

**COMPUTATIONAL STUDIES ON SUBSTRATE
SPECIFICITY IN
LYTIC POLYSACCHARIDE MONOOXYGENASES**

*Dissertation submitted in partial fulfillment of the requirement for the
degree of*

BACHELOR OF TECHNOLOGY

IN

BIOINFORMATICS

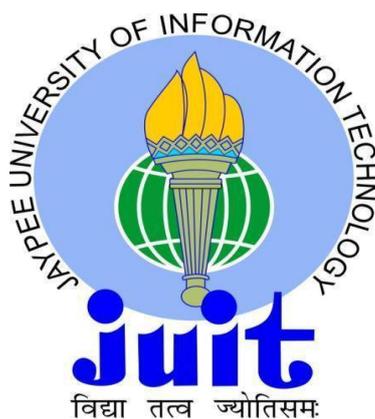
By

RADHIKA ARORA

141503

UNDER THE GUIDANCE OF

DR. RAGOTHAMAN M. YENNAMALLI



JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

May 2018

TABLE OF CONTENTS

	Page Number
DECLARATION BY THE SCHOLAR	iv
SUPERVISOR'S CERTIFICATE	v
ACKNOWLEDGEMENT	vi
LIST OF ACRONYMS AND ABBREVIATIONS	vii
LIST OF FIGURES	viii
LIST OF FLOWCHARTS	ix
LIST OF TABLES	x
ABSTRACT	xi
CHAPTER-1	
INTRODUCTION	1
1.1 Relevance of this Research	1
1.2 Cellulose	1
1.3 Lytic Polysaccharide Monooxygenases	2
1.3.1 Types of LPMO	3
1.3.1.1 Auxiliary Activity	3
1.4 LPMO and substrate interactions	5
Chapter-2	
Hypothesis	7
Chapter-3	
Materials and Methods	8
3.1 CAZy	9
3.2 PDB	9
3.3 cellulose-builder	9
3.4 AutoDock	9

3.4.1 Steps of AutoDock	10
3.5 PyMOL	10
Chapter-4	
Results and Discussion	12
4.1 CAZy and PDB results	12
4.2 Result of cellulose-builder	22
4.3 Results of AutoDock	22
Chapter-5	
Conclusion	24
Chapter-6	
References	25
Appendix A	29
A1: cellulose-builder	

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the B.Tech. project report entitled **“Computational Studies on Substrate Specificity in Lytic Polysaccharide Monooxygenases”** submitted at **Jaypee University of Information Technology, Wagnaghat, India**, is an authentic record of my work carried out under the supervision of **Dr. Ragothaman M. Yennamalli**. I have not submitted this work elsewhere for any other degree or diploma.

(Signature of the Scholar)

Radhika Arora (141503)

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology, Wagnaghat, India

Date:

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the B.Tech. project report entitled “**Computational Studies on Substrate Specificity in Lytic Polysaccharide Monooxygenases**”, submitted by **Radhika Arora (141503)** at **Jaypee University of Information Technology, Wagnaghat, India** is a bonafide record of her original work carried out under my supervision. The work has not been submitted elsewhere for any other degree or diploma.

(Signature of Supervisor)

Dr. Ragothaman M. Yennamalli

Assistant Professor (Grade II)

Date:

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to all those who provided me the possibility to complete this report. A special gratitude I give to my final year project supervisor, Dr. Ragothaman M. Yennamalli whose contribution in stimulating suggestions and encouragement helped me to coordinate our project especially in writing this report. Furthermore, i would also like to acknowledge with much appreciation the crucial role of the staff of Bioinformatics lab, who gave the permission to use all the required equipments and the necessary material to complete the task. Last but not least, I have to appreciate the guidance given by other faculty member of Bioinformatics and the evaluation panels especially in our project presentation that has improved our communication skills.

List of Acronyms and Abbreviations

AA – Auxiliary Activities

CAZy – Carbohydrate Active Enzymes

CBM33 – Carbohydrate Binding Module family 33

CDH – cellobiose dehydrogenase

GH61 – Glycoside Hydrolases family 61

LPMO – Lytic Polysaccharide Monooxygenases

MD – Molecular Dynamics

PDB – Protein Data Bank

Tcl – Tool Command Language

LIST OF FIGURES

Figure Number	Caption
1.1	Schematic representation of microfibril and cellulose
1.2	Schematic representation of Cellulose showing the hydrogen bonds (dashed)
1.3	Schematic representation of crystal structure of Cellulose. (a) side view (b) top view. Red, cyan and gray spheres represent the oxygen, hydrogen and carbon atoms
1.4	Three types of LPMOs classified based on the site of attack
1.5	LPMOs and other cellulolytic enzyme's synergistic activity on breaking down the crystalline cellulose
1.6	LPMO substrate binding and their conserved residues
4.1	Structure of Cellulose was obtained from cellulose-builder
4.2	Top and side view of 4B5Q (<i>Phanerochaete chrysosporium</i>) and cellulose
4.3	Top and side view of 5ACF (<i>Lentinus similis</i>) and cellulose
4.4	Top and side view of 2YET (<i>Thermoascus aurantiacus</i>) and cellulose

LIST OF FLOWCHARTS

Flowchart Number	Caption	Page Number
3.1	Methodology	8
3.2	Steps of AutoDock	11

LIST OF TABLES

Table Number	Caption	Page Number
4.1	PDB ID for AA9, AA10, AA11, and AA13	12
4.2	LPMO's substrate	14
4.3	Selected substrate on the basis of type 1, 2, and 3	22
4.4	Time taken by a AutoDock for Receptor and Ligand docking	23

ABSTRACT

Plant cell wall is composed of cellulose, hemicellulose, and lignin in tight covalent linkages and non-covalent bonds that are difficult to degrade in its crystalline form. A class of enzymes, namely Lytic Polysaccharide Monooxygenases (LPMOs) produced by bacteria, fungi, and viruses, was discovered that can break an internal glycosidic bond by binding to crystalline surface of cellulose. Thus, it helps to degrade the cell wall. However, there is regiospecificity in LPMOs as to which carbon (C1 and/or C4) is under attack by LPMOs. This regiospecificity is thought to be encoded in the substrate binding residues of LPMOs. Thus, it becomes necessary for a polysaccharide to bind with the specific site of LPMO.

Here, using computations methods we have carried out studies on substrate specificity and structural changes in LPMO. A starting optimized conformation of cellulose and LPMO is used in order for evaluating conformational changes using protein-carbohydrate force field parameters.

CHAPTER 1

INTRODUCTION

1.1 Relevance of this research

Many countries are taking initiatives for decreased use of fossil fuels, as fossil fuels is a non-renewable resource, not eco friendly, and proposed to have a limited existence in our earth. The most important aspect is that their consumption leads to emission of carbon dioxide results in global warming and climate change. Therefore, it is necessary to take alternate fuels, one of which is plant biomass that can provide a good fuel. Biomass is a renewable resource and it is obtained from dead or living plants. It can be made available from the forest waste, agricultural waste, and agro- industrial waste.

Biomass is considered as the major source of energy and their contribution towards supply of energy being around 10-14% in the earth [1]. Plant cell wall is mainly composed of middle lamella, primary, and secondary cell wall. The lignocellulosic biomass components are: cellulose (40–50%), hemicelluloses (20-40%), lignin (20-30%), proteins, pectin, lipids, soluble sugar, and minerals [1]. It is an alternative for petrol to obtain fuels due to its low price, and obtained from agricultural waste [1]. The lignocellulosic biomass is composed of covalent linkages and non-covalent bonds that are difficult to degrade in its crystalline form.

1.2 Cellulose

Cellulose ($C_6H_{10}O_5$)_n is the most abundant polymer in the world and helps in the support and protection of the plant cell wall. Cellulose is a linear polymer of β -(1 \rightarrow 4)-linked glucan chains that are aggregated by hydrogen bonds and van der Waals forces to form microfibrils [2] (Figure 1.1 and 1.2) [3, 4]. Many glucan chains are aligned in parallel and form one sheet. The cellulose structure shows a high degree of polymorphism. The interaction between the glucan chains and the packing is not completely uniform within cellulose, which leads to the formation of amorphous and crystalline regions (Figure 1.3) [5]. As crystalline part of a molecule is difficult to break, they need to be either converted to amorphous form or degraded in its crystalline form.

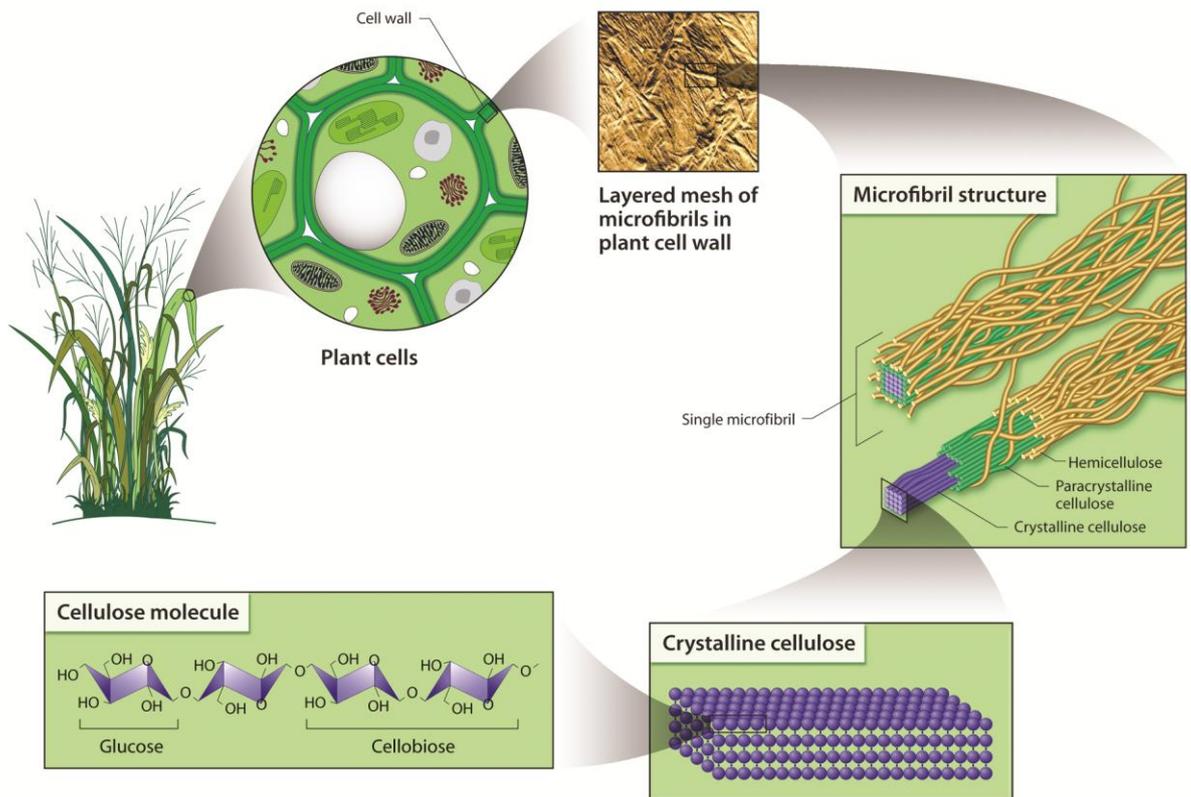


Figure 1.1: Schematic representation of Microfibril and Cellulose. Image adapted from [3].

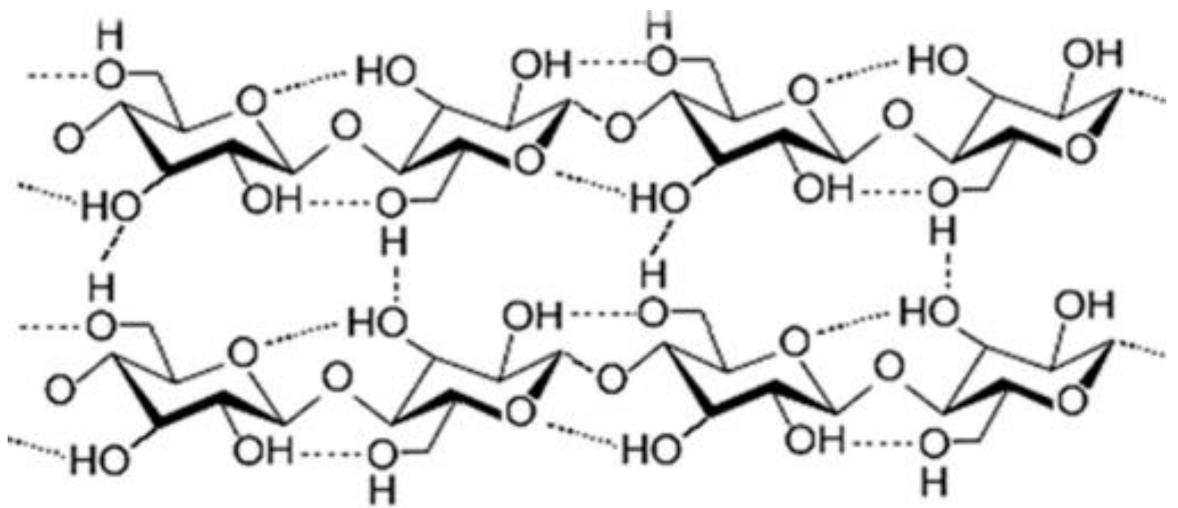


Figure 1.2: Schematic representation of Cellulose showing the hydrogen bonds (dashed). Image adapted from [4].

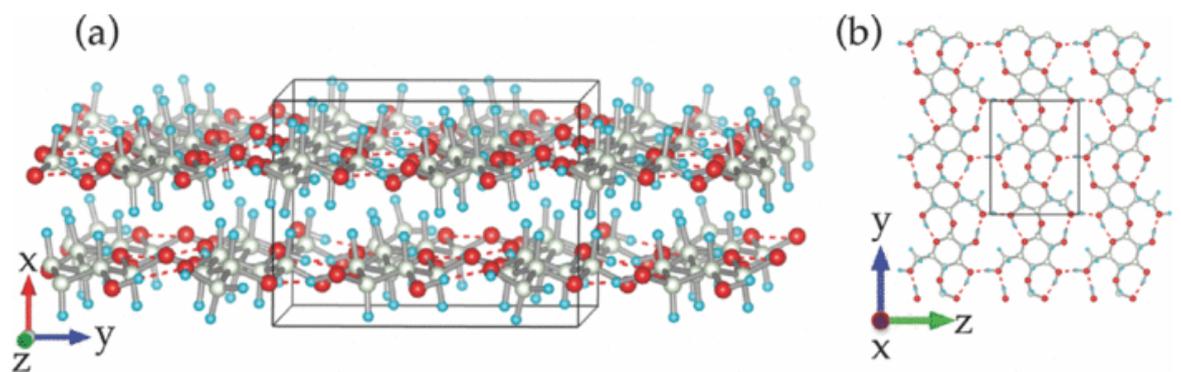


Figure 1.3: Schematic representation of crystal structure of Cellulose. (a) side view (b) top view. Red, cyan and gray spheres represent the oxygen, hydrogen and carbon atoms. Image adapted from [5].

1.3 Lytic Polysaccharide Monooxygenases

Lytic Polysaccharide Monooxygenases (LPMO) whereby Lytic means destruction and polysaccharide means (poly = many, saccharide = sugar) carbohydrates formed by the repeated units connected with the help of glycosidic bonds. LPMOs are enzymes that initiate the insertion of a single oxygen atom from O₂ into a substrate and thus break in the glycosidic bond. The redox reaction which occurs is a copper dependent reaction. LPMO is produced by microorganisms that degrade biomass and assist in the degradation of recalcitrant cellulose and/or chitin by carrying out the oxidative cleavage of glycosidic bonds [6]. LPMOs main role is to create new chain termini in the interior of the cellulose fibril [7].

In 2005, *Serratia marcescens* was reported as the first bacteria that secretes LPMO. *Lentinus similis* was reported as the first fungi that express LPMO with bound oligosaccharides. In 2016, *Bacillus thuringiensis* was the first structure of bacterial LPMOs that shows the properties of antifungal whereby it kills or inactivate fungi [8]. In 2010, the outcome of polysaccharide degradation by LPMO was discovered [9]. The first LPMO structure of wood-degrading model organism, *Phanerochaete chrysosporium* was solved from a basidiomycete fungus. This fungus contains up to 9 LPMO genes that are known to express LPMOs when grown on lignocellulosic substrates. Expression of these LPMO genes in *P. chrysosporium* was observed when the fungus grown on the lignocelluloses substrate [10].

Apart from these, there are other LPMOs whose structures are known. Fungal AA9 LPMO whose structures has been solved are: *H. jecorina* GH61B (PDB ID: 2VTC), *Thermoascus aurantiacus* GH61A (PDB ID: 2YET and 3ZUD), *Thielavia terrestris* GH61E (PDB ID: 3EII and 3EJA), and *Neurospora crassa* LPMO2 and LPMO3 (PDB ID: 4EIR and 4EIS). LPMO structures of CBM33 enzymes are: *Serratia marcescens* CBP21(PDB ID: 2LHS, 2BEM, and 2BEN), *Vibrio cholerae* CBM33 (PDB ID: 2XWX), *Enterococcus faecalis* CBM33 (PDB ID: 4A02, 4ALC, 4ALE, 4ALQ, 4ALR, 4ALS, 4ALT), *Burkholderia pseudomallei* CBM33 (PDB ID: 3UAM) [10], *Bacillus amyloliquefaciens* CBM33 (PDB ID: 2YOW, 2YOX, 2YOY, 5IJU), *Streptomyces coelicolor* CBM2 (PDB ID: 4OY7), *Cellvibrio japonicas* CBP33 (PDB ID: 5FJQ).

The recent research of Voshol et al identified a new family in LPMO i.e. LPMO14, in which the species of *Aspergillus* was examined to contain LPMO14 encoding genes [11]. LPMO14 have been predicted to have substrate specificity with pectin and glucan [11]. This result has not been yet confirmed experimentally, yet and the existence of AA14 is still debated.

1.3.1 Types of LPMO

LPMOs are classified into different types on the basis of their site of attack [12]. There are two types of reactions based on the hydrogen abstraction at the different carbon atom positions (either C1 or C4) followed by glycosidic (C-O) bond cleavage, where enzymes specifically attacking C1 are called as type 1 LPMOs forming Lactone and Aldonic acid as products and enzymes specifically attacking C4 are called as type 2 LPMOs forming Ketoaldose and Geminal idol as products [12]. If some enzymes do not have specificity to either C1 or C4, they are termed as type 3 LPMOs (Figure 1.4).

The method of High performance Anion Exchange with Pulsed Amperometric Detection (HPAEC-PAD) is used for the detection of C1 and C4-oxidized sugar, where HPAEC is porous graphitized carbon (PCG) chromatography that allows the separation of C1 and C4, and detection is done by using Mass Spectrometry (MS) [9].

1.3.1.1 Auxiliary Activity

Carbohydrate-Active enZYmes (CAZy) is a database that describes the enzymes families of structurally related catalytic and carbohydrate-binding modules. It gives information about enzymes that modify, degrade or helps in the creation of glycosidic bonds. CAZy is a database that analysis the structural, genomic, and biochemical information on CAZymes. CAZy decided to categorize the families of LPMO to start a new CAZy class named as “Auxiliary Activities (AA)” [4]. Glycoside Hydrolases family 61 (GH61) was renamed as AA9 and Carbohydrate Binding Module family 33 (CBM33) was renamed as AA10 [14]. Not all CBM33 structures contain a metal ion in the putative catalytic center of enzymes, whereas in GH61 all contain a metal ion [10].

Auxiliary Activity (AA) cover redox enzymes that act in conjunction with CAZymes. Four LPMO families are: AA9, AA10, AA11, and AA13. Total structures for LPMO families are AA9 (14), AA10 (18), AA11 (1), and AA13 (1), respectively in the CAZy

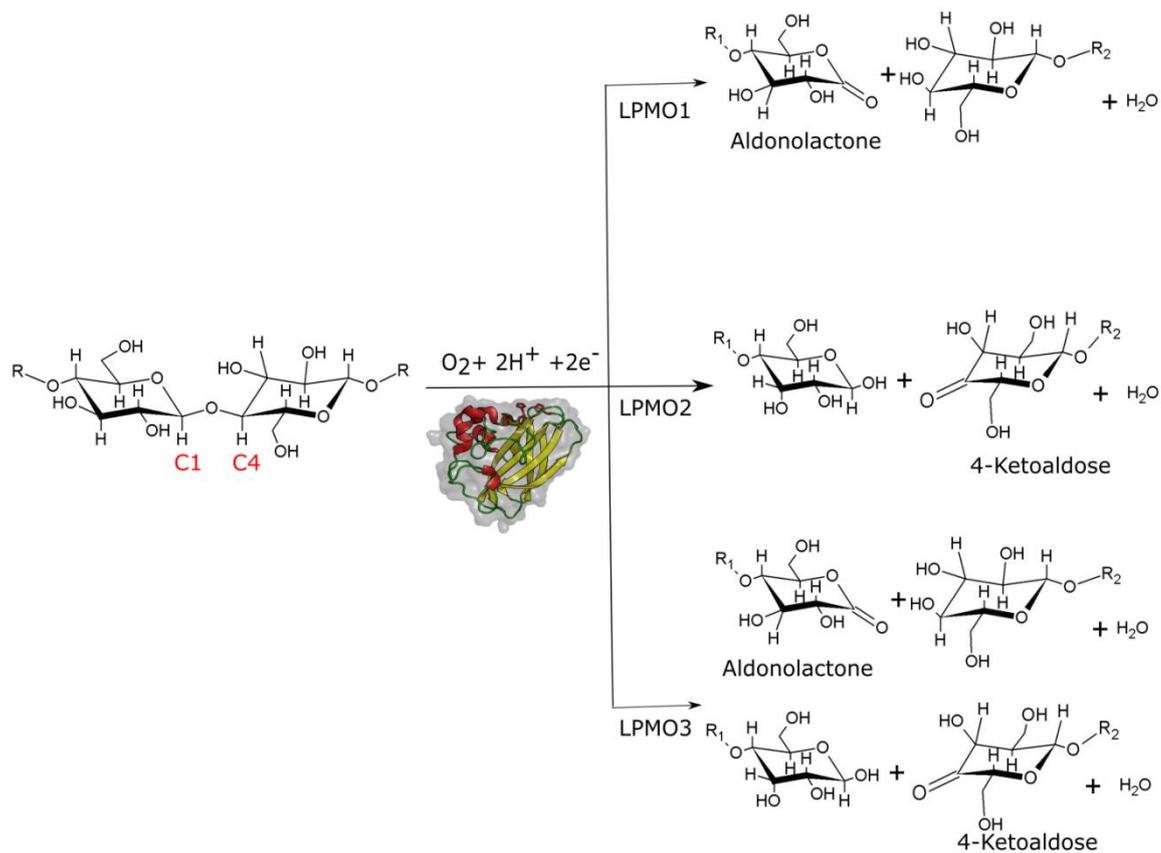


Figure 1.4: Three types of LPMOs classified based on the site of attack. C1 attacking site are known as type 1 LPMOs; C4 attacking site are known as type 2 LPMOs; and C1/C4 both attacking site are known as type 3 LPMOs. Image adapted from [12].

database. As of May 2018, the AA9, AA10, AA11, and AA13 families consist respectively of 385, 3337, 99, and 18 sequences in the CAZy database [4]. All fungal LPMOs belong to AA9, AA11, and AA13 while the AA10 family is found only in bacteria. AA9 of fungal LPMOs share a common ancestor with bacterial LPMOs [8].

Structure of LPMO shows the state of reduction, where three ligands of nitrogen, coordinated by the copper atom in a geometry of T-shaped [15]. InterPro database determine families of PMOs on the basis of correlated sequence [8]. The absence of sequence similarity but the presence of structural similarity of AA9 and AA10, suggests that both share the ancestral protein [16]. In general, very less number is conserved of amino acids among the AA9, AA10, AA11, and AA13 [11].

β -sandwich fold is present in LPMO structure and there is an area of flat binding surface of substrate with polar and aromatic residues [9]. In 2016, Moses et al showed that the type 1 LPMOs displayed highest aromaticity while type 2 LPMOs showed the lowest aromaticity [17].

Modular proteins of AA9 contain AA9 domain coupled with different CBMs. Substrate binding area lacks in the AA9 domain lacks substrate binding area. The binding site's unavailability results into the reclassification of the AA enzymes as opposed to GHs [17].

The main chain amino group of the N terminal histidine (His1) and imidazole side chain contribute two of its nitrogens and the second conserved histidine contributes the third nitrogen. Copper ions lessen dioxygen, which take electrons from an external electron donor, such as a reducing agent that is either provided by the substrate or by a co-secreted enzyme called cellobiose dehydrogenase (CDH) [12]. CDH consider as the electron donor for bacterial or fungal PMOs. N-terminal His1, hydrogen bonds, and water, act as a pathway that transfer proton and further it helps in the stabilization of Cu-O₂ [8]. The rate of activity of LPMO increases with the help of a light. Figure 1.5 depicts the schematic representation of LPMO's synergism with other enzymes.

Cellulose is a crystalline flat surface and LPMO is the one who attacks in crystalline cellulose and break the bonds of glycosidic and then the other enzymes come and bind on an amorphous region and break it into smaller parts [9]. The polymers contain

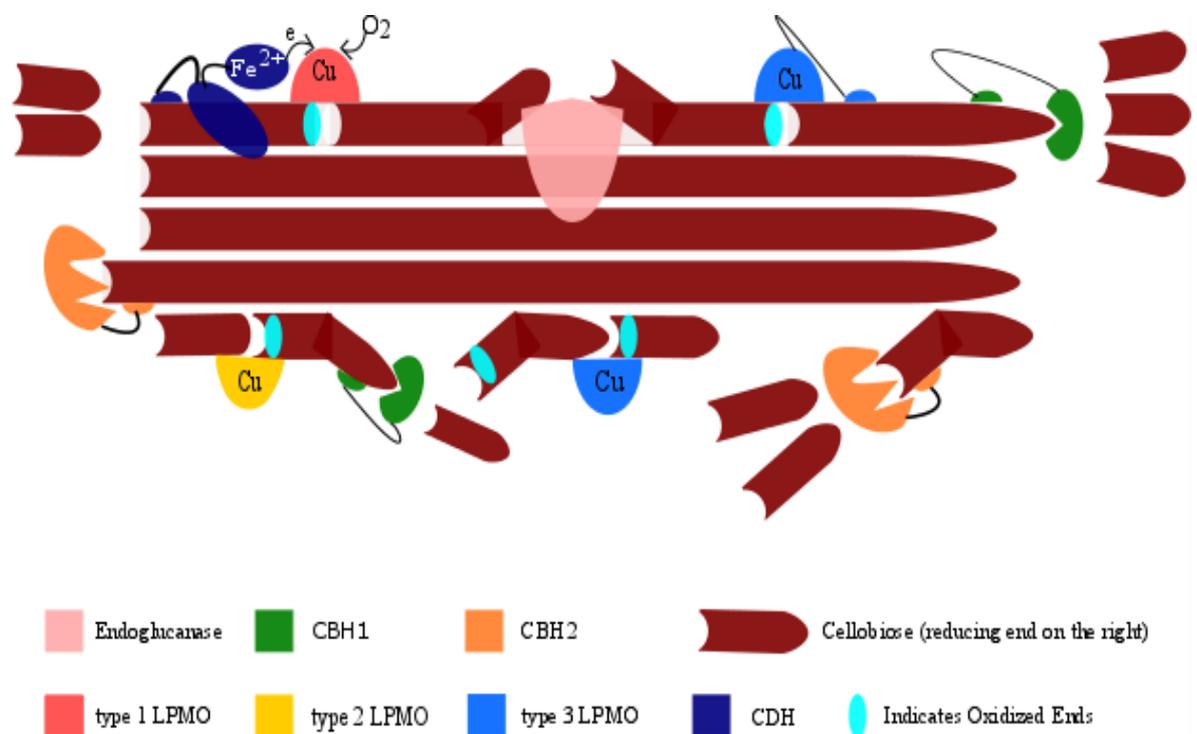


Figure 1.5: LPMOs and other cellulolytic enzyme's synergistic activity on breaking down the crystalline cellulose. Removal of hydrogen atom (bound to either from C1 or C4) creates an electron imbalance in the glycosidic bond. For maintaining the balance, it leads to the release of oxygen and results in the breakage of the glycosidic bond.

domains, include CBM and GH. The increase in the substrate binding promote by the protein domain that contain aromatic residues. The mechanism behind this is not determined [8].

In 2017, Agostoni et al work shows that chitin and cellulose can be sensed by the bacteria with the help of DasR and CebR regulator. Both regulates LPMO's expression. *Xylanimonas cellulositytica* secrete two LPMOs, isolated from decomposed plant parts, which help in the degradation of xylan and cellulose. Many bacteria get nutrient by degradation of polysaccharide. Interaction of host organism and such bacteria implement a feeding strategy. By attacking to zooplankton, *Vibrio cholera* degrades the chitin and produces excess levels of ammonia, resulting in toxic levels among cattles. The interaction of both helps in the larva development, (in the case of insect that act as the host organism) as it takes the energy form the plant cell wall during degradation process [8].

1.5 LPMO and substrate interactions

Due to the binding of copper LPMO structure gets stabilized. Residues involves with the substrate interaction were recognized by NMR, site-directed mutagenesis and X-ray crystallography [15].

Substrate specificity means substrate binds with a specific enzyme to give a specific product [18]. Structural determinants leading to LPMO's substrate specificity are still not determined. There is still no evidence against the changed substrate specificity of LPMOs [15].

Aromatic residues play a significant role in terms of substrate binding. A very small change in position of the orientation of the substrate could impact the balance of C1/C4-oxidation. More aromaticity is shown by the LPMO that oxidize C1. LPMO regioselectivity is determined by the composition of aromatic residues that resides on the surface [9]. Figure 1.6 represents the LPMO substrate binding and their conserved residues [15].

In 2013, Wu et al work showed that MD simulation on PchGH61D conducted for 100 ns, from where the role of the three tyrosine residues (Tyr 28, Tyr 75, and Tyr 198) was

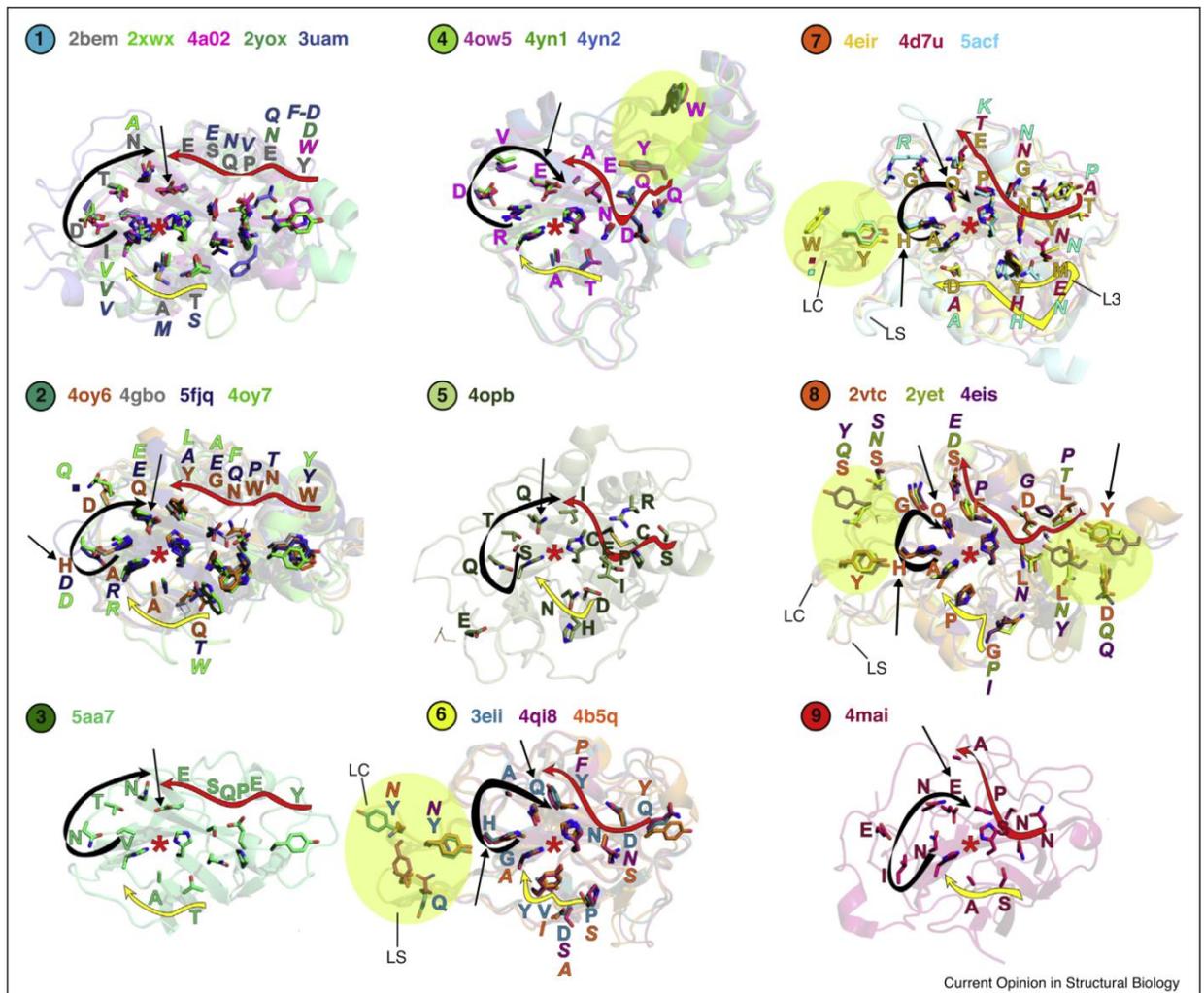


Figure 1.6: LPMO substrate binding and their conserved residues. Red star depicts the structures that were taken on account of histidine-brace/copper center. Sticks represent the side chain. Yellow, black and red arrows indicates environment of catalytic center and it also contain some specific residues that are conserved in nature. The yellow area connects beta-strands 3 and 4 in the core beta-sandwich and contains the second catalytic histidine; some LPMOs have an insertion here, mention as L3. The red area is part of the L2 loop. The black area connects the 2 last β -strands of the β -sandwich in the 9 clusters. Image adapted from [15].

examined and these may be considered as important residues for binding. Active site of PchGH61D is over the surface of cellulose. PchGH61D on the surface of cellulose is docked, Tyr 75 is bound with the chain of cellulose, and this conformation retains until 100 ns whereas Tyr 28 and Tyr 198 align over the same chain, and during the simulation, the location of residues was quite stable [10].

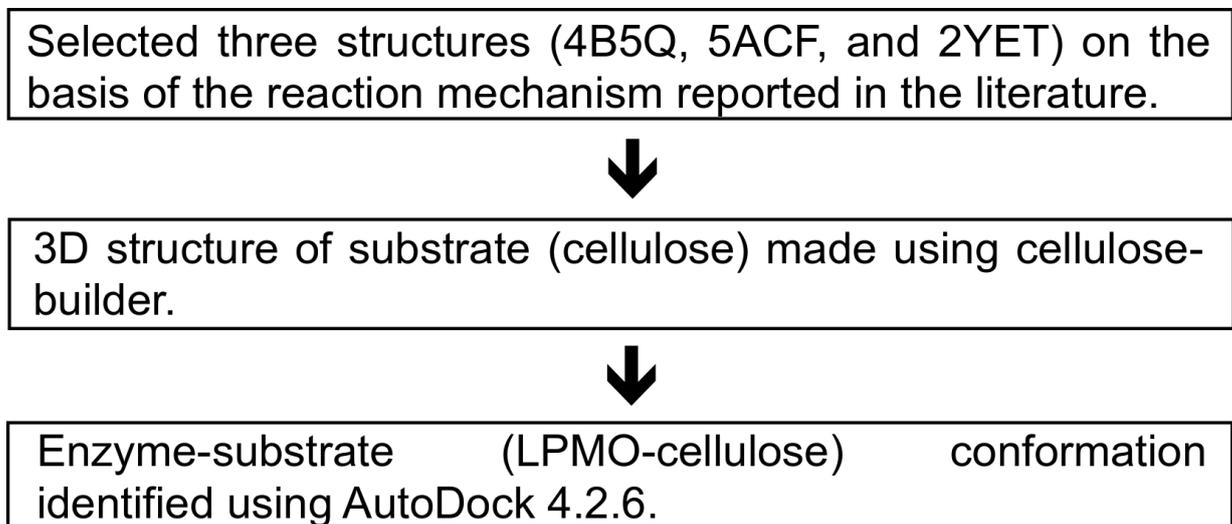
CHAPTER 2

HYPOTHESIS

LPMO and its interaction with chitin and cellulose is known, but the mechanism of how LPMO specifically binds to its substrate is still not clear. Specifically, binding of cello-oligosaccharide with the help of CH- π interaction and hydrogen bond are reported but their substrate specificity of structural determinants is still not determined. In this study, I hypothesize that the following residues bind with the substrates, some previously reported in the literature. These are histidine (His), tyrosine (Tyr), glutamine (Gln), and phenylalanine (Phe). There could be other residues in LPMO that are interacting with the substrate resulting in substrate binding specificity. Thus, in this study, I hope to identify residues that are important for the substrate binding and predict which residues driving substrate specificity.

CHAPTER 3

MATERIALS AND METHODS



Flowchart 3.1: Methodology. Methods used to carry out the substrate specificity in LPMO in this study.

3.1 CAZy

CAZy provides information about the enzymes involved in the synthesis, transport and metabolism of carbohydrate [4]. The database mainly includes glycoside hydrolase, carbohydrate binding families, etc. CAZy was used to retrieve the information of the four different families of LPMOs (AA9, AA10, AA11, and AA13).

3.2 PDB

It gives information about the 3D shapes of proteins, nucleic acids. It helps to understand all aspects of health, disease, biomedicine, and protein synthesis [19]. From the Protein Data Bank, download LPMOs structures for their substrate.

3.3 cellulose-builder

“cellulose-builder” is a toolkit used for building cellulose crystalline structures in PDB format [20]. The prerequisite for cellulose-builder tools were Octave, VMD, and psfgen, and all were saved in the same directory (path). Octave is a tool with maths-oriented syntax that helps in the visualization and plotting. VMD is a visualization program for analyzing and displaying the molecules [21]. psfgen contain a library of structure and an interface of Tcl. A 3D structure of cellulose was built using cellulose-builder, of any size in PDB formats [20].

3.4 AutoDock

Molecular Docking is a technique that uses computational approach to see the conformations of ligand in receptor. The objective is to obtain an optimized conformation and relative orientation between ligand and protein such that the free energy force field to evaluate the conformations. Docking method has components, the search algorithms, used to predict the conformations and scoring function that assigns a fitness value to conformation.

It is designed to predict how substrate binds to a receptor of known 3D structure. AutoDock4 consists of two main programs: autogrid calculates the grids whereas autodock performs the docking of the ligand to a set of grids describing the receptor [22, 23].

3.4.1 Steps of AutoDock

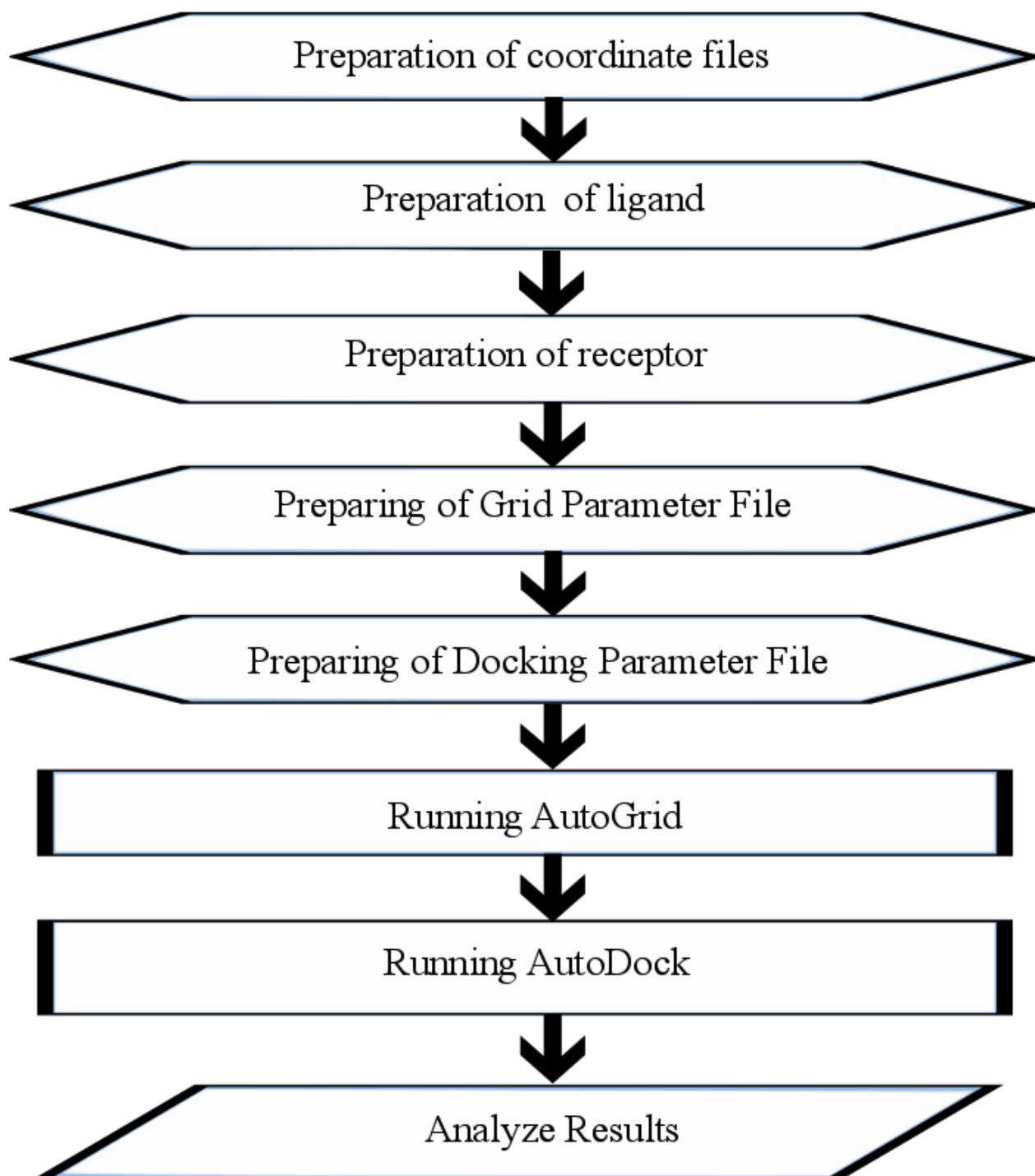
Cellulose (or any other polysaccharide which degrades by the LPMO) was taken as the input file for the ligand and 4B5Q (or any other PDB ID of LPMO) was taken as the input file for the receptor. Flowchart 2 describes the AutoDock step was used to carry out the interaction of LPMO and Cellulose.

Step 1 was to prepare the coordinate files of ligand and receptor. Ligand file was prepared from cellulose-builder whereas Receptor file was taken from the PDB, and only one chain of PDB was taken. Step 2 was to prepare the ligand file. Open Babel was used to convert the pdb file to pdbqt file where the addition of charges and atom type was taken. Step 3 was to prepare the macromolecular file. Rigid docking was done by the addition of hydrogen atom and by computing the Gasteiger charges. Step 4 was to prepare the grid parameter file, by setting the map types of the ligand. Coordinates of histidine was taken as the parameter of grid box and the file was saved. Step 5 was to prepare the docking parameter file by selecting the Genetic Algorithm as the search parameter, and the file was saved. Step 6 was to run the AutoGrid. Step 7 was to run the AutoDock. The last step was to analyze the results of docking by playing the conformations.

3.5 PyMOL

PyMOL is a python based molecular visualization tool. It is used to visualize the 3D images of small molecules and macromolecules [24].

The tool was used for the visualization of LPMO and Cellulose. It was also used for checking the conformations whether they reside on the flat surface of LPMO. Further on, the best conformation was chosen and the session file of that particular conformation with the LPMO was saved.



Flowchart 3.2: Steps of AutoDock. AutoDock was used to carry out the interaction of LPMO and Cellulose.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 CAZy and PDB Results

CAZy was used to retrieve the information of the four different families of LPMOs (AA9, AA10, AA11, and AA13). From the PDB, download LPMOs structures for their substrate. Total PDB IDs for AA9, AA10, AA11, and AA13 were 24, 32, 2, and 4, respectively. A total number of 62 PDB IDs were present for the LPMO (Table 4.1).

Table 4.1: PDB ID for AA9, AA10, AA11, and AA13.

AA9	AA10	AA11	AA13
2VTC	2BEM	4MAH	4OPB
2YET	2BEN	4MAI	5LSV
3EII	2LHS		5T7J
3EJA	2XWX		5T7N
3ZUD	2YOY		
4B5Q	2YOW		
4D7U	2YOX		
4D7V	3UAM		
4EIR	4A02		
4EIS	4ALC		
4QI8	4ALE		
5ACF	4ALQ		
5ACG	4ALR		
5ACH	4ALS		
5ACI	4ALT		
5ACJ	4GBO		
5FOH	4OW5		
5N04	4OY6		
5N05	4OY7		

5TKF	4OY8
5TKG	4X27
5TKH	4X29
5TKI	4YN1
5UFV	4YN2
	5AA7
	5FJQ
	5FTZ
	5L2V
	5IJU
	5VG0
	5VG1
	5SWZ

The detailed description of Table 1 is given in Table 2.

Table 4.2: LPMO's and their Substrate. It gives information about the substrate, their site of attack (C1, or C3, or C1/C4, or Not Determined), which family they belong, organism name. Substrates that are mainly for the LPMO are Cellulose (Phosphoric Acid Swollen Cellulose, Avicel, cellooligosaccharides, lignocellulose, hemicellulose), glucan (xyloglucan, glucomannan), chitin (α , β), and Starch.

Substrates	Site of attack	Protein Name	PDB code	Family	Organism	Reference
Cellulose	ND	HjAA9_B, HjGH61B, GH61B, Cel61B, EG7, TrAA9_B	2vtc	AA9	<i>Tricoderma reesei</i>	[14]
PASC, PCS	C1/C4	TaAA9_A, TaAA9A, TaGH61, TaGH61A, TaLPMO9A	2yet	AA9	<i>Thermoascus aurantiacus</i>	[14]
PASC, Avicel	C1	TtAA9_E, TrGH61E, GH61E, 131562	3eii	AA9	<i>Thielavia terrestris</i>	[14]
PASC, Avicel	C1	TtAA9_E, TrGH61E, GH61E, 131562	3eja	AA9	<i>Thielavia terrestris</i>	[14]
PASC, PCS	C1/C4	TaAA9_A, TaAA9A, TaGH61, TaGH61A, TaLPMO9A	3zud	AA9	<i>Thermoascus aurantiacus</i>	[14]
PASC, Avicel	C1	PcAA9_D, PcLPMO9D, PcGH61D,	4b5q	AA9	<i>Phanerochaete chrysosporii</i>	[14]

		GH61D			<i>m</i>	
PASC, cellooligosaccharides, xyloglucan, glucomannan, β -glucan	C4	NcAA9_C, NcLPMO9C, NCU02916, PMO-02916, GH61-3	4d7u	AA9	<i>Neurospora crassa</i>	[14]
PASC, cellooligosaccharides, xyloglucan, glucomannan, β -glucan	C4	NcAA9_C, NcLPMO9C, NCU02916, PMO-02916, GH61-3	4d7v	AA9	<i>Neurospora crassa</i>	[14]
PASC	C4	NcAA9_D, NcLPMO9D, PMO-2, NCU01050, GH61-4	4eir	AA9	<i>Neurospora crassa</i>	[14]
PASC	C1/C4	NcAA9_M, NcLPMO9M, PMO-3, NCU07898, GH61-13	4eis	AA9	<i>Neurospora crassa</i>	[14]
PASC	C1	NcAA9_F, NcLPMO9F, PMO-03328, NCU03328, GH61-6	4qi8	AA9	<i>Neurospora crassa</i>	[14]
PASC, cellooligosaccharides	C4	LsAA9_A	5acf	AA9	<i>Lentinus similis</i>	[14]
PASC,	C4	LsAA9_A	5acg	AA9	<i>Lentinus</i>	[14]

cellooligosacc harides					<i>similis</i>	
PASC, cellooligosacc harides	C4	LsAA9_A	5ach	AA9	<i>Lentinus similis</i>	[14]
PASC, cellooligosacc harides	C4	LsAA9_A	5aci	AA9	<i>Lentinus similis</i>	[14]
PASC, cellooligosacc harides	C4	LsAA9_A	5acj	AA9	<i>Lentinus similis</i>	[14]
ND	ND	NcLPMO9A	5foh	AA9	<i>Neurospora crassa</i>	
Cellooligosac charides	C4	LsAA9A	5n04	AA9	<i>Lentinus similis</i>	[25]
Cellooligosac charides	C4	LsAA9A	5n05	AA9	<i>Lentinus similis</i>	[25]
Glycan	C4	NcPMO-2, NcLPMO9D, PMO-2, NCU01050, GH61-4	5tkf	AA9	<i>Neurospora crassa</i>	[26]
Cellulose	C1	NcPMO-2, NcLPMO9D, PMO-2, NCU01050, GH61-4	5tkg	AA9	<i>Neurospora crassa</i>	[27]
Cellulose	C1	NcPMO-2, NcLPMO9D, PMO-2, NCU01050, GH61-4	5tkh	AA9	<i>Neurospora crassa</i>	[27]

Cellulose	C1	NcPMO-2, NcLPMO9D, PMO-2, NCU01050, GH61-4	5tki	AA9	<i>Neurospora crassa</i>	[27]
PASC	C1	MtPMO- 3,MYCTH_9266 8	5ufv	AA9	<i>Thermothelo myces thermophila</i>	[28]

α -chitin, chitin	β - C1	SmAA10_A, SmLPMO10A, CBP21, Cbp21, Cbp	2be m	AA10	<i>Serratia marcescens</i>	[14]
α -chitin, chitin	β - C1	SmAA10_A, SmLPMO10A, CBP21, Cbp21, Cbp	2ben	AA10	<i>Serratia marcescens</i>	[14]
α -chitin, chitin	β - C1	SmAA10_A, SmLPMO10A, CBP21, Cbp21, Cbp	2lhs	AA10	<i>Serratia marcescens</i>	[14]
ND	ND	VcAA10_B, VCA0811, VcGbpA, GbpA	2xw x	AA10	<i>Vibrio cholera</i>	[14]
ND	ND	BaAA10_A, BaCBM33, ChbB, Rbam17540	2yo w	AA10	<i>Bacillus amyloliquefa ciens</i>	[14]
ND	ND	BaAA10_A, BaCBM33, ChbB, Rbam17540	2yox	AA10	<i>Bacillus amyloliquefa ciens</i>	[14]

ND	ND	BaAA10_A, BaCBM33, ChbB, Rbam17540	2yoy	AA10	<i>Bacillus amyloliquefa ciens</i>	[14]
ND	ND	BpAA10_A	3uam	AA10	<i>Burkholderia pseudomallei</i>	[14]
α -chitin, chitin	β - C1	EfAA10_A, EfCBM33A, EfaCBM33, EF0362	4a02	AA10	<i>Enterococcus faecalis</i>	[14]
α -chitin, chitin	β - C1	EfAA10_A, EfCBM33A, EfaCBM33, EF0362	4alc	AA10	<i>Enterococcus faecalis</i>	[14]
α -chitin, chitin	β - C1	EfAA10_A, EfCBM33A, EfaCBM33, EF0362	4ale	AA10	<i>Enterococcus faecalis</i>	[14]
α -chitin, chitin	β - C1	EfAA10_A, EfCBM33A, EfaCBM33, EF0362	4alq	AA10	<i>Enterococcus faecalis</i>	[14]
α -chitin, chitin	β - C1	EfAA10_A, EfCBM33A, EfaCBM33, EF0363	4alr	AA10	<i>Enterococcus faecalis</i>	[14]
α -chitin, chitin	β - C1	EfAA10_A, EfCBM33A, EfaCBM33, EF0364	4als	AA10	<i>Enterococcus faecalis</i>	[14]
α -chitin, chitin	β - C1	EfAA10_A, EfCBM33A,	4alt	AA10	<i>Enterococcus faecalis</i>	[14]

		EfaCBM33, EF0365				
PASC, Avicel, chitin	β - C1/C4 (C1 on chitin)	TfAA10_A, TfLPMO10A, E7, Tfu_1268	4gbo	AA10	<i>Thermobifidi a fusca</i>	[14]
ND	ND	Fusolin	4ow 5	AA10	<i>unidentified entomopoxvi rus</i>	[14]
PASC, Avicel, chitin	β - C1/C4 (C1 on chitin)	ScAA10_B, ScLPMO10B, SCO0643, SCF91.03c	4oy6	AA10	<i>Streptomyces coelicolor</i>	[14]
PASC, Avicel	C1	ScAA10_C, ScLPMO10C, LPMO10C, CelS2, SCO1188, SCG11A.19	4oy7	AA10	<i>Streptomyces coelicolor</i>	[14]
PASC, Avicel, chitin	β - C1/C4 (C1 on chitin)	ScAA10_B, ScLPMO10B, SCO0643, SCF91.03c	4oy8	AA10	<i>Streptomyces coelicolor</i>	[14]
ND	ND	Fusolin	4x27	AA10	<i>Entomopoxvi rinae</i>	[14]
ND	ND	Fusolin	4x29	AA10	<i>Entomopoxvi rinae</i>	[14]
ND	ND	Fusolin (ACV034)	4yn1	AA10	<i>Anomala cuprea entomopoxvi rus</i>	[14]
ND	ND	Fusolin (partial)	4yn2	AA10	<i>unidentified entomopoxvi</i>	[14]

						<i>rus</i>	
α -chitin, β -chitin	C1	JdAA10_A, JdLPMO10A, Jden_1381	5aa7	AA10	<i>Jonesia denitrificans</i>	[14]	
ND	ND	LMRG_01781	5l2v	AA10	<i>Listeria monocytogenes</i>		
α -chitin, β -chitin	C1	CjAA10_A, CjLPMO10A, CJA_2191, Cbp33A, Lpmo10A	5fjq	AA10	<i>Cellvibrio japonicas</i>	[14]	
β -Chitin	C1/(C4)	SlAA10_E, SliLPMO10E, SLI_3182	5ftz	AA10	<i>Streptomyces lividans</i>	[14]	
(α , β) chitin	ND	ChbB, BaAA10A, BaCBM33, Rbam17540, BAMF_1859	5iju	AA10	<i>Bacillus amyloliquefaciens</i>	[29]	
Chitin	C1	JdLPMO10A, Jden_1381	5vg0	AA10	<i>Jonesia denitrificans</i>	[30]	
Chitin	C1	JdLPMO10A, Jden_1381	5vg1	AA10	<i>Jonesia denitrificans</i>	[30]	
Chitin	ND	LPMO10A,BiLPMO10A	5wsz	AA10	<i>Bacillus thuringiensis</i>		

β -Chitin	C1	AoAA11 Ao(LPMO11) (AO090102000501)	4mah	AA11	<i>Aspergillus oryzae</i>	[14]
β -Chitin	C1	AoAA11	4mai	AA11	<i>Aspergillus</i>	[14]

		Ao(LPMO11) (AO0901020005 01)			<i>oryzae</i>	
--	--	------------------------------------	--	--	---------------	--

ND (starch)	C1	AoAA13 (AO0907010002 46) (AOR_1_454114)	4opb	AA13	<i>Aspergillus oryzae</i>	[14]
chitin, lignocellulose , hemicellulose, starch-derived	ND	AoAA13, AO09070100024 6, AOR_1_454114	5lsv	AA13	<i>Aspergillus oryzae</i>	[31]
ND	ND	AoAA13, AO09070100024 6, AOR_1_454114	5t7j	AA13	<i>Aspergillus oryzae</i>	[31]
ND	ND	AoAA13, AO09070100024 6, AOR_1_454114	5t7n	AA13	<i>Aspergillus oryzae</i>	[31]

Based on the site of attack and reaction mechanism from Table 2, I select three PDB IDs. The details of the structure are shown in Table 3.

Table 4.3: Selected Substrate on the basis of type 1, 2 and 3. I select each site of attack for substrate like cellulose and perform AutoDock for the given PDB ID.

Substrates	Site of attack	Protein Name	PDB code	Family	Organism
PASC, PCS	C1/C4	TaAA9_A, TaAA9A, TaGH61, TaGH61A, TaLPMO9A	2yet	AA9	<i>Thermoascus aurantiacus</i>
PASC, Avicel	C1	PcAA9_D, PcLPMO9D, PcGH61D, GH61D	4b5q	AA9	<i>Phanerochaete chrysosporium</i>
PASC, cellooligosaccharides	C4	LsAA9_A	5acf	AA9	<i>Lentinus similis</i>

4.2 Result of cellulose-builder

Figure 4.1 represents a structure of cellulose consisting of 6 strands, where each strand has 20 glucose units. A 3D structure of cellulose was built using cellulose-builder, of any size in PDB formats.

4.3 Results of AutoDock

- LPMO's active site is present in a relatively flat/planar surface which has been observed to bind to crystalline surface of polymers such as cellulose and chitin.
- For each enzyme and substrate 50 rigid body docking were performed and one conformation was selected after clustering the results.
- Figure 8 represents top and side view of 4B5Q.
- Figure 9 represents top and side view of 5ACF.
- Figure 10 represents top and side view of 2YET.

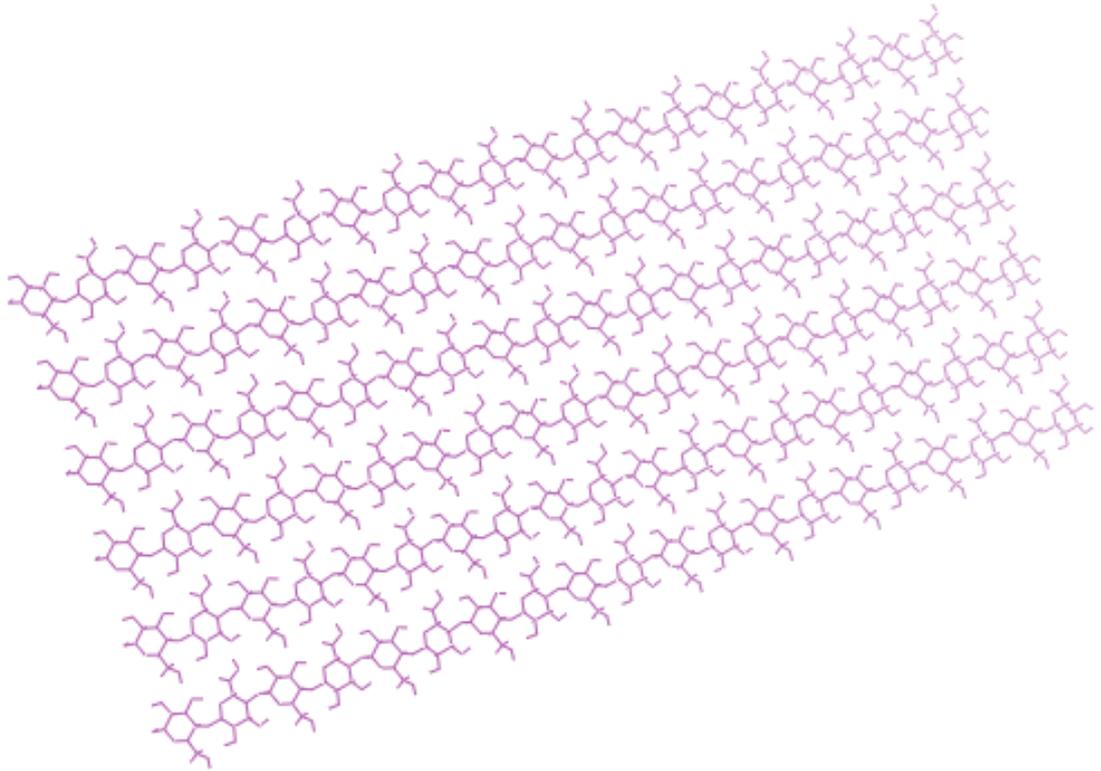


Figure 4.1: Structure of Cellulose was obtained from cellulose-builder. A sheet of cellulose, consisting of 6 strands, where each strand has 20 glucose units, then total glucose units is 120. A 3D structure of cellulose was built using cellulose-builder, of any size in PDB formats.

Table 4.4: Time taken by an Autodock for Receptor and Ligand docking.

Receptor	Ligand	Total Conformations Selected	Selected Conformation	Time taken (Real)	Processor Used
4B5Q	Cellulose	12	4 th	5h 10m 01.75s	i5
5ACF	Cellulose	9	3 rd	5h 15m 46.45s	i5
2YET	Cellulose	17	39 th	70h 58m 56.97s	i3

Ligand-protein interaction energies were pre-calculated and then used as the look-up table during simulation. Molecular Dynamics simulation will be used further to examine the effect of oxidation on the structure of crystalline cellulose.

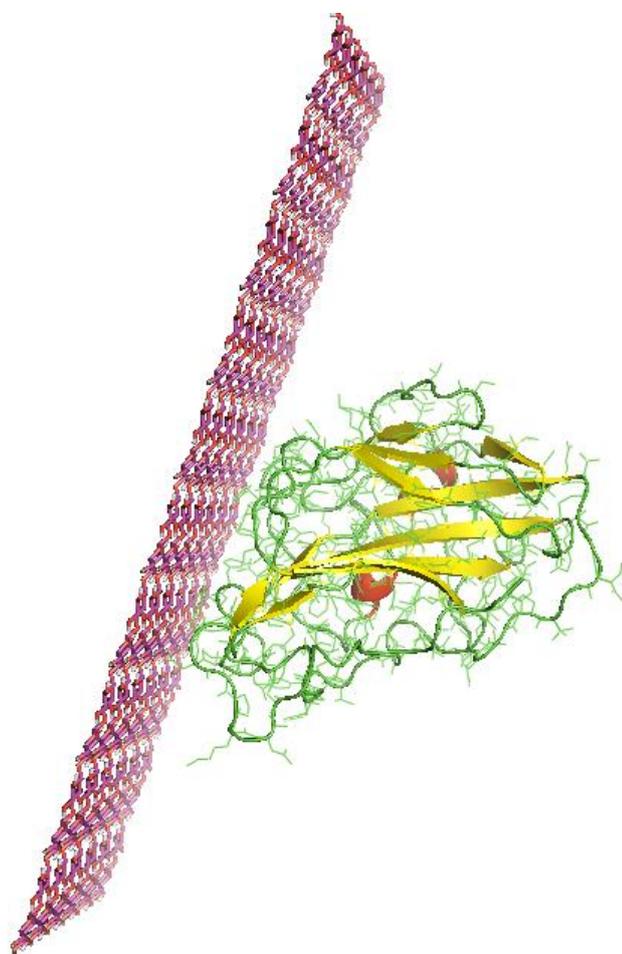
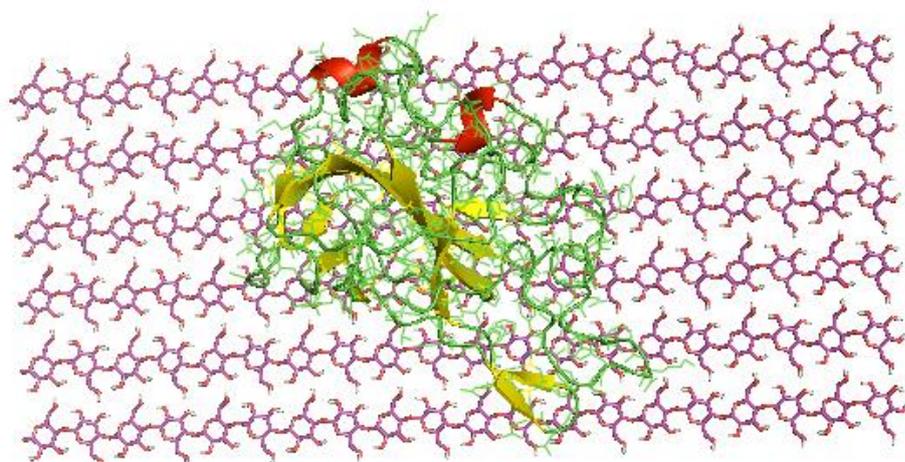


Figure 4.2: Top and side view of 4B5Q (*Phanerochaete chrysosporium*) and cellulose. Site of attack for 4B5Q is C1. After docking, their 4th conformation was selected.

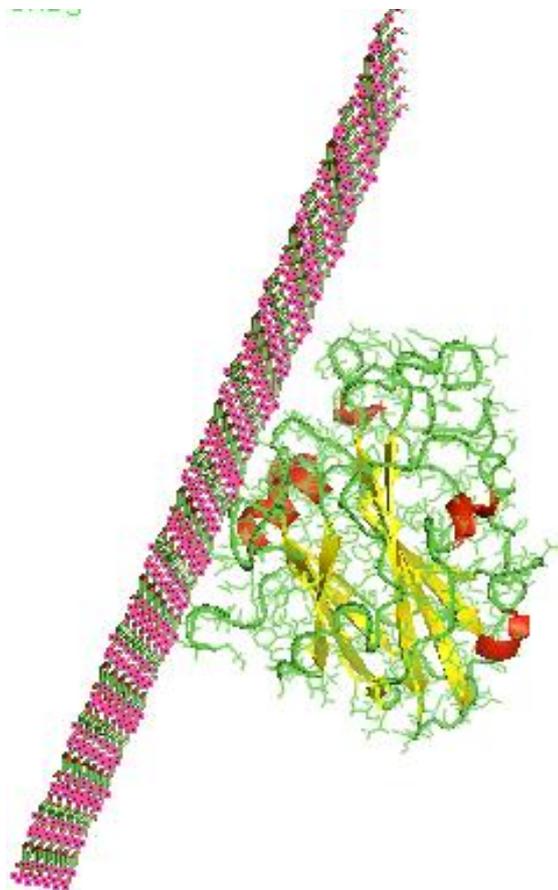
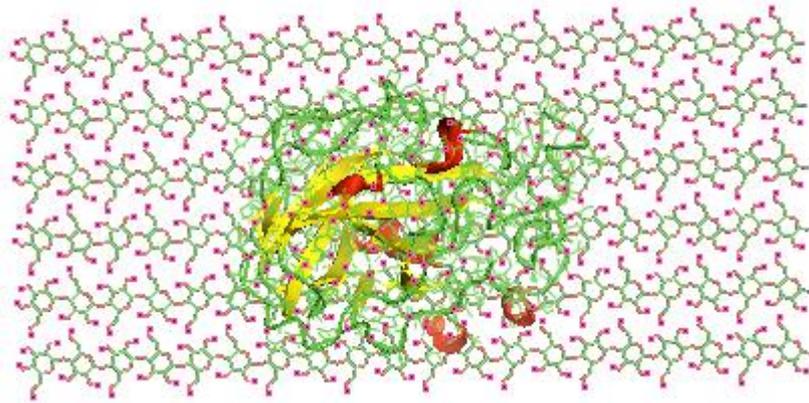


Figure 4.3: Top and side view of 5ACF (*Lentinus similis*) and cellulose. Site of attack for 5ACF is C4. After docking, their 3rd conformation was selected.

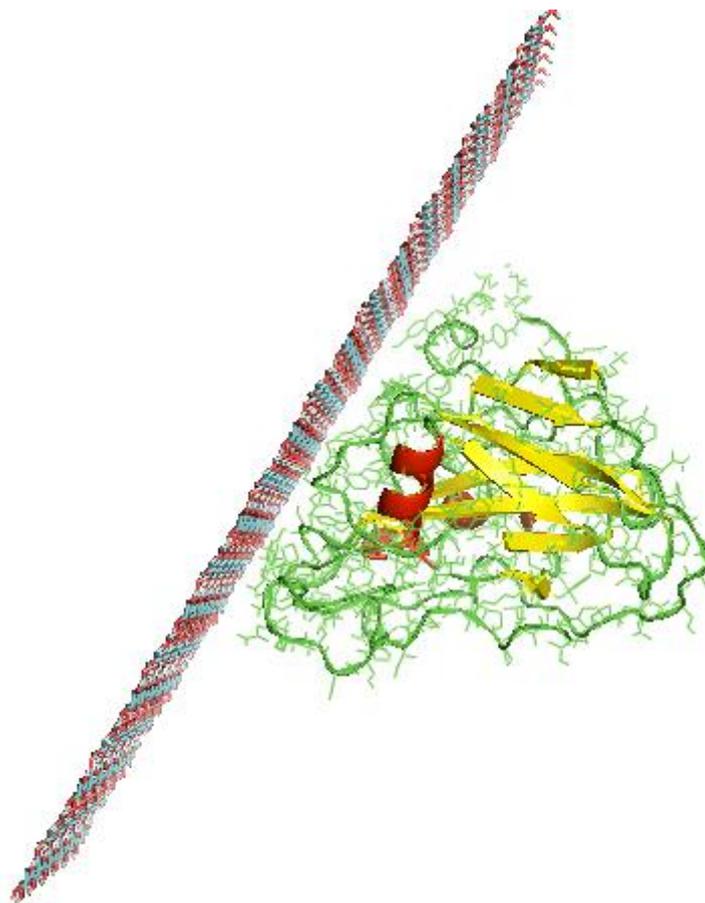
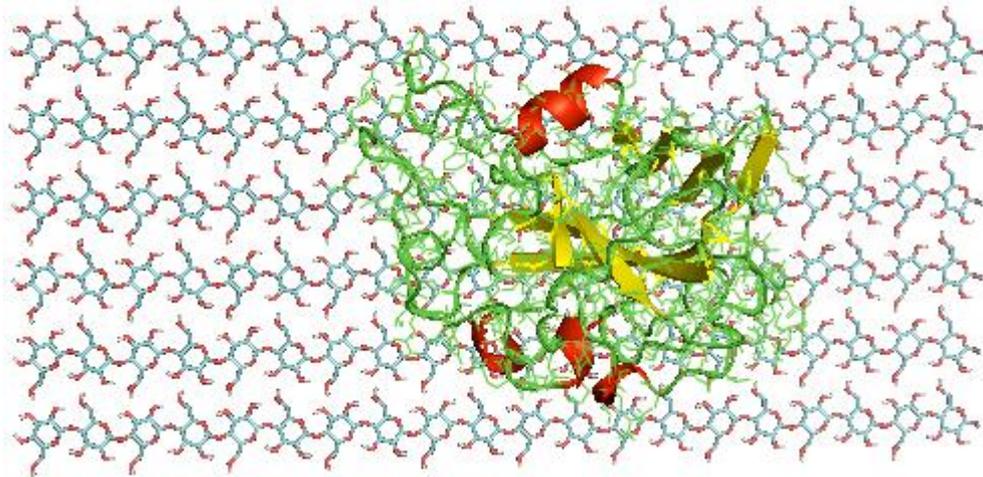


Figure 4.4: Top and side view of 2YET (*Thermoascus aurantiacus*) and cellulose. Site of attack for 2YET is C1/C4. After docking, their 39th conformation was selected.

CHAPTER 5

CONCLUSION

LPMO does not cut the chain of polymer but it cleaves the glycosidic bond. In general, specific residues of LPMO are known, but it is still unknown that how the substrate bound with the respective polysaccharide. In summary, this study provides a better understanding of the substrate specificity in LPMO.

Two types of conversion are mainly used for the degradation that includes, thermochemical conversion and the biochemical process help the biomass waste to convert into the bio-fuels. Biochemical conversion is a favourable approach over thermochemical conversion does not emerge their result into the destruction of polysaccharides [17]. A good model for bioconversion of lignocellulose, takes low concentration of enzyme loading with increase in their hydrolysis rate. Performance of hydrolysis can be improved if the enzymes use distinct reaction mechanism [32].

Stabilization of LPMO structure is due to the binding of copper. If two LPMO families have high structural similarity then the fold level and active site are conserved. AA9 and AA10 reduce the costs of the hydrolysis step in the production of ethanol [1].

If the substrate specificity of LPMO is clearer, then we can increase the production rate, where LPMO helps more in the biomass degradation for getting a high yield of the product i.e. bio-fuels. These bio-fuels would help in decreasing the rate of fossil fuel consumption, leading to lessen the effects of harmful climatic changes in the environment.

Starting conformation of three different enzyme-substrate complexes were obtained where 4B5Q and cellulose complex represent C1 site of attack (type 1), 5ACF and cellulose complex represent C4 site of attack (type 2), and 2YET and cellulose complex C1/C4 site of attack (type 3).

This study will throw some light on the regiospecificity of LPMO.

CHAPTER-6

REFERENCES

1. T. Corrêa, L. dos Santos and G. Pereira, "AA9 and AA10: from enigmatic to essential enzymes", *Applied Microbiology and Biotechnology*, vol. 100, no. 1, pp. 9-16, 2015.
2. Y. Nishiyama, P. Langan and H. Chanzy, "Crystal Structure and Hydrogen-Bonding System in Cellulose I β from Synchrotron X-ray and Neutron Fiber Diffraction", *Journal of the American Chemical Society*, vol. 124, no. 31, pp. 9074-9082, 2002.
3. "Genomic Science Program: U.S. Department of Energy", *Genomicscience.energy.gov*, 2005. [Online]. Available: <https://genomicscience.energy.gov/roadmap/>.
4. C. Zhou, Q. Wu, "Recent Development in Applications of Cellulose Nanocrystals for Advanced Polymer-Based Nanocomposites by Novel Fabrication Strategies".
5. Q. Zhu, V. Sharma, A. Oganov and R. Ramprasad, "Predicting polymeric crystal structures by evolutionary algorithms", *The Journal of Chemical Physics*, vol. 141, no. 15, p. 154102, 2014.
6. J. Agger, T. Isaksen, A. Varnai, S. Vidal-Melgosa, W. Willats, R. Ludwig, S. Horn, V. Eijsink and B. Westereng, "Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation", *Proceedings of the National Academy of Sciences*, vol. 111, no. 17, pp. 6287-6292, 2014.
7. J. Vermaas, M. Crowley, G. Beckham and C. Payne, "Effects of Lytic Polysaccharide Monooxygenase Oxidation on Cellulose Structure and Binding of Oxidized Cellulose Oligomers to Cellulases", *The Journal of Physical Chemistry B*, vol. 119, no. 20, pp. 6129-6143, 2015.
8. M. Agostoni, J. Hangasky and M. Marletta, "Physiological and Molecular Understanding of Bacterial Polysaccharide Monooxygenases", *Microbiology and Molecular Biology Reviews*, vol. 81, no. 3, pp. e00015-17, 2017.
9. B. Danneels, M. Tanghe, H. Joosten, T. Gundinger, O. Spadiut, I. Stals and T. Desmet, "A quantitative indicator diagram for lytic polysaccharide monooxygenases

- reveals the role of aromatic surface residues in HjLPMO9A regioselectivity", *PLOS ONE*, vol. 12, no. 5, p. e0178446, 2017.
10. M. Wu, G. Beckham, A. Larsson, T. Ishida, S. Kim, C. Payne, M. Himmel, M. Crowley, S. Horn, B. Westereng, K. Igarashi, M. Samejima, J. Ståhlberg, V. Eijsink and M. Sandgren, "Crystal Structure and Computational Characterization of the Lytic Polysaccharide Monooxygenase GH61D from the Basidiomycota Fungus *Phanerochaete chrysosporium*", *Journal of Biological Chemistry*, vol. 288, no. 18, pp. 12828-12839, 2013.
 11. G. Voshol, E. Vijgenboom and P. Punt, "The discovery of novel LPMO families with a new Hidden Markov model", *BMC Research Notes*, vol. 10, no. 1, 2017.
 12. W. Beeson, V. Vu, E. Span, C. Phillips and M. Marletta, "Cellulose Degradation by Polysaccharide Monooxygenases", *Annual Review of Biochemistry*, vol. 84, no. 1, pp. 923-946, 2015.
 13. V. Lombard, H. Golaconda Ramulu, E. Drula, P. Coutinho and B. Henrissat, "The carbohydrate-active enzymes database (CAZy) in 2013", *Nucleic Acids Research*, vol. 42, no. 1, pp. D490-D495, 2013.
 14. K. Frandsen and L. Lo Leggio, "Lytic polysaccharide monooxygenases: a crystallographer's view on a new class of biomass-degrading enzymes", *IUCrJ*, vol. 3, no. 6, pp. 448-467, 2016.
 15. G. Vaaje-Kolstad, Z. Forsberg, J. Loose, B. Bissaro and V. Eijsink, "Structural diversity of lytic polysaccharide monooxygenases", *Current Opinion in Structural Biology*, vol. 44, pp. 67-76, 2017.
 16. A. Book, R. Yennamalli, T. Takasuka, C. Currie, G. Phillips and B. Fox, "Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases", *Biotechnology for Biofuels*, vol. 7, no. 1, p. 109, 2014.
 17. V. Moses, R. Hatherley and Ö. Tastan Bishop, "Bioinformatic characterization of type-specific sequence and structural features in auxiliary activity family 9 proteins", *Biotechnology for Biofuels*, vol. 9, no. 1, 2016.
 18. R. Byerrum, "Introduction to modern biochemistry (Karlson, P.)", *Journal of Chemical Education*, vol. 43, no. 1, p. A76, 1966.
 19. H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov and P. Bourne, "The Protein Data Bank", *Nucleic Acids Research*, vol. 28, no. 1, pp. 235-242, 2000.

20. T. Gomes and M. Skaf, "Cellulose-Builder: A toolkit for building crystalline structures of cellulose", *Journal of Computational Chemistry*, vol. 33, no. 14, pp. 1338-1346, 2012.
21. W. Humphrey, A. Dalke and K. Schulten, "VMD: Visual molecular dynamics", *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33-38, 1996.
22. M. Sanner, "Python: a programming language for software integration and development", *J. Mol. Graphics Mod.*, vol. 17, pp. 57-61, 1999.
23. G. Morris, R. Huey, W. Lindstrom, M. Sanner, R. Belew, D. Goodsell and A. Olson, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility", *Journal of Computational Chemistry*, vol. 30, no. 16, pp. 2785-2791, 2009.
24. Schrodinger, "The PyMOL Molecular Graphics System", Version 1.3, 2010.
25. K. Frandsen, J. Poulsen, T. Tandrup and L. Lo Leggio, "Unliganded and substrate bound structures of the cellobiosaccharide active lytic polysaccharide monooxygenase Ls AA9A at low pH", *Carbohydrate Research*, vol. 448, pp. 187-190, 2017.
26. W. O'Dell, P. Swartz, K. Weiss and F. Meilleur, "Crystallization of a fungal lytic polysaccharide monooxygenase expressed from glycoengineered *Pichia pastoris* for X-ray and neutron diffraction", *Acta Crystallographica Section F Structural Biology Communications*, vol. 73, no. 2, pp. 70-78, 2017.
27. W. O'Dell, P. Agarwal and F. Meilleur, "Oxygen Activation at the Active Site of a Fungal Lytic Polysaccharide Monooxygenase", *Angewandte Chemie*, vol. 129, no. 3, pp. 785-788, 2016.
28. E. Span, D. Suess, M. Deller, R. Britt and M. Marletta, "The Role of the Secondary Coordination Sphere in a Fungal Polysaccharide Monooxygenase", *ACS Chemical Biology*, vol. 12, no. 4, pp. 1095-1103, 2017.
29. R. Gregory, G. Hemsworth, J. Turkenburg, S. Hart, P. Walton and G. Davies, "Activity, stability and 3-D structure of the Cu(ii) form of a chitin-active lytic polysaccharide monooxygenase from *Bacillus amyloliquefaciens*", *Dalton Transactions*, vol. 45, no. 42, pp. 16904-16912, 2016.
30. J. Bacik, S. Mekasha, Z. Forsberg, A. Kovalevsky, G. Vaaje-Kolstad, V. Eijsink, J. Nix, L. Coates, M. Cuneo, C. Unkefer and J. Chen, "Neutron and Atomic Resolution X-ray Structures of a Lytic Polysaccharide Monooxygenase Reveal Copper-

Mediated Dioxygen Binding and Evidence for N-Terminal Deprotonation", *Biochemistry*, vol. 56, no. 20, pp. 2529-2532, 2017.

31. K. Frandsen, J. Poulsen, M. Tovborg, K. Johansen and L. Lo Leggio, "Learning from oligosaccharide soaks of crystals of an AA13 lytic polysaccharide monooxygenase: crystal packing, ligand binding and active-site disorder", *Acta Crystallographica Section D Structural Biology*, vol. 73, no. 1, pp. 64-76, 2017.
32. S. Jung, Y. Song, H. Kim and H. Bae, "Enhanced lignocellulosic biomass hydrolysis by oxidative lytic polysaccharide monooxygenases (LPMOs) GH61 from *Gloeophyllum trabeum*", *Enzyme and Microbial Technology*, vol. 77, pp. 38-45, 2015.

APPENDIX A

A.1: cellulose-builder

“./cellulose-builder.sh fibril 6”

In Linux operating system, open terminal window and input the directory, that has the necessary builder files and set the path accordingly. The above-mentioned command was taken as “input command” in the terminal, where fibril 6 means that the disaccharide was 6 and in short, there were total 6 strands that can be taken.

After successfully running a tool, five output files were generated that contain-crystal.pdb, crystal.psf, crystal.xyz, psfgen.sh, and psfgen.log. The main concern is crystal.pdb file, which is further implemented in docking and MD simulations [20].