzymes known till now. Out of total 28 arginine residues, 22 directly form salt linkages with the oppositely charged amino acids. 16 of 21 lysine residue and 10 from 14 histidine residue participate in salt bridge interactions. Of 34 aspartate residues, only 1 does not participate in salt bridge interaction. 23 of 30 glutamate residues also form salt bridges. This is in contrast to the Rice Os4Bglu12 and Os3Bglu7. In Rice Os4Bglu12, 18 of 25 arginine residues, 15 of 30 lysine, 3 of 10 histidine, 15 of 27 aspartate and 15 of 20 glutamate participate in salt bridge interaction with 2 disulphide bonds rendering it stability at 20-50 °C for 60 minutes [204] while Os3Bglu7 loses its activity after 40 °C owing to lesser ionic interactions with only 1 disulphide bond [35]. In Os3Bglu7, only 8 out of 21 arginine residue, 5 out of 25 lysine residues, 3 out of 11 histidines, 14 from total of 24 aspartate residues and 4 out of 12 glutamate residues form salt interactions. These factors contribute to the stability of PRGH1 with its maximal activity at 65°C for 20 minutes and further activity is retained up to 70°C.

2.3.15. Active Site analysis

The active site of PRGH1 is deep and narrow situated at the bottom of a cleft bounded by carboxy terminal ends of β strands. The catalytic acid/base Glu173 and nucleophile residue Glu389 are placed on the opposed sides of the catalytic pocket approximately 5 Å apart. Exceptionally, the active site of PRGH1 has a restriction in the cleft as seen in the surface view of the protein-ligand complex (Figure 2.15 A). This is due to the longer side chains of Met361 and Arg180 being present close together at the entrance of the active site. On superimposition of PRGH1 on 3PTQ and also through MSA results, it is clear that Trp365 (in 3PTQ) has been replaced by Met361 and Asn186 by Arg180 [204]. Trp365 is conserved residue in all O- Glycosyl hydrolases and play important role in glycone and aglycone moiety binding. In fact, this Trp is so important that it can be used for the structural classification of glycosyl hydrolases depending upon its orientation at the active site [10]. However, PRGH1 lacking this conserved Trp is similar to S-glycosidases i.e. Myrosinases as they also lack this

Trp [23]. This modification in the shape of active site might play a crucial role in substrate selectivity. Although PRGH1 has a wider active site but due to this cleft there is a hindrance at the entry of the catalytic site which restricts entry of bulky oligosaccharides or oligosaccharides with a large aglycone moiety.

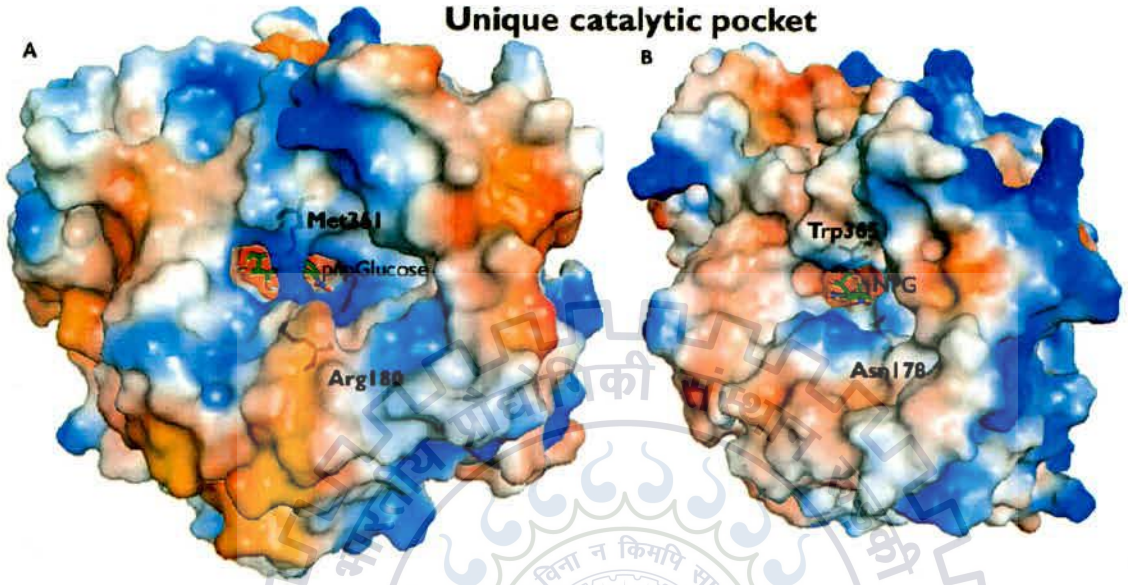


Figure 2.15. **(A)** Surface view of PRGH1 showing the active site pocket with its ligand 4-nitrophenyl beta-D-glucopyranoside (*p*NP-Glc) **(B)** and its comparison to Os4bglu12 from Rice (3PTQ) with its ligand 2, 4-Dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (NPG).

2.3.16. Docking Studies

There is no crystal structure of PRGH1 alone or in complex with substrates reported till date. Therefore, to know the structural details of ligand protein binding, 4-nitrophenyl beta-D-fucopyranoside (*p*NP-Fuc), 4-nitrophenyl beta-D-glucopyranoside (*p*NP-Glc), 4-nitrophenyl beta-D-galactopyranoside (*p*NP-Gal), 4- D-(+)-Cellobiose, D-(+)-Cellotriose and D-(+)-Cellotetraose were docked on to the active site of enzyme. Standard relaxed chair conformations of these ligands retrieved from PubChem were used for docking. It was found that non-reducing glycosyl end of these ligands docked at the same position as the glucosyl residue at subsite -1 in 2RGM which further confirms the accuracy of docking studies as shown in Figure 2.16.

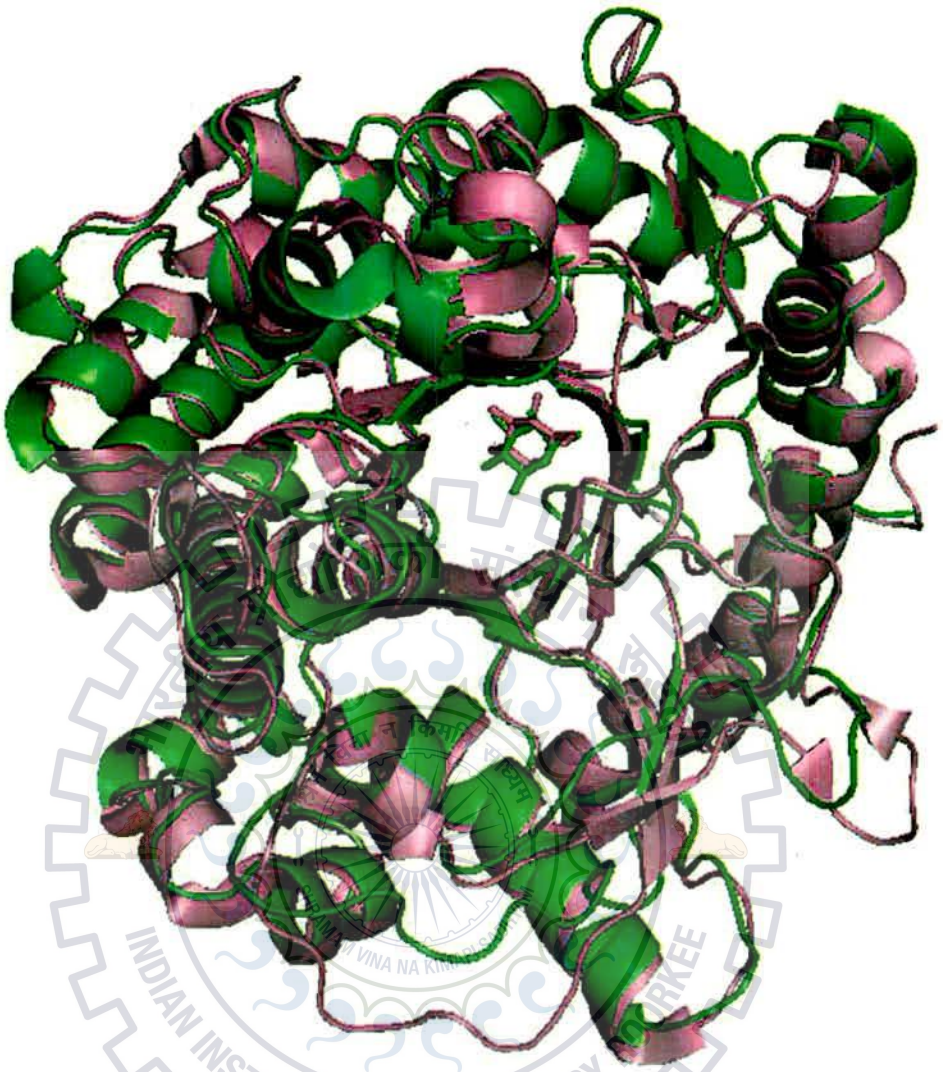


Figure 2.16. Docking studies have been validated via structural superimposition of PRGH1-Glucose complex (green), with Rice BGlu1 G2F (2-deoxy-2-fluoro-glucose) complex (2RGM; pink).

On docking *p*NP-Fuc to the active site of the enzyme, the fucose moiety occupied the subsite-1 surrounded by 6 residues namely Gln23, His127, Tyr316, Glu389, Glu446 and Trp447. The *p*NP-Fuc was found in strained conformation hydrogen bonded to catalytic acid/base Glu173. The distance of Oε1 of Glu173 to O1 and O2 of fucose was 2.67 Å and 2.94 Å and which can be reasonably a good distance for protonation by the catalytic acid. Oε2 of Glu173 hydrogen bonds with O1 and O2 at 3.38 Å and 3.43 Å respectively. Like myrosinases, O2 group is recognized by Asn172 ND2 (3.27 Å). O3 of fucose was hydrogen bonded to Gln23 Oε1 (2.8 Å), His127 Nε2 (3.04 Å) and Trp447 Nε1 (3.2 Å). Further, O4

group linked with the Gln23 Nε2 (3.04 Å) via hydrogen bond, Glu446 Oε1 (3.34 Å) and Trp447 Nε1 (3.07 Å). TheTrp439 Nε1 binds to the O6 group (3.09 Å). Trp447 indole ring also forms a hydrophobic stacking interaction with the sugar as seen in other glycosyl hydrolases as well. The carbonyl carbon C6 of the fucose lies at a hydrogen bonding distance from the Glu446 Oε2 (2.79 Å), suggesting it would hydrogen bond to the O6 if it would have been there in fucose. The docking score for this conformation was -6.42 kcal/ mol. Many of these interactions are conserved with other glycosyl hydrolases.

With the same docking parameters, *p*NP-Glc and *p*NP-Gal were also docked onto PRGH1 active site. The docking scores were -6.06 kcal/ mol and -5.86 kcal/ mol respectively for *p*NP-Glc and *p*NP-Gal. In contrast to *p*NP-Fuc, *p*NP-Glc occupied the active site of PRGH1 in a relaxed chair configuration. This transition from a strained to a bit relaxed conformation at subsite -1 might be responsible for less activity towards the glucose moiety. In *p*NP-Glc, Oε1 and Nε2 of Gln23 hydrogen bonds to the O3 (2.77 Å) and O4 group (2.88 Å). The Glu446 Oε1 and Oε2 hydrogen bonds to the O4 (3.32 Å) and O6 group (2.81 Å) of glucose moiety. In addition, Nε1 of Trp439 hydrogen bonds to O4 (2.93 Å) and that of Trp447 interacts with O3 (2.82 Å). The Trp447 Nε1 does not interact with the O4 group as in case of *p*NP-Fuc. The O3 group points towards the Nε2 of His127 at a distance of 3.71 Å, incapable of forming a hydrogen bond. In the complex, Glu173 Oε1 hydrogen bonds to 1-OH group at 2.91 Å and Glu389 Oε1 and Oε2 hydrogen bonds to O1 group (2.57 Å and 2.90 Å) placing the glucose moiety ready for protonation and nucleophilic attack. Glu173 Oε2 linked by hydrogen bonds to O2 at a distance of 3.21 Å.

On docking *p*NP-Gal, *p*NP-Gal occupies the same space and conformation as of *p*NP-Glc. The O2 group interacts with the Asn172 ND2 (3.3 Å). The O3 group is in hydrogen bonding with Gln23 Oε1 (2.8 Å), His127 Nε2 (2.99 Å) and Trp447 NE1 (3.31 Å). The O4 group interacts with Trp447 Nε1 (2.98 Å) and Gln23 Nε2 (2.89 Å). The Glu446 Oε1 and Oε2 hydrogen bonds to the O4 (3.47 Å) and O6 group (2.28 Å) of galactose moiety. Unlike the O4 in *p*NP-Glc, O6 group binds to the Trp439 Nε1 (2.77 Å). The Glu173 Oε1 interacts with O1 and O2 at 2.8 Å and 2.96 Å and Oε2 with O2 at 3.47 Å respectively.

The relative activity of the enzyme towards these three substrates is determined by the interactions between Glu446 and 4OH and 6OH. This glutamate forms a bidentate hydrogen bonding network with the 4-OH and 6-OH group at subsite -1. This glutamate side chain can adopt its position to an axial O4 retaining the ability to recognize galacto-configured substrates due to its conformational freedom. The change of OH4 from equatorial to axial

position decreases the interaction with Glu446 while the hydrogen bonding interaction with Gln23 is almost unaffected. This explains the fact that PRGH1 prefers *p*NP-Glc substrate than the *p*NP-Gal. However it may be interesting to note that *p*NP-Fuc although galacto-configured but is still preferred on *p*NP-Glc by the enzyme. It may be thought that the interaction with Glu446 and 6-OH destabilizes the enzyme-substrate complex which might be the basis for high activity on fucosides which lack the 6OH group. The binding energies of the *p*Np-Fuc, *p*Np-Glc and *p*Np-Gal further proves that PRGH1 has higher activity on fucose as compared to the glucose and galactose moiety as shown in figure 2.17. This has also been reported earlier in Rice Os3Bglu6 [210].

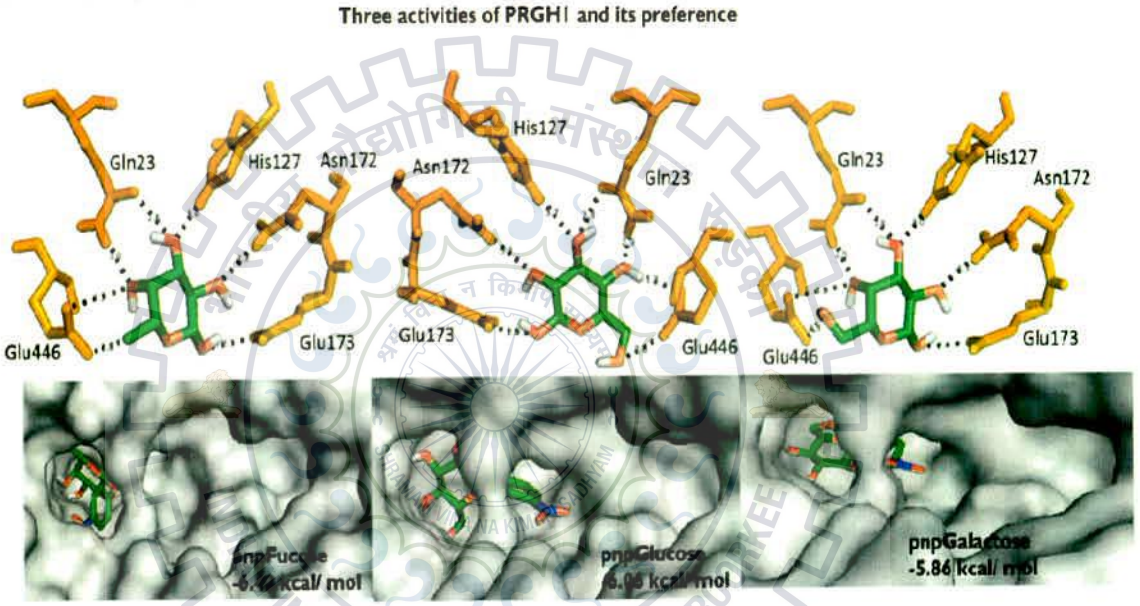


Figure 2.17. The docking of the pNP-Glucose, pNp-Fucose and pNp-Galactose with PRGH1 along with their binding energies and interaction studies.

Cellobiose, Cellotriose and Cellotetraose were also docked to know the basis of decreased activity of the enzyme with increase in chain length and also to know the locations of β-1, 4- connected glucosyl residues at subsites +1, +2 and +3 (Figure 2.18). Cellobiose docked with docking energy of -5.63 kcal/ mol while cellotriose with a docking score of -2.12 kcal/mol. The highest total docking score of cellotetraose (+5.19 kcal/mol) further indicates its poor binding with the enzyme. Interactions with the non-reducing glycosyl residue remains almost conserved and few new hydrogen bonding interactions are observed at subsites +1, +2 and +3. The active site of the enzyme is not very efficient to accommodate the long oligosaccharides like Cellotetraose as shown in figure 2.18.

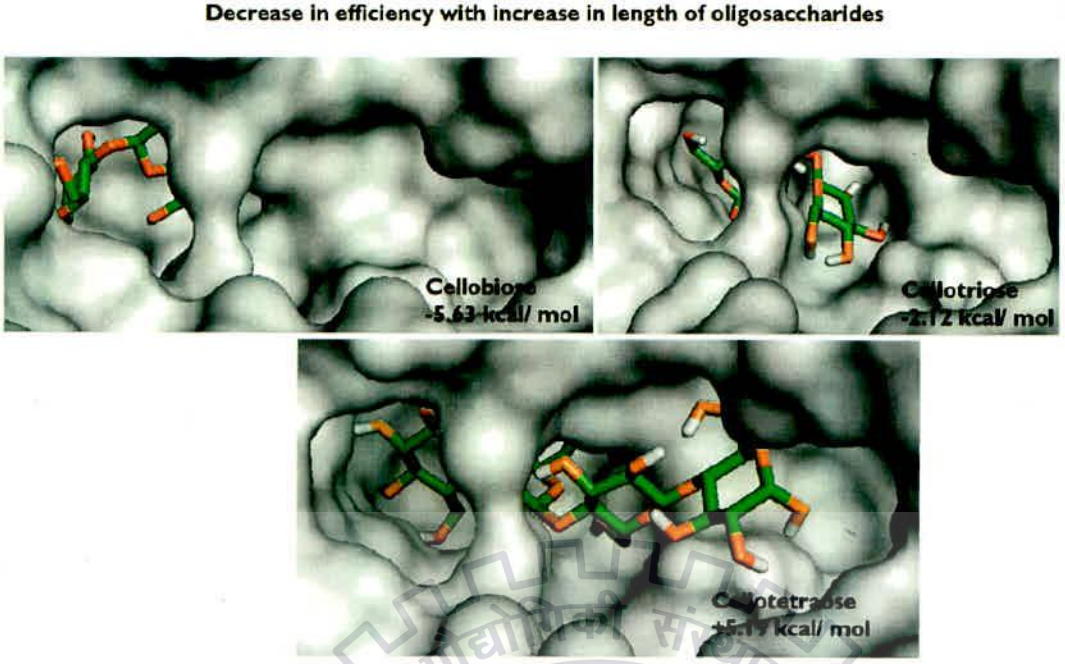


Figure 2.18. Cellobiose, Cellotriose and Cellotetraose were docked onto the PRGH1 to know their fitting into the active site.

2.4. Conclusion

The *Putranjiva roxburghii* β-glucosidase gene was cloned and submitted to NCBI with the accession no. KF006311. The β-glucosidase (PRGH1) enzyme was expressed in prokaryotic system and purified to homogeneity by employing Ni-NTA affinity chromatography. Oligomerization study showed that at higher concentration this enzyme forms a mixture of different oligomers with molecular weight around ~660 kDa, which corresponds to the decameric mass of the enzyme. This enzyme shows activity towards aryl as well as glucooligosaccharides and natural β-D-Glycosides. Mixed substrate kinetics proves that all the three chromogenic substrate are catalyzed in a single active site. Mutational studies were done to find out the catalytic acid/base and catalytic nucleophile. Expression of GH1 in seed environment is favored by the massive glycosylation, reduced exposed surface owing to oligomerization, a disulphide bond, salt bridges and main chain stabilization by hydrogen bonds between main chain and acidic/basic residues. These factors make it extremely stable suitable for the dehydrated environment of the seed. The presence of various salt bridges and a conserved disulphide bond makes it an extremely thermostable enzyme. Multiple sequence alignment shows the conservation of glycone binding residues in PRGH1 i.e. Gln23, His127, Trp128, Asn172, Glu173, Glu389, Gln446 and Trp447. Some of the aglycone binding residues like Asn314, Tyr316 and Trp439 are also conserved. ESPript and modelling studies shows that it is an α/β protein with an $(\beta/\alpha)_8$ barrel structural fold similar to other glycosyl hydrolases. Glycosyl hydrolases act via retaining double-displacement reaction mechanism including two carboxylate residues where Glu173 acts like a catalytic acid and Glu389 acts as a nucleophile. Glu173 acts as a protonating agent for the aglycone leaving group in the glycosylation step and and as a deprotonating agent of water in deglycosylation step while Glu389 attacks on the anomeric carbon for displacing the aglycone moiety and forming a covalent intermediate. Interestingly, Glu446 acts as an important catalytic residue involved in hydrogen bonding in all the complexes. The Glu446 is conserved among all glycosyl hydrolases and this further suggests that it might play an essential role in hydrolysis mechanism. The active site of PRGH1 is unique in having a restriction in the active site due to the longer side chains of Met361 and Arg180 being present close together at the mouth of the active site which might be responsible for the distinct substrate specificity of PRGH1 and also its decreased affinity towards longer oligosaccharides. Biochemical data revealed that PRGH1 shows preference for *p*NP-Fuc followed by *p*NP-Glc and then by *p*NP-Gal. This has been proved through bioinformatics analysis also. These three substrates recognized by the enzymes differ in their OH positioning and it can be proposed that the enzyme tolerates

equatorial position of the 4-OH group more than the axial position giving *p*NP-Glc preference over the *p*NP-Gal. However, in case *p*NP-Fuc, since it lacks the 6-OH group and thus the hydrogen bonding interaction with Glu446 Oε2. PRGH1 lacks the conserved Trp365 (3PTQ) present in all the O-glycosyl hydrolases. S-glycosidases also lack this conserved Trp. The closeness of PRGH1 to *S. alba* myrosinase in phylogenetic cluster along with conservation of some myrosinase aglycone residues like Arg180, Trp128 suggests that PRGH1 might also show mechanisitic similarity to myrosinase and can hydrolyze thioglucosides. PRGH1 might be evolving evolutionary from an O-glycosidase towards an S-glycosidase for the plant defence mechanism against the toxic products absorbed by the plants and also generated during mastication.

CHAPTER 3

ROLE OF N-LINKED GLYCOSYLATION ON THE STABILITY AND ACTIVITY OF PRGH1 ENZYME

3. ROLE OF N-LINKED GLYCOSYLATION ON THE STABILITY AND ACTIVITY OF PRGH1 ENZYME

3.1. Introduction

Nascent proteins undergo various post translational modifications like proteolysis, acetylation, lipidation, formylation, methylation, ubiquitination, phosphorylation and glycosylation before transforming into a mature protein. Post translational modification (PTM) involves the attachment of various functional groups like phosphate, acetate or biomolecules like carbohydrates, lipids to increase the protein diversity and its function. Glycosylation is one of the main PTM which involves the addition of sugar moiety to proteins. Carbohydrates act as the glycosyl donor while the acceptors are proteins whose hydroxyl or other functional groups accept the carbohydrate units ranging from monosaccharides to complex polysaccharides. Glycosylation occurs in the Endoplasmic reticulum (ER) or Golgi (*trans*-, *medial*-, or *cis*-) cisternae lumen. Glycosylation is a highly ordered reaction occurring in almost each organism from archaea, eubacteria and eukaryotes.

The carbohydrate-peptide bond is divided into various groups depending on the oligosaccharide attachment and the nature of bond. They are N-linked glycosylation, O-linked glycosylation, glypiation, C- linked glycosylation, phosphoglycosylation. N-linked i.e. Asparagine linked or the O-linked (Serine/Threonine hydroxyl group) are the most common modes of glycosylation commonly seen in secretory and surface proteins. Both the N- and the O-linked glycosylation are considerably different from each other like the O-linked glycans are short and are added one by one in contrast to the N-linked glycans which involves the addition of a 14 residue preformed glycan.

The carbohydrate units added act as signals for the transport of proteins from ER to the targeted organelle. In addition, glycosylation has lot of impact on antibodies, help in recognition of proteins, cell-cell adhesion. Antibody glycosylation is a common post translational modification which plays a crucial role in its effector function [187, 270]. It was found that deglycosylated antibodies in comparison to the glycosylated forms have altered intrinsic properties like the molecular hydrodynamic radii increased after deglycosylation. Moreover, the deglycosylated protein is more prone to proteolytic cleavage and has less thermal stability than its glycosylated form. Deglycosylated antibodies also had higher aggregation rates. Therefore, glycol-engineering to produce antibodies with specific glycoforms is a trendy field nowadays to increase the therapeutic value of the antibody as aglycosyl antibodies are only studied when effectors function is not desired.

N-linked glycosylation in plants play important roles for proteins right from their conformation, folding, stability from proteases, location to their biological function. In fact glycosylation is such an important aspect of protein maturation that many studies are undergoing to know the effect of alterations of glycosylation mechanism on plants with the help of N-linked glycosylation and its processing inhibitors, mutation of Asn residue whose amino group participate in N-glycosylation.

N-linked glycosylation help in attaining the biological activity of the protein like in case of jack bean lectin. It is produced as an inactive glycosylated pro-Con A protein [33] which can be converted into its active functional form having lectin activity called as the Con A protein. This activation process involves the hydrolysis of the glycopeptide bond of pro-Con A and can be done with the help of deglycosylation enzymes called as EndoH [190]. In fact, N-glycosylation is also needed for the transportation of Con A to the storage vacuole as well as for its regulation of lectin activity. They favour the functional stability of the enzyme as seen in case of stem bromelain from *Ananas comosus*, where deglycosylated bromelain show decreased enzyme activity and less stability in the organic solvents as compared to the glycosylated forms [115]. The oligosaccharide side-chains in the N-linked glycans are also involved in the secretion of glycoproteins into the extracellular part as seen in the case of cell wall β-fructosidase in carrot [64, 65]. Due to the remarkable N-linked glycosylation capacity of plants, they can be considered as an alternative host for the production of recombinant therapeutic proteins as seen in case of N-linked glycosylation in phytohaemagglutinin i.e. PHA from common bean called *Phaseolus vulgaris*. N-glycans of PHA are found to mature in a natural manner in transgenic and other different expression system of plants as is clear from the glycans retrieved from bean and other recombinant sources [191]. Plants can also be used as other organisms for therapeutic glycoproteins production by knocking out the N-glycan processing genes specific to plant or by inserting the machinery required for the addition, synthesis of human sugars and their transport [81].

Our earlier study targeted on the detection and partial characterization of a heat resistance family 1 glycoside hydrolase enzyme (PRGH1) having both β- galactosidase and β-glucosidase activities [183]. This native form of this enzyme was sourced from the seeds of *Putranjiva roxburghii* plant. It is a medicinal plant belonging to *Euphorbiaceae* family. The main purpose of this study is to make a comparison between the native (glycosylated) and deglycosylated form of this enzyme and to elucidate the effect of glycosylation on the stability of the enzyme.

3.2. Materials and Methods

3.2.1. Purification of enzyme

For the purification of native glycosylated PRGH1, affinity chromatography using concanavalin A-agarose resin was employed. The detailed procedure was described in the previous chapter.

3.2.2. Removal of N-linked glycosylation

N-linked glycans were removed by using PNGase F enzyme (NEB). The deglycosylation of native enzyme was carried out according to the manufacturer's instruction with slight modifications. 20 µl of PNGase F enzyme (500,000 units/ml) was incubated with 10 mg of PRGH1 at 37 °C for 36 hrs. Fresh PNGase F enzyme (500 units/ml) was added in every 12 hrs. Native PRGH1 without PNGase F served as control. At the end of the incubation, the reaction mixtures were analysed on 12% SDS-PAGE. After deglycosylation reaction the protein sample was applied to gel filtration chromatography to obtain purified deglycosylated PRGH1. The fractions containing deglycosylated protein was collected and concentrated.

3.2.3. Carbohydrate estimation

To determine the carbohydrate content of both the glycosylated and deglycosylated protein preparation, the phenol-sulphuric acid method as described by Dubois and co-workers was followed [54].

3.2.4. Enzyme activity assays and kinetics parameters

The enzyme activity of deglycosylated PRGH1 enzyme on the aryl glycosides such as *p*-nitrophenyl-β-D-glucopyranoside (*p*NP-Glc), *p*-nitrophenyl-β-D-galactopyranoside (*p*NP-Gal) and *p*-nitrophenyl-β-D-fucopyranoside (*p*NP-Fuc) were assayed at same condition as described previously [183]. The kinetics parameters were determined from Lineweaver-Burk plot by using these substrates. The activity of deglycosylated PRGH1 enzyme was compared with the glycosylated form of the enzyme.

3.2.5. Effect of temperature and pH

The temperature optima of deglycosylated PRGH1 was determined by the activity on *p*NP-Glc as substrate. The enzymatic activity was assayed by incubating the reaction mixtures at different temperature ranging from 10 °C to 90 °C with 5 °C increments. The

thermostability of glycosylated and deglycosylated PRGH1 was studied by incubating the reaction mixture for 30 min, with temperatures ranging from 10 °C to 90 °C. The standard assay procedure was followed by using *p*NP-Glc as substrate. All the above experiments presented were performed in triplicate.

For optimal pH determination, deglycosylated PRGH1 was assayed in 10 mM Britton-Robinson buffer, with pH values ranging from 3.0 to 11.0 at 37 °C using *p*NP-Glc as substrate. The pH stability was determined by calculating the residual enzyme activity after incubation at 37 °C for 1 h in suitable buffer.

3.2.6. Effect of alcohols

The effects of alcohols (methanol, ethanol and 1-propanol) on the enzymatic activity of both glycosylated and deglycosylated PRGH1 were studied. The enzymes were co-incubated with different concentrations (0-30%) of alcohol for 30 min and residual activity was measured by using *p*NP-Glc as substrate.

3.2.7. Proteolysis studies

Glycosylated and deglycosylated PRGH1 were used for proteolysis studies. Both the forms of this enzyme were incubated with two different proteases separately with molar ratio of 50:1 for 1 hr at 37 °C. Trypsin and papain were incubated with different forms of PRGH1 in 50 mM Tris-HCl; pH 8.0, 10 mM $CaCl_2$. SDS-PAGE analysis was performed to examine the extent of proteolysis in different forms of PRGH1.

3.2.8. Circular dichroism spectroscopy

Circular dichroism studies of both glycosylated and deglycosylated PRGH1 were carried out on a Chirascan CD Spectrometer (Applied Photophysics Ltd., Surrey KT22 7PB, UK). Far-UV CD spectra (190-260 nm) were recorded using 1 mm path length quartz cell at 25 °C with band width of 1 nm and time per point was 0.5 s. The secondary structure content of both the form of enzyme was evaluated in 20 mM potassium phosphate buffer at pH 7.4. The CD spectra were interpreted using online DICHROWEB programme. The CD results were expressed in terms of mean residue ellipticity.

3.2.9. Fluorescence measurements

For the intrinsic fluorescence experiments, a spectrofluorometer (Horiba Jobin Yvon, Model FL3-21) controlled by peltier thermal accessories (Model LF1-3751) was used. For

this experiment, both the glycosylated and deglycosylated protein samples (5 μM) were excited using an excitation wavelength of 295 nm and emission spectra were measured in the range of 300-450 nm using 5 nm slit widths for both excitation and emission. The protein samples were prepared in buffer containing 20 mM Tris-HCl; pH 8.0 and background signal obtained with buffer was subtracted for data evaluation.

3.2.10 Recombinant PRGH1 production in Saccharomyces cerevisiae

For the production of glycosylated recombinant PRGH1 enzyme, the *prgh1* gene was cloned and overexpressed in *Saccharomyces cerevisiae.* The detail procedure is described in next chapter.

3.2.11. Generation of site-specific glycosylation mutant constructs

Site directed mutagenesis was performed using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) by following the instructions of manufacturer. The yXYNSEC-PRGH1 plasmid (described in the next chapter) was used for mutations. Mutations and respective synthetic mutagenic oligonucleotide primers are listed in table 3.1.

Table 3.1. Mutations and their respective primers and the bold alphabets denote the mutated codon.

Mutation	Forward Primer (5'-3')	Reverse primer (5'-3')
N5Q	AACAATTCCTTCC**AG**AGAAGTGCCTT TCCCGATGAC	GTCATCGGGAAAGGCACTTCT**CTG**GAAG GAATTGTT
N29Q	CGAAGGTGAAGCA**CAG**AAAAGTGGC AGAGGCCCAAGC	CTGCCACTTTT**CTG**TGCTTCACCTTCGTA CTGGTAAGC
N253Q	GCCTTACTCC**CAG**AATACGGTTGATG AAGAAGCTGCCC	CTTCATCAACCGTATT**CTG**GGAGTAAGG CACGTACCAG
N374Q	CGATATCTTTTG**CAG**TACACCAAAGA TGCGTACAGAAATCCAAC	GTACGCATCTTTGGTGTA**CTG**CAAAAGA TATCGAATCCCCTC
N397Q	CAAGATGAT**CAG**GGCACCGTACCGAT GAGCATACTCC	GCTCATCGGTACGGTGCC**CTG**ATCATCTT GATTGTATCC
N407Q	GAGCATACTCCTT**CAG**GATACTCGTC GGATAATATACTATGAAAC	GTATATTATCCGACGAGTATC**CTG**AAGG AGTATGCTCATCGGTAC
N472Q	GACATGCTAAA**CAG**TCAACAGAGTG GTTCACAAATTTTCTGC	GTGAACCACTCTGTTGA**CTG**TTTAGCATG TCTTTCCAAATG

PCR was carried out with Phusion polymerase, a pair of primers designed for a particular mutation, and yXYNSEC-PRGH1 plasmid as template. The PCR product was subjected to *DpnI* restriction enzyme digestion to digest the methylated template. Digested product was used to transform XL1-Blue competent cells. The successful introduction of the desired mutation was verified by DNA sequencing of mutate constructs. Recombinant PRGH1 mutant

proteins were expressed, purified and assayed as described above for the wild type recombinant PRGH1.

3.2.12. Bioinformatics analysis

Since there is no crystal structure available for PRGH1 enzyme in PDB, homology modeling was done to find the tertiary structure of PRGH1 using Rice (PDB Id: 3PTK) Os4bglu12 as a template as described in Chapter-2. Molecular docking of the oligosaccharide was done using Autodock 4.2.5.1. The structure of the ligand molecule Man3GlcNAc2 i.e. (D-Manpα1-3[D-Manpα1-6]D-Manpβ1-4DGlcpNAcβ1-4DGlcpNAc1-OH) was made using Glycam Carbohydrate builder (http://glycam.org/tools/molecular-dynamics/oligosaccharide-builder/build-glycan?id=1). The energy minimized model of GH1 and the ligands were prepared for docking studies using MGL tools 1.5.6. Grid maps were computed using Autogrid4 having grid dimensions 66 x 72 x 62 Å with grid spacing 0.375 Å centered around the active site. Lamarckian genetic algorithm was used for docking with a population size of 150 and maximum no. of 250000 energy evaluations with maximum no. of 27000 generations. No. of GA runs were set to be 50. After successful completion of docking, results were sorted on the basis of their energy profiles and the conformation with the lowest energy was analyzed for hydrogen bonding interactions with the protein using PyMOL.

3.3. Results and Discussion

3.3.1. Protein purification and deglycosylation

The native glycosylated PRGH1 was purified to homogeneity by employing concanavalin A affinity chromatography (discussed in the previous chapter). This result shows that native PRGH1 is a glycosylated protein. Prolong incubation of PRGH1 with PNGase F enzyme resulted the deglycosyaltion of native PRGH1. SDS-PAGE analysis under reducing condition indicated in that deglycosylated PRGH1 migrated with an apparent molecular mass (Mr) of 60 kDa (Figure 3.1). This difference of molecular masses between the glycosylated and deglycosylated PRGH1 suggested that PNGase F enzyme successfully cleave the N-linked sugar moieties. This result was further strengthened by the fact that deglycosylated PRGH1 didn't bind to concanavalin A matrix. Deglycosylation was further confirmed by estimating the carbohydrate fraction by Phenol-sulphuric acid method. Previously many researchers have used PNGase F enzyme to remove the glycan moieties of glycoprotein successfully [85, 226].

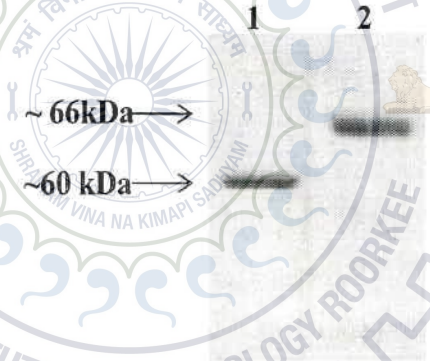


Figure 3.1. Deglycosylation of native PRGH1 enzyme by using PNGase F. Lane1, deglycosylated PRGH1; Lane 2, Glycosylated PRGH1.

3.3.2. Differences in the catalytic activities

Table 3.2 shows both forms of this protein followed the same pattern in substrate preferences. The glycosylated form of this protein shows a slight higher affinity towards substrates as deglycosylated form has higher Km than the glycosylated form. The values of V_{max} for glycosylated PRGH1 are significantly higher than the corresponding V_{max} values for deglycosylated PRGH1. These results suggest that the active site architecture of the enzyme mightn't have changed so much by virtue of deglycosylation. However, the catalytic activity may be changed owing to the topological effect.

Table 3.2. Kinetic parameters of glycosylated and deglycosylated PRGH1 enzyme.

Substrate	K_m (mM)	V_{max}(μkat/mg)	K_{cat} (s^{-1})	K_{cat}/K_m ($M^{-1}s^{-1}$)
Glycosylated (Native) PRGH1				
*p*NP-Glc	0.53	0.181	12	2.27×10^4
*p*NP-Gal	0.64	0.112	7.39	1.15×10^4
*p*NP-Fuc	0.46	0.216	14.32	3.11×10^4
Deglycosylated PRGH1				
*p*NP-Glc	0.57	0.139	9.26	1.62×10^4
*p*NP-Gal	0.69	0.080	5.30	7.68×10^3
*p*NP-Fuc	0.49	0.171	11.33	2.31×10^4

3.3.3. Effect of temperature

PRGH1 is a thermostable enzyme having optimum temperature at 65 °C. Deglycosylated PRGH1 follows the same pattern with an overall less activity at each temperature (Figure 3.2 A). At the optimum temperature, the deglycosylated enzyme showed 1.4 fold less activity than the glycosylated enzyme. Here glycosylation is not required for the activity of the enzyme but required for optimal activation. Thermostability experiments showed that the glycosylated enzyme was fairly stable over a broad temperature range of 20-70 °C, and the enzyme activity reduced above 70 °C. However, the activity of the deglycosylated enzyme reduced sharply above 45 °C. The result clearly demonstrated that deglycosylation preparation is more sensitive towards the high temperature. Similar study with fungal β-glucanase [180] and human aquaporin 10 protein [178] showed that glycosylation affects the thermostability of the proteins. Thermal stability of proteins were also studied by using differential scanning calorimetry [16].

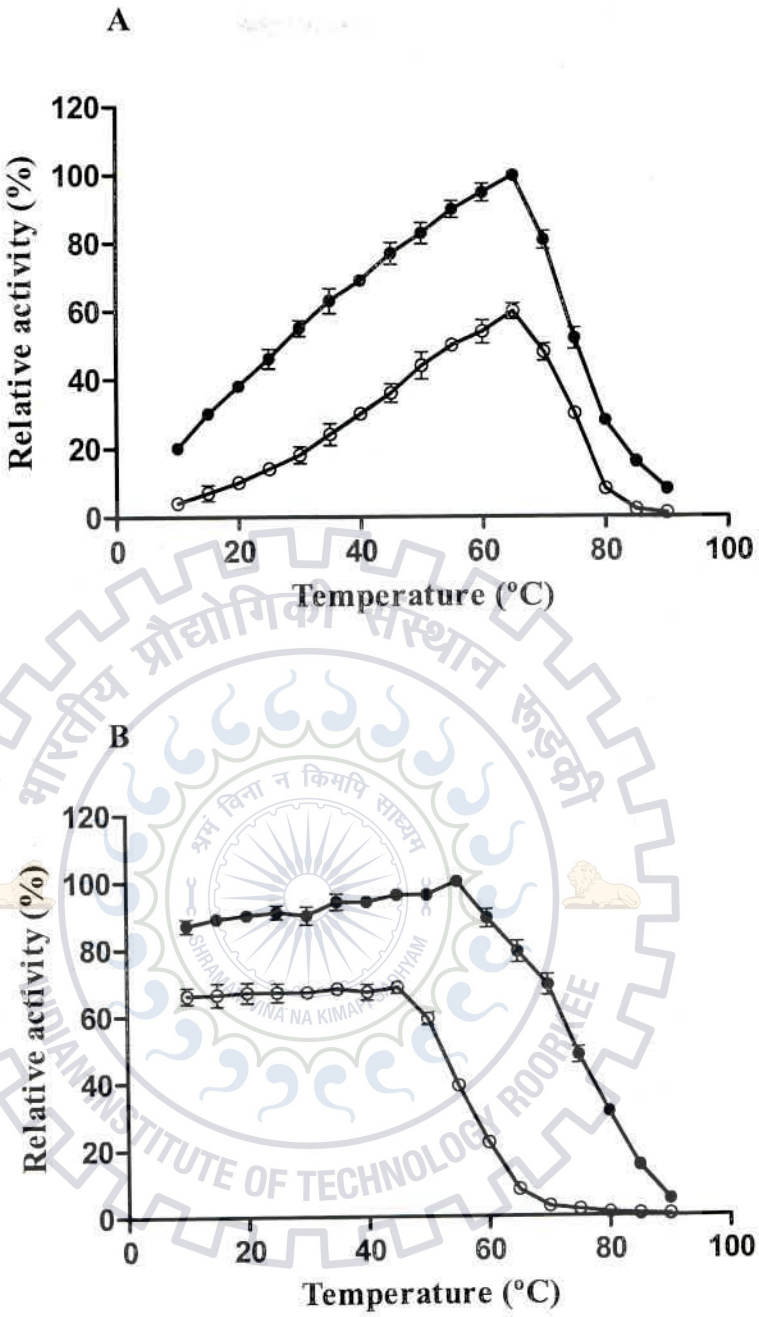


Figure 3.2. The effect of temperature on the glycosylated and deglycosylated PRGH1 activity (A) and stability (B). Solid circles denote glycosylated PRGH1 and open circle denotes deglycosylated PRGH1. For stability assay, the purified enzymes were incubated at different temperatures at pH 4.6 for 30 min and residual activity was measured. Results are the average of triplicate and the bars indicate the standard error.

3.3.4. Effect of pH

The optimum pH for the *p*NP-Glc activity of glycosylated PRGH1 was studied over a pH range of 3.0 to 11.0 at 37 °C. As shown in figure, the enzyme retained most of its original activity at pH 5.0-6.0 with optimum at pH 4.6 (Figure 3.3 A). At optimum pH the glycosylated PRGH1 showed 1.5 fold more activity than the deglycosylated PRGH1. However, at the extreme ends of the pH i.e. pH 3.0 and 9.0 both forms of the enzyme showed similar activity. The deglycosylated PRGH1 enzyme shows a similar trend with an overall less activity at every pH. In addition to this at higher pH the activity of deglycosylated PRGH1 decreased sharply. The effects of various pHs on the stability of recombinant PRGH1 were checked (Figure 3.3 B). Both the forms of this protein were relatively stable over a pH range of 4.0 to 8.0.

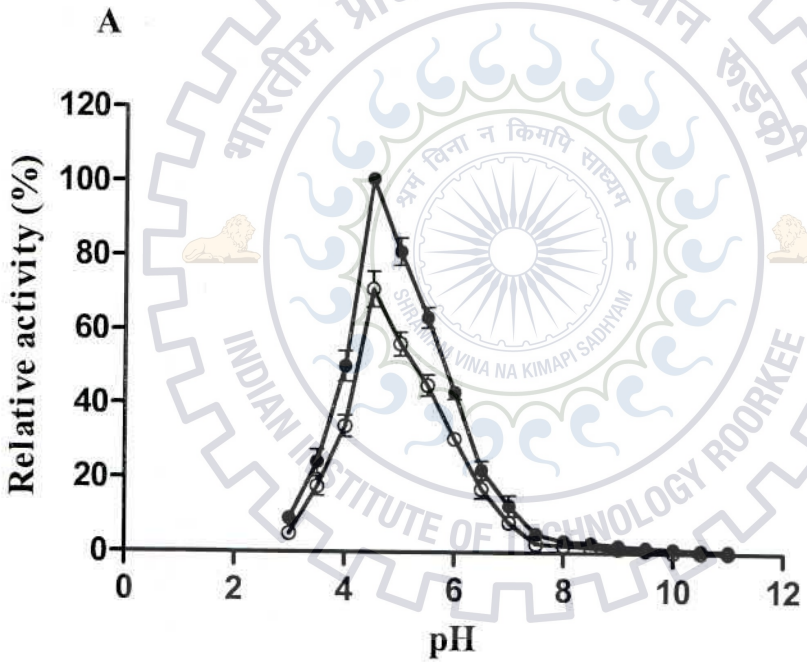


Figure 3.3. (A) The effect of pH on the glycosylated and deglycosylated PRGH1 activity. Solid circles denote glycosylated PRGH1 and open circle denotes deglycosylated PRGH1.

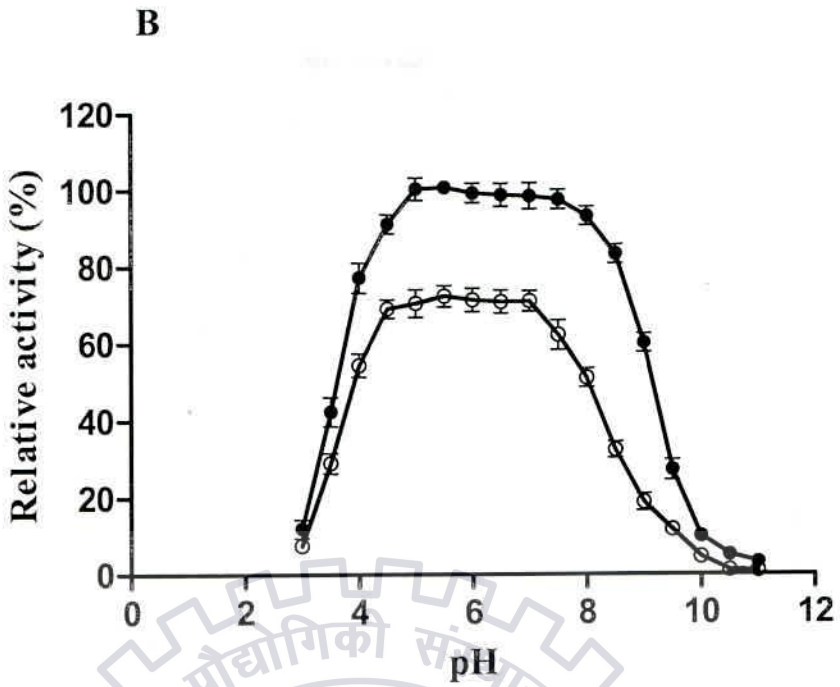


Figure 3.3. (B) The effect of pH on the glycosylated and deglycosylated PRGH1 stability. Solid circles denote glycosylated PRGH1 and open circle denotes deglycosylated PRGH1. For stability assay, the purified enzymes were incubated at different pH at 37 °C for 1 h and residual activity was measured. Results are the average of triplicate and the bars indicate the standard error.

3.3.5. Effects of alcohols

The effect of lower chain alcohols was studied by incubating both the glycosylated and deglycosylated PRGH1 enzyme with various concentrations of alcohols. The result showed that at a fixed concentration, the activity of both the form of enzyme decreases with increase in the length of the hydrocarbon chain of the alcohol (Figure 3.4). It can be seen that low concentration of methanol has stimulatory effect on both the preparation. From all the above experiment, it can be noticed that the deglycosylated form of the enzyme has relatively lower activity than the glycosylated form. Likewise, this same trend is repeated here. So, it can be concluded that alcohols have no drastic effect upon deglycosylation unlike the stem bromelain enzyme sourced from *Ananas comosus* [115].

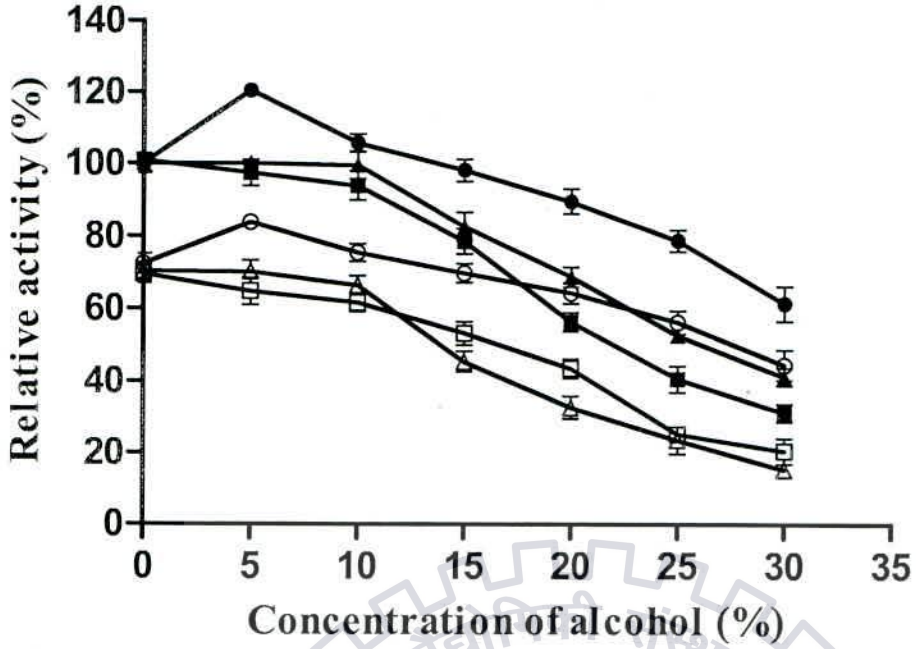


Figure 3.4. Effect of alcohols on relative activity of both glycosylated (solid structure) and deglycosylated (empty structure): methanol (circle), ethanol (triangle), 1-propanol (square). Results are the average of triplicate and the bars indicate the standard error.

3.3.6. Effect of glycosylation on protease susceptibility

We have carried out limited proteolysis of both the glycosylated and deglycosylated PRGH1 in order to determine if unglycosylated PRGH1 is more susceptible to proteolytic degradation than the glycosylated counterpart. The result showed that the unglycosylated PRGH1 was quite sensitive to proteolysis against the two proteases we have used at 1:50 molar ratios. However, native glycosylated PRGH1 was found to be completely resistance to proteolysis (Figure 3.5). Proteolysis is an important mean to study the structural instability of the proteins imposed due to the deglycosylation process. Studies reported that 'protein resistance to proteolytic attack increases with its conformational rigidity and vulnerability to proteolysis reflects the segmental mobility' [68, 69, 103]. Proteolysis study showed that glycosylated PRGH1 is highly resistance to proteolytic degradation which precisely suggests the structure of native glycosylated PRGH1 is quite compact and rigid than the deglycosylated counterpart. Further, mutational studies showed that the glycosylation sites are present on the loop, which is believed to be the most susceptible region for protease attack. The glycosylation might prevent the proteases to access the proteolysis sites. Thus it can be concluded that the conformation of both forms of this protein is different as showed by differences in proteolysis sensitivity. This result was further strengthened by the CD study.

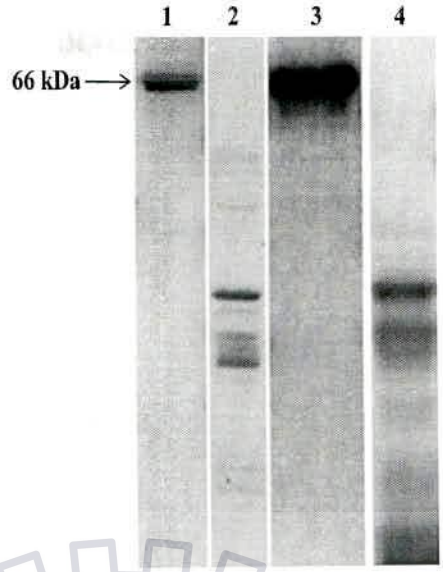


Figure 3.5. Proteolysis studies of both glycosylated and deglycosylated PRGH1 in the presence of protease. Lane 1, glycosylated PRGH1 with trypsin; Lane 2, deglycosylated PRGH1 with trypsin; Lane 3, glycosylated PRGH1 with papain, Lane 4, glycosylated PRGH1 with papain.

3.3.7. CD spectra of glycosylated and deglycosylated enzyme

The effect of deglycosylation on the secondary structure of PRGH1 was examined by far UV CD studies. The far UV CD spectrum showed there is minor difference between the two forms of this protein (Figure 3.6). It was observed that deglycosylation process resulted in a decrease of the ellipticity suggesting little conformational changes upon deglycosylation. This study is supported by other reports; they stated glycosylation didn't alter the secondary conformation of the protein in a significant manner [6, 37, 115].

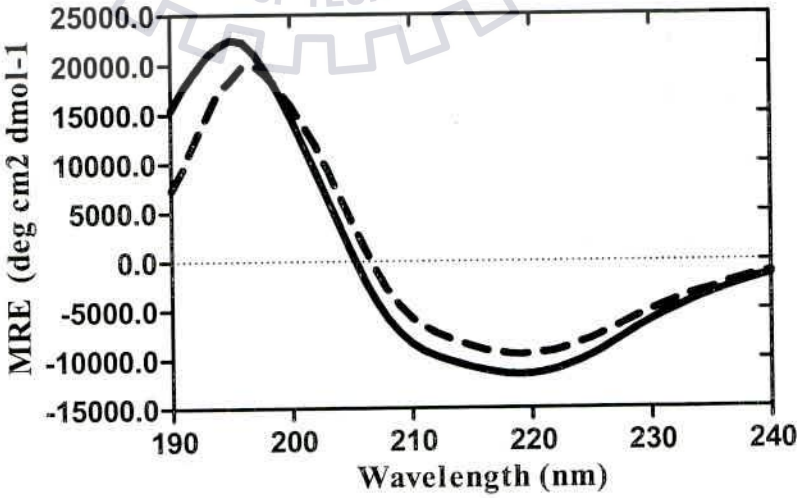


Figure 3.6. Far UV, CD spectra of glycosylated (——) and deglycosylated (– – –) PRGH1.

3.3.8. Intrinsic fluorescence of glycosylated and deglycosylated PRGH1

We measured the fluorescence of the intrinsic tryptophan residues of both glycosylated and deglycosylated PRGH1 to know the conformational changes induced by deglycosylation. Compared to the glycosylated form, the deglycosylated form of PRGH1 displayed a 1.13 fold increase in emission intensity at 342 nm and a red shift of ~3 nm in the emission maximum (Figure 3.7). The shift and the increase in the emission peak may be due to change in the micro-environment of the chromophoric groups because of the deglycosylation. This result showed that tryptophan residues got exposure upon deglycosylation and they migrate from a more hydrophobic core to a less hydrophobic environment. This result strongly suggests that a significant conformational change of PRGH1 enzyme follows the deglycosylation of the protein which precisely suggests the compact and rigid structure of native glycosylated may be changed to a comparatively flexible structure upon glycosylation. Fluorescence studies on human thyroglobulin showed a significant change in conformation upon deglycosylation [82].

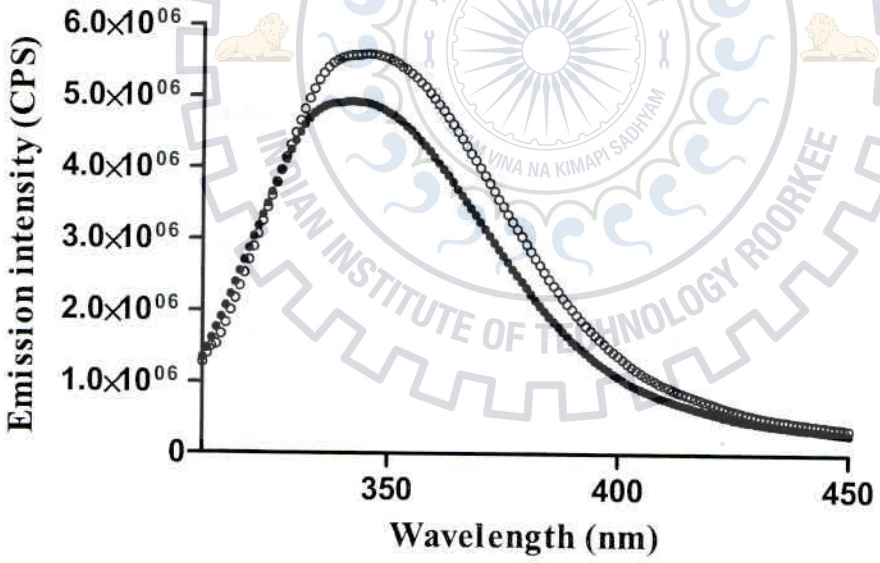


Figure 3.7. Effect of glycosylation on the intrinsic fluorescence emission spectra of glycosylated (solid circle) and deglycosylated (open circle) PRGH1. The experimental conditions are described under materials and methods. The emission intensity unit is counts per second (CPS).

3.3.9. Identification of N-linked glycosylation sites on PRGH1 by mutagenesis

The PRGH1 enzyme was successfully expressed in *Saccharomyces cerevisiae* and purified to homogeneity (detail in the next chapter). The molecular weight of purified recombinant PRGH1 is ~66 kDa (similar to native PRGH1), hence contain N-glycosylated sugars. PRGH1 contains 7 potential N-linked glycosylation sites (Asn-X-Ser/Thr) as predicted by using online tools. These include the asparagines at positions 5, 29, 253, 374, 397, 407 and 472. These sites are mostly conserved in plant β-glucosidases. Individual mutant constructs were generated at each putative N-linked glycosylation site. These mutants allow for accuracy in determining whether an individual site is utilized for N-linked glycosylation. The asparagine residues were mutated to glutamine and transformed into *Saccharomyces cerevisiae.* Each mutated protein was purified and subjected to analysis by SDS-PAGE. The assumption was that there would be a little increase in electrophoretic mobility in a mutant PRGH1 as compared to the wild PRGH1 if the mutated sites were N-linked glycosylated in the wild PRGH1. Result shows that three of seven mutants exhibited a slight greater mobility i.e. decrease in molecular mass (Figure 3.8). Each purified protein was assayed by using *p*NP-Glc as substrate to further verify that the purified protein was our target. Several such mutational studies were carried out to trace the N-linked sites [30, 145].

Figure 3.8. Effect of site specific N-linked glycosylation mutants on glycosylation. Lane 1, 4 and 6 are wild type; Lane 2, 3, 5, 6-10 are mutants with mutations at 29, 253, 397, 374, 407, 472 and 5 sites respectively.

3.3.10. Bioinformatics analysis

The biochemical studies on PRGH1 revealed three N-glycosylation sites with Asn-X-Ser/Thr motif are glycosylated. These sites are Asn29, Asn253 and Asn397. Therefore to know their location on the model and further study the glycosylation behavior and how it helps from proteolysis, modeling of PRGH1 and its docking with the high mannose structures found in plants is done (Figure 3.9). All the three glycosylated asparagine residues are seen on the loop regions and the surface view also indicates that they are exposed rather than buried inside which makes them excellent candidate for glycosylation. The glycan structure i.e. Man3GlcNAc2 i.e. (D-Manpα1-3[D-Manpα1-6]D-Manpβ1-4DGlcpNAcβ1-4DGlcpNAc1-

OH) forms hydrogen bonding interactions with the exposed asparagines. There are reports that glycosylation hinders the biological activity of the protein, however none of the potential sites mentioned here interferes with the active site residues (Figure 3.9). However, glycosylation is of great importance in term of stabilizing the PRGH1 structure and preventing it from proteolysis. Presence of the sugar moieties at the surface of the enzyme may prevent the proteases to cleave the glycosylated PRGH1 enzyme.

Figure 3.9. Docking of the glycan ($Man_3GlcNAc_2$) ($Man_{5-9}GlcNAc_2$) (D-Manpα1-3[D-Manpα1-6]D-Manpβ1-4DGlcpNAcβ1-4DGlcpNAc1-OH) with PRGH1 on the glycosylation sites Asn29, Asn253 and Asn397. The glycan structure and the glycosylation sites are shown in sticks with carbon colored green, Nitrogen blue and oxygen red color. The catalytic acid/base Glu173 and catalytic nucleophilic residue Glu389 are shown in sticks in orange color.

3.4. Conclusion

Glycosylation, one of the most complex and common PTMs, is of immense interest for its role in many biological functions such as cell signaling, molecular recognition, and immune defence. In addition to this, it has subtle effects on the various physiological activities like immunogenicity, solubility, circulatory half life and thermostability. The aim of this investigation was to examine the role of glycosylation for the stability and its consequence effect on the activity of a very efficient thermostable β-glucosidase from *Putranjiva roxburghii* plant (PRGH1). We successfully produced deglycosylated PRGH1 by using PNGase F. We compared the activities of both forms of this enzyme under various conditions like different pH, temperature and alcohols. The result showed that deglycosylation decreases the overall activity of the PRGH1. At higher pH the deglycosylated PRGH1 showed a sharp decrease in activity. The temperature profile of both the glycosylated and deglycosylated enzyme clearly reflect that glycosylated form of this enzyme have greater stability at higher temperature. Altogether, these data reflects that the carbohydrate moieties contribute to the stability of the native protein making it more resistance to different denaturants. Spectral properties of both glycosylated and deglycosylated enzymes showed that the conformation of the native protein changed to a certain extent after removal of the N-linked sugars indicating the possible effect of glycosylation on local and/or global conformational dynamics of proteins. Proteolysis study along with the spectral studies suggests that the structure of native glycosylated PRGH1 is quite compact and rigid than the deglycosylated counterpart. Mutagenesis studied shows that out of seven potential glycosylation site three sites were glycosylated. Additional studies will be required to determine how each mutation affects the stability and activity of the enzyme. Overall, our results suggest that N-linked glycosylation is not essential for β-glucosidase enzyme activity but it is important for stability of the enzyme for the optimal activity.

CHAPTER 4

CLONING, EXPRESSION, PURIFICATION OF PRGH1 ENZYME BY USING YEAST SYSTEM AND ITS APPLICATION FOR BIOETHANOL PRODUCTION

4. CLONING, EXPRESSION, PURIFICATION OF PRGH1 ENZYME BY USING YEAST SYSTEM AND ITS APPLICATION FOR BIOETHANOL PRODUCTION

4.1. Introduction

Cellulose is the most abundant polysaccharide on the earth and virtually an unlimited source of renewable bioenergy. However, in the absence of suitable treatment, a huge amount of industrial and agricultural insoluble cellulosic waste has accumulated, becoming a cause of environmental pollution [31]. The production of fuel ethanol from this cellulosic biomass remains a highly attractive one in terms of environmental, commercial and social sustainability [63]. Sugars for the fermentation process are accessed from cellulosic material through chemical or enzymatic hydrolysis. The enzymatic hydrolysis of cellulosic biomass involves conversion of biomass to reducing sugars, and the subsequent conversion of the reducing sugars to ethanol. However, this process is very costly owing to the recalcitrance of cellulose, resulting the low yield and high cost of the enzymatic hydrolysis process [157]. Microbial cellulolytic enzyme complex consists of three basic types of enzymes, including cellobiohydrolases (EC 3.2.1.91), endoglucanases (EC 3.2.1.4) and β-glucosidases (EC 3.1.2.21), which work synergistically to degrade cellulose to glucose. Cellobiohydrolases and endoglucanases synergistically degrade native cellulose to generate cellobiose, which is a strong product inhibitor of both the enzymes [126]. β-glucosidases cleave the β-1,4-glycosidic linkage of cellobiose to generate D-glucose. Therefore, β-glucosidases not only catalyze the final step of cellulose degradation, but also allow the cellulolytic enzymes to function more efficiently by relieving the cellobiose mediated inhibition [216]. However, cellobiose is consistently accumulated owing to the weak β-glucosidases activity of most microbial cellulases making cellobiose hydrolyzation the rate-limiting step during this enzymatic hydrolysis process of the cellulosic biomass [217].

One-step conversion of cellulosic biomass to bioethanol with an organism capable of cellulose degradation and efficient fermentation (consolidated bioprocessing-CBP) may offer cost reductions of bioethanol production [157]. One of the most effective ethanol-producing yeasts, *Saccharomyces cerevisiae*, has several advantages including (i) natural robustness in industrial process (ii) simplicity in genetic manipulation (iii) larger cell size, which simplify their separation from the culture broth (iv) resistance to viral infection and (v) commonly regarded as safe (GRAS) status due to its long association with the food and beverage industries [87]. Despite of these advantages *S. cerevisiae* has a major drawback due to its inability to degrade cellobiose and longer chain cello-oligosaccharides efficiently and these

are the dominant soluble by-products of cellulose hydrolysis. Therefore, construction of β-glucosidases overproducing strain is an important strategy to enhance the efficient utilization of cellobiose. Great efforts have been made to genetically engineer *S. cerevisiae* to grow on cellobiose by expressing heterologous β-glucosidases genes. Mostly, β-glucosidase from bacterial and fungal origin have been transferred to *S. cerevisiae* enabling the growth on cellobiose [86, 159, 236], but no effort has been made with the efficient plant β-glucosidases till date. Enzyme thermostability is crucial during the saccharification step because steam is always used to make the substrates more appropriate for enzymatic hydrolysis process [150]. Thermostable β-glucosidases can be used in the saccharification procedure without a pre-cooling process.

Our previous study focused on the identification and partial characterization of a thermostable family 1 glycosyl hydrolase enzyme (PRGH1) which was sourced from *Putranjiva roxburghii* plant [183]. *P. roxburghii* is a medicinal plant from *Euphorbiaceae* family. In this study, the *prgh1* gene was overexpressed under the control of the PGK1 promoter and terminator in *S. cerevisiae*. The enzymatic properties of the purified recombinant PRGH1 were characterized. The ability to the sustain growth of the obtained strain on cellobiose as sole carbon source was studied under aerobic conditions. To the best of our knowledge, it is the first report where a plant β-glucosidase gene was used to engineer *S. cerevisiae* for better utilization of cellobiose.

4.2. Materials and Methods

4.2.1. Microorganism strains and culture conditions

E. coli strain XL1-Blue (Stratagene) was used as host strain for maintenance and amplification of constructs. Luria-Bertani medium [202] was used to cultivate the bacterial strain, and 100 μg/ml of ampicillin was added for selecting transformant. *S. cerevisiae* Y294 was cultured in YPD medium (1% yeast extract, 2% peptone, 2% glucose) at 30 °C. Selective synthetic complete (SC) medium (2% glucose. 0.67% yeast nitrogen base [HiMedia, India] containing amino acid supplements) was used to select and maintain *S. cerevisiae* transformant.

4.2.2. Recombinant vector construction

The enzymes for amplification, restriction digestion and ligation were purchased from NEB. The yXYNSEC vector, pDF1 plasmid and *S. cerevisiae* Y294 strain were kindly provided by Prof. Willem Heber van Zyl, Stellenbosch University, South Africa. Previously, we have cloned the 1617 bp β-glucosidase gene *prgh1*, into the pGEM-T vector and the gene sequence was submitted in the NCBI Gene Databank with the Accession no KF006311. The ORF of PRGH1 gene without native signal sequence was amplified by PCR from the recombinant plasmid pGEM-T-PRGH1 by using the primers set 5'-GAGCTCGCGAAATTCCTTCAACAGAAGTG-3' and 5'-CGGAAGATCTTAAGCGGCTGCTGATCTAATAG-3' having overhanging regions encoding NruI and BglII restriction sites respectively. The purified PCR product was digested with NruI and BglII and was ligated to NruI/BglII site of yXYNSEC vector, which enabled the fusion of *prgh1* gene to xyn2 secretion signal from *Trichoderma reesei*. The detail construction of the multi-copy, yeast expression vector yXYNSEC was described by Rooyen *et al.* [236]. Correct construction was confirmed by restriction digestion and subsequent DNA sequencing. The resulting construct was designated yXYNSEC-PRGH1.

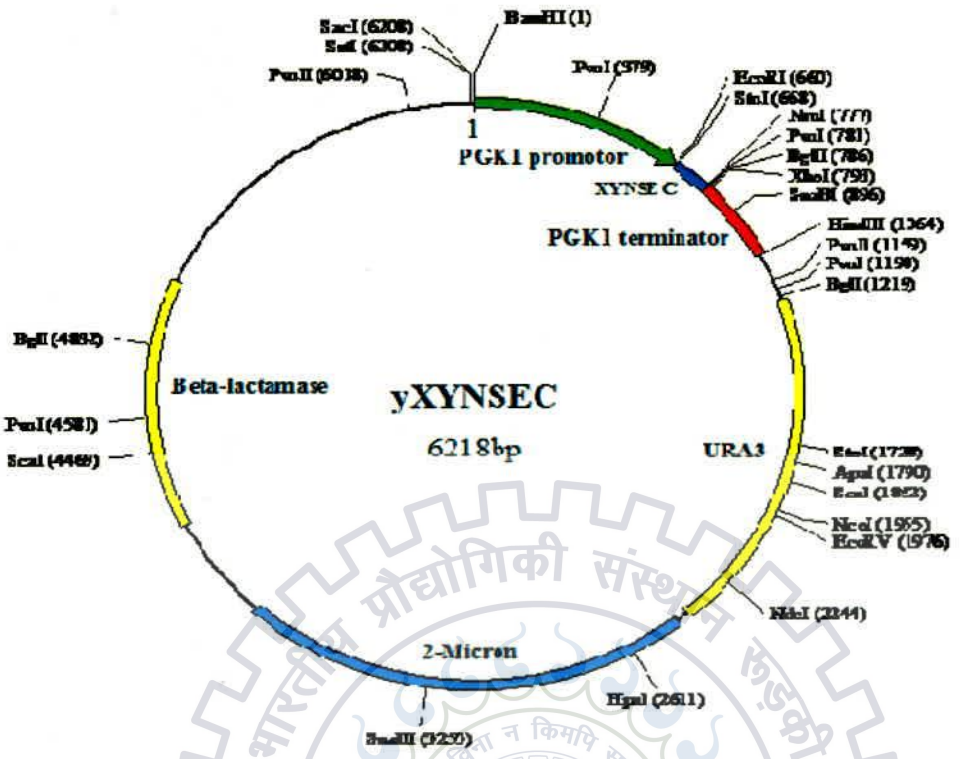


Figure 4.1. Plasmid map of yXYNSEC [236].

4.2.3. Yeast transformation and FUR1 gene disruption

S. cerevisiae Y294 was transformed with the yXYNSEC-PRGH1 plasmid by the dimethyl sulfoxide-lithium acetate method described by Hill et al. [98]. Transformants were screened and maintained on SC^{-Ura} medium. Further confirmation of the transformation was done with PCR. To ensure autoselection of the URA3-bearing yXYNSEC-PRGH1 plasmid in non-selective medium, disruption of the uracil phosphoribosyltransferase (FUR1) gene in the *S. cerevisiae* Y294 transformants was performed by using pDF1 plasmid [133]. Fur1::Leu2 autoselective transformants were screened on $SC^{-Ura-Leu}$ medium.

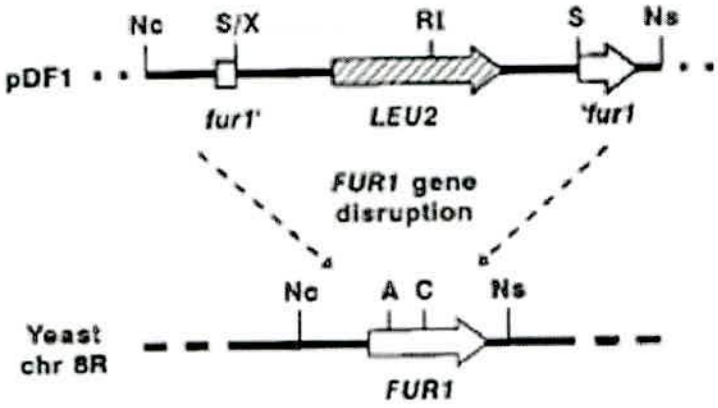


Figure 4.2. Schematic representation of the disruption of the FUR1 gene in the genome of β-glucosidase producing *S. cerevisiae* strain, figure adopted from reference no. 133.

4.2.4. Purification of recombinant PRGH1

The autoselective *S. cerevisiae* Y294 transformants were inoculated onto a defined solid medium with cellobiose as the sole carbon source. The plate was incubated for 2 days and then overlaid with 0.8% agar containing 5 mM 4-methylumbelliferyl-β-D-glucopyranoside (MUG). After that, the plate was incubated at 50 °C for 15 min and the colonies with β-glucosidase activity was monitored under UV light (365 nm).

For protein purification, autoselective *S. cerevisiae* Y294 cells with strong β-glucosidase activity were grown in 1 L of defined medium with cellobiose as the sole carbon source at 30 °C for 48 h. The supernatant of the culture was separated from the cellular mass by centrifugation for 5 min at 5000 X g. The supernatant was utilized as crude enzyme for the purification. The supernatant was percolated through filter paper and concentrated up to 50 mL using a 10 kDa cutoff Amicon Ultra-15 concentrator (Millipore, Bedford, Massachusetts, USA). The concentrated supernatant was precipitated at increasing concentration of ammonium sulfate at 4 °C. The resulting precipitates containing most amount of enzyme were collected by centrifugation, dissolved in distilled water and dialyzed against 50 mM Tris-HCl; pH 8.0 and loaded on a DEAE-Sepharose column pre-equilibrated with the same buffer. The column was washed extensively with 50 mM Tris-HCl; pH 8.0 to remove unbound proteins and bound proteins were eluted with step gradient from 0 to 0.3 M NaCl. Homogeneity of the each fraction was analyzed by 12% SDS-PAGE. The pure and active fractions were pooled, dialyzed and concentrated by Amicon Ultra (10 kDa MWCO).

4.2.5. SDS-PAGE and MUG-zymogram analysis

SDS-PAGE was performed with a 12% polyacrylamide gel using the method described by Laemmli [134]. The samples were incubated for 5 min at 95 °C in the presence of 10% SDS and 0.5% β-ME before subjected to SDS-PAGE. The gel was stained with Coomassie Brilliant Blue R-250 and subsequently destained with destaining solution.

The β-glucosidase activity of the purified enzyme was observed by the zymogram assay with 4-methylumbelliferyl β-D-glucopyranoside (MUG) as substrate by using 6% native PAGE. The gel was run for 3 h at a constant current of 10 mA at 4 °C. After electrophoresis, the gel was repeatedly rinsed with distilled water and 50 mM citrate phosphate buffer; pH 5.0 before being overlaid with 0.5 mM MUG in same buffer, and incubated at 50 °C for 10 min. The fluorescence product was visualised under UV 365 nm, and subsequently gel was stained with Coomassie Brilliant Blue R-250.

4.2.6. Circular dichroism spectroscopy

Circular dichroism study of purified recombinant PRGH1 was carried out with a Chirascan Circular Dichroism Spectrometer (Applied Photophysics Ltd., Surrey KT22 7PB, United Kingdom). Far-UV CD spectra (190-260 nm) were collected using 1 mm path length quartz cell at 25 °C with spectral band-width of 1 nm and time per point was 0.5 s. Protein concentration was kept 0.2 mg/ml and the protein sample was filtered using 0.45 μM filter prior to use. The CD spectra were analyzed using online DICHROWEB programme. The CD results were expressed in terms of mean residue ellipticity.

4.2.7. Substrate specificity

After optimising the purification of recombinant PRGH1 enzyme, the substrate specificity of the purified enzyme was determined. The enzyme activity on the aryl glycosides such as *p*-nitrophenyl-β-D-glucopyranoside, *p*-nitrophenyl-β-D-galactopyranoside, *p*-nitrophenyl-β-D-fucopyranoside and *p*-nitrophenyl-β-D-cellobioside were assayed at same condition as described previously [183]. The kinetics parameters were determined from Lineweaver-Burk plot by using these substrates. The substrate specificity of recombinant PRGH1 towards different glucooligosaccharides like cellobiose, cellotriose, cellotetrose, cellopentaose, trehalose, lactose and maltose were analyzed by quantifying the released glucose using glucose (HK) assay kit (Sigma). Furthermore, the natural substrates like

amygdalin, arbutin, laminarin, genistin, daidzin, β-gentiobiose and quercetin 3-β-D-glucoside were also assayed at the same condition.

4.2.8. Effect of temperature and pH on enzyme activity

The temperature optima of recombinant PRGH1 was determined by the activity on cellobiose as substrate. The enzymatic activity was assayed by incubating the reaction mixtures at different temperature ranging from 10 °C to 90 °C with 5 °C increments. The thermostability was determined by measuring the residual activity of recombinant PRGH1 after incubating the enzyme at different temperatures (10 °C-90 °C) in 50 mM citrate phosphate buffer; pH 5.0 for 1 h. For optimal pH determination, recombinant PRGH1 was assayed in 10 mM Britton-Robinson buffer, with pH values ranging from 3.0 to 11.0 at 65 °C using cellobiose as substrate. The pH stability was determined by calculating the residual enzyme activity after incubation at 65 °C for 1 h in suitable buffer.

4.2.9. Effect of inhibitors and additives

The inhibitory effect of N-bromosuccinimide (NBS) and δ-gluconolactone on the activity of the recombinant PRGH1 towards the substrate cellobiose were determined by incubating 1 mM inhibitor with 1 μmol of enzyme at 37 °C for 10 min. Similarly, the effects of additives were assessed by evaluating enzyme activity in the presence of various metal ions, EDTA, DMSO, Triton X-100, Urea, DTT, Guanidine hydrochloride, β-mercaptoethanol, and SDS. The effects of alcohols were studied by co-incubating the enzymes with different concentrations of alcohol for 30 min and residual activity was measured by using cellobiose as substrate.

4.2.10. Shake-flask fermentation of cellobiose

Autoselective *S. cerevisiae* Y294 transformants were used in SSF experiments. First, the transformants were pre-cultured overnight in YPD medium, harvested by centrifugation for 5 min at 3000 X g at room temperature, resulted cell pellet was twice washed with sterilized distilled water, resuspended in a medium containing cellobiose as single carbon source and finally inoculated into fermentation medium. Initial cell density was adjusted to an OD_{600} of 0.05. Flask fermentation experiments were performed using 50 mL of medium containing 20 g/L of cellobiose as sole carbon source in 250 mL flask at 30 °C with an agitation speed of 200 rpm under an aerobic condition. The medium was supplemented with different amino acids according to the auxotrophic requirements and to increase the

heterologous protein expression as per Rooyen et al. [236]. The growth curve, cellobiose consumption, and ethanol production of this culture was determined at 0, 6, 12, 18, 24, 30, 36, 42 and 48 hrs. The cell density was measured by using spectrophotometer at 600_{nm}. Cellobiose and ethanol concentration were determined as described below.

4.2.11. Simultaneous saccharification and fermentation (SSF) assays with cellulosic substrates

Equivalent SSF experiments were carried out batch wise in 250 mL conical flasks with CMC/rice straw/sugarcane bagasse as single carbon source. Both rice straw and sugarcane bagasse were alkali pre-treated with diluted NaOH (0.5%) for 30 minutes at 100 °C. After NaOH pre-treatment the solid mass was separated from the liquid fraction, then washed several time with running water to neutralise the pH followed by a final rinse with distilled water and dried to constant weight at 60 °C. The conical flasks were loaded with 10 g of dry cellulosic mass (both alkali treated/non-treated) and 70 mL of 50 mM Sodium phosphate buffer; pH 5.0 and the media sterilized at 121 °C for 20 min. Prior to the SSF process the cellulosic masses were pre-treated with a cellulosic load (20 FPU/g of dry cellulosic mass) and incubated at 45 °C for 24 h in a rotary shaker at 100 rpm. Then the temperature of the media was cooled down to 30 °C and *S. cerevisiae* Y294 transformants suspension was added aseptically to the medium and the SSF was run for 72 h. Samples were taken at 0, 4, 8, 12, 18, 24, 30, 36, 42, 48, and 72 h, and analyzed for cellobiose, glucose and ethanol. In all SSF experiment, parental *S. cerevisiae* Y294 was used in corresponding experiment as a control. Cellobiose, glucose and ethanol concentration were determined with a Shimadzu HPLC chromatograph equipped with a refractive index detector using a Bio-Rad Aminex HPX-87H column. The column temperature was maintained at 50 °C and water was used as mobile phase at a flow rate of 0.5 ml min^{-1}. Filter paper unit was calculated by the protocol described by Ghose [75].

4.3. Results and discussion

4.3.1. Construction of recombinant S. cerevisiae strain expressing β-glucosidase

The β-glucosidase gene from plant source (*Putranjiva roxburghii*) was amplified excluding the native secretion signal. The primers were designed on the basis of gene sequence data available in Genbank with the Accession no KF006311. The β-glucosidase gene (*prgh1*) was cloned in frame with the *T. reesei xyn2* secretion signal into yXYNSEC vector under the transcriptional control of PGK1 promoter and terminator for constitutive expression. This obtained yXYNSEC-PRGH1 plasmid was used to transform *S. cerevisiae* Y294 to Uracil prototrophy (Ura^+). The *S. cerevisiae* Y294 (Ura^+) transformants showed strong luminescence around the colonies under UV light on SC^{-ura} selection plate when MUG was used as substrate, but the native *S. cerevisiae* Y294 strain did not show a clear zone (Figure 4.3). PCR with DNA from transformants confirmed the presence of yXYNSEC-PRGH1 plasmid. The FUR1 gene of transformants was disrupted to generate autoselective strains. This deletion was confirmed with PCR and the parental *S. cerevisiae* Y294 was used as reference. Therefore, the PRGH1 gene was attached to the xyn2 secretion signal sequence [133] and expressed constitutively from a high copy number yeast expression vector. The expression is under the transcriptional control of the PGK1 promoter and terminator of S. cerevisiae. The xyn2 signal was attached for successful secretion of the target enzyme to the growth medium [197].

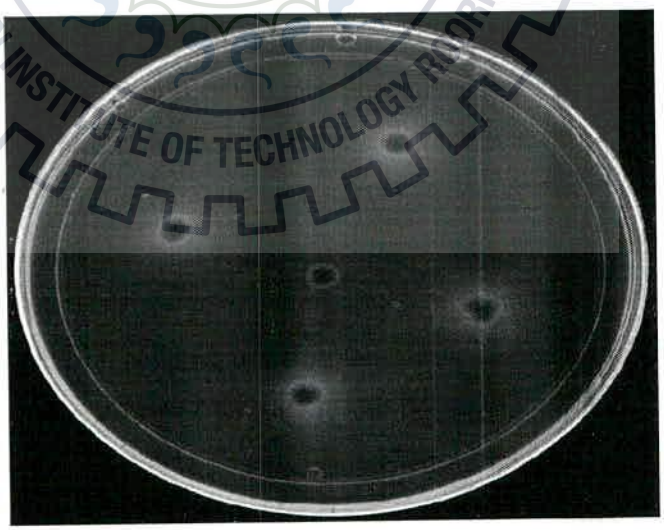


Figure 4.3. Expression of *prgh1* gene in *S. cerevisiae* Y294. *S. cerevisiae* Y294 transformants (peripheral colonies) showed the β-glucosidase activity when MUG was used as substrate. Native *S. cerevisiae* Y294 (center colony) didn't show any significance β-glucosidase activity.

4.3.2. Purification of recombinant PRGH1 expressed in S. cerevisiae

The recombinant PRGH1 was expressed constitutively under PGK1 promoter in *S. cerevisiae* Y294. The protein was secreted to extracellular fraction of culture. At 60% ammonium sulphate, most of the PRGH1 enzyme was precipitated out. These resulting precipitates were used for further purification as described in materials and method section and the recombinant PRGH1 was purified to homogeneity. The recombinant PRGH1 enzyme was purified 25.23 fold with a 54.34% yield against cellobiose as the substrate. The purification results were summarized in Table 4.1 and the SDS-PAGE pattern of each purification step was shown in figure 4.4 A. Like this enzyme, several β-glucosidases purified by using several complicated chromatographic steps [260].

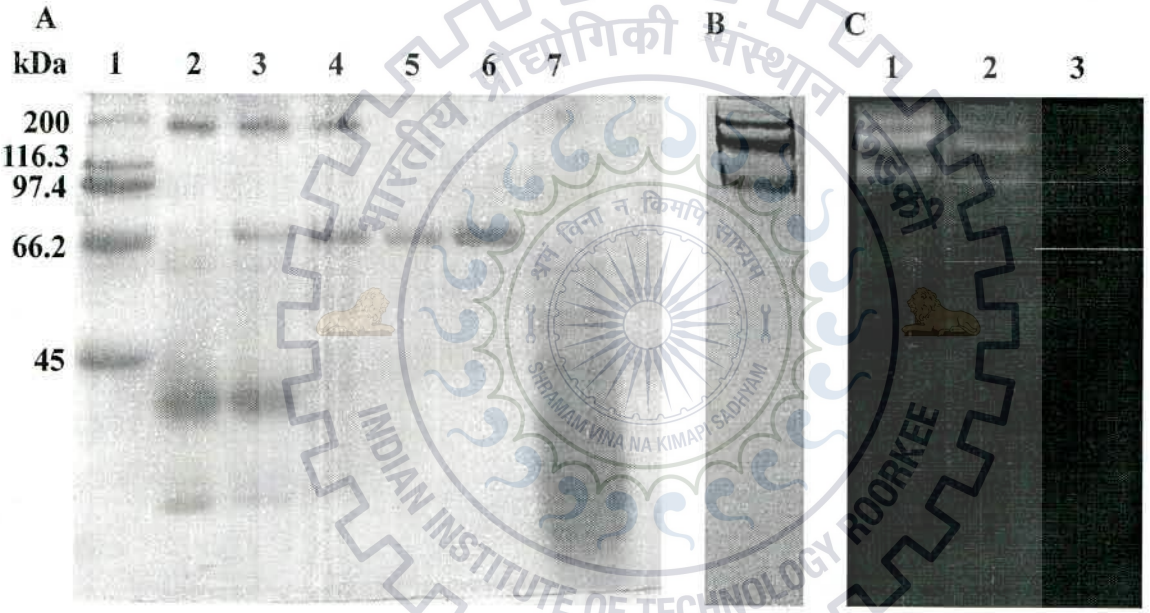


Figure 4.4 (A) SDS-PAGE (12% *w/v*) analysis of recombinant PRGH1 purification. Lane 1, molecular weight marker; Lane 2, concentrated culture supernatant of native *S. cerevisiae* Y294; Lane 3, concentrated culture supernatant of recombinant *S. cerevisiae* Y294; Lane 4, last fraction of gradient ammonium sulphate precipitation (60%); Lane 5 and Lane 6, 200 mM NaCl fraction from DEAE; Lane 7, cell lysates of recombinant *S. cerevisiae* Y294. **(B)** Electrophoretogram of the purified recombinant PRGH1 in the native page. **(C)** Zymogram of the purified recombinant PRGH1 in the native page with different fluorescence substrates. Lane 1, MUG as substrate; Lane 2, MUGA as substrate; Lane 3, concentrated culture supernatant of native *S. cerevisiae* Y294.

Table 4.1 Summary of purification of recombinant PRGH1 secreted by recombinant *S. cerevisiae* Y294.

Purification Steps	Total protein (mg)	Total activity* (U)	Specific activity * (U/mg)	Recovery rate (%)	Purification (fold)
Supernatant	148.3	194.7	1.31	100	1.0
Concentrated supernatant	117.3	179.6	1.53	92.24	1.16
Ammonium sulphate precipitation	22.8	143.6	6.29	73.75	4.80
DEAE-Sepharose chromatography	3.2	105.8	33.06	54.34	25.23

*The activity was given for the β-glucosidase only.

4.3.3. SDS-PAGE and zymogram analysis

SDS-PAGE analysis under reducing condition indicated that recombinant PRGH1 was migrated with an apparent molecular mass (Mr) of 66 kDa. However, Native-PAGE analysis showed that recombinant PRGH1 exists in higher order oligomeric states (Figure 4.4 B). Our previous study with native PRGH1 (purified from plant source) confirmed that at high concentration of this enzyme it exists in different oligomeric forms (Second chapter). After Native-PAGE, the β-glucosidase and β-galactosidase activity of the purified recombinant PRGH1 was observed with 4-methylumbelliferyl β-D-glucopyranoside (MUG) and 4-methylumbelliferyl β-D-galactopyranoside (MUGA) respectively used as substrates. The result showed that the enzyme was able to hydrolyze both the substrates, since clear bands were observed under UV 365 nm after incubation (Figure 4.4 C), but the fluorescence intensity was higher when MUG was used as substrate indicating it has more β-glucosidase activity than β-galactosidase. However, no fluorescent band developed when crude extract of native *S. cerevisiae* Y294 was used in this zymogram assay (Figure 4.4 C, lane 3).

The secreted recombinant PRGH1 has a molecular mass of ~66 kDa, which is similar to the protein sourced from the *Putranjiva roxburghii*, but a 7 kDa difference was observed when compared to the protein overexpressed in the bacterial host (Second chapter). These results clearly suggest that in the *S. cerevisiae* host the protein was able to glycosylate as a part of post translational modifications. High-level secretion of heterologous gene products in *S. cerevisiae* is mediated by the hydrophobic N-terminal extension of the polypeptide [133]. This leader peptide translocates the protein from Golgi complex to culture medium via plasma

membrane. Studies suggest that the hydrophobic XYN2 secretion was cleaved on the carboxylic side of Lys-Arg dibasic cleaving site by the KEX2 protease of *S. cerevisiae* [197].

4.3.4. Circular dichroism study

The CD spectrum between 190 and 240 nm indicated that α-helix is the principal type of secondary structure in the recombinant PRGH1 enzyme and structural integrity is well maintained throughout purification (Figure 4.4). The deconvolution of CD data with DichroWeb program indicates that the α-helical and β-sheet contents were 35% and 21% respectively. The percentage of the secondary structure elements well support our sequence based model. This result was also coinciding with CD spectrum result of plant sourced protein [183]. The CD result also suggested that the enzyme maintained its integrity during the purification process.

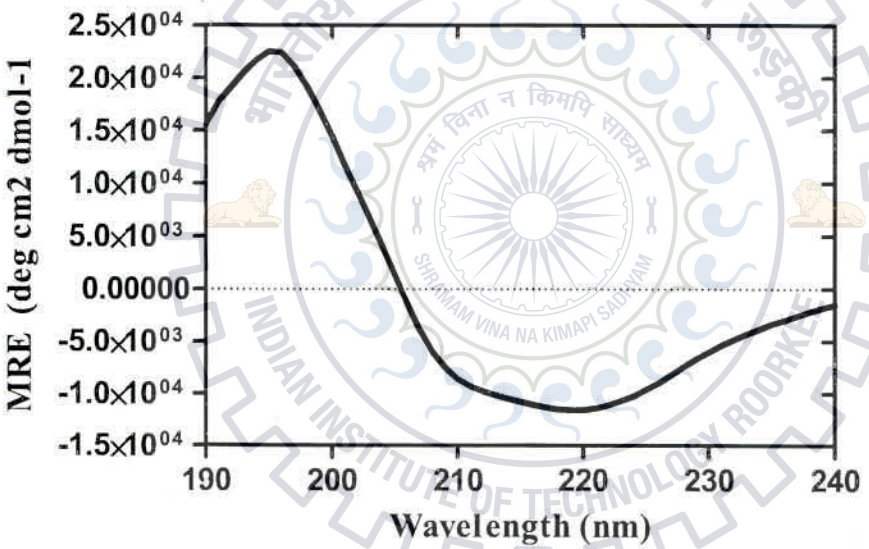


Figure 4.4. Far-UV CD spectrum (190-240 nm) of purified recombinant PRGH1 in 20mM sodium phosphate buffer (pH 8.0).

4.3.5. Substrate specificity and kinetic parameters of recombinant PRGH1

Glycosyl hydrolase family 1 enzymes show a wide range of enzyme activity, so that we have used various substrates to characterize this enzyme. Purified enzyme was used in enzyme assays. The specificities of recombinant PRGH1 against different substrates are shown in table 4.2 and 4.3. The result showed that the enzyme was active against *p*NP substrates as well as natural substrates. This enzyme was best active against *p*-nitrophenyl β-D-fucopyranoside (*p*NP-Fuc), with relative activity of 132% and also very active against β-D-

glucopyranoside (*p*NP-Glc) and *p*-nitrophenyl β-D-galactopyranoside (*p*NP-Gal) with relative activity of 100 and 80% respectively. Both *p*-nitrophenyl-β-D-cellobioside (*p*NP-Cel) and cellobiose also efficiently hydrolyzed by this enzyme. But, the recombinant PRGH1 showed more activity against *p*NP-Fuc and *p*NP-Glc (analogue of cellobiose, consisting of two glucose moiety linked by the β-1, 4-glycosidic linkage) than the *p*NP-Cel (analogue of cellotriose, consisting of three glucose moiety linked by the β-1, 4-glycosidic linkage). The recombinant PRGH1 had no significant activity on CMC, maltose, sucrose, trehalose, melibiose, nigerose, xylan, laminarin or lichenan as substrates. The enzyme could hydrolyze the cellooligosaccharides, but the activity decreased gradually as the chain length increased. Salicin was hydrolyzed efficiently by this enzyme.

The hydrolysis rate at different concentrations of various *p*NP substrates confirmed that the recombinant enzyme follows Michaelis-Menten kinetics. The Km and Vmax values for different *p*NP substrates and cellobiose obtained from Lineweaver-Burk plot and the results are shown in table 4.4. The values were obtained under standard assay conditions. The K_{cat}/K_m values clearly demonstrate that the enzyme has slightly higher affinity towards *p*NP substrates than cellobiose.

With respect to their specificities of substrates, most of the β-glucosidase characterized till date can be catagorized into three groups: (i) showed high specificity towards aryl β-glucosides (ii) known as cellobiases, hydrolyzes cello-oligosaccharides and cellobiose (iii) hydrolyzes both type of substrates [89]. Our study showed that PRGH1 is a broad specific β-glucosidase enzyme. The analysis of kinetic parameters showed that PRGH1 is much more competent at the hydrolysis of β-glucosidase and β-fucosidase substrate than the β-galactosidase substrate. This enzyme able to hydrolyzes cellobiose at a significant rate, which is most important from the commercial point of view.

Table 4.2 Relative hydrolytic activity of purified recombinant PRGH1 enzyme with various chromogenic substrates.

Substrate	Relative activity (%)
p-nitrophenyl β-D-glucopyranoside	100
p-nitrophenyl β-D-galactopyranoside	78
p-nitrophenyl β-D-fucopyranoside	132
p-nitrophenyl β-D-cellobioside	36
p-nitrophenyl β-D-xylopyranoside	3
o-nitrophenyl β-D-galactopyranoside	< 0.001
p-nitrophenyl β-D-manopyranoside	< 0.001
p-nitrophenyl α-D-glucopyranoside	< 0.001
p-nitrophenyl α-D-galactopyranoside	< 0.001

Table 4.3 Relative hydrolytic activity of purified recombinant PRGH1 enzyme with various disaccharides and natural substrates.

Substrate	Relative activity (%)
Cellobiose	100
Cellotriose	66
Cellotetrose	54
Cellopentaose	42
Trehalose	< 0.001
Maltose	< 0.001
Sucrose	< 0.001
Trehalose	< 0.001
Melibiose	< 0.001
Nigerose	< 0.001
Xylan	< 0.001
Laminarin	29
Lichenan	19
Amygdalin	54
Arbutin	47
Salicin	38
Gentiobiose	47
Quercetin 3-β-D-glucoside	28
Esculin	43

Table 4.4 Kinetic parameters of recombinant PRGH1 enzyme.

Substrate	K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m ($M^{-1}s^{-1}$)
*p*NP-Glc	0.52	11.73	2.26×10^4
*p*NP-Gal	0.65	7.95	1.22×10^4
*p*NP-Fuc	0.47	14.05	2.99×10^4
*p*NP-Cel	3.11	7.3	2.34×10^3
Cellobiose	1.26	7.88	6.25×10^3

4.3.6. Effect of metal ions, inhibitors and other reagents on recombinant PRGH1 activity

Most of the divalent metal ions and EDTA had no significance effect on enzyme activity. However, Zn^{2+} and Hg^{2+} ions have the inhibitory effect on β-glucosidase activity. NBS completely abolished the function of the enzyme and a significance loss in activity was observed in case of δ-gluconolactone. The addition of SDS resulted in a significant improvement on the activity, whereas DMSO and Triton X-100 resulted in slight activity enhancement. The effect of lower chain alcohols was studied by incubating the enzyme with various concentrations of alcohols. The result showed that the enzyme was very stable against ethanol and 1-propanol and at low concentration methanol showed stimulatory effect (Table 4.5). In the presence of the metal ions the activity of the enzyme remains unaltered except some. This seems to be an advantage for the industrial process as the presence of various metal ions in the natural feedstocks can't hinder the activity of the enzyme. In the presence of reducing agent the activity of the enzyme didn't decreased significantly indicating that Cys don't play any role in the catalysis. Two well established inhibitors i.e. N-bromosuccinimide and δ-gluconolactone inhibited the activity of PRGH1, indicates that this enzyme is a β-glucosidase rather than an exoglucanase. Different chemical reagents altered the activity of the enzyme. This enzyme showed resistance towards ethanol. DMSO and 10% ethanol enhanced PRGH1 activity. Some other study also reported that alcohols act as an activator owing their glycosyltransferase activity [101, 194].

Table 4.5. Effects of metal ions, chelating agent, chemical agents and organic solvents on the enzyme activity of recombinant PRGH1 enzyme.

Effectors	Concentration	Relative activity (%)
No additive	-	100 ± 4.5
Metal ions		
Mg^{2+}	1mM	98.5 ± 5.1
Mn^{2+}	1mM	103.7 ± 4.9
Ca^{2+}	1mM	101.3 ± 6.1
Ni^{2+}	1mM	90.0 ± 3.1
Co^{2+}	1mM	97.1 ± 2.8
Zn^{2+}	1mM	38.5 ± 3.5
Ba^{2+}	1mM	102.5 ± 3.7
Hg^{2+}	1mM	48.8 ± 7.6
Fe^{2+}	1mM	115.2 ± 1.5
Pb^{2+}	1mM	107.7 ± 2.1
Surfactants		
SDS	1mg ml^{-1}	117.6 ± 5.8
Triton X-100	10mg ml^{-1}	110.0 ± 1.3
Chelating agent		
EDTA	10mM	100.2 ± 4.9
Chemical agents		
β-mercaptoethanol	1mM	94.3 ± 4.7
DTT	1mM	96 .2± 3.8
Urea	5mM	98.4 ±3.3
Hydrochloride guanidine	10mM	91.7 ± 6.2
DMSO	10% (v/v)	114.1 ± 3.7
Inhibitors		
N-bromosuccinimide	1mM	19.9 ± 5.3
δ-gluconolactone	5mM	53.0 ± 3.7
Organic solvents		
Ethanol	10% (v/v)	99.4 ± 4.4
Ethanol	25% (v/v)	52.3 ± 2.7
1-propanol	10% (v/v)	93.7 ± 1.8
1-propanol	25% (v/v)	40.4 ± 4.3
Methanol	5% (v/v)	119.2 ± 3.8
Methanol	10% (v/v)	103 ± 5.7

4.3.7. Effect of pH and temperature on enzyme activity

The optimum pH for the cellobiose activity of recombinant PRGH1 was studied over a pH range of 3.0 to 11.0 at 65 °C. As shown in figure, the enzyme retained most of its original activity at pH 5.0-6.0 with optimum at pH 5.0. The effects of various pHs on the stability of recombinant PRGH1 were checked (Figure 4.5 B). After incubating the enzyme at different pH for 1h the residual activity was measured. The enzyme was quite stable over a broad pH range of 4.0 to 8.0; retained 70% of original catalytic activity against cellobiose but below pH 4.0, the activity declined quickly. Under optimal pH condition, the purified recombinant PRGH1 had an optimum temperature of 65 °C. The effect of temperature on enzyme activity and stability are shown in Figure 4.6. Thermostability experiments showed that the enzyme was fairly stable over a broad temperature range of 20-70 °C, and the enzyme activity reduced above 70 °C. Unlike to this enzyme several cold active β-glucosidases were reported earlier [15, 52, 251, 252].

Result showed that it was active in wide pH and temperature range. Thermostability is a significant character of β-glucosidase during the enzymatic conversion of cellobiose to reducing sugars. Generally, during saccharification step steam is applied on the biomass to make it easier for degradation. A thermostable enzyme like PRGH1 along with other themostable enzymes can be used after the heating step avoiding the pre-cooling process, hence reducing processing time, decreasing contamination, saves energy, and enhancing the quality and yield of fermentation [150].

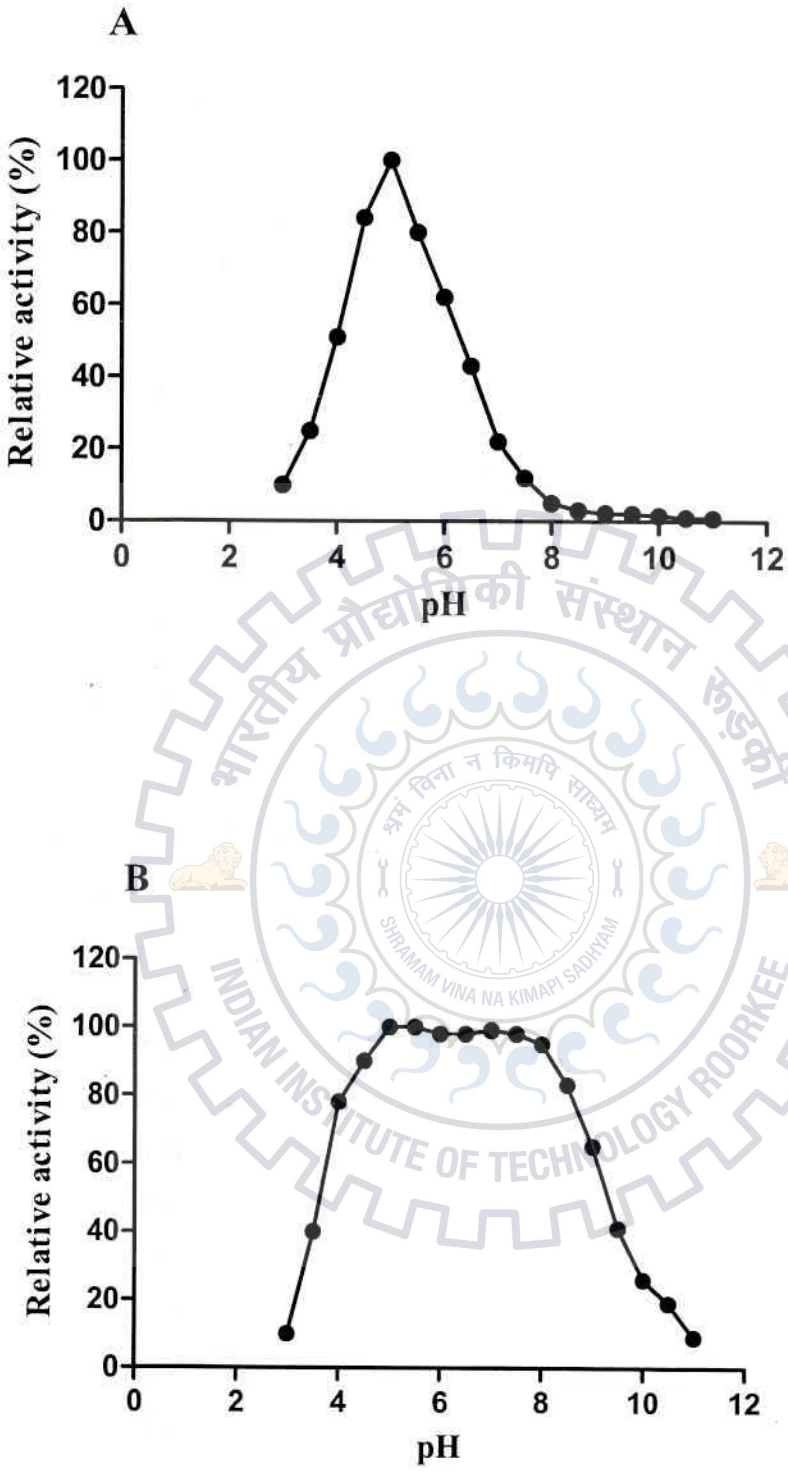


Figure 4.5. Optimal pH (A) and pH stability (B) of purified recombinant PRGH1 enzyme. The relative activities of β-glucosidase against cellobiose are expressed as percentage normalized to the sample with the highest activity in each test.

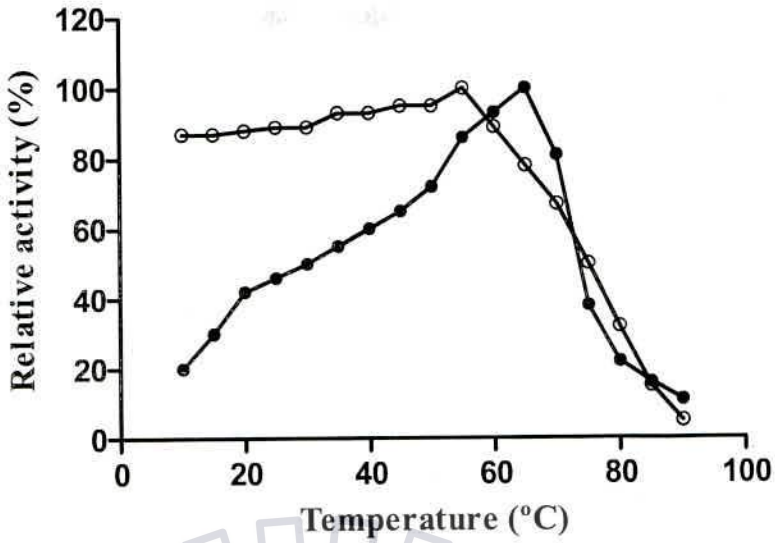


Figure 4.6. The effect of temperature on the purified recombinant PRGH1 activity (solid circles) and stability (open circles). For stability assay, the purified enzymes were incubated at different temperatures at pH 5.0 for 1 hour and residual activity was measured.

4.3.8. Fermentation of cellobiose by recombinant S. cerevisiae

Under aerobic condition, flask fermentation was carried out to investigate the ability of recombinant *S. cerevisiae* Y294 to produce ethanol from the minimal media containing 2% (w/v) cellobiose as a sole carbon source. In this culture condition, the growth profile of both native and recombinant *S. cerevisiae* Y294 was compared and the result confirmed that the recombinant *S. cerevisiae* Y294 culture showed a better growth profile in cell density (Figure 4.7.). Figure 4.8 and 4.9 demonstrates the rate of cellobiose consumption, and subsequently ethanol production respectively. The concentrations of glucose decreased with time and consequently the ethanol concentration increased. In case of recombinant *S. cerevisiae* Y294 culture, the highest ethanol concentration was 4.05 g/L, which was 2.9 fold higher when compared to native *S. cerevisiae* Y294 (used as control), grown in the same condition. Total cellobiose consumption was observed for the recombinant strain after 36 h, with consumption rate of 0.55 g cellobiose/L h, yielding 0.20 g ethanol/g cellobiose. β-glucosidase are strongly inhibited by glucose which is the end product of cellobiose hydrolysis. In this cellobiose fermentation experiment, no significant amount of glucose accumulated during fermentation suggesting the efficient conversion of glucose by the yeast.

The growth curve suggest that the recombinant *S. cerevisiae* Y294 was able to grow in cellobiose medium owing to the overexpression of β-glucosidase gene. *S. cerevisiae* have

very fast carbon flux in the glycolysis process, when glucose was used as the carbon source. Cellobiose, which is a dimer of glucose, was predicted to be used in similar manner as glucose by the host. But, the utilization of cellobiose seems to be slower than the glucose. It was also found that in the cellobiose medium the percentage of higher metabolites of glycolysis were at elevated level [116]. In shale flask culture, the recombinant *S. cerevisiae* Y294 was able to ferment the cellobiose more efficiently. This result suggested the possible use of this new construct in SSF process. *S. cerevisiae* is well characterized previously, where large scale bioreactor were scaled up from shake flask [28].

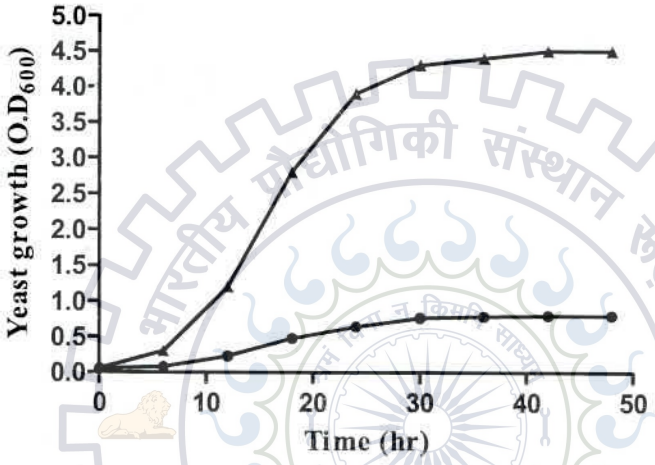


Figure 4.7. The growth profile of native (circle) and recombinant (triangle) *S. cerevisiae* Y294 on defined medium with cellobiose as sole carbon source.

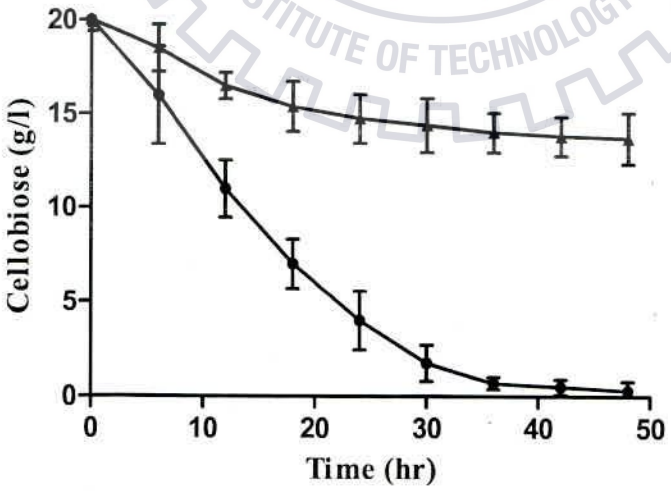


Figure 4.8. Comparison of cellobiose consumption between native (circle) and recombinant (triangle) *S. cerevisiae* Y294. Results are the average of triplicate and the bars indicate the standard error.

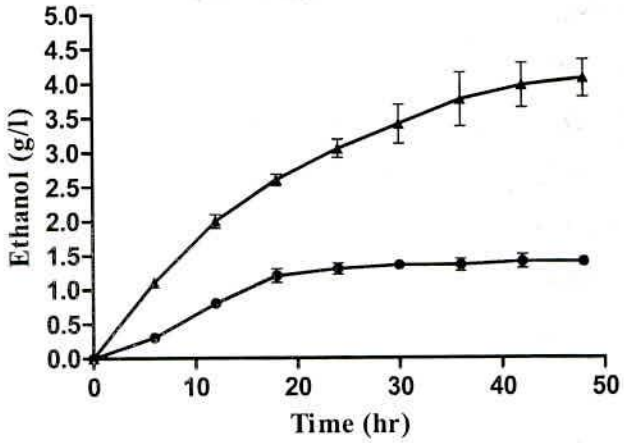


Figure 4.9. Comparison and time course of ethanol production between native (circle) and recombinant (triangle) *S. cerevisiae* Y294. Results are the average of triplicate and the bars indicate the standard error.

4.3.9. Simultaneous saccharification and fermentation of cellulosic substrates with the recombinant S. cerevisiae Y294

In the current SSF experiments, both treated and untreated different cellulosic biomasses were applied to 12 h enzymatic prehydrolysis, after which the concentrations of cellobiose and glucose were evaluated. NaOH pre-treatment was found to be more effective for cellulase activity in case of rice straw and sugarcane bagasse, while cellulase activity was independent of this alkali treatment when CMC was used as substrate (Table 4.6).

Table 4.6. Concentration of cellobiose and glucose after enzymatic prehydrolysis of both untreated and alkali treated cellulosic substrates. Temperature: 45 °C; agitation speed: 100 rpm; time duration: 24 hrs.

Cellulosic substrate	Cellobiose (g L^{-1})	Glucose (g L^{-1})
CMC	0.11 ± 0.02	23.39 ± 0.90
Untreated rice straw	7.61 ± 0.25	31.68 ± 2.83
Alkali treated rice straw	13.78 ± 0.47	49.20 ± 3.41
Untreated sugarcane bagasse	12.62 ± 0.68	48.93 ± 3.65
Alkali treated sugarcane bagasse	22.58 ± 0.21	66.29 ± 2.56

Commercial cellulases are usually used for converting cellulose to cellobiose or other oligosaccharides. Despite the fact that these cellulases possesses high exo-and endo-glucanase activity, but show very low β-glucosidase activity. Since, prehydrolysis of rice straw and sugarcane bagesse generate glucose and significant amount of cellobiose, recombinant *S. cerevisiae* Y294 was used for efficient consumption of cellobiose and efficient bioethanol production. Cell density, concentrations of glucose and ethanol were used to estimate the efficiency of SSF. The recombinant *S. cerevisiae* Y294 produced more ethanol from the cellulosic biomass (of rice straw/sugarcane bagasse) as compared to the native *S. cerevisiae* Y294. Pre-treatment of the biomasses had significance influence on ethanol production was due to high concentration of glucose in fermenting medium. As expected, sugarcane bagasse was the best substrate for SSF as the pre-treatment and hydrolysis process synergistically produced more fermentable sugar and subsequently more production of ethanol (Table 4.7). The time course of ethanol production from CMC and alkali treated feedstocks were shown in Figure 4.9.

Table 4.7. Remaining concentration of cellobiose and glucose and ethanol production at the end of SSF process.

Cellulosic substrate	Cellobiose (g L^{-1})	Glucose (g L^{-1})	Ethanol (g L^{-1})
CMC	0.00	0.28	13.69 ± 1.43
Alkali treated rice straw	0.00	0.39	32.48 ± 2.64
Alkali treated sugarcane bagasse	0.00	0.46	53.95 ± 2.72

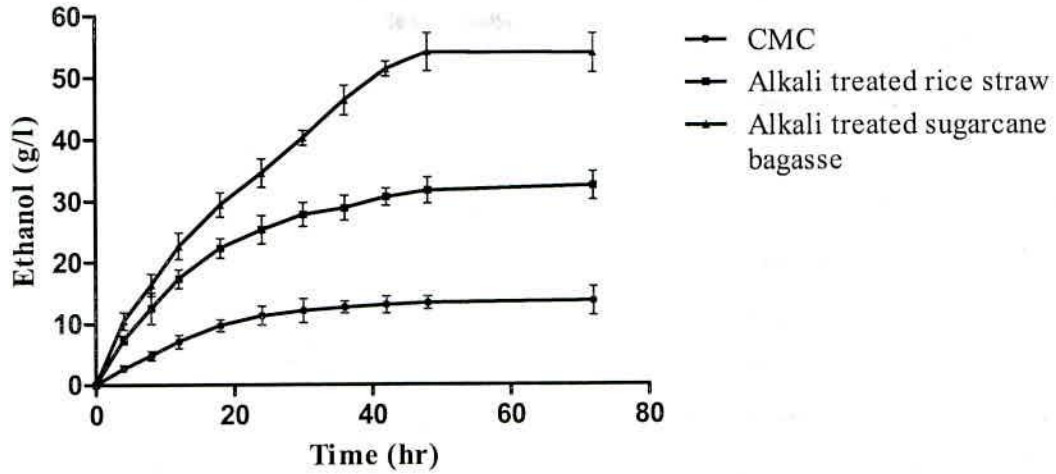


Figure 4.10. Time course of second generation ethanol production from different cellulosic substrates in shaken flask. Results are the average of triplicate and the bars indicate the standard error.

For all of the cellulosic substrates, the cellobiose was completely converted into the glucose after 40 h of inoculation, indicating the efficiency of β-glucosidase. The overall statistics of cellobiose consumption and ethanol production is shown in Table 4.7. Alkali treated sugarcane bagasse was found to be the best substrate for the SSF. Figure 4.11 showed the time course of cellobiose and glucose consumption and the production of ethanol in case of the SSF using sugarcane bagasse.

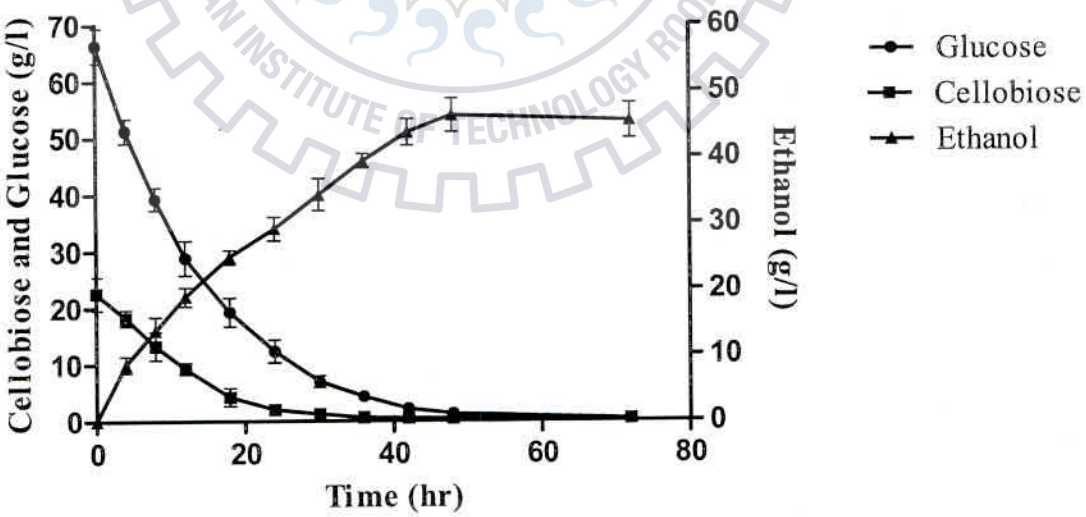


Figure 4.10. SSF of sugarcane bagasse by the recombinant *S. cerevisiae* Y294 harbouring *Putranjiva roxburghii* β-glucosidase gene for second generation ethanol production in shaken flask. Results are the average of triplicate and the bars indicate the standard error.

In the current SSF process we have used commercial cellulose enzyme for the conversion of cellulose to cellooligosaccharides and cellobiose. Though this cellulose enzyme has significant endo- and exo-glucanase activity but has very low β-glucosidase activity. This is the bottle neck of the whole process. Hence, it is necessary to add commercial β-glucosidase or transform β-glucosidase gene into the fermenting host. For this reason, we have successfully expressed a plant β-glucosidase gene in the *S. cerevisiae* Y294. The result showed that almost all cellobiose and glucose were consumed. The efficiency of the production of ethanol can be improved by employing anaerobic system, immobilizing the genes in the yeast cell surface and increasing the copy number of *prgh1* gene in the host cell. Prehydrolysis of cellulosic materials has generated glucose as well as cellobiose. Cellobiose was consumed during SSF process. This confirmed that β-glucosidase enzyme is relatively resistance to the presence of glucose, which is unlike for a β-glucosidase as most of the β-glucosidase enzyme is hindered by the presence of glucose [179]. The difference in glucose production between these three cellulosic substrates was due to the architecture of substrates [100, 218, 219]. Overall, the SSF result point out that, the recombinant *S. cerevisiae* Y294 strain expressing the β-glucosidase gene has great potential for bioethanol production from the natural wastes by the SSF process.

4.4. Conclusion

The hydrolysis of cellulosic biomass and its effective conversion to bio-ethanol by using engineered microorganisms is a very potent and cost effective strategy. The hydrolysis of cellobiose and cellooligosaccharides to glucose by β-glucosidase enzyme is the key step towards the saccharification of cellulose. The β-glucosidase enzyme relieves the feedback inhibition on cellobiohydrolase and endoglucanases imposed by cellobiose. We successfully expressed a plant β-glucosidase gene (PRGH1) in the *S. cerevisiae* Y294 and studied the properties of the enzyme. The optimal pH and temperature of the enzyme activity was found to be 5.0 and 65 °C respectively indicating this enzyme is a thermostable enzyme with preference towards moderate acidic condition. The enzyme showed broad substrate specificity and able to hydrolyze cellobiose significantly. The enzyme showed resistance towards alcohols, suggesting the enzyme can be used in fermentation industry more efficiently. The recombinant *S. cerevisiae* Y294 harbouring *prgh1* gene showed better growth profile, cellobiose consumption and ethanol production. In addition to this, complementing with commercial cellulose enzyme the recombinant *S. cerevisiae* Y294 was used in SSF experiment using CMC, rice straw, sugarcane bagesse as sole carbon source. Although different β-glucosidase genes were used to improve the efficiency of SSF, this is the first report where a plant thermostable efficient β-glucosidase gene was used for SSF process. Using complex cellulosic substrates significant amount of second generation ethanol was produced in shaken flask fermentation process. The yield of the bioethanol can be improved by scaling up the process for bioreactors. The study demonstrated the feasibility of using the β-glucosidase gene to enhance the second generation cellulosic ethanol production. In addition to this the enzyme shows better activity when compared to the rPRGH1 (produced by using *E. coli*), although the order of substrate preferences remain same.

CHAPTER 5

CHARACTERIZATION OF A HAD-FAMILY PHOSPHATASE ENZYME FROM *STAPHYLOCOCCUS LUGDUNENSIS*

5. CHARACTERIZATION OF A HAD-FAMILY PHOSPHATASE ENZYME FROM *STAPHYLOCOCCUS LUGDUNENSIS*

5.1. Introduction

The haloalkanoic acid dehalogenase superfamily (HADSF) named after its first archeal enzyme member; 2-haloacid dehalogenase is of enormous size with its members present in all the living organisms [102]. Dehalogenases are the enzymes used for detoxification of the halogenated compounds. They hydrolyse the carbon-halogen bond using water for the nucleophilic substitution and constitute an important reservoir for detoxifying these halogenated pollutants from the environment. HADSF constitute different enzymes like dehalogenase, phosphatase, phosphonatase, β-phosphoglucomutase and ATPases. Phosphatase forming the majority in HADSF can vary in figure from 30 in prokaryotes to 200-300 in the eukaryotes [3]. They help in phosphorous transfer reactions by transferring the phosphate group to an Asp as an active site residue. Numerous HAD superfamily phosphatases share the same function while their origin is from different species, however some members act differently [102].

HADSF phosphatases have a typical Rossmann fold with a central β-sheet. Like other Rossmann folds, the active site is formed by the C-terminus end residues of the central β-sheet. The location of the substrate binding site is also common among the enzymes having this Rossmann fold and is present in a loop between strand S1 and the helix next to it. However HADSF has two distinct structural motifs from Rossmann fold family enzymes called as Squiggle and a Flap and are used for acquiring different conformations during catalysis [3, 24, 153].

There is very less homology between the family members but still they have four conserved motifs with five catalytic residues which form the active site [230]. These motifs in the Rossmann fold form a common catalytic core and also stabilize the reaction phospho-intermediates with the help of electrostatic interactions. HADSF members have a shared reaction mechanism with aspartate present in Motif 1 used as a nucleophile [4]. They require Mg2+ ion as a cofactor for its activity and forms an aspartyl-phosphate intermediate during the reaction. This phosphoryl group is then transferred to a water molecule resulting in the formation of an alcohol [153].

HAD superfamily members have a broad substrate specificity range due to the presence of an additional region called Cap domain which provides extra surface for interacting with the substrates. HAD members can be divided into three subfamilies based on the location and nature of fold of Cap domain. Subfamily I i.e. C1 is present between the

motif I and II, while subfamily II i.e. C2 is present between the motif II and III. Subfamily III are called as C0 as they have no cap domain. The rearrangement of the cap domain results in two different conformations called the open and the closed states with solvent exposure in the open state and exclusion in the closed state [24].

Staphylococcus lugdunensis, first described in 1988 by Freney et al, is coagulase-negative staphylococci. Lugdunensis species has been named after the French city, called Lyon where it first originated [71]. It occurs naturally as part of human skin flora but can cause various serious infections like osteomyelitis, septicaemia, peritonitis, arthritis and violent endocarditis [8, 146, 198]. Unlike other coagulase negative species, it can also cause a serious infection after cataract surgery called endopthalmitis. Major infections are found on skin or in blood stream, soft tissues and prosthetic devices. This chapter discusses cloning, expression and characterization of a HAD family hydrolase enzyme of *Staphylococcus lugdunensis* (SHFD) and it's *in silico* studies to construct its three dimensional model and trace its evolution. Crystallization studies of SHFD were also done using the Hampton crystal screens series and also the manual prepared solution. SHFD belongs to subfamily IIB of HADSF and possesses phosphatase activity.

5.2. Materials and Methods

5.2.1. Chemicals and Enzymes

Isopropyl β-D-1-thiogalactopyranoside (IPTG), Phenylmethylsulphonyl fluoride (PMSF), Inhibitor cocktail, Nickel NTA affinity bead and Imidazole were purchased from Sigma-Aldrich. Gel filtration protein marker was purchased from Biorad. dNTPs mix, Restriction enzymes (NheI and XhoI), TaqDNA polymerase and Ligase were purchased from NEB.

5.2.2. Cloning of SHFD gene in expression vector

SHFD gene was amplified from genomic DNA of *S. lugdunensis* using PCR. Gene specific primers flanked with restriction sites were used for the amplification primers. The forward primer having NheI restriction site and the reverse primer having XhoI site were designed to code for 280 amino acids. The primers used are as following:

SHFD Forward-5'-CAGCGCTAGCATTAAAGCAATATTTTTAGATATGG-3'
SHFD Reverse-5'-CGCACTCGAGTCAAAATCTATCTGCTAAATAAC-3'
PCR condition used for the amplification were initial denaturation for 4 minutes at 94 °C, followed by 30 cycles as per 94 °C for 30 seconds / 58°C for 1 minute/ 72 °C for 1 minute trailed with a final extension cycle at 72 °C for 10 minutes. PCR product was analysed on 1% agarose gel and sharp expected bands were purified from gel using Invitrogen gel extraction kit. The purified fragment and the vector, pET-28c(+) were digested with restriction enzymes NheI and XhoI and digested DNA were again gel purified. Ligation was done using the digested gene and the vector. The ligation product was transformed into DH5α host cells and the positive clone was screened using plasmid DNA as a template for PCR with gene specific primers and also by restriction digestion. Cloning was further confirmed by sequencing using T7 promoter and T7 terminator as universal primers.

5.2.3. Overexpression and purification of recombinant SHFD

The recombinant plasmid pET-28c(+) having SHFD insert was transformed into *E. coli* Rosetta cells for expression of SHFD protein in the presence of antibiotics i.e. kanamycin (30 μg/ml) and chloroamphenicol (35 μg/ml). In order to optimize SHFD overexpression, different concentrations of IPTG (0.1-1.0 mM) with different temperatures (16-37 °C) and different time intervals were used. Optimized overexpression condition was found to be 16 °C with 200 μM IPTG. Rosetta cells harbouring pET-28c(+)-SHFD were harvested by

centrifuging at 5,000 rpm for 10 min at 4 °C and cell pellets were stored at -80 °C until further processing.

In the course of purification of SHFD, frozen cell pellets were thawed on ice and resuspended in 50 mM Tris-HCl; pH 8.0 buffer having 400 mM NaCl, 10mM imidazole and 5% glycerol(v/v). Lysozyme (0.5 mg/ml), Phenylmethylsulphonyl fluoride (0.3 mM) and protease inhibitor cocktail were also added after resuspension of harvested frozen cells. Cells were lysed on ice using sonicator with a pulse of 10 seconds on and 30 seconds off for 12 times and lysate was clarified by centrifuging at 14,000 rpm for 60 minutes at 4 °C and supernatant was loaded onto Ni-NTA column pre-equilibrated with the above lysis buffer. Protein was eluted using an imidazole gradient of 50 mM and 250 mM and eluted fractions were analysed on 12% SDS-PAGE. The fractions having the pure protein were pooled together for dialysis against buffer containing 25 mM Tris-HCl; pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM DTT. N-terminal His-tag on the purified SHFD protein was also removed along the dialysis by subjecting it to TEV protease digestion at 4 °C for 16 hours. Reverse Ni-NTA chromatography was done in order to remove the uncleaved SHFD, cleaved His-tags and the His-tagged TEV protease. The flow through containing the untagged SHFD was collected for further use. The flow through and the eluted fractions were also run on 12% SDS gel for confirmation of His-tag cleavage.

5.2.4. Gel filtration chromatography

For further purity and oligomeric state determination, the concentrated untagged SHFD 3mg/ml) was loaded on to the HiLoad 16/600 Superdex 200 pg size gel filtration column (GE Healthcare) equilibrated with running buffer (25 mM Tris-HCl; pH 7.4, 150 mM NaCl, and 5% glycerol). Flow rate of 0.5 ml min^{-1} was used for protein elution. Proteins were collected in 2 ml fraction and homogeneity of each fraction was analyzed on 12% SDS PAGE. Standard marker of Bio-Rad gel filtration containing bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa) and vitamin B12 (1.35 kDa) was applied to the column under similar conditions in order to generate a standard curve for determination of the accurate molecular mass of SHFD.

5.2.5. Circular dichroism spectroscopy

Circular dichroism (CD) study was carried out on a Chirascan Circular Dichroism spectromemter (Applied Photophysics Ltd., Surrey KT22 7PB, United Kingdom). Far-UV CD spectra (190-260 nm), at 25 °C were recorded in 1 mm path length quartz cell with band width

of 1 nm and time per point was 0.5 s. The purified protein (4 μM) was buffer exchanged into 20 mM sodium phosphate buffer (pH 7.4) for use in CD experiments. Three consecutive scans were collected in order to reduce the random error. The buffer baseline was subtracted from the protein spectra in order to obtain the final spectra. The secondary structure predictions were done using DICHROWEB program [152]. The results of CD measurements were expressed as mean residue ellipticity (MRE) in deg cm^2 $dmol^{-1}$.

5.2.6. Kinetic studies

Phosphatase activity of SHFD was checked using p-nitrophenyl phosphate (*p*NPP) as substrate. All the reactions were carried out thrice and mean was taken as the data point. The assay mixture consisting of 15 mM $MgCl_2$, 20 μg diluted SHFD enzyme in 50 mM Sodium acetate buffer (pH 5.0) was kept for incubation at 25 °C for 5 minutes. The substrate i.e. 10 mM *p*NPP was added to initiate the reaction and after 5 minutes incubation at 25 °C, the reaction was terminated using 0.2 M NaOH (1 ml). The eppendorfs containing the reaction mixture were centrifuged and supernatant was used to measure the liberated *p*-nitrophenol at 405 nm using Beckman UV-Vis spectrophotometer. In addition, enzymatic activity was also checked in the presence of EDTA (15 mM) using the same assay conditions as mentioned. One unit of phosphatase activity was termed as amount of SHFD enzyme needed to convert 1 millimole of the substrate pNPP to its product i.e. *p*-nitrophenol per second under the above mentioned experimental conditions. In order to calculate the K_m and V_{max}, different concentrations of the substrate pNPP (0 to 20 mM) were used for assay for a particular enzyme concentration (20 μg diluted SHFD).

5.2.7. Effect of temperature and pH on enzyme activity

The optimum temperature of SHFD was determined by the activity on p-nitrophenyl phosphate (*p*NPP) as substrate. The enzymatic activity was assayed by incubating the reaction mixtures at different temperature ranging from 10 °C to 60 °C with 5 °C increments. The thermostability was determined by measuring the residual activity of SHFD after incubating the enzyme at different temperatures (10 °C-60 °C) in 50 mM Sodium acetate buffer; pH 5.0 for 1h. For optimal pH determination, SHFD was assayed in 10 mM Britton-Robinson buffer, with pH values ranging from 3.0 to 10.0 at 25 °C using p-Nitrophenyl phosphate (*p*NPP) as substrate. The pH stability was determined by calculating the residual enzyme activity after incubation at 25 °C for 1 h in suitable buffers.

5.2.8 Crystallization of SHFD

SHFD crystallization experiments were performed using the sitting-drop vapor diffusion method with a series of crystal screens (Hampton Research, USA). For this 1 µl of untagged SHFD protein (10 mg ml^{-1}) and 1 µl of reservoir solution, were placed into the wells and mixed well as a drop without bubble formation. Further, the plates were covered with sticky tapes to protect from dryness, dust and microbial contamination. The crystal plates were placed at 4 °C and 20 °C in a vibration free chamber for crystal development and were checked for crystal formation at the regular intervals of 5 days using a microscope (Leica, Singapore).

5.2.9. Multiple sequence alignment

SHFD amino acid sequence retrieved from NCBI was used as a FASTA sequence for BLASTp homology search. Sequences showing significant similarities to SHFD hydrolase enzyme were aligned using ClustalW [227] taking default parameters. This alignment was enriched with its secondary structure using ESPript [195].

5.2.10. Phylogenetic analysis

MEGA version 6 was used to construct phylogenetic tree from the amino acid alignments using the Maximum Likelihood method based on the JTT matrix-based model [225]. The bootstrap statistical analysis (1,000 replications) was used to test the reliability of the branching tree.

5.2.11 Structural modeling

The three dimensional model of SHFD was constructed with the help of Modeller9v10 program using the crystal structure of a sugar phosphate phosphatase BT4131 from the bacterium *Bacteroides thetaiotaomicron VPI-5482* (PDB Id: 1YMQ) as a template. The model structures were visualized by PyMol [45]. The best model was analysed on the basis of its stereochemistry using PROCHECK [140].

5.3. Results and discussion

5.3.1. Cloning of SHFD gene

Genomic DNA of *S. lugdunensis* has been used for the amplification of SHFD gene using gene specific forward and reverse primers. PCR product of 840 bp encoding 280 amino acids was obtained as shown below in figure 5.1 A. SHFD gene was cloned in pET-28c(+) with the help of sticky end ligation created by restriction digestion. The colonies obtained through antibiotic resistance were screened with the help of PCR and restriction digestion. Both PCR and restriction digestion show the presence of an appropriate band at 840 bp as shown in figure 5.2 B.

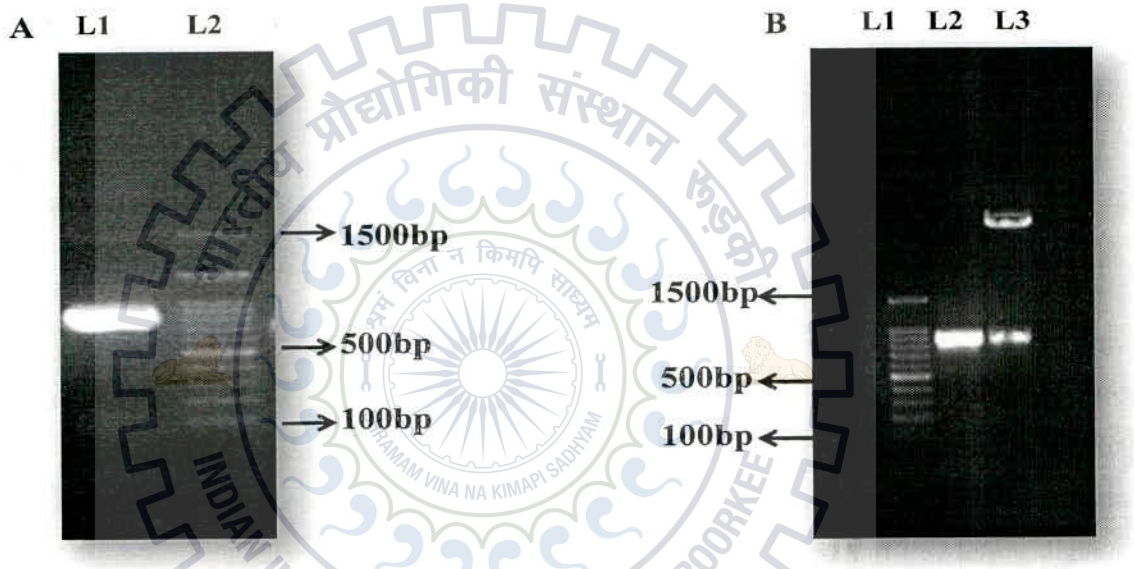


Figure 5.1. (A) PCR amplification of SHFD from the genomic DNA of *Staphylococcus lugdunensis*. L1, DNA amplification product; L2, 100 bp molecular weight marker **(B)** Cloning of SHFD. L1, 100 bp molecular weight marker; L2, PCR amplification of the one of the positive clone; L3, NheI and XhoI digestion of the same positive clone.

5.3.2. Purification of SHFD

A ~34 kDa protein was expressed in Rosetta *E.coli* strain and purified using Nickel-NTA column. Cells were lysed by sonicator and the protein was found to be in the soluble fraction after centrifugation. The supernatant with the protein was loaded onto pre-equilibrated nickel affinity column and was purified using 50 mM and 250 mM imidazole gradient. The purified protein was analysed on 12% SDS which further confirms its molecular weight i.e. ~34 kDa as shown in figure 5.4. There are several reports that HADSF phosphatase members like BT2127, BT4131, AraL possess similar molecular masses [78,

102, 154]. His-tagged SHFD was also subjected to reverse nickel chromatography to get the untagged SHFD which is used for assay and crystallization experiments.

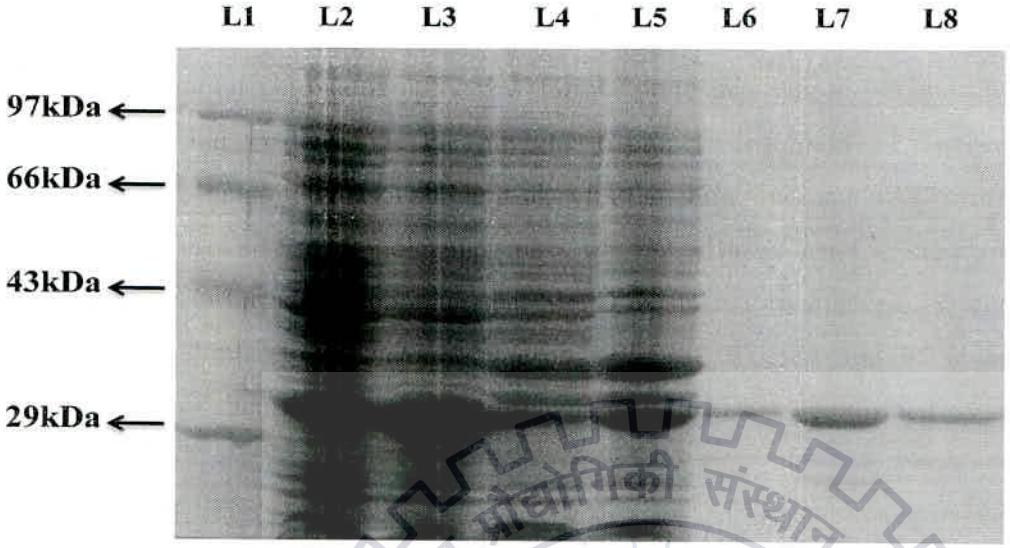


Figure 5.2. SDS-PAGE (12% *w/v*) analysis of SHFD affinity purification. L1, molecular weight marker; L2, flowthrough; L3, supernatant; L4 and L5, wash; L6; L7 and L8, purified $SHFD_{His}$ (~34kDa).

5.3.3. Gel filtration chromatography

Gel filtration chromatography was further done in order to get the pure untagged SHFD protein and to determine the oligomeric property of SHFD. Untagged SHFD on loading to the column eluted as a single peak with the elution volume of 87 ml (Figure 5.4.). This peak corresponds to a molecular mass of ~32 kDa as per the standards used for calibration. It also further corroborates that SHFD exists as a monomeric protein in solution state which is similar to *B. thetaiotaomicron VPI-5482* BT4131 and *B. thetaiotaomicron* BT1666. Both of them are monomeric proteins of ~30 kDa as confirmed by gel filtration studies [153, 154].

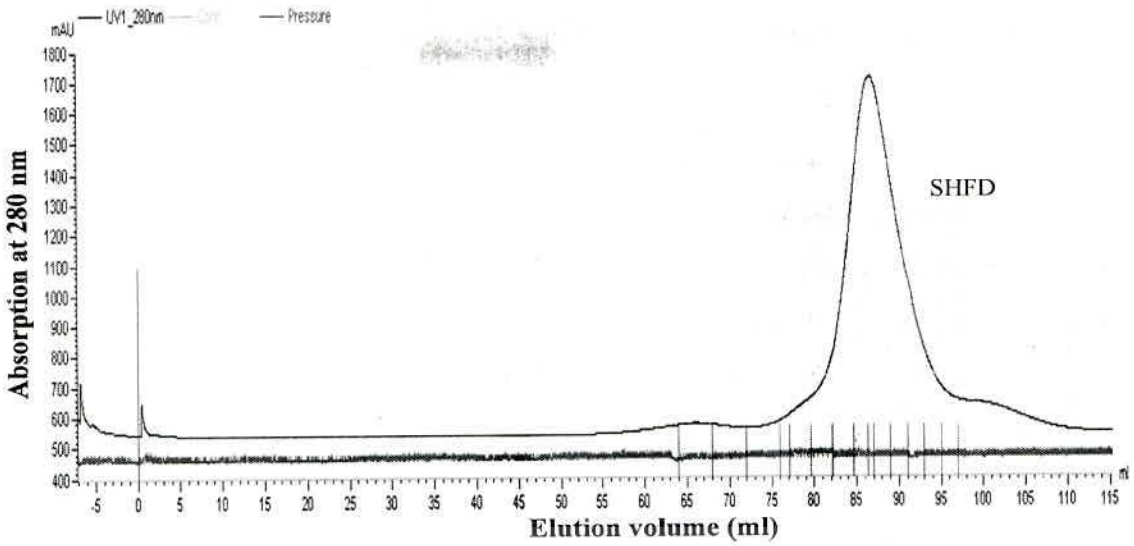


Figure 5.3. Gel filtration chromatogram showing the elution profile of SHFD on FPLC gel filtration column (HiLoad 16/60 Superdex 200).

5.3.4. Circular Dichroism study

Far-UV CD spectrum in the range of 190-240nm was used to determine the secondary structure content of SHFD in 20mM phosphate buffer (pH 7.4). Exploration of CD data showed that SHFD is an α/β protein with 37% α helix and 23% β sheets. These results are consistent with the already reported HAD family hydrolase crystal structures [153, 154, 246].

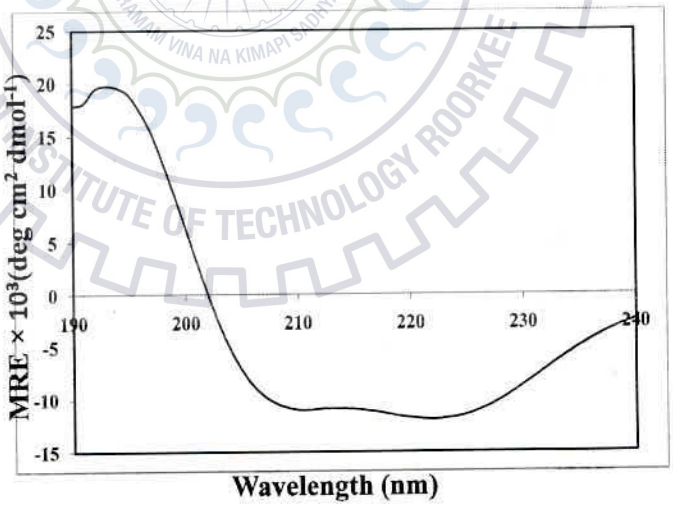


Figure 5.4. Far-UVCD spectrum (190-240 nm) of purified SHFD in 20mM sodium phosphate buffer (pH 7.4). The SHFD spectrum shows characteristics of an αβ protein.

5.3.5. Kinetic studies using pNPP as the substrate

The phosphatase activity of SHFD was measured using pNPP as the substrate. All the reactions were done in triplicate to avoid any error and further average was taken for perfection. SHFD follows Michaelis-Menten kinetics for hydrolysing its substrate pNPP at different concentrations. The K_m and V_{max} values for hydrolysing the pNPP by SHFD was found using 0 to 10 mM concentrations of pNPP from Lineweaver-Burk plot. The values obtained for K_m and V_{max} are 0.32 mM and 0.36 U/mg respectively. The k_{cat} for pNPP was found to be 21.43 ± 0.85s^{-1} and the k_{cat}/K_m value is 66.96 $mM^{-1}s^{-1}$. The kinetics parameters showed that it is a very active enzyme which can hydrolyze the phosphate linked bond efficiently. This enzyme is much faster than already reported phosphatises. The K_m and k_{cat} for pNPP in AraL phosphatase are 50.00 ± 23.32 mM and 0.012 ± 0.0006 s^{-1}. The k_{cat}/K_m of AraL towards pNPP is 0.24 $s^{-1}M^{-1}$ [78]. In addition, it was seen that magnesium ion is essential for SHFD phosphatase activity as addition of EDTA, a chelator for Mg^{2+} ion did not permit the phosphatase activity as also seen in case of AraL [78].

5.3.6. Effect of pH on enzyme activity

High intensities of the phosphatase activities were found in the pH range of 4.0 to 6.0. At pH 3.5 and 6.5 the enzyme retains 20% of its activity. The enzyme activity increases from pH 3 to 5 with peak activity at pH 5.0 and it decreases beyond pH 5.0 at 25 °C. The enzyme lost its activity in the pH range of 7.0 to 10.0. SHFD can also be called as an acid phosphatase due to its maximal activity at pH 5.0 i.e. in the acidic range like the acid phosphatase from *Vignaradiate* (mung bean) seeds [221]. This optimal pH has also been similar to the earlier isolated three acid phosphatases from Wheat Germ. These three isoenzymes have pH optima as 5.5, 4.5 and 4.0 [240]. However, it is in contrast to the neutral phosphatase AraL from *B. subtilis* which has optimum activity at pH 7.0 [78]. The pH stability study shows that this enzyme is relatively stable in the pH range of 4.0 to 9.0.

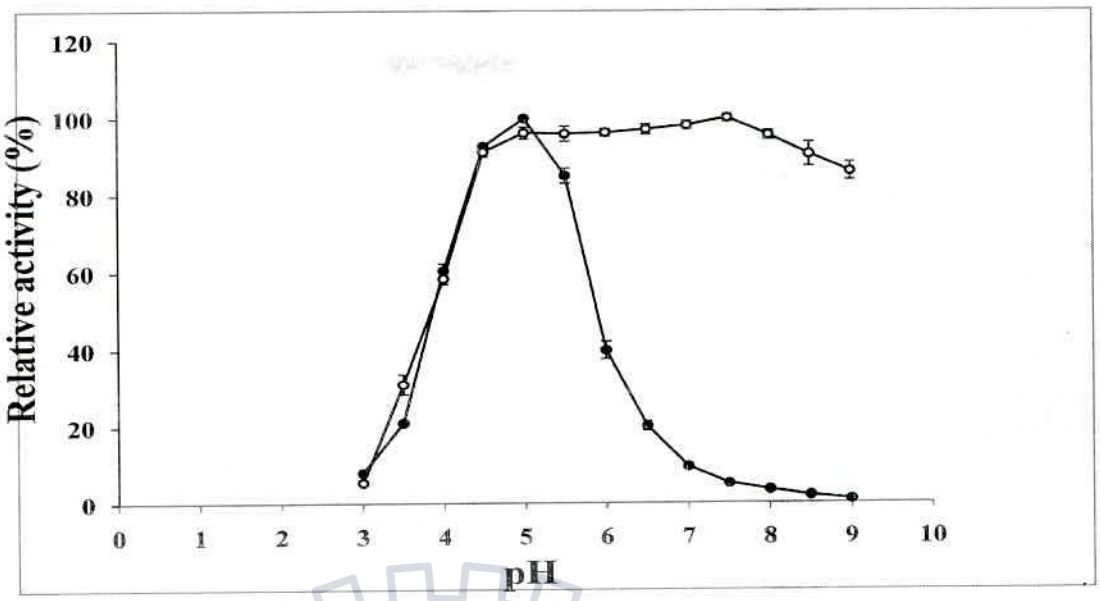


Figure 5.5. Optimal pH (solid circles) and pH stability (open circles) of purified SHFD. The relative activities of the enzyme against PNPP are expressed as percentage normalized to the sample with the highest activity in each test. Results are the average of triplicate and the bars indicate the standard error.

5.3.7. Effect of temperature on enzyme activity

Enzyme activity was noted at different temperatures ranging from 10-60 °C in Sodium acetate buffer pH 5.0. The enzyme was found to be active at broad range i.e. 20-40 °C. SHFD retains 20% residual activity at 10 °C which further increases with the temperature till 25 °C and thereafter, the enzyme activity decreased continuously and abolishes almost completely at 55 °C. The optimal temperature for enzyme activity was found to be 25 °C at pH 5.0. However, this result is in contrast to the AraL neutral phosphatase where no activity is detected at 25 °C. AraL and the mung bean acid phosphatase have significantly higher temperature optima than SHFD and are thus considered as thermolabile enzyme [78, 221]. Result clearly shows it is relatively more active in low temperatures which make it a novel acid phosphatase. Thermostability assay showed that at lower temperature the enzyme is stable.

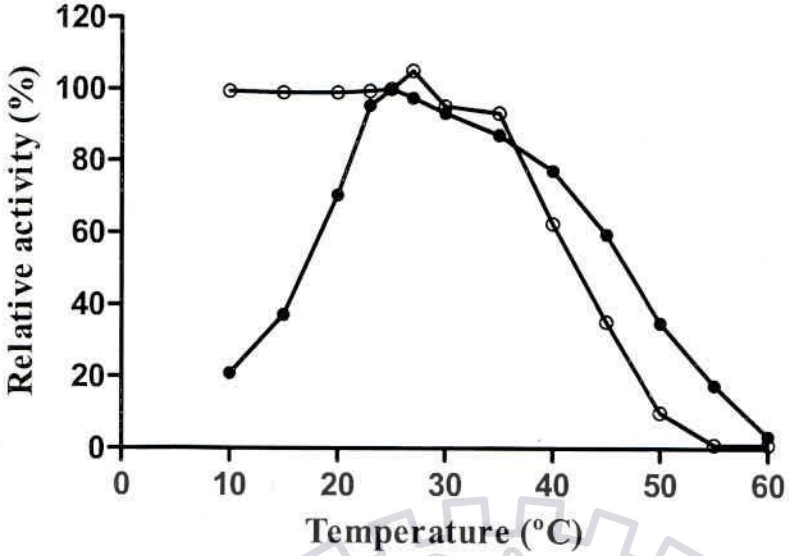


Figure 5.6. The effect of temperature on the purified SHFD activity (solid circles) and stability (open circles). Results are the average of triplicate and the bars indicate the standard error. For stability assay, the purified enzymes were incubated at different temperatures at pH 5.0 for 20min and residual activity was measured.

5.3.8. X-ray crystallography study

To get the crystal of SHFD, a number of the Hampton series crystal screens and manually made buffered salts were used. The crystals of SHFD were developed by sitting drop method in 0.02 M calcium chloride dihydrate, 0.1 M sodium acetate trihydrate and 15% v/v MPD after 30 days of incubation at 4 °C (Figure 5.7). X-ray diffraction was not successful due to the problem in indexing. The crystallization process to get good quality diffractable crystals of the SHFD is still in progress.

Figure 5.7. Single crystal of SHFD.

5.3.9. Homology search using NCBI

Amino acid sequence similarity search was carried out using NCBI BLASTp. BLAST results showed that protein belong to the HAD-like superfamily i.e. Haloacid dehalogenase-

like superfamily. SHFD showed significant similarity to the HAD subclass II sugar phosphatase Bt4131 from *Bacteriodes thetaiotaomicron VPI-5482*, HAD-like phosphatase Yida from *Escherichia coli*, Haloacid Dehalogenase-Like Hydrolase from *Klebsiella pneumonia* and *Bacillus subtilis*, hydrolase from *Lactobacillus brevis*. It showed 28% identity with HAD subclass II sugar phosphatase Bt4131 from *B. thetaiotaomicron VPI-5482*, 29% with Yida from *E. coli*, 24% with Haloacid Dehalogenase-Like Hydrolase from *K. pneumonia*, 22% with Haloacid Dehalogenase-Like hydrolase from *B. subtilis* and 25% with hydrolase from *L. brevis*.

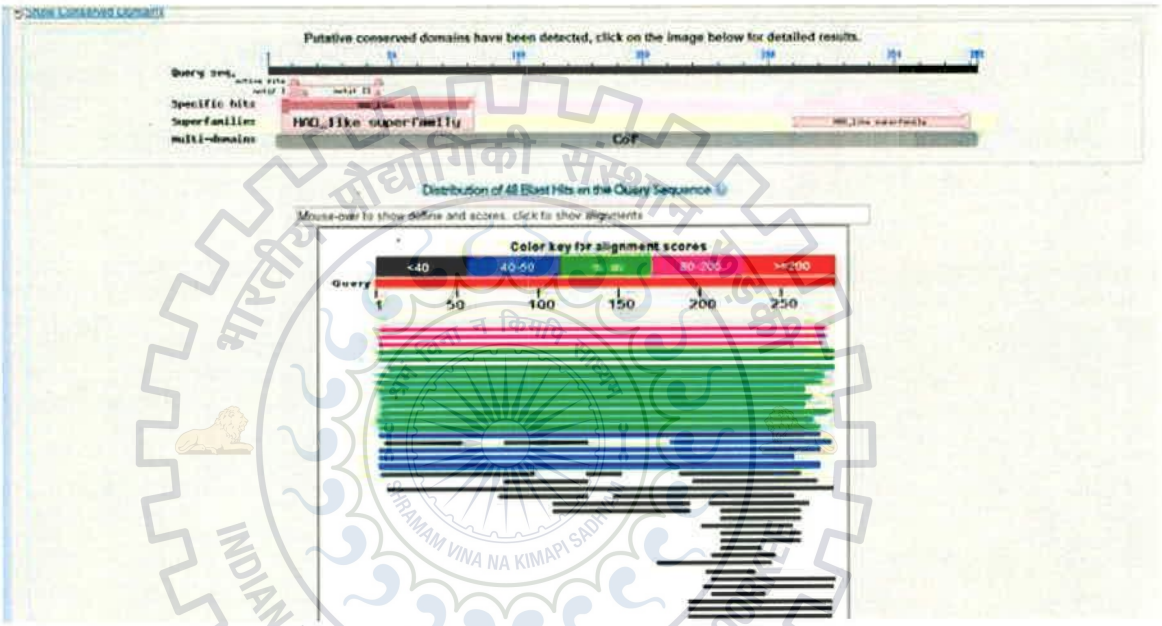


Figure 5.8. Homology search done against Protein Data Bank using NCBI BLASTp.

5.3.10. Multiple sequence alignment

Mutiple sequence alignment (MSA) with homologous HAD family phosphatases and hydrolases show significant similarities and differences. SHFD possess the four conserved loops containing the four motifs of HAD superfamily. These motifs have certain signature sequences which are boxed in figure 5.9. The conserved catalytic residues in SHFD are aspartate10 and aspartate12 in loop I, threonine43, arginine45 in loop II, lysine210 in loop III and aspartate233, asparagine 236 and aspartate237 in loop IV. It is proposed that Asp10 will act as a nucleophilic residue resulting in the formation of an aspartyl-phosphate intermediate while the second Asp i.e. Asp12 acts as an acid/base catalytic residue positioned by loop II residue Arg45 as seen in case of *B. thetaiotaomicron VPI-5482* [153]. Thr43 is said to help in binding the phosphoryl group whereas Lys210 helps in orienting the Asp for nucleophilic

attack as well as for shielding the charge on the negatively charged phosphate moiety. The motif IV conserved residues helps in the binding the cofactor metal ion necessary for the SHFD activity. Thus these four loops present in the core domain are conserved in the HADSF and are responsible for orientation, binding and activation of the phosphate moiety for transfer. The core domain also prepares the nucleophile for attack on phosphoryl moiety.

Figure 5.9. Multiple sequence alignment of SHFD with HAD superfamily members. Four motifs (I, II, III, IV) present in the core domain are boxed along with the catalytic conserved residues highlighted in yellow colour.

5.3.11. Secondary structure analysis

Secondary structure of SHFD was predicted using ESPript server as seen in Figure 5.10. SHFD contains 37% α-helices, 23% β-stranded sheets and 36% coil region. SHFD is an α/β protein with 10 α-helices and 13 β-sheets which further corroborates with the CD results.

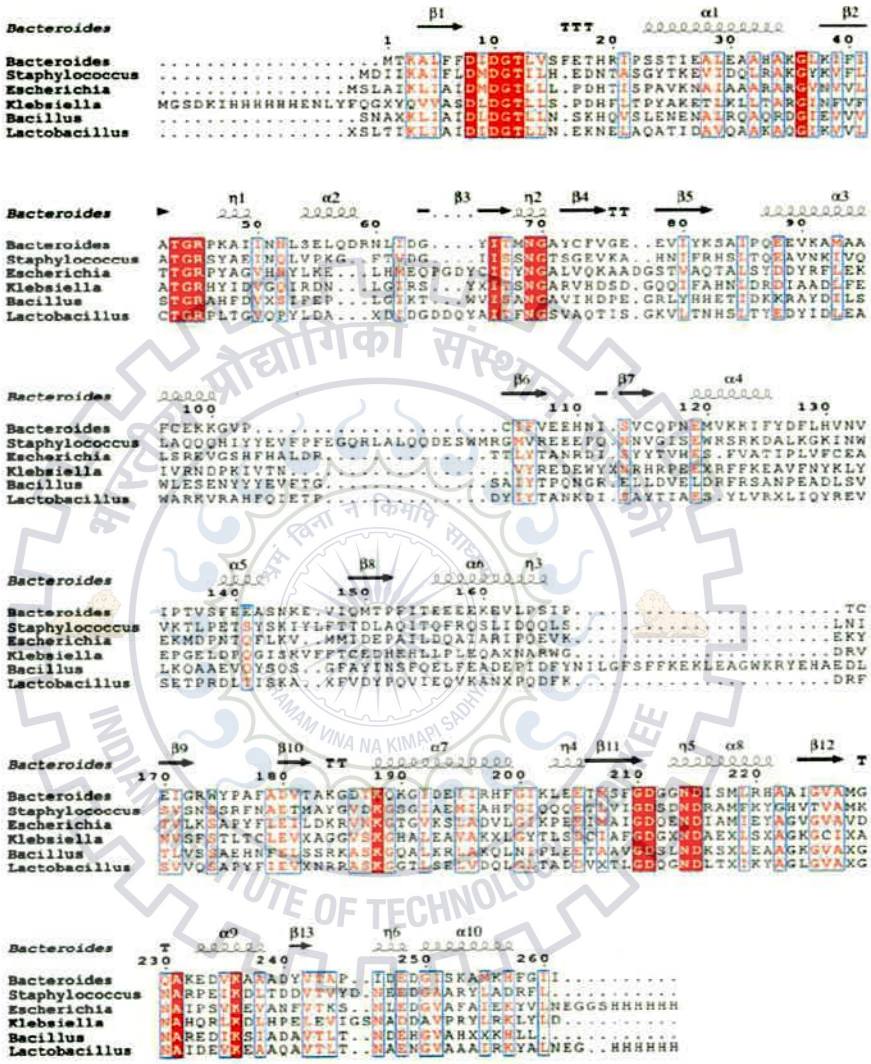


Figure 5.10. Showing secondary structure analysis by ESPript. α-helices are represented as squiggles while β-sheet are represented as arrows, TT represents β-turn.

5.3.12. Phylogenetic analysis

Phylogenetic analysis of SHFD along with the other HAD family hydrolases and phosphatases was done using MEGA6 server to trace their evolutionary among them. Phylogenetic tree dissected them into two major groups which have a common ancestral evolution. SHFD was found to cluster closely with the sugar phosphatase Bt4131 from *B.*

thetaiotaomicronVPI-5482 (Figure 5.11). However, Haloacid Dehalogenase-Like Hydrolase from *K. pneumonia* and *B. subtilis* form a distinct group.

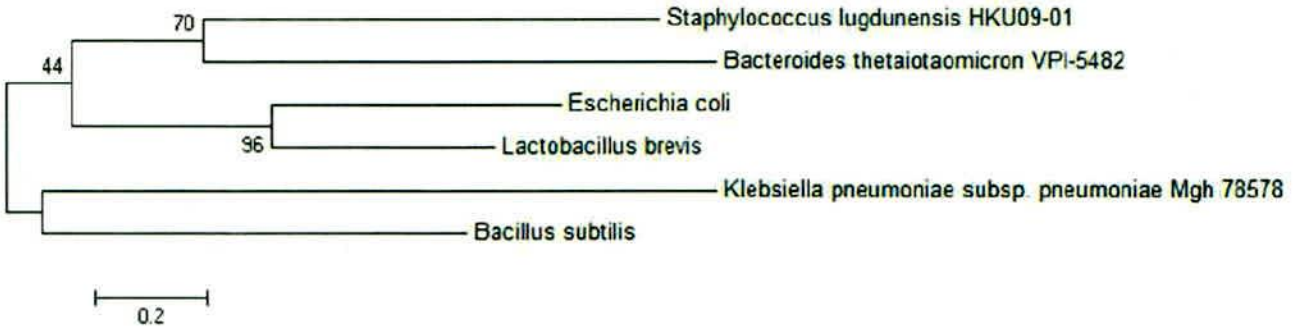


Figure 5.11. Phylogenetic tree of SHFD and other related sequences was constructed by the maximum likelihood method using MEGA6. The numbers overhead and under the branches are determined by bootstrap analysis to point out the confidence levels of the relationship among these paired sequences. The tree is drawn to scale, with branch lengths being measured in the number of substitutions per site.

5.3.13. Three dimensional model and its verification

The three dimensional model of SHFD was constructed using crystal structure of a sugar phosphate phosphatase BT4131 from the bacterium *Bacteroides thetaiotaomicron VPI-5482* (PDB Id: 1YMQ) as a template by Modeller9.14. Out of 10 models, the best model was preferred using PROCHECK program of SAVES. The predicted model consists of 10 α-helices and 13 β-sheets. Like other HADSF members, the SHFD protein is composed of two domains i.e. a bigger and a smaller domain. The bigger domain represents the conserved core region comprising four loops that surrounds the active site. It is having a modified Rossmann fold with six stranded β-sheets surrounded by six α-helices. The smaller domain covers the active site and is linked to this core domain with the help of two additional β-sheets. This smaller domain also called as the Cap domain is responsible for the substrate specificity of HAD family. The Cap domain is a mixed αβ domain with αββ (αβαβ) αββ topology which is also seen in subfamily IIB of HADSF. Thus SHFD can be classified as a subfamily IIB phosphatase.

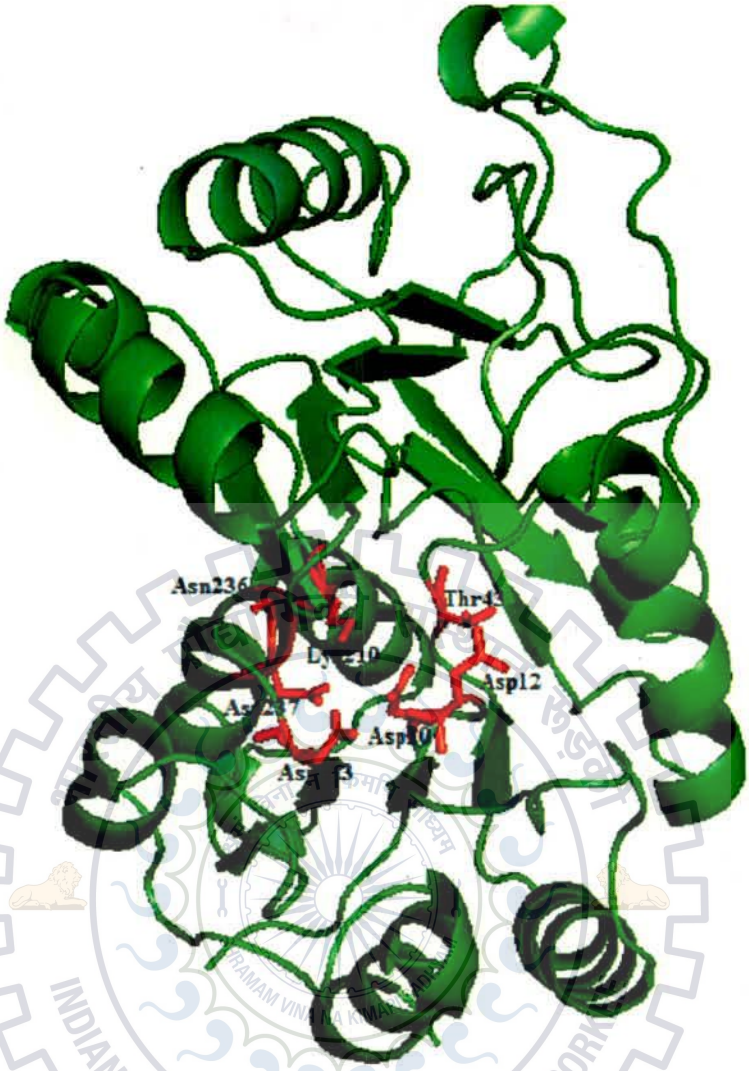


Figure 5.12. Three dimensional model of SHFD. The catalytic key residues are shown in red sticks.

5.4. Conclusion

SHFD from *Staphylococcus lugdunensis* belongs to a Haloacid dehalogenase superfamily. SHFD gene with ~840 bp has been cloned and overexpressed in *E. coli*. The enzyme with a molecular mass of ~32 kDa has been purified using Ni^{2+}-NTA affinity and gel filtration chromatography. SHFD shows phosphatase activity with an optimum temperature of 25.0 °C. SHFD is an acid phosphatase with an optimum pH of 5.0. The kinetic parameters (K_m= 0.32mM, V_{max} = 0.36 U/mg, k_{cat} = 21.43 ± 0.85 $^{-1}$ and k_{cat}/K_m = 66.96 $mM^{-1}s^{-1}$) indicate that it is a very efficient enzyme. SHFD is a mixed α/β protein as predicted by the ESPript and CD spectrum. Multiple sequence alignment shows the conservation of nucleophilic Asp10, acid/base catalyst Asp12, phosphate binding Ser43, and many other catalytic residues like Arg45, Lys210, Asp233, Asn236 and Asp237. SHFD is a two domain protein, with a larger core domain comprising four conserved loops surrounding the active site. The core domain also has a modified Rossmann fold with six stranded β-sheets surrounded by six α-helices. It is responsible for binding, reorienting of the phosphate group along with the co-factor Mg^{2+} and also in preparing the Asp10 for the nucleophilic attack. The smaller Cap domain has a αββ(αβαβ)αββ topology which further classifies it as an subfamily IIB phosphatase.

CONCLUSIONS

CONCLUSIONS

CLONING, EXPRESSION, CHARACTERIZATION AND BIOINFORMATIC STUDIES OF *PUTRANJIVA ROXBURGHII* FAMILY 1 GLYCOSYL HYDROLASE (PRGH1)

- A 66 kDa native protein, with β-fucosidase, β-glucosidase, and β-galactosidase activities, was purified and characterized from seeds of *Putranjiva roxburghii*. The β-fucosidase activity of PRGH1 was found to be one of the highest among related plant β-glucosidase enzymes.
- The *Putranjiva roxburghii* β-glucosidase gene was cloned and submitted to NCBI with the accession no. AHN8564.
- The β-glucosidase (PRGH1) enzyme was expressed in prokaryotic system and purified to homogeneity by employing Ni-NTA affinity chromatography.
- Oligomerization study showed that at higher concentration this enzyme forms a mixture of different oligomers with molecular weight similar to the decameric mass of the enzyme.
- This enzyme shows activity towards aryl as well as glucooligosaccharides and natural β-D-Glycosides.
- The pH optima values of both the native and recombinant enzymes towards β-fucosidase and β-glucosidase activity was found to be 4.6 whereas the pH optima for β-galactosidase activity was found to be 4.4.
- The optimum temperature was found to be 65 °C and the thermostability studies showed that recombinant protein was more susceptible to higher temperature than native protein.
- Mixed substrate kinetics proved that all the three *p*NP substrates were catalyzed in a single active site.
- Mutational studies showed that Glu173 and Glu389 were catalytic acid/base and catalytic nucleophile respectively.
- Bioinformatic studies revealed that the presence of various salt bridges and a conserved disulphide bond makes it an extremely thermostable enzyme.
- Multiple sequence alignment shows the conservation of glycone binding residues in PRGH1.
- The active site of PRGH1 is unique in having a restriction in the active site due to the longer side chains of Met361 and Arg180 being present close together at the entrance

of the active site which might be responsible for the distinct substrate specificity of PRGH1 and also its decreased affinity towards longer oligosaccharides.

- Bioinformatic studies revealed the cause substrate preferences of PRGH1 towards *p*NP substrates.
- Bioinformatic studies proposed that PRGH1 might be evolving evolutionary from an O-glycosidase towards an S-glycosidase for the plant defence mechanism against the toxic products absorbed by the plants and also generated during mastication.
- PRGH1 can be utilized for different valuable applications in the development of new biofuel feed stock, processing of cellulose and lactose containing byproducts in food industry, agriculture, beverages and in medicinal industry as an enzyme additives.

ROLE OF N-LINKED GLYCOSYLATION ON THE STABILITY AND ACTIVITY OF PRGH1 ENZYME

- Glycosylation, one of the most complex and common PTM, is of immense interest for its role in many biological functions such as cell signaling, molecular recognition, and immune defence.
- Deglycosylated PRGH1 was produced by using PNGase F enzyme and sugar was estimated by using phenol-sulphuric acid method.
- Deglycosylation decreases the overall activity of the native PRGH1 enzyme.
- Glycosylation contributes to the stability of the native protein making it more resistance to different denaturants i.e. temperature, pH and alcohols.
- Proteolysis study along with the spectral studies suggests that the structure of native glycosylated PRGH1 is quite compact and rigid than the deglycosylated counterpart.
- Mutagenesis studied showed that out of seven potential glycosylation sites, three sites were glycosylated and *in silico* studies predicts that all the glycosylation sites were present at the surface of the enzyme.
- Bioinformatic studies reveled that glycosylation do not interfere with the active site residues of PRGH1.
- Overall, our results suggest that, N-linked glycosylation is not essential for β-glucosidase enzyme activity but it is important for stability of the enzyme for the optimal activity.

CLONING, EXPRESSION, PURIFICATION OF PRGH1 ENZYME BY USING YEAST SYSTEM AND ITS APPLICATION FOR BIOETHANOL PRODUCTION

- The hydrolysis of cellulosic biomass and its effective conversion to bio-ethanol by using engineered microorganisms is a very potent and cost effective strategy.
- The *prgh1* gene was cloned in yXYNSEC vector under the transcriptional control of PGK1 promoter and terminator for constitutive expression.
- To ensure autoselection of the URA3-bearing yXYNSEC-PRGH1 plasmid in non-selective medium, disruption of the uracil phosphoribosyltransferase (FUR1) gene in the *S. cerevisiae* Y294 transformants was performed by using pDF1 plasmid.
- Recombinant PRGH1 was overexpressed in *S. cerevisiae* Y294, purified and biochemically characterized.
- The optimal pH and temperature of the enzyme activity was found to be 5.0 and 65 °C respectively indicating this enzyme is a thermostable enzyme with preference towards moderate acidic condition.
- The enzyme showed broad substrate specificity and able to hydrolyze cellobiose significantly.
- The enzyme showed resistance towards alcohols, suggesting the enzyme can be used in fermentation industry more efficiently.
- Inhibitors like N-bromosuccinamide and δ-gluconolactone inhibits the activity of the PRGH1 enzyme.
- The recombinant *S. cerevisiae* Y294 harbouring *prgh1* gene showed better growth profile, cellobiose consumption and ethanol production than the native *S. cerevisiae* Y294, when cellobiose was used as sole carbon source. The highest ethanol production was 4.05 g/L for recombinant *S. cerevisiae* Y294 culture, which was 2.9 fold higher than the native *S. cerevisiae* Y294.
- Complementing with commercial cellulose enzyme, the recombinant *S. cerevisiae* Y294 was used in SSF experiments using CMC, rice straw, sugarcane bagesse as carbon source. Sugarcane bagesse was the best substrate for SSF as the pre-treatment and hydrolysis process synergistically produced more fermentable sugars and subsequently more production of ethanol.

CHARACTERIZATION OF A HAD-FAMILY PHOSPHATASE ENZYME FROM STAPHYLOCOCCUS LUGDUNENSIS

- SHFD gene having ~840 bp has been cloned, overexpressed in *E. coli.*
- The enzyme with a molecular mass of ~32 kDa has been purified using Ni^{2+}-NTA affinity and gel filtration chromatography and biochemically characterized.
- SHFD shows phosphatase activity with an optimum temperature of 25.0°C and an optimum pH of 5.0.
- SHFD is a Mg^{2+} dependant enzyme having kinetic parameters (K_m= 0.32mM, V_{max} = 0.36U/mg, k_{cat} = 21.43 ± 0.85 s^{-1} and k_{cat}/K_m = 66.96 $mM^{-1}s^{-1}$) indicate that it is a very efficient enzyme.
- SHFD is a mixed α/β protein as predicted by the ESPript and CD spectrum.
- Multiple sequence alignment shows the conservation of nucleophilic Asp10, acid/base catalyst Asp12, phosphate binding Ser43, and many other catalytic residues like Arg45, Lys210, Asp233, Asn236 and Asp237.
- SHFD is a two domain protein, with a larger core domain comprising four conserved loops surrounding the active site.
- The core domain also has a modified Rossmann fold with six stranded β-sheets surrounded by six α-helices.

REFERENCES

REFERENCES

1. Ahmad, E., Naeem, A., Javed, S., Yadav, S., & Khan, R. H. The minimal structural requirement of concanavalin A that retains its functional aspects. *Journal of biochemistry*, *142*(3), 307-315 (2007).

2. Ahmadian, M. R., Stege, P., Scheffzek, K., & Wittinghofer, A. Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nature structural biology*, *4*(9), 686-689 (1997).

3. Allen, K. N., & Dunaway-Mariano, D. Markers of fitness in a successful enzyme superfamily. *Current opinion in structural biology*, *19*(6), 658-665 (2009).

4. Allen, K. N., & Dunaway-Mariano, D. Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends in biochemical sciences*, *29*(9), 495-503 (2004).

5. Aravind, L., Galperin, M. Y., & Koonin, E. V. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends in biochemical sciences*, *23*(4), 127-129 (1998).

6. Arroyo-Reyna, A., & Hernández-Arana, A. The thermal denaturation of stem bromelain is consistent with an irreversible two-state model. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *1248*(2), 123-128 (1995).

7. Arumugam, A., Agullo, P., Boopalan, T., Nandy, S., Lopez, R., Gutierrez, C., Narayan, M., & Rajkumar, L. Neem leaf extract inhibits mammary carcinogenesis by altering cell proliferation, apoptosis, and angiogenesis. *Cancer biology & therapy*, *15*(1), 26-34 (2014).

8. Babu, E., & Oropello, J. *Staphylococcus lugdunensis*: the coagulase-negative staphylococcus you don't want to ignore (2011).

9. Baker, A. S., Ciocci, M. J., Metcalf, W. W., Kim, J., Babbitt, P. C., Wanner, B. L., ... & Dunaway-Mariano, D. Insights into the mechanism of catalysis by the PC bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis. *Biochemistry*, *37*(26), 9305-9315 (1998).

10. Barleben, L., Panjikar, S., Ruppert, M., Koepke, J., & Stöckigt, J. Molecular architecture of strictosidine glucosidase: the gateway to the biosynthesis of the monoterpenoid indole alkaloid family. *The Plant Cell Online*,*19*(9), 2886-2897 (2007).

11. Barnett, C. C., Berka, R. M., & Fowler, T. Cloning and amplification of the gene encoding an extracellular β-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. *Nature Biotechnology*, *9*(6), 562-567 (1991).

12. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A. & Eddy, S. R. The Pfam protein families database. *Nucleic acids research*, *32*(suppl 1), D138-D141 (2004).

13. Bhatia, Y., Mishra, S., & Bisaria, V. S. Microbial β-glucosidases: cloning, properties, and applications. *Critical reviews in biotechnology*, *22*(4), 375-407 (2002).

14. Bhattacharya, A., Prajapati, R., Chatterjee, S., & Mukherjee, T. K. Concentration-Dependent Reversible Self-Oligomerization of Serum Albumins through Intermolecular β-Sheet Formation. *Langmuir*, *30*(49), 14894-14904 (2014).

15. Białkowska, A. M., Cieśliński, H., Nowakowska, K. M., Kur, J., & Turkiewicz, M. A new β-galactosidase with a low temperature optimum isolated from the Antarctic Arthrobacter sp. 20B: gene cloning, purification and characterization. *Archives of microbiology*, *191*(11), 825-835 (2009).

16. Biswas, S., & Kayastha, A. M. Thermal stability of Phaseolus vulgaris leucoagglutinin: a differential scanning calorimetry study. *Journal of biochemistry and molecular biology*, *35*(5), 472-475 (2002).

17. Boisson, M., Gomord, V., Audran, C., Berger, N., Dubreucq, B., Granier, F., ... & Lepiniec, L. Arabidopsis glucosidase I mutants reveal a critical role of N-glycan trimming in seed development. *The EMBO journal*, *20*(5), 1010-1019 (2001).

18. Bosch, D., Castilho, A., Loos, A., Schots, A., & Steinkellner, H. N-glycosylation of plant-produced recombinant proteins. *Current pharmaceutical design*, *19*(31), 5503-5512 (2013).

19. Bowles, D. J.. Post-translational processing of concanavalin A. In *SEMINAR SERIES-SOCIETY FOR EXPERIMENTAL BIOLOGY* (Vol. 53, pp. 257-257). Cambridge University Press (1993).

20. Bruins, M. E., Janssen, A. E., & Boom, R. M. Thermozymes and their applications. *Applied biochemistry and biotechnology*, *90*(2), 155-186 (2001).

21. Brzobohaty, B., Moore, I., Kristoffersen, P., Bako, L., Campos, N., Schell, J., Palme, K. Release of active cytokinin by a *β*-glucosidase localized to the maize root meristem. Science. 262, 1051–1054 (1993).

22. Burda, P., & Aebi, M. (1999). The dolichol pathway of N-linked glycosylation.*Biochimica et Biophysica Acta (BBA)-General Subjects*, *1426*(2), 239-257 (2003).

23. Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., & Henrissat, B. The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl–enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure*, *5*(5), 663-676 (1997).

24. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., & Aravind, L. Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic

diversity in a superfamily of phosphoesterases and allied enzymes. *Journal of molecular biology*, *361*(5), 1003-1034 (2006).

25. Cabezs, J. A., and Vazquer-Pernas, R. Separation and properties of α and β-galactosidase and other glycosidases from jack bean meal. Rev. Esp. Fisiol. 25, 147-152 (1969).

26. Calvo, P., Santamaria, M. G., Melgar, M. J., & Cabezas, J. A. Kinetic evidence for two active sites in β-D-fucosidase of Helicella ericetorum.*International Journal of Biochemistry*, *15*(5), 685-693 (1983).

27. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic acids research*, *37* (suppl 1), D233-D238 (2009).

28. Cavka, A., Alriksson, B., Rose, S. H., van Zyl, W. H., & Jönsson, L. J. Production of cellulosic ethanol and enzyme from waste fiber sludge using SSF, recycling of hydrolytic enzymes and yeast, and recombinant cellulase-producing Aspergillus niger. *Journal of industrial microbiology & biotechnology*,*41*(8), 1191-1200 (2014).

29. Ceriotti, A., Duranti, M., and Bollini, R. Effect of N-glycosylation on the folding and structure of plant proteins. Journal of Experimental Botany. 49, 1091-1103 (1998).

30. Chan, C. P., Mak, T. Y., Chin, K. T., Ng, I. O. L., & Jin, D. Y. N-linked glycosylation is required for optimal proteolytic activation of membrane-bound transcription factor CREB-H. *Journal of cell science*, *123*(9), 1438-1448 (2010).

31. CHEOL, K. K., Yoo, S. S., Oh, Y. A., & KIM, S. J. Isolation and characteristics of Trichoderma harzianum FJ1 producing cellulases and xylanase. *Journal of microbiology and biotechnology*, *13*(1), 1-8 (2003).

32. Chinchetru, M. A., Cabezas, J. A., and Calvo, P. Characterization and kinetics of β-D-Gluco/Fuco/Galactosidase from sheep liver. Comp. Biochem. Physiol. 75B, 719-728 (1983).

33. Chrispeels, M. J., Hartl, P. M., Sturm, A., & Faye, L. Characterization of the endoplasmic reticulum-associated precursor of concanavalin A. Partial amino acid sequence and lectin activity. *Journal of Biological Chemistry*, *261*(22), 10021-10024 (1986).

34. Chuankhayan, P., Rimlumduan, T., Tantanuch, W., Mothong, N., Kongsaeree, P. T., Metheenukul, P., Svasti, J., Jensen, O.N., and Ketudat Cairns, J. Functional and structural differences between isoflavonoid β-glycosidases from *Dalbergia* sp. Arch Biochem Biophys, 468, 205-216 (2007).

35. Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Oonanant, W., Bevan, D. R., ... & Cairns, J. R. K. Structural insights into rice BGlu1 β-glucosidase

oligosaccharide hydrolysis and transglycosylation. *Journal of molecular biology*, *377*(4), 1200-1215 (2008).

36. Collet, J. F., Gerin, I., Rider, M. H., Veiga-da-Cunha, M., & Van Schaftingen, E. Human L-3-phosphoserine phosphatase: sequence, expression and evidence for a phosphoenzyme intermediate. *FEBS letters*, *408*(3), 281-284 (1997).

37. Côté, J. P., Charbonneau, M. È., & Mourez, M. Glycosylation of the Escherichia coli TibA self-associating autotransporter influences the conformation and the functionality of the protein. *PloS one*, *8*(11), e80739 (2013).

38. Coutinho, P.M., and Henrissat, B. CAZY: Carbohydrate Active Enzymes. http://www.cazy.org/ (January 2, 2014) (1998).

39. Coyne, S., Litomska, A., Chizzali, C., Khalil, M. N., Richter, K., Beerhues, L., & Hertweck, C. Control of Plant Defense Mechanisms and Fire Blight Pathogenesis through the Regulation of 6-Thioguanine Biosynthesis in Erwinia amylovora. *ChemBioChem*, *15*(3), 373-376 (2014).

40. Dai, J., Finci, L., Zhang, C., Lahiri, S., Zhang, G., Peisach, E., ... & Dunaway-Mariano, D. Analysis of the Structural Determinants Underlying Discrimination between Substrate and Solvent in β-Phosphoglucomutase Catalysis†‡. *Biochemistry*, *48*(9), 1984-1995 (2009).

41. Dan, S., Marton, I., Dekel, M., Bravdo, B. A., He, S., Withers, S. G., & Shoseyov, O. Cloning, expression, characterization, and nucleophile identification of family 3, Aspergillus nigerβ-glucosidase. *Journal of Biological Chemistry*, *275*(7), 4973-4980 (2000).

42. Davies, G., & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure*, *3*(9), 853-859 (1995).

43. Dean, N. Asparagine-linked glycosylation in the yeast Golgi. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1426*(2), 309-322 (1999).

44. Dekker, R. F. Kinetic, inhibition, and stability properties of a commercial β-D-glucosidase (cellobiase) preparation from aspergillus niger and its suitability in the hydrolysis of lignocellulose. *Biotechnology and Bioengineering*, *28*(9), 1438-1442 (1986).

45. DeLano, W. L. The PyMOL molecular graphics system (2002).

46. Dell, A., & Morris, H. R. Glycoprotein structure determination by mass spectrometry. *Science*, *291*(5512), 2351-2356 (2001).

47. Dempski, R. E., & Imperiali, B. Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Current opinion in chemical biology*, *6*(6), 844-850 (2002).

48. Dhaked, R. K., Alam, S. I., & Singh, L. Characterization of beta-galactosidase from an Antarctic Bacillus sp. *Indian Journal of Biotechnology*,*4*(2), 227-231 (2005).

49. Dhaked, R. K., Sharma, S. K., Parida, M. M., & Singh, L. Isolation and characterization of Clostridium botulinum type E from soil of Gwalior, India.*Journal of natural toxins*, *11*(1), 49-56 (2002).

50. Dharmawardhana, D. P., Ellis, B. E., & Carlson, J. E. A [beta]-Glucosidase from lodgepole pine xylem specific for the lignin precursor coniferin. *Plant physiology*, *107*(2), 331-339 (1995).

51. Dixit, A., Alam, S. I., Dhaked, R. K., & Singh, L. Development of an immunodetection test for a botulinum-like neurotoxin produced by Clostridium sp. RKD. *Indian J Med Res*, *124*(3), 355-362 (2006).

52. Długołecka, A., Cieśliński, H., Bruździak, P., Gottfried, K., Turkiewicz, M., & Kur, J. Purification and biochemical characteristic of a cold-active recombinant esterase from Pseudoalteromonas sp. 643A under denaturing conditions. *Polish journal of microbiology/Polskie Towarzystwo Mikrobiologow. The Polish Society of Microbiologists*, *58*(3), 211-218 (2008).

53. Driouich, A., Gonnet, P., Makkie, M., Laine, A. C., & Faye, L. The role of high-mannose and complex asparagine-linked glycans in the secretion and stability of glycoproteins. *Planta*, *180*(1), 96-104 (1989).

54. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P., & Smith, F. Colorimetric method for determination of sugars and related substances. *Analytical chemistry*, *28*(3), 350-356 (1956).

55. Dwevedi, A., & Kayastha, A. M. Optimal immobilization of β-galactosidase from Pea (PsBGAL) onto Sephadex and chitosan beads using response surface methodology and its applications. *Bioresource technology*, *100* (10), 2667-2675 (2009).

56. Dzeja, P. P., & Terzic, A. Phosphotransfer networks and cellular energetics. *Journal of Experimental Biology*, *206*(12), 2039-2047 (2003).

57. Ellgaard, L., & Helenius, A. Quality control in the endoplasmic reticulum. *Nature reviews Molecular cell biology*, *4*(3), 181-191 (2003).

58. Ermonval, M., Kitzmüller, C., Mir, A. M., Cacan, R., & Ivessa, N. E. N-glycan structure of a short-lived variant of ribophorin I expressed in the MadIA214 glycosylation-defective cell line reveals the role of a mannosidase that is not ER mannosidase I in the process of glycoprotein degradation.*Glycobiology*, *11*(7), 565-576 (2001).

59. Escamilla-Treviño, L. L., Chen, W., Card, M. L., Shih, M. C., Cheng, C. L., & Poulton, J. E. Arabidopsis thaliana β-glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. *Phytochemistry*, *67*(15), 1651-1660 (2006).

60. Eyzaguirre, J., Hidalgo, M., & Leschot, A. 23 β-Glucosidases from Filamentous Fungi: Properties, Structure, and Applications (2005).

61. Falk, A., & Rask, L. Expression of a zeatin-O-glucoside-degrading beta-glucosidase in Brassica napus. *Plant Physiology*, *108*(4), 1369 (1995).

62. Faraldos, J. A., Gonzalez, V., Li, A., Yu, F., Köksal, M., Christianson, D. W., & Allemann, R. K. Probing the Mechanism of 1, 4-Conjugate Elimination Reactions Catalyzed by Terpene Synthases. *Journal of the American Chemical Society*, *134*(51), 20844-20848 (2012).

63. Farrell, A. E., Plevin, R. J., Turner, B. T., Jones, A. D., O'hare, M., & Kammen, D. M. Ethanol can contribute to energy and environmental goals. *Science*, *311*(5760), 506-508 (2006).

64. Faye, L., & Chrispeels, M. J. Apparent inhibition of β-fructosidase secretion by tunicamycin may be explained by breakdown of the unglycosylated protein during secretion. *Plant physiology*, *89*(3), 845-851 (1989).

65. Faye, L., & Chrispeels, M. J. Transport and processing of the glycosylated precursor of concanavalin A in jack-bean. *Planta*, *170*(2), 217-224 (1987).

66. Fengel, D., & Wegener, G. (Eds.). *Wood: chemistry, ultrastructure, reactions*. Walter de Gruyter (1983).

67. Fetzner, S., & Lingens, F. Bacterial dehalogenases: biochemistry, genetics, and biotechnological applications. *Microbiological Reviews*, *58*(4), 641 (1994).

68. Fontana, A. Structure and stability of thermophilic enzymes studies on thermolysin. *Biophysical chemistry*, *29*(1), 181-193 (1988).

69. Fontana, A., de Laureto, P. P., De Filippis, V., Scaramella, E., & Zambonin, M. Probing the partly folded states of proteins by limited proteolysis. *Folding and Design*, *2*(2), R17-R26 (1997).

70. Fötisch, K., & Vieths, S. N-and O-linked oligosaccharides of allergenic glycoproteins. *Glycoconjugate journal*, *18*(5), 373-390 (2001).

71. Freney, J., Brun, Y., Bes, M., Meugnier, H., Grimont, F., Grimont, P. A., Nervi, C. & Fleurette, J. Staphylococcus lugdunensis sp. nov. and Staphylococcus schleiferi sp. nov., two species from human clinical specimens. *International Journal of Systematic Bacteriology*, *38*(2), 168-172 (1988).

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

73. Gangwar, S. P., Meena, S. R., & Saxena, A. K. Cloning, purification, crystallization and preliminary X-ray analysis of ESX-1-secreted protein regulator (EspR) from Mycobacterium tuberculosis. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *67*(1), 83-86 (2010).

74. Gemmill, T. R., & Trimble, R. B. Overview of N-and O-linked oligosaccharide structures found in various yeast species. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1426*(2), 227-237(1999).

75. Ghose, T. K. Measurement of cellulase activities. *Pure and applied Chemistry*, *59*(2), 257-268 (1987).

76. Ghosh, M., Grunden, A. M., Dunn, D. M., Weiss, R., and Adams, M. W. W. Characterization of Recombinant and Native Forms of an Unusual Cobalt-Dependent Proline Dipeptidase (Prolidase) from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. J. Bacteriology. 180, 4781-4789 (1998).

77. Giordani, R., and Noat, G. Isolation, molecular properties and kinetic studies of a strict of a β-fucosidase from *Latuca savita* latex. Eur. J. Biochem. 175, 619-625 (1998).

78. Godinho, L. M., & de Sá-Nogueira, I. Characterization and regulation of a bacterial sugar phosphatase of the haloalkanoate dehalogenase superfamily, AraL, from Bacillus subtilis. *FEBS Journal*, *278*(14), 2511-2524 (2011).

79. Golden, K. A., Wong, C. S., Jeremiason, J. D., Eisenreich, S. J., Sanders, G., Hallgren, J, Swackhamer, D. L., Engstrom, D. R. & Long, D. T. Accumulation and preliminary inventory of organochlorines in Great Lakes sediments. *Water Science & Technology*, *28*(8-9), 19-31 (1993).

80. Goldman, P., Milne, G. W. A., & Keister, D. B. Carbon-halogen bond cleavage III. Studies on bacterial halidohydrolases. *Journal of Biological Chemistry*, *243*(2), 428-434 (1968).

81. Gomord, V., Fitchette, A. C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., & Faye, L. Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant biotechnology journal*, *8*(5), 564-587(2010).

82. Grimaldi, S., Pozzi, D., Verna, R., Lio, S., Giganti, G., De Pirro, R., & Monaco, F. Enzymatic deglycosylation of human thyroglobulin: fluorescence studies. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *957*(1), 105-110 (1988).

83. Grohmann, K., & Baldwin, E. A. Hydrolysis of orange peel with pectinase and cellulase enzymes. *Biotechnology letters*, *14*(12), 1169-1174 (1992).

84. Grohmann, K., Baldwin, E. A., & Buslig, B. S. Production of ethanol from enzymatically hydrolyzed orange peel by the yeast Saccharomyces cerevisiae. *Applied Biochemistry and biotechnology*, *45*(1), 315-327 (1994).

85. Grueninger-Leitch, F., D'Arcy, A., D'Arcy, B., & Chène, C. Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. *Protein science*, *5*(12), 2617-2622 (1996).

86. Guo, Z. P., Zhang, L., Ding, Z. Y., Gu, Z. H., & Shi, G. Y. Development of an industrial ethanol-producing yeast strain for efficient utilization of cellobiose. *Enzyme and microbial technology*, *49*(1), 105-112 (2011).

87. Hahn-Hägerdal, B., Wahlbom, C. F., Gárdonyi, M., van Zyl, W. H., Otero, R. R. C., & Jönsson, L. J. Metabolic engineering of Saccharomyces cerevisiae for xylose utilization. In *Metabolic Engineering* (pp. 53-84). Springer Berlin Heidelberg (2001).

88. Hardiman, E., Gibbs, M., Reeves, R., & Bergquist, P. Directed evolution of a thermophilic β-glucosidase for cellulosic bioethanol production.*Applied biochemistry and biotechnology*, *161*(1-8), 301-312 (2010).

89. Harnpicharnchai, P., Champreda, V., Sornlake, W., & Eurwilaichitr, L. A thermotolerant β-glucosidase isolated from an endophytic fungi, Periconia sp., with a possible use for biomass conversion to sugars. *Protein expression and purification*, *67*(2), 61-69 (2009).

90. Hartmann-Schreier, J., and Schreier, P. Properties of β-glucosidase from *Carica papaya* fruit. Food. Chem. 26, 201-212 (1987).

91. Hassan, M., Kumar, V., Singh, T. P., & Yadav, S. Proteomic analysis of human amniotic fluid from Rh− pregnancy. *Prenatal diagnosis*, *28*(2), 102-108 (2008).

92. Hasunuma, T., & Kondo, A. Development of yeast cell factories for consolidated bioprocessing of lignocellulose to bioethanol through cell surface engineering. *Biotechnology advances*, *30*(6), 1207-1218 (2012).

93. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J*, *280*, 309-316 (1991).

94. Henrissat, B., & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J, 293*, 781-788 (1993).

95. Henrissat, B., and Davies, G. Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. 7, 637-644 (1997).

96. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., & Davies, G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences*, *92*(15), 7090-7094 (1995).

97. Hileman, B. Concerns broaden over chlorine and chlorinated hydrocarbons. *Chemical and Engineering News(USA)*, *71*(16), 11-20 (1993).

98. Hill, J., Donald, K. I. G., & Griffiths, D. E. DMSO-enhanced whole cell yeast transformation. *Nucleic acids research*, *19*(20), 5791-5791 (1991).

99. Hisano, T., Hata, Y., Fujii, T., Liu, J. Q., Kurihara, T., Esaki, N., & Soda, K. Crystal Structure of L-2-Haloacid Dehalogenase from Pseudomonas sp. YL AN α/β HYDROLASE STRUCTURE THAT IS DIFFERENT FROM THE α/β HYDROLASE FOLD. *Journal of Biological Chemistry*, *271*(34), 20322-20330 (1996).

100. Hodge, D. B., Karim, M. N., Schell, D. J., & McMillan, J. D. Soluble and insoluble solids contributions to high-solids enzymatic hydrolysis of lignocellulose. *Bioresource Technology*, *99*(18), 8940-8948 (2008).

101. Hong, J., Tamaki, H., & Kumagai, H. Unusual hydrophobic linker region of β-glucosidase (BGLII) from Thermoascus aurantiacus is required for hyper-activation by organic solvents. *Applied microbiology and biotechnology*, *73*(1), 80-88 (2006).

102. Huang, H., Patskovsky, Y., Toro, R., Farelli, J. D., Pandya, C., Almo, S. C., Allen, K. N. & Dunaway-Mariano, D. Divergence of structure and function in the haloacid dehalogenase enzyme superfamily: Bacteroides thetaiotaomicron BT2127 is an inorganic pyrophosphatase. *Biochemistry*, *50*(41), 8937-8949 (2011).

103. Hubbard, S. J. The structural aspects of limited proteolysis of native proteins. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *1382*(2), 191-206 (1998).

104. Iwashita, K., Nagahara, T., Kimura, H., Takano, M., Shimoi, H., & Ito, K. The bglA gene of Aspergillus kawachii encodes both extracellular and cell wall-bound β-glucosidases. *Applied and environmental microbiology*, *65*(12), 5546-5553 (1999).

105. Jangir, D. K., Charak, S., Mehrotra, R., & Kundu, S. FTIR and circular dichroism spectroscopic study of interaction of 5-fluorouracil with DNA. *Journal of Photochemistry and Photobiology B: Biology*, *105*(2), 143-148 (2011).

106. Janjanam, J., Singh, S., Choudhary, S., Pradeep, M. A., Kumar, S., Kumaresan, A., ... & Mohanty, A. K. Molecular cloning, sequence characterization and heterologous expression of buffalo (Bubalus bubalis) oviduct-specific glycoprotein in E. coli. *Molecular biology reports*, *39*(12), 10031-10043 (2012).

107. Janjanam, JJ., Jamwal, M., Singh, S., Kumar, S., Panigrahi , A. K., Hariprasad, G., Jena, M. K1, Anand , V., Kumar, S., Kaushik, J. K., Dang, A. K., Mukesh, M., Mishra, B. P., Srinivasan, A., Reddy, V. S and Mohanty A. K. Proteome analysis of functionally

differentiated bovine (Bos indicus) mammary epithelial cells isolated from milk. Proteomics 13 (21): 3093–3242 (2013).

108. Jenkins, J., Leggio, L. L., Harris, G., & Pickersgill, R. β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes wit 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS letters*, *362*(3), 281-285 (1995).

109. Jin, Z., Katsumata, K. S., Lam, T. B. T., & Iiyama, K. Covalent linkages between cellulose and lignin in cell walls of coniferous and nonconiferous woods. *Biopolymers*, *83*(2), 103-110 (2006).

110. Jiratchariyakul, W., Beerhues, L., Mahady, G. B., Kummalue, T., & Vongsakul, M. Botanicals in Dietary Supplements. *Evidence-Based Complementary and Alternative Medicine*, *2013* (2013).

111. Johansen, P. G., Marshall, R. D., & Neuberger, A. Carbohydrates in protein. 3. The preparation and some of the properties of a glycopeptide from hen's-egg albumin. *Biochemical Journal*, *78*(3), 518 (1961).

112. Jung, S. T., Kang, T. H., Kelton, W., & Georgiou, G. Bypassing glycosylation: engineering aglycosylated full-length IgG antibodies for human therapy. *Current opinion in biotechnology*, *22*(6), 858-867 (2011).

113. Kapoor, A., Shandilya, M., & Kundu, S. Structural insight of dopamine β-hydroxylase, a drug target for complex traits, and functional significance of exonic single nucleotide polymorphisms. *PloS one*, *6*(10), e26509 (2011).

114. Kelly, J., Logan, S. M., Jarrell, K. F., VanDyke, D. J., & Vinogradov, E. A novel N-linked flagellar glycan from Methanococcus maripaludis.*Carbohydrate research*, *344*(5), 648-653 (2009).

115. Khan, R. H., Rasheedi, S., & Haq, S. K. Effect of pH, temperature and alcohols on the stability of glycosylated and deglycosylated stem bromelain.*Journal of biosciences*, *28*(6), 709-714 (2003).

116. Kim, B., Du, J., Eriksen, D. T., & Zhao, H. Combinatorial design of a highly efficient xylose-utilizing pathway in Saccharomyces cerevisiae for the production of cellulosic biofuels. *Applied and environmental microbiology*, *79*(3), 931-941 (2013).

117. Kim, H. S., Kim, N. R., Yang, J., & Choi, W. Identification of novel genes responsible for ethanol and/or thermotolerance by transposon mutagenesis in Saccharomyces cerevisiae. *Applied microbiology and biotechnology*, *91*(4), 1159-1172 (2011).

118. Kim, T. Y., Oh, E. J., Jin, Y. S., & Oh, M. K. Improved resistance against oxidative stress of engineered cellobiose-fermenting Saccharomyces cerevisiae revealed by metabolite profiling. *Biotechnology and Bioprocess Engineering*, *19*(6), 951-957 (2014).

119. Kimata-Ariga, Y., Kubota-Kawai, H., Lee, Y. H., Muraki, N., Ikegami, T., Kurisu, G., & Hase, T. Concentration-dependent oligomerization of cross-linked complexes between ferredoxin and ferredoxin–NADP+ reductase.*Biochemical and biophysical research communications*, *434*(4), 867-872 (2013).

120. Kishore, D., Kundu, S., & Kayastha, A. M. Thermal, chemical and pH induced denaturation of a multimeric β-galactosidase reveals multiple unfolding pathways. *PloS one*, *7*(11), e50380 (2012).

121. Klundt, T., Bocola, M., Lütge, M., Beuerle, T., Liu, B., & Beerhues, L. A single amino acid substitution converts benzophenone synthase into phenylpyrone synthase. *Journal of Biological Chemistry*, *284*(45), 30957-30964 (2009).

122. Knowles, J. R. Enzyme-catalyzed phosphoryl transfer reactions. *Annual review of biochemistry*, *49*(1), 877-919 (1980).

123. Köksal, M., Potter, K., Peters, R. J., & Christianson, D. W. 1.55 Å-resolution structure of ent-copalyl diphosphate synthase and exploration of general acid function by site-directed mutagenesis. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1840*(1), 184-190 (2014).

124. Kong, Q. X., Cao, L. M., Zhang, A. L., & Chen, X. Overexpressing GLT1 in gpd1Δ mutant to improve the production of ethanol of Saccharomyces cerevisiae. *Applied microbiology and biotechnology*, *73*(6), 1382-1386 (2007).

125. Koonin, E. V., & Tatusov, R. L. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity: application of an iterative approach to database search. *Journal of molecular biology*, *244*(1), 125-132 (1994).

126. Krogh, K. B. R. M., Kastberg, H., Jørgensen, C. I., Berlin, A., Harris, P. V., & Olsson, L. Cloning of a GH5 endoglucanase from genus Penicillium and its binding to different lignins. *Enzyme and Microbial Technology*, *44*(6), 359-367 (2009).

127. Krogh, K. B., Mørkeberg, A., Jørgensen, H., Frisvad, J. C., & Olsson, L. Screening genus Penicillium for producers of cellulolytic and xylanolytic enzymes. In *Proceedings of the Twenty-Fifth Symposium on Biotechnology for Fuels and Chemicals Held May 4–7, 2003, in Breckenridge, CO* (pp. 389-401). Humana Press (2004).

128. Kumar, R., Singh, S., & Singh, O. V. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of industrial microbiology & biotechnology*, *35*(5), 377-391 (2008).

129. Kumar, S. B., Venkateswaran, K., & Kundu, S. Alternative conformational model of a seed protein DeK1 for better understanding of structure-function relationship. *Journal of Proteins & Proteomics*, *1*(2) (2013).

130. Kumari, N., Tajmul, M., & Yadav, S. Proteomic Analysis of Mature *Lagenaria siceraria* Seed. *Applied biochemistry and biotechnology*, *175*(8), 3643-3656 (2015).

131. Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., ... & Yakunin, A. F. Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. *Journal of Biological Chemistry*, *281*(47), 36149-36161 (2006).

132. Kwak, J., Grigsby, C. C., Rizki, M. M., Preti, G., Köksal, M., Josue, J., ... & Beauchamp, G. K. Differential binding between volatile ligands and major urinary proteins due to genetic variation in mice. *Physiology & behavior*, *107*(1), 112-120 (2012).

133. La Grange, D. C., Pretorius, I. S., & Van Zyl, W. H. Expression of a Trichoderma reesei beta-xylanase gene (XYN2) in Saccharomyces cerevisiae. *Applied and environmental microbiology*, *62*(3), 1036-1044 (1996).

134. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*(5259), 680-685 (1970).

135. Lahiri, S. D., Zhang, G., Dai, J., Dunaway-Mariano, D., & Allen, K. N. Analysis of the substrate specificity loop of the HAD superfamily cap domain.*Biochemistry*, *43*(10), 2812-2820 (2004).

136. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. Caught in the act: The structure of phosphorylated β-phosphoglucomutase from Lactococcus lactis. *Biochemistry*, *41*(26), 8351-8359 (2002).

137. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. Diversification of function in the haloacid dehalogenase enzyme superfamily: The role of the cap domain in hydrolytic phosphorus carbon bond cleavage.*Bioorganic chemistry*, *34*(6), 394-409 (2006).

138. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction.*Science*, *299*(5615), 2067-2071 (2003).

139. Lamarre-Vincent, N., & Hsieh-Wilson, L. C. Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *Journal of the American chemical society*, *125*(22), 6612-6613 (2003).

140. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, *26*(2), 283-291 (1993).

141. Leah, R., Kigel, J., Svendsen, I., and Mundy, J. Biochemical and molecular characterization of a barley seed β-glucosidase. J. Biol. Chem. 270, 15789–15797 (1995).

142. Leavvy, G. A., and McAllan, A. Mammalian fucosidases. 3. B-D-Fucosidase activity and its relation to β-D-galactosidase. Biochem. J. 87, 361-367 (1963).

143. Lechner, J., & Wieland, F. Structure and biosynthesis of prokaryotic glycoproteins. *Annual review of biochemistry*, *58*(1), 173-194 (1989).

144. Lee, J. Biological conversion of lignocellulosic biomass to ethanol.*Journal of biotechnology*, *56*(1), 1-24 (1997).

145. Lenertz, L. Y., Wang, Z., Guadarrama, A., Hill, L. M., Gavala, M. L., & Bertics, P. J. Mutation of putative N-linked glycosylation sites on the human nucleotide receptor P2X7 reveals a key residue important for receptor function.*Biochemistry*, *49*(22), 4611-4619 (2010).

146. Leung, M. J., Nuttall, N., Mazur, M., Taddei, T. L., McComish, M., & Pearman, J. W. Case of Staphylococcus schleiferi endocarditis and a simple scheme to identify clumping factor-positive staphylococci. *Journal of clinical microbiology*, *37*(10), 3353-3356 (1999).

147. Lin, Y., & Tanaka, S. Ethanol fermentation from biomass resources: current state and prospects. *Applied microbiology and biotechnology*, *69*(6), 627-642 (2006).

148. Lindman, B., Karlström, G., & Stigsson, L. On the mechanism of dissolution of cellulose. *Journal of molecular liquids*, *156*(1), 76-81 (2010).

149. Little, M., & Williams, P. A. A bacterial halidohydrolase. *European Journal of Biochemistry*, *21*(1), 99-109 (1971).

150. Liu, D., Zhang, R., Yang, X., Xu, Y., Tang, Z., Tian, W., & Shen, Q. Expression, purification and characterization of two thermostable endoglucanases cloned from a lignocellulosic decomposing fungi Aspergillus fumigatus Z5 isolated from compost. *Protein expression and purification*, *79*(2), 176-186 (2011).

151. Liu, E., & Hu, Y. Construction of a xylose-fermenting Saccharomyces cerevisiae strain by combined approaches of genetic engineering, chemical mutagenesis and evolutionary adaptation. *Biochemical Engineering Journal*, *48*(2), 204-210 (2010).

152. Lobley, A., Whitmore, L., & Wallace, B. A. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*, *18*(1), 211-212 (2002).

153. Lu, Z., Dunaway-Mariano, D., & Allen, K. N. The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state. *Proceedings of the National Academy of Sciences*, *105*(15), 5687-5692 (2008).

154. Lu, Z., Dunaway-Mariano, D., & Allen, K. N. The X-ray crystallographic structure and specificity profile of HAD superfamily phosphohydrolase BT1666: Comparison of paralogous functions in B. thetaiotaomicron. *Proteins: Structure, Function, and Bioinformatics*, *79*(11), 3099-3107 (2011).

155. Lu, Z., Wang, L., Dunaway-Mariano, D., & Allen, K. N. Structure-function analysis of 2-keto-3-deoxy-D-glycero-D-galactonononate-9-phosphate phosphatase defines specificity elements in type C0 haloalkanoate dehalogenase family members. *Journal of Biological Chemistry*, *284*(2), 1224-1233 (2009).

156. Lukowitz, W., Nickle, T. C., Meinke, D. W., Last, R. L., Conklin, P. L., & Somerville, C. R. Arabidopsis cyt1 mutants are deficient in a mannose-1-phosphate guanylyltransferase and point to a requirement of N-linked glycosylation for cellulose biosynthesis. *Proceedings of the National Academy of Sciences*, *98*(5), 2262-2267 (2001).

157. Lynd, L. R., Weimer, P. J., Van Zyl, W. H., & Pretorius, I. S. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and molecular biology reviews*, *66*(3), 506-577 (2002).

158. Matsutani, K., Fukuda, Y., Murata, K., Kimura, A., & Yajima, N. Adaptation mechanism of yeast to extreme environments: construction of salt-tolerance mutants of the yeast Saccharomyces cerevisiae. *Journal of fermentation and bioengineering*, *73*(3), 228-229 (1992).

159. McBride, J. E., Zietsman, J. J., Van Zyl, W. H., & Lynd, L. R. Utilization of cellobiose by recombinant β-glucosidase-expressing strains of Saccharomyces cerevisiae: characterization and evaluation of the sufficiency of expression. *Enzyme and microbial technology*, *37*(1), 93-101 (2005).

160. McCarter, J. D., & Withers, G. S. Mechanisms of enzymatic glycoside hydrolysis. *Current opinion in structural biology*, *4*(6), 885-892 (1994).

161. Meng, L., and Feldman, L. A rapid TRIzol-based two step method for DNA-free RNA extraction from *Arabodopsis siliques* and dry seeds. Biotechnology J. 5, 183-186 (2010).

162. Merten, J. A., Schultz, K. M., & Klug, C. S. Concentration-dependent oligomerization and oligomeric arrangement of LptA. *Protein Science*, *21*(2), 211-218 (2012).

163. Messner, P. Bacterial glycoproteins. *Glycoconjugate journal*, *14*(1), 3-11 (1997).

164. Minic, Z. Physiological roles of plant glycosyl hydrolases. Planta. 227, 723-740 (2008).

165. Mishima, D., Tateda, M., Ike, M., & Fujita, M. Comparative study on chemical pretreatments to accelerate enzymatic hydrolysis of aquatic macrophyte biomass used in water purification processes. *Bioresource technology*, *97*(16), 2166-2172 (2006).

166. Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., ... & Finn, R. D. The InterPro protein families database: the classification resource after 15 years. *Nucleic acids research*, gku1243 (2014).

167. Monteiro, G., Pereira, G. A. G., & Netto, L. E. S. Regulation of mitochondrial thioredoxin peroxidase I expression by two different pathways: one dependent on cAMP and the other on heme. *Free Radical Biology and Medicine*, *32*(3), 278-288 (2002).

168. Montreuil, J., Schachter, H., & Vliegenthart, J. F. G. Primary structure of glycoprotein glycans. *Glycoproteins I*, *29*, 13 (1995).

169. Morais, M. C., Zhang, G., Zhang, W., Olsen, D. B., Dunaway-Mariano, D., & Allen, K. N. X-ray crystallographic and site-directed mutagenesis analysis of the mechanism of Schiff-base formation in phosphonoacetaldehyde hydrolase catalysis. *Journal of Biological Chemistry*, *279*(10), 9353-9361 (2004).

170. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., & Allen, K. N. The crystal structure of Bacillus cereus phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily.*Biochemistry*, *39*(34), 10385-10396 (2000).

171. Morant, A.V., Jorgensen, K., Jorgensen, C., Paquette, S.M., Sanchez-Perez, R., Moller B.L., and Bak, S. β-Glucosidases as detonators of plant chemical defense. Phytochemistry. 69, 1795–1813 (2008).

172. Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., & Ladisch, M. Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource technology*, *96*(6), 673-686 (2005).

173. Mutter, M., Gassmann, R., Buttkus, U., & Altmann, K. H. Switch Peptides: pH-Induced α-Helix to β-Sheet Transitions of Bis-amphiphilic Oligopeptides. *Angewandte Chemie International Edition in English*, *30*(11), 1514-1516 (1991).

174. Nagai, K., Ihara, Y., Wada, Y., and Taniguchi, N. N-linked Glycosylation is requisite for the enzyme activity and Golgi retention of N-acetylglucosaminyltransferase III. Glycobiology. 7, 769-776 (1997).

175. Narayan, M. Disulfide bonds: protein folding and subcellular protein trafficking. *FEBS Journal*, *279*(13), 2272-2282 (2012).

176. Nevoigt, E. Progress in metabolic engineering of Saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews*, *72*(3), 379-412 (2008).

177. Nomura, Y., Onda, Y., Ohno, S., Taniguchi, H., Ando, K., Oka, N., Nishikawa, K., and Yokogawa, T. Purification and comparision of native and recombinant tRNA-guanine transglycosylases from *Methanosarcina acetivorans*. Protein Expression and Purification. 88, 13-19 (2013).

178. Öberg, F., Sjöhamn, J., Fischer, G., Moberg, A., Pedersen, A., Neutze, R., & Hedfalk, K. Glycosylation increases the thermostability of human aquaporin 10 protein. *Journal of Biological Chemistry*, *286*(36), 31915-31923 (2011).

179. Öhgren, K., Bura, R., Lesnicki, G., Saddler, J., & Zacchi, G. A comparison between simultaneous saccharification and fermentation and separate hydrolysis and fermentation using steam-pretreated corn stover.*Process Biochemistry*, *42*(5), 834-839 (2007).

180. Olsen, O., & Thomsen, K. K. Improvement of bacterial β-glucanase thermostability by glycosylation. *Journal of general microbiology*, *137*(3), 579-585 (1991).

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

182. Pal, R., Cristan, E. A., Schnittker, K., & Narayan, M. Rescue of ER oxidoreductase function through polyphenolic phytochemical intervention: implications for subcellular traffic and neurodegenerative disorders.*Biochemical and biophysical research communications*, *392*(4), 567-571 (2010).

183. Patel, G. K., Kar, B., & Sharma, A. K. Characterization of a thermostable family 1 Glycosyl Hydrolase enzyme from Putranjiva roxburghii seeds. *Applied biochemistry and biotechnology*, *166*(3), 523-535 (2012).

184. Pazy, Y., Motaleb, M. A., Guarnieri, M. T., Charon, N. W., Zhao, R., & Silversmith, R. E. Identical phosphatase mechanisms achieved through distinct modes of binding phosphoprotein substrate. *Proceedings of the National Academy of Sciences*, *107*(5), 1924-1929 (2010).

185. Peisach, E., Selengut, J. D., Dunaway-Mariano, D., & Allen, K. N. X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily. *Biochemistry*, *43*(40), 12770-12779 (2004).

186. Pena, P. V., Davrazou F., Shi X., Walter K., Verkhusha V. V., Gozani O., Zhao R. Kutateladze T. G. Molecular mechanism of H3K4Me3 recognition by Plant Homeodomain of Inhibitor of Growth 2 tumor suppressor. Nature 442, 100-103 (2006).

187. Pradeep, M. A., Jagadeesh, J., De, A. K., Kaushik, J. K., Malakar, D., Kumar, S., ... & Mohanty, A. K. Purification, sequence characterization and effect of goat oviduct-specific glycoprotein on in vitro embryo development. *Theriogenology*, *75*(6), 1005-1015 (2011).

188. Prasad, S., Singh, A., & Joshi, H. C. Ethanol as an alternative fuel from agricultural, industrial and urban residues. *Resources, Conservation and Recycling*, *50*(1), 1-39 (2007).

189. Qian, N. Y., Stanley, G. A., Hahn-Hägerdal, B., & Rådström, P. Purification and characterization of two phosphoglucomutases from Lactococcus lactis subsp. lactis and their regulation in maltose-and glucose-utilizing cells. *Journal of bacteriology*, *176*(17), 5304-5311 (1994).

190. Ramis, C., Gomord, V., Lerouge, P., & Faye, L. Deglycosylation is necessary but not sufficient for activation of proconcanavalin A. *Journal of experimental botany*, *52*(358), 911-917 (2001).

191. Rayon, C., Lerouge, P., & Faye, L. The protein N-glycosylation in plants. *Journal of Experimental Botany*, *49*(326), 1463-1472 (1998).

192. Reed, J., & Kinzel, V. Primary structure elements responsible for the conformational switch in the envelope glycoprotein gp120 from human immunodeficiency virus type 1: LPCR is a motif governing folding. *Proceedings of the National Academy of Sciences*, *90*(14), 6761-6765 (1993).

193. Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Janssen, D. B., & Dijkstra, B. W. Three-dimensional structure of L-2-haloacid dehalogenase from Xanthobacter autotrophicus GJ10 complexed with the substrate-analogue formate. *Journal of Biological Chemistry*, *272*(52), 33015-33022 (1997).

194. Riou, C., Salmon, J. M., Vallier, M. J., Günata, Z., & Barre, P. Purification, characterization, and substrate specificity of a novel highly glucose-tolerant β-glucosidase fromAspergillus oryzae. *Applied and Environmental Microbiology*, *64*(10), 3607-3614 (1998).

195. Robert, X., & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*, *42*(W1), W320-W324 (2014).

196. Rodríguez, M., Pujol, M., Pérez, L., Gavilondo, J. V., Garrido, G., Ayala, M., ... & Borroto, C. Plantas transgénicas de Nicotiana tabacum L. expresan anticuerpo monoclonal aglicosilado con actividad antitumoral. *Biotecnología Aplicada*, *30*(2), 157-161 (2013).

197. Romanos, M. A., Scorer, C. A., & Clare, J. J. Foreign gene expression in yeast: a review. *Yeast*, *8*(6), 423-488 (1992).

198. Ros, M. J., Ramirez, A., Arteaga, E., Alberto, C., Gil, J., & Reina, J. Infection by Staphylococcus lugdunensis: clinico-microbiologic characterization of 25 cases]. *Enfermedades infecciosas y microbiologia clinica*, *17*(5), 223-226 (1999).

199. Rosenheck, K., & Doty, P. The far ultraviolet absorption spectra of polypeptide and protein solutions and their dependence on conformation.*Proceedings of the National Academy of Sciences of the United States of America*, *47*(11), 1775 (1961).

200. Roth, J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chemical reviews*, *102*(2), 285-304 (2002).

201. Rouyi, C., Baiya, S., Lee, S. K., Mahong, B., Jeon, J. S., Ketudat-Cairns, J. R., & Ketudat-Cairns, M. Recombinant Expression and Characterization of the Cytoplasmic Rice β-Glucosidase Os1BGlu4. *PloS one*, *9*(5), e96712 (2014).

202. Sambrook, J. E. F. F., Fritsch, E. F., & Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd edn Cold Spring Harbor. *NY Cold Spring Harbor Laboratory* (1989).

203. Sambrook, J., and Russell, D. W. Molecular Cloning: A Laboratory manual, second and third ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001).

204. Sansenya, S., Opassiri, R., Kuaprasert, B., Chen, C. J., & Cairns, J. R. K. The crystal structure of rice (Oryza sativa L.) Os4BGlu12, an oligosaccharide and tuberonic acid glucoside-hydrolyzing β-glucosidase with significant thioglucohydrolase activity. *Archives of biochemistry and biophysics*, *510*(1), 62-72 (2011).

205. Sanz-Aparicio, J., Hermoso, J.A., Martinez-Ripoll, M., Lequerica, J. L., and Polaina, J. Crystal structure of beta-glucosidase A from Bacillus polymyxa; insights into the catalytic activity in family 1 glycosyl hydrolases. J. Mol. Biol. 175 (3), 491-502 (1998).

206. Sawkar, A. R., Cheng, W. C., Beutler, E., Wong, C. H., Balch, W. E., & Kelly, J. W. Chemical chaperones increase the cellular activity of N370S β-glucosidase: a therapeutic strategy for Gaucher disease. *Proceedings of the National Academy of Sciences*, *99*(24), 15428-15433 (2002).

207. Schwarz, F., & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. Current opinion in structural biology, 21(5), 576-582 (2011).

208. Seal, S. N., & Rose, Z. B. Characterization of a phosphoenzyme intermediate in the reaction of phosphoglycolate phosphatase. *Journal of Biological Chemistry*, *262*(28), 13496-13500 (1987).

209. Sehnem, N. T., da Silva Machado, A., Leite, F. C. B., de Barros Pita, W., de Morais, M. A., & Ayub, M. A. Z. 5-Hydroxymethylfurfural induces ADH7 and ARI1 expression in tolerant industrial Saccharomyces cerevisiae strain P6H9 during bioethanol production. *Bioresource technology*, *133*, 190-196 (2013).

210. Seshadri, S., Akiyama, T., Opassiri, R., Kuaprasert, B., & Cairns, J. K. Structural and enzymatic characterization of Os3BGlu6, a rice β-glucosidase hydrolyzing hydrophobic glycosides and (1→ 3)-and (1→ 2)-linked disaccharides. *Plant physiology*, *151*(1), 47-58 (2009).

211. Shewale, J. G. β-Glucosidase: its role in cellulase synthesis and hydrolysis of cellulose. *International Journal of Biochemistry*, *14*(6), 435-443 (1982).

212. Sjostrom, E. *Wood chemistry: fundamentals and applications*. Elsevier. (1993).

213. Smith, D.L., Starrett, D.A., and Gross, K.C. A gene coding for tomato fruit *β*-glucosidase II is expressed during fruit ripening. Plant Physiol. 117,417– 423 (1998).

214. Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, *12*(4), 43R-56R (2002).

215. Stepanova, E. V., Landesman, E. O., Fedorova, T. V., Yakovleva, K. E., & Koroleva, O. V. Comparison of the efficiency of polygalacturonase and β-glucosidase enzyme preparations in stabilization of cherry plum wine material. *Applied Biochemistry and Microbiology*, *42*(6), 609-615 (2006).

216. Sternberg, D., Vuayakumar, P., & Reese, E. T. β-Glucosidase: microbial production and effect on enzymatic hydrolysis of cellulose. *Canadian Journal of Microbiology*, *23*(2), 139-147 (1977).

217. Stockton, B. C., Mitchell, D. J., Grohmann, K., & Himmel, M. E. Optimumβ-D-glucosidase supplementation of cellulase for efficient conversion of cellulose to glucose. *Biotechnology letters*, *13*(1), 57-62 (1991).

218. Sukumaran, R. K., Singhania, R. R., & Pandey, A. (2005). Microbial cellulases-production, applications and challenges. *Journal of Scientific and Industrial Research*, *64*(11), 832 (1993).

219. Sukumaran, R. K., Singhania, R. R., Mathew, G. M., & Pandey, A. Cellulase production using biomass feed stock and its application in lignocellulose saccharification for bio-ethanol production. *Renewable Energy*, *34*(2), 421-424 (2009).

220. Sun, L., Gurnon, J. R., Adams, B. J., Graves, M. V., & Van Etten, J. L. Characterization of a β-1, 3-glucanase encoded by chlorella virus PBCV-1.*Virology*, *276*(1), 27-36 (2000).

221. Surchandra, T., Roy, S. S., Singh, N. R., Sahoo, M. R., & Prakash, N. Partial purification and biochemical characterization of acid phosphatase from germinated mung bean (Vigna radiata) seeds (2012).

222. Swanson, P. E. Dehalogenases applied to industrial-scale biocatalysis.*Current opinion in biotechnology*, *10*(4), 365-369 (1999).

223. Sylvestre, M., Macek, T., & Mackova, M. Transgenic plants to improve rhizoremediation of polychlorinated biphenyls (PCBs). *Current opinion in biotechnology*, *20*(2), 242-247 (2009).

224. Takashima, S., Nakamura, A., Hidaka, M., Masaki, H., & Uozumi, T. Molecular cloning and expression of the novel fungal β-glucosidase genes from Humicola grisea and Trichoderma reesei. *Journal of biochemistry*, *125*(4), 728-736 (1999).

225. Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, *30*(12), 2725-2729 (2013).

226. Tarentino, A. L., Gomez, C. M., & Plummer Jr, T. H. Deglycosylation of asparagine-linked glycans by peptide: N-glycosidase F. *Biochemistry*, *24*(17), 4665-4671 (1985).

227. Thompson, J. D., Higgins, D. G., & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.*Nucleic acids research*, *22*(22), 4673-4680 (1994).

228. Tomar, A. K., Sooch, B. S., Raj, I., Singh, S., Singh, T. P., & Yadav, S. Isolation and identification of Concanavalin A binding glycoproteins from human seminal plasma: a step towards identification of male infertility marker proteins. *Disease markers*, *31*(6), 379-386 (2011).

229. Tremblay, L. W., Dunaway-Mariano, D., & Allen, K. N. Structure and activity analyses of Escherichia coli K-12 NagD provide insight into the evolution of biochemical function in the haloalkanoic acid dehalogenase superfamily. *Biochemistry*, *45*(4), 1183-1193 (2006).

230. Tribble, G. D., Mao, S., James, C. E., & Lamont, R. J. A Porphyromonas gingivalis haloacid dehalogenase family phosphatase interacts with human phosphoproteins and is important for invasion. *Proceedings of the National Academy of Sciences*, *103*(29), 11027-11032 (2006).

231. Tripathi, P., Kumari, A., Rath, P., & Kayastha, A. M. Immobilization of α-amylase from mung beans (Vigna radiata) on Amberlite MB 150 and chitosan beads: A comparative study. *Journal of Molecular Catalysis B: Enzymatic*, *49*(1), 69-74 (2007).

232. Trombetta, E. S. The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. *Glycobiology*, *13*(9), 77R-91R (2003).

233. Uchiyama, T., Miyazaki, K., & Yaoi, K. Characterization of a novel β-glucosidase from a compost microbial metagenome with strong transglycosylation activity. *Journal of Biological Chemistry*, *288*(25), 18325-18334 (2013).

234. Vallmitjana, M., Ferrer-Navarro, M., Planell, R., Abel, M., Ausín, C., Querol, E., Planas., A., & Pérez-Pons, J. A. Mechanism of the family 1 β-glucosidase from Streptomyces sp: catalytic residues and kinetic studies. *Biochemistry*, *40*(20), 5975-5982 (2001).

235. Van der Ploeg, J., Van Hall, G., & Janssen, D. B. Characterization of the haloacid dehalogenase from Xanthobacter autotrophicus GJ10 and sequencing of the dhlB gene. *Journal of bacteriology*, *173*(24), 7925-7933 (1991).

236. Van Rooyen, R., Hahn-Hägerdal, B., La Grange, D. C., & Van Zyl, W. H. Construction of cellobiose-growing and fermenting Saccharomyces cerevisiae strains. *Journal of biotechnology*, *120*(3), 284-295 (2005).

237. van Zyl, W. H., Lynd, L. R., den Haan, R., & McBride, J. E. Consolidated bioprocessing for bioethanol production using Saccharomyces cerevisiae. In *Biofuels* (pp. 205-235). Springer Berlin Heidelberg (2007).

238. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, *3*(2), 97-130 (1993).

239. Verdoucq, L., Morinière, J., Bevan, D. R., Esen, A., Vasella, A., Henrissat, B., & Czjze, M. Structural Determinants of Substrate Specificity in Family 1 β-Glucosidases novel insights from the crystal structure of sorghum dhurrinase-1, a plant β-glucosidase with strict specificity, in complex with its natural substrate. *Journal of Biological Chemistry*, *279*(30), 31796-31803 (2004).

240. Verjee, Z. H. M. Isolation of three acid phosphatases from wheat germ.*European Journal of Biochemistry*, *9*(3), 439-444 (1969).

241. Voorhorst, W. G., Eggen, R. I., Luesink, E. J., & De Vos, W. M. Characterization of the celB gene coding for beta-glucosidase from the hyperthermophilic archaeon Pyrococcus furiosus and its expression and site-directed mutation in Escherichia coli. *Journal of bacteriology*, *177*(24), 7105-7111(1995).

242. Wakuta, S., Hamada, S., Ito, H., Matsuura, H., Nabeta, K., & Matsui, H. Identification of a β-glucosidase hydrolyzing tuberonic acid glucoside in rice (Oryza sativa L.). *Phytochemistry*, *71*(11), 1280-1288 (2010).

243. Walker, D. E., and Axelrod, B. Evidence for a single catalytic site on the "beta-D-glucosidase-beta-D-galactosidase" of almond emulsion. Arch Biochem Biophys. 187 (1), 102-107 (1978).

244. Wallnfels, K., and Malhotra, O.P. Galactosidases. Adv. Carbohydrate. Chem. 16, 239-298 (1961).

245. Walsh, C., & Walsh, C. *Posttranslational modification of proteins: expanding nature's inventory* (pp. 1-490). Englewood, CO: Roberts and Company Publishers (2006).

246. Wang, Q., Trimbur, D., Graham, R., Warren, R. A. J., & Withers, S. G. Identification of the Acid/Base Catalyst in Agrobacterium faecalis. beta.-Glucosidase by Kinetic Analysis of Mutants. *Biochemistry*, *34*(44), 14554-14562 (1995).

247. Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H., ... & Kim, S. H. Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states. *Journal of molecular biology*, *319*(2), 421-431 (2002).

248. Wang, W., Kim, R., Jancarik, J., Yokota, H., & Kim, S. H. Crystal structure of phosphoserine phosphatase from Methanococcus jannaschii, a hyperthermophile, at 1.8 Å resolution. *Structure*, *9*(1), 65-71 (2001).

249. Waterhous, D. V., & Johnson Jr, W. C. Importance of environment in determining secondary structure in proteins. *Biochemistry*, *33*(8), 2121-2128 (1994).

250. Wati, L., Dhamija, S. S., Singh, D., Singh-Nee Nigam, P., & Marchant, R. Characterisation of genetic control of thermotolerance in mutants of Saccharomyces cerevisiae. *GENETIC ENGINEER AND BIOTECHNOLOGIST*, *16*(1), 19-26 (1996).

251. Wierzbicka-Woś, A., Bartasun, P., Cieśliński, H., & Kur, J. Cloning and characterization of a novel cold-active glycoside hydrolase family 1 enzyme with β-glucosidase, β-fucosidase and β-galactosidase activities. *BMC biotechnology*, *13*(1), 22 (2013).

252. Wierzbicka-Wos, A., Cieslinski, H., Wanarska, M., Kozłowska-Tylingo, K., Hildebrandt, P., & Kur, J. A novel cold-active bD-galactosidase from the Paracoccus sp. 32d-gene cloning, purification and characterization. *Microb Cell Fact*, *10*, 108 (2011).

253. Wilkins, M. R., Suryawati, L., Maness, N. O., & Chrz, D. Ethanol production by Saccharomyces cerevisiae and Kluyveromyces marxianus in the presence of orange-peel oil. *World Journal of Microbiology and Biotechnology*, *23*(8), 1161-1168 (2007).

254. Withers, S. G., Rupitz, K., Trimbur, D., & Warren, R. A. J. Mechanistic consequences of mutation of the active site nucleophile Glu 358 in Agrobacterium. beta.-glucosidase. *Biochemistry*, *31*(41), 9979-9985 (1992).

255. Withers, S. G., Warren, R. A. J., Street, I. P., Rupitz, K., Kempton, J. B., & Aebersold, R. Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a retaining glycosidase. *Journal of the American Chemical Society*, *112*(15), 5887-5889 (1990).

256. Wolfenden, R., Lu, X., & Young, G. Spontaneous hydrolysis of glycosides. *Journal of the American Chemical Society*, *120*(27), 6814-6815 (1998).

257. Wyman, C. E. Biomass ethanol: technical progress, opportunities, and commercial challenges. *Annual Review of Energy and the Environment*, *24*(1), 189-226 (1999).

258. Yadav, S. C., Jagannadham, M. V., & Kundu, S. Equilibrium unfolding of kinetically stable serine protease milin: the presence of various active and inactive dimeric intermediates. *European Biophysics Journal*, *39*(10), 1385-1396 (2010).

259. Yamaguchi, T., Nakayama, K., Hayashi, T., TANAKA, Y., & KOIKE, S. Molecular cloning and characterization of a novel β-1, 3-glucanase gene from rice. *Bioscience, biotechnology, and biochemistry*, *66*(6), 1403-1406 (2002).

260. Yan, Q., Zhou, W., Li, X., Feng, M., & Zhou, P. Purification method improvement and characterization of a novel ginsenoside-hydrolyzing β-glucosidase from *Paecilomyces bainier* sp. 229. *Bioscience, biotechnology, and biochemistry*, *72*(2), 352-359 (2008).

261. Yang, J., Bae, J. Y., Lee, Y. M., Kwon, H., Moon, H. Y., Kang, H. A., ... & Choi, W. Construction of *Saccharomyces cerevisiae* strains with enhanced ethanol tolerance by mutagenesis of the TATA-binding protein gene and identification of novel genes associated with ethanol tolerance. *Biotechnology and bioengineering*, *108*(8), 1776-1787 (2011).

262. Yang, S., Wang, L., Yan, Q., Jiang, Z., & Li, L. Hydrolysis of soybean isoflavone glycosides by a thermostable β-glucosidase from *Paecilomyces thermophila*. *Food Chemistry*, *115*(4), 1247-1252 (2009).

263. Yazawa, H., Iwahashi, H., & Uemura, H. Disruption of URA7 and GAL6 improves the ethanol tolerance and fermentation capacity of *Saccharomyces cerevisiae*. *Yeast*, *24*(7), 551-560 (2007).

264. Young, N. M., Brisson, J. R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., ... & Szymanski, C. M. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *Journal of Biological Chemistry*, *277*(45), 42530-42539 (2002).

265. Zechel, D. L., & Withers, S. G. Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Accounts of Chemical Research*, *33*(1), 11-18 (2000).

266. Zhang, G., Mazurkie, A. S., Dunaway-Mariano, D., & Allen, K. N. Kinetic evidence for a substrate-induced fit in phosphonoacetaldehyde hydrolase catalysis. *Biochemistry*, *41*(45), 13370-13377 (2002).

267. Zhang, G., Morais, M. C., Dai, J., Zhang, W., Dunaway-Mariano, D., & Allen, K. N. Investigation of metal ion binding in phosphonoacetaldehyde hydrolase identifies scquence markers for metal-activated enzymes of the HAD enzyme superfamily. *Biochemistry*, *43*(17), 4990-4997 (2004).

268. Zhao, R., Collins, E. J., Bourret, R. B. & Silversmith, R. E. Structure and catalytic mechanism of the *E. Coli* chemotaxis phosphatase chez. *Nature structural & molecular biology*, *9*(8), 570-575 (2002).

269. Zhao, S., Zhang, Z., & Lee, P. C. Comparison of the enzymatic activities of native and recombinant protein phosphatise-1 toward histone. *Biochemistry and Molecular Biology International*, *34*(5), 1027-1033 (1994).

270. Zheng, K., Bantog, C., & Bayer, R. The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs*, *3*(6), 568-576 (2011).