Structural Dynamics, Mechanism, and Interaction of Lytic Polysaccharide Monooxygenases with Cellobiose dehydrogenase

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

BACHELOR OF TECHNOLOGY

IN

BIOINFORMATICS

By

MANSI ARORA

UNDER THE GUIDANCE OF

Dr. Ragothaman Yennamalli

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
May 2018

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the B.Tech thesis entitled **"Structural Dynamics, Mechanism, and Interaction of Lytic Polysaccharide Monooxygenases and Cellobiose Dehydrogenase"** submitted at **Jaypee University of Information Technology,Waknaghat India**, is an authentic record of my work carried out under the supervision of **Dr. Ragothaman Yennamalli**. I have not submitted this work elsewhere for any other degree or diploma.

Mansi Arora

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology, Waknaghat

Date:

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the B.Tech thesis entitled, **"Structural Dynamics, Mechanism, and Interaction of Lytic Polysaccharide Monooxygenases and Cellobiose Dehydrogenase"**, submitted by **Mansi Arora** at **Jaypee University of Information Technology, Waknaghat**, **India,** is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Supervisor: Dr. Ragothaman Yennamalli

Designation: Assistant Professor

Affiliation: Jaypee University of Information Technology, Waknaghat

Date:

ACKNOWLEDGEMENT

Being a final year B.Tech Bioinformatics student, I feel privileged to have successfully completed this undertaking, which provided me the exposure to a real-time research project to add on to my future goals as well. It helped me in developing a scientific temperament for developing a problem and working on it in a systematic and chronological way.

I express heartiest gratitude towards Dr. Ragothaman Yennamalli for his immense support and guidance throughout in supervising me with finite patience, and without whose invaluable suggestions and unstinted cooperation, the present desertion would not have been possible. I am also thankful to Ms. Somlata for the lab support, and Mr. Arvind Yadav and my colleague Ms. Radhika Arora, whose continuous motivation kept me moving in the right direction.

Besides, I would also thank Jaypee University of Information Technology for the help and cooperation provided.

Last but not the least; it was a worthwhile experience for me, taking me to a yet another level of research work.

CONTENTS

1. Chapter 1 – Introduction

1.1 Lytic Polysaccharide Monooxygenases

1.2 Simulation Studies on LPMO

1.3 Cellobiose dehydrogenase

1.4 LPMO-CDH Interaction

2. Chapter 2 – Hypothesis

3. Chapter 3 – Materials and Methods

3.1 CAZy

3.2 Normal Mode Analysis

3.3 PyMOL

3.4 Molecular Docking

4. Chapter 4 – Results and Discussion

4.1 CDH and LPMO – Structure and Mechanism

4.2 Normal Mode Analysis

4.2.1 B-factor cross correlation map of CDH

4.2.2 Individual slow modes for first 10 GNM modes of CDH

4.2.3 Identification of Open-Closed conformations of CDH

4,3 Molecular Docking

4.4 Protein-Protein Interaction Map

5. Chapter 5 – Conclusion

Appendix 1

6. Chapter 6 – References

LIST OF ABBREVIATIONS

LPMO- Lytic Polysaccharide Monooxygenases

CDH- Cellobiose dehydrogenase

GH- Glycoside hydrolases

QM/MM- Quantum Mechanical/Molecular Mechanics

CBM- carbohydrate binding module

PL- Polysaccharide Lyases

CE- Carbohydrate esterases

CAZy- Carbohydrate Active Enzymes

NMR- Nuclear Magnetic Resonance

MD- Molecular Dynamics

FAD- Flavin adeneine dinucleotide

CYT- Cytochrome domain

DH- Dehydrogenase

NMA- Normal Mode Analysis

ANM- Anisotropic Network Model

GNM- Gaussian Network Model

ENM- Elastic Network Model

CG- Coarse Grained

PPI- Protein-Protein Interaction

IET- Interdomain electron transfer

VMD- Visual Molecular Dynamics

LIST OF FIGURES

1) Figure 1.1: Schematic representation of C1 (Type1) and C4 (Type2) mechanism of action identified in Lytic Polysaccharide Monoxygenases (LPMO).
2) Figure 1.2: Electron Flow and LPMO activation
3) Figure 3.1: The overall method used in this study
4) Figure 4.1: Structure of Cellobiose Dehydrogenase
5) Figure 4.2: Structure of LPMO
6) Figure 4.3 : Theoretical and experimental B-factor cross correlation map for GNM
7) Figure 4.4 : Individual slowest modes for first 10 modes for GNM
8) Figure 4.5: First 10 slowest ANM modes showing the open-closed conformation states of CDH
9) Figure 4.6: Step-wise schematic representation for molecular docking
10) Figure 4.7: Cartoon representation of models selected after docking
    A) Model02
    B) Model08
    C) Model09
    D) Res06
11) Figure 4.8: Protein interaction map for
    A) Model02 of ClusPro
    B) Model08 of ClusPro
    C) Model09 of ClusPro
    D) Res06 of PatchDock

LIST OF TABLES

Table 4.1: Slowest modes with their corresponding open or closed conformational states.

ABSTRACT

For bioethanol production from lignocellulosic biomass, an enzyme cocktail is used, where these enzymes act synergistically to breakdown the polymeric structure of cellulose to simple sugars. Among the cocktail, is Lytic Polysaccharide Monooxygenase (LPMO) that has the ability to act on crystalline form of cellulose. Study on LPMOs is interesting due to this property and for its optimal activity, a reducing agent is required, which is usually Cellobiose dehydrogenase (CDH). The activation of LPMO occurs by means of electron transfer within CDH and then from CDH to LPMO. The structural and dynamical duet of LPMO and CDH and their interactions are unexplored till date. In this study, I have identified the conformational changes using computational approaches. My hypothesis is that there are key residues that are responsible for LPMO binding with CDH, and for transfer of the electron to LPMO. Also, there is a highly likely event of an allosteric change occurring in either CDH or in LPMO during binding, where the conformational change between the flavin and heme domains of CDH may possibly influence the electron transfer process. I discuss the conformational sampling using Elastic Network Models.

CHAPTER 1

INTRODUCTION

Enzymes are able to function optimally by physical and chemical arrangement of its structure, where the change can be local, such as breaking of existing bonds to create new ones, or global, such as a conformational change causing distinct apoenzyme and holoenzyme structures. While allostery is the driving factor for the enzyme to interact with its substrate, the magnitude of allosteric change can be quantified either through experimental or computational methods. In the case of experimental methods, the rate limiting steps are the purity of protein and its substrate, longevity of the interaction, and product release. All or any of these rate limiting steps can lead to delay in gathering data. In contrast, computational approaches can reduce this time and generate large amounts of data that can be analyzed, which allows significant conclusions to be drawn that can enhance the strategy of performing the same experiment. In the case of biofuels, a diverse set of enzymes needs to be used during industrial processes, which act synergistically to breakdown the polymeric structure of cellulose to simple sugars. One such set of enzymes are Lytic Polysaccharide Monooxygenase (LPMO) and Cellobiose dehydrogenase (CDH), which work in collaboration to increase the biomass degradation.

The plant cell wall is composed of a middle lamella, a primary and a secondary cell wall. Polysaccharide is the most abundant of all the components. Plant biomass consists of materials from the cell wall which are tough to be degraded by enzymes because of the complex lignocellulose structure [46,47]. LPMOs have shown to increase the activity of hydrolytic enzymes in degrading these cell wall compounds. Such an enzyme is seen as an effective green catalyst to degrade non-edible plant biomass for production of biofuels and biomaterials. They are expected to alter the composition of future commercial enzyme cocktails.

1.1 Lytic Polysaccharide Monooxygenases

Lytic Polysaccharide Monooxygenases (LPMO) whereby ‘Lytic’ means destruction and ‘polysaccharide’ means (poly = many + saccharide = sugar), i.e. carbohydrates formed by the repeated units connected with the help of glycosidic

bonds. ‘Polysaccharide Monooxygenases’ refer to the the enzymes that initiate the insertion of a single O atom from O_2 into a substrate.

Lytic polysaccharide monooxygenases (LPMOs) are recently discovered enzymes that have shown immense industrial application for degrading crystalline form of cellulose, as they boost the degradation process significantly [1]. Their importance and their relevance have been described in detail in the reviews [2-4]. It is now well established that LPMOs can be classified into three subtypes [41], based on the site of attack namely 1) LPMO1 when oxidation occurs at C1 carbon, 2) LPMO2 when oxidation occurs at C4 carbon, and 3) LPMO3 if either C1 or C4 carbons are attacked (Figure 1.1). Additionally, there are four CAZy families to which LPMOs are classified as auxiliary activity enzymes (AA9, AA10, AA11, and AA13) on the basis of their potential abilities to help the originally classified enzymes, glycoside hydrolases (GH), the polysaccharide lyases (PL), the carbohydrate esterases (CE), in gaining access to the carbohydrates encrusted in the plant cell wall [5].

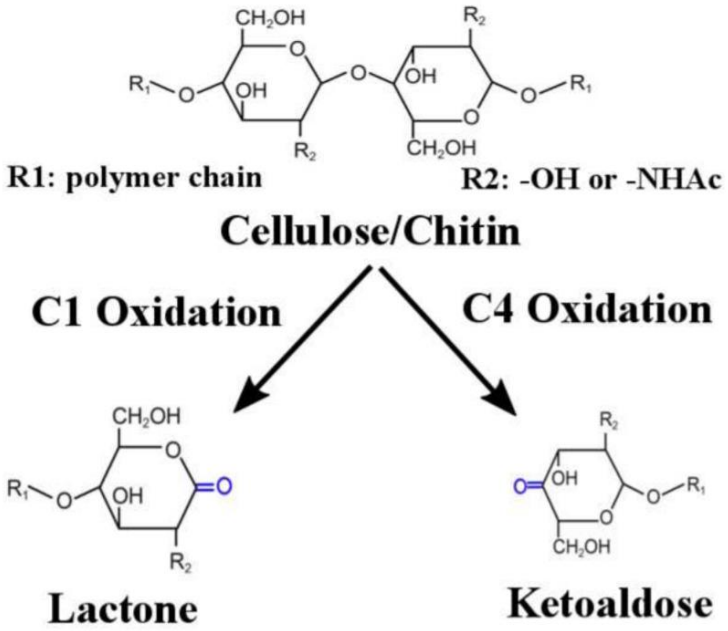

Figure 1.1: Schematic representation of C1 (Type1) and C4 (Type2) mechanism of action identified in Lytic Polysaccharide Monoxygenases (LPMO). LPMOs oxidize either on C1 or C4 carbon giving rise to specific products, lactone or ketoaldose, respectively.

1.2 Simulation Studies on LPMO

Molecular insights into LPMO's mechanism of action have been significantly improved by computational studies from QM/MM models to multiscale models [6]. The QM/MM models that were built using the density functional theory on LPMO belonging to the AA9 family from *Thermoascus auranticus* shed light on the geometry and coordination chemistry of the reactive oxygen with CuII atom [6]. The results indicated that the formation of the complex (copper-oxyl reactive oxygen species) drives the catalytic activity with a rebound step for oxygen to complete the cycle [6].

Another similar study informs us about the four-coordinate tetragonal structure of *T.aurenticus* in oxidised state and a three-coordinate T-shaped structure in reduced state [7]. The O2 reactivity of the Cu(I) site has been evaluated computationally by means of experimentally calibrated DFT calculations. To determine the number and type of coordinating ligands in Cu-AA9, extended X-ray absorption fine structure (EXAFS) was performed on the oxidized and reduced enzyme forms [7]. MD simulations of LPMO from the AA9 family have revealed that the loop regions undergo conformational changes that make the enzyme flexible during substrate binding. These findings are in agreement with the QM/MM results where the distance between the active site copper and C1 carbon is around 5Å, where a superoxide intermediate of the reaction (a product of the reactive oxygen with CuII atom) can be easily accommodated. The tyrosines (Tyr28, Tyr75, and Tyr198) were computationally observed to form local hydrophobic interactions stabilizing the active site during substrate binding [8].

In order to accurately capture the enzyme dynamics using MD simulations, the use of accurate force fields for a given system is required. A recent study probed potential energy landscape for the AA9 family to create a specific set of force-field parameter [9]. Use of such accurate force fields that can represent metallo-proteins consists of single point energy evaluations over a rectangular grid involving selected internal coordinates, and incorporate the generation of energy profiles for the bond stretch, angle bend and torsions, and will thus enable more realistic simulations and ultimately enhance our knowledgebase of LPMOs.

Recently, the reduction of the LPMO active site of AA9 enzyme from *Thermoascus auranticus* from state 1 (resting state) to 2 (reduced state) and two isomers of state 3 (Copper superoxide intermediate) has been investigated [10]. The results of combined

quantum mechanics and molecular mechanics (QM/MM) methodology proved that the computational protocol that were followed in this study can reproduce the observed decrease in the coordination number when CuII is reduced to CuI. Among the two isomers that are observed in Cu−superoxide complex, the multiscale modeling revealed that there is a preference for one isomer over the other for energetic stability. Further work on the enzyme substrate complex from the same group has led to validation of four enzyme-substrate intermediate models in terms of calculating bond-dissociation energy (BDE) [11]. BDE calculations are time consuming in an experimental setup and thus the alternative method of calculating BDEs from computational methods are quicker and sensitive. Specifically, in the LPMO studied (pdb id: 2yet [12]) the bond-dissociation energy for the four intermediates $[Cu–OH]^{3+}$ $[Cu–OH]^{2+}$, $[Cu–O]^{2+}$, and $[Cu–O]^{+}$ are comparable, however the intermediate $[Cu–OH]^{3+}$ is not favorable compared to the other three. The study also highlighted the non-dependency of the aromatic residue in the active site, as many LPMOs have either a Tyr or a Phe at the same position [11].

The MD simulations once again prove their worth for identifying key areas that deviate from the crystal structures of ScLPMO10B and ScLPMO10C LPMO as a starting place for surface charge modifications to increase stability in ILs. The MD has been performed for 250 ns in three ILs at 10 and 20 wt% in water and in pure water. The IL effects of dynamic fluctuations for specific regions of the enzyme on exposure to ionic liquid, the effects on enzyme's overall structure as well as the structure of enzyme's active site have been comprehensively and comparatively studied for both the LPMOs. The results clearly indicate that they both show structural similarity, and the fluctuations in IL and water are nearly same. Therefore both the LPMOs are unaffected by the influence of ionic liquids [13].

To study the functional aspects of CBP21, a chitin-active member of carbohydrate binding module family, NMR techniques, and isothermal titration calorimetry have been used, which proved that CBP21 is a compact and rigid molecule, except at its catalytic metal binding site. CBP21 depends on Cu ion for catalysis, and binding of cyanide to the metal indicates that it is involved in oxidative cleavage of the substrate. Further, the comparisons with GH61 LPMO show that their metal binding sites are significantly different despite the fact that both catalyse the same reaction. An approach

that uses the pH dependency of both chitin-CBP21 interaction and the 1H exchange rate led to the identification of the residues involved in binding CBP21 to the chitin surface based on the first NMR structure ever resolved for an LPMO [14].

1.3 Cellobiose dehydrogenase

The copper ion contained in LPMO needs to be reduced by an external electron donor before the activation of dioxygen, for the catalysis to proceed. The previous laboratory experiments have revealed that functional electron donors include small molecule reductants like ascorbic acid or gallic acid, lignin present in plant cell walls, certain redoxactive proteins, such as cellobiose dehydrogenase (CDH) and phenolic compounds that undergo redox cycling by glucose-methanol-choline oxidoreductases. CDH has only been found in fungi. They have two domains, flavin adenine dinucleotide (FAD)-binding dehydrogenase (DH) domain coupled to a heme-binding cytochrome (CYT) domain. The in vivo role of CDHs is uncertain since their discovery. Many hypotheses have been proposed on their functions due to widespread occurrence. One such function is their ability to produce H_2O_2, which is one of the strongest oxidizing agents in the aqueous systems and creates hydroxyl radicals to degrade or alter cellulose. This indicates a role of CDH in unspecific plant cell wall degradation. [15]

1.4 LPMO-CDH interaction

LPMOs are copper dependent and play a significant role in the conversion of biomass. They are responsible for catalyzing oxidative cleavage of glycosidic bonds in a process involving molecular oxygen and an electron donor, such as cellobiose dehydrogenase (CDH)(Figure 1.2).

CDH oxidizes disaccharides or oligosaccharides to their corresponding aldonic acids. In this process, the two electrons obtained from the substrate are stored in the DH domain by reduction of the FAD. One electron can be transferred from the DH domain to the CYT domain (by reduction of the heme group), by internal electron transfer [15]. CDHs can thereby perform two electron reduction reactions (via the DH domain) or one-electron reductions (via the CYT domain). CDHs are capable of efficient transfer of electrons to both small chemical compounds and proteins [15].

Using protein NMR and isothermal titration calorimetry, the interactions between a fungal LPMO and three soluble substrates and CDH have been studied [16]. The results

indicate the areas on the LPMO surface which interact with the varying substrates and this shows that both the substrate and CDH bind to a region that is centered around the copper site. Further, it is inferred from the data generated, that the electron transfer occurs before substrate binding, suggesting important new leads for understanding the reaction mechanism of LPMOs. [16]

ELECTRON FLOW

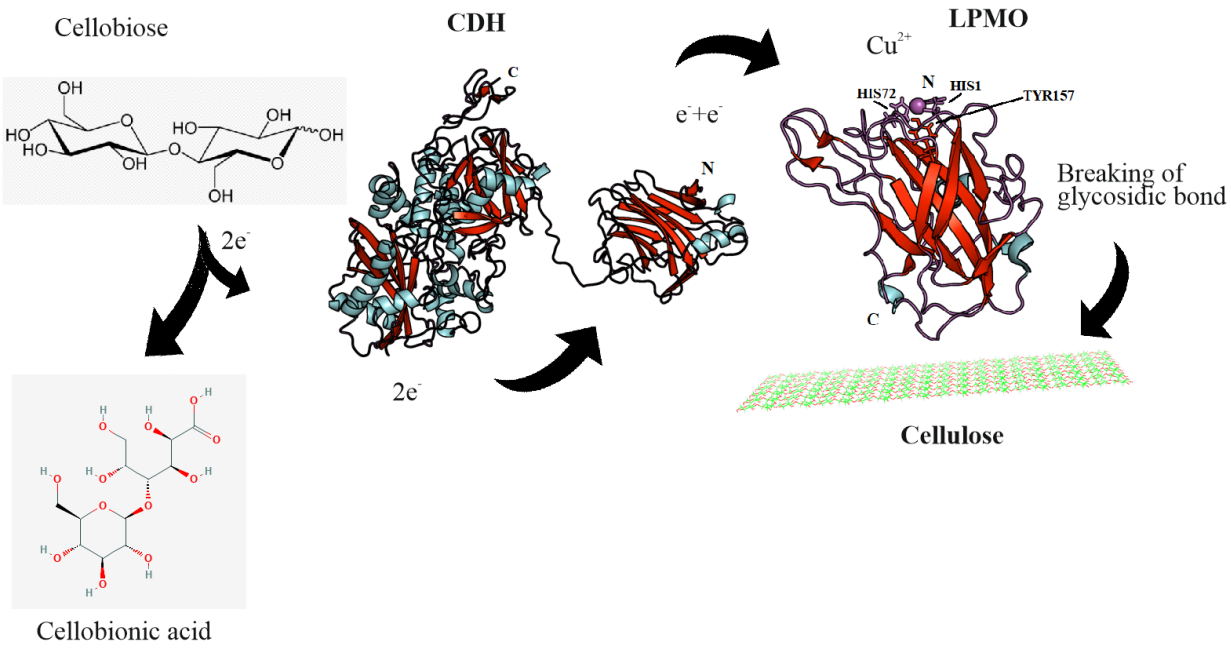

Figure 1.2: Electron Flow and LPMO activation. Schematic diagram for the activation of LPMO due to donation of electrons from CDH, via the interdomain transfer between the two domains of CDH.

CHAPTER 2

HYPOTHESIS

LPMO and CDH are known to interact where CDH donates the electron and LPMO utilizes the electron to oxidize the polysaccharide. However, the mechanism of action as to have LPMO and CDH interact, and their dynamics is still unclear.

In this study, I hypothesize that using specific residues in LPMO that are interacting with the CDH, CDH undergoes an open-close conformational change, giving rise to a synergistic behavior of the LPMO-CDH interaction. This oxidative CDH/LPMO system, where CYT domain of CDH is mobile and reduces the copper site of LPMO, enhances the degradation rate of crystalline cellulose. This CDH-LPMO system is widespread throughout the fungal kingdom together with the well-known hydrolytic cellulases [42,43]. The residues are selected from previous report of Li *et al.* where they are conserved. Ultimately, in this study, I find the possible conformational changes that CDH undergoes before and after binding to LPMO.

CHAPTER 3

MATERIALS AND METHODS

3.1 CAZy

CAZy is a database of Carbohydrate Active enzymes (CAZy), which provides information about the enzymes involved in the synthesis, transport and metabolism of carbohydrates. The database mainly includes glycoside hydrolases (which assist in hydrolysis of glycosidic bonds in complex sugars), glycosyl transferases (which catalyse the transfer of saccharide moieties from an activated nucleotide sugar to a nucleophilic glycosyl acceptor molecule), polysaccharide lyases (which catalyse the breaking of polysaccharide), carbohydrate esterase (which split esters into an acid and an alcohol) and carbohydrate binding families. The accuracy of CAZy lies in the fact that the major contribution of the database has been the dissemination of a stable nomenclature for these enzymes [17]. As LPMOs are also carbohydrate active enzymes, they are included under CAZy. To cross-check the different LPMO enzymes corresponding to their pdb ids and further, selected LPMO (pdb id: 4qi8) and CDH (pdb id: 4qi7) of *Neurospora crassa* for the interaction study (Figure 3.1). Specifically *N.crassa* was taken into consideration because there are such large amounts of purified PMOs, which are difficult to obtain from the native fungal producers. Each enzyme studied or referred hereby, has been accurately verified by CAZy for their corresponding families.

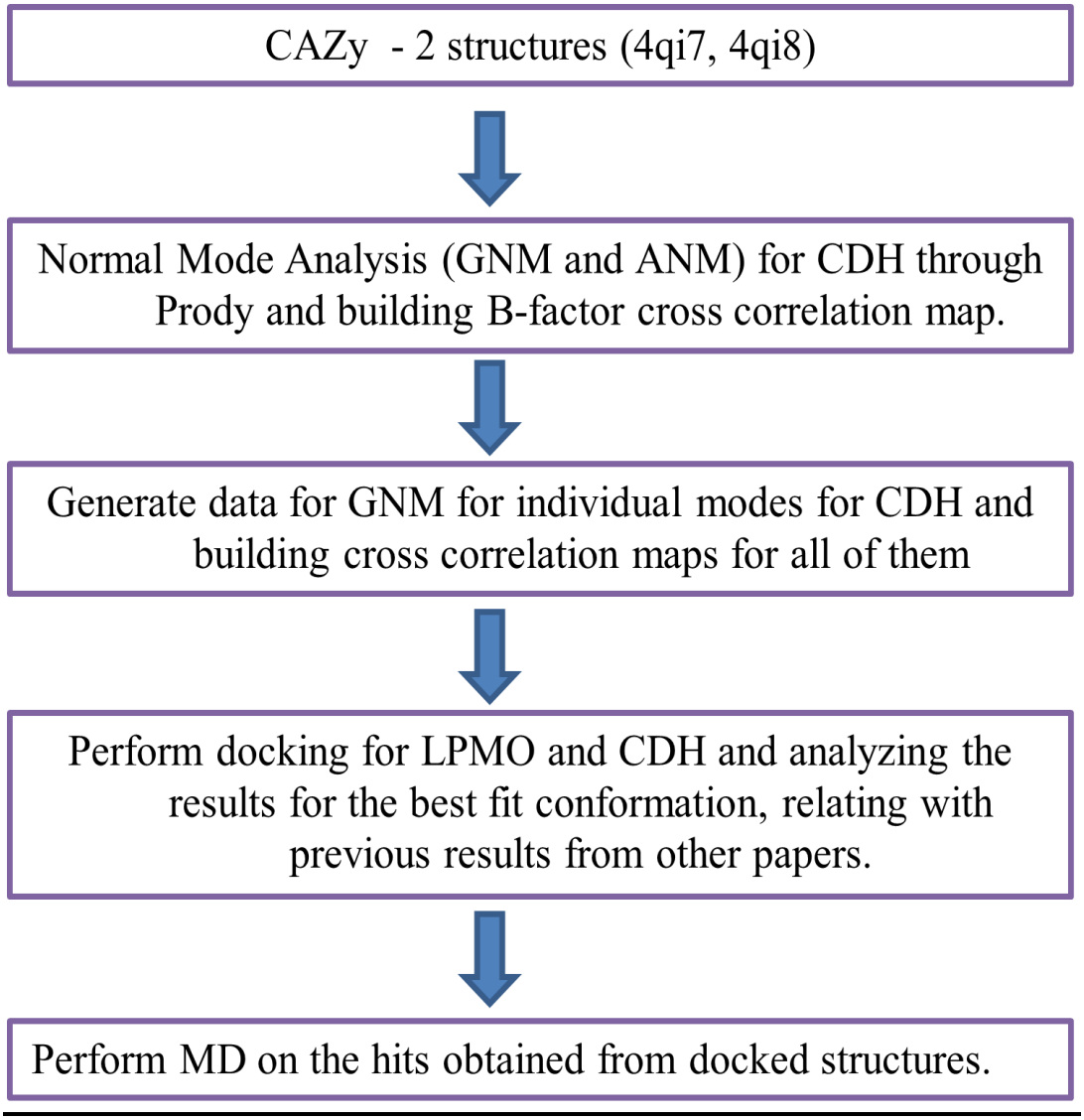

Figure 3.1: The overall method used in this study. A broader view of the workflow is described.

3.2 Normal Mode Analysis, ProDy and VMD

Normal Mode Analysis (NMA) is used for the representation of both the fast and slowest modes of a highly complex networked structure, such as a protein or any biomolecules [18]. GNM was developed to study the contribution of topological constraints on the collective protein dynamics [21]. ANM provides motions along three coordinates, and also gives rise to excessively high fluctuations because GNM is penalized against any inter-residue fluctuation [23].

While, the MD simulations focus more on exact forces between the atoms and then solve the equations in an appropriate manner, NMA approximates the equation for motion of the molecule that can be solved in a more exact manner. NMA can be simply described as study of harmonic potential wells by using analytical means and within a short time can provide the insight to more important dynamics of protein structures [19]. Two types of NMA, which are frequently used, are the Gaussian Network Model (GNM) and Anisotropic Network Model (ANM).

GNM stores the normal mode data describing the intrinsic dynamics of the protein structure by generating a Kirchoff matrix. Tirion's "single parameter model" for NMA made the energy potential more simplified, which explained NMA with uniform harmonic motion [20]. Later, Bahar and co-workers produced a much simpler version of NMA, i.e., Elastic Network Model (ENM). GNM was developed to study the contribution of topological constraints on the collective protein dynamics [21]. Thus, GNM is used for observation of global dynamic behavior. In GNM, the Cα carbon residues are represented as nodes in a network connected by springs (if within 7Å radius), undergoing Gaussian distributed fluctuations [22]. According to GNM, these fluctuations are assumed to be influenced by neighboring atoms and their influence can be measured by the local packing density of residues around every single C^{α} residue [22]. Theoretical fluctuations are computed and compared against the experimental B-factors for the validation of GNM.

ANM is a modification of GNM, where the distance cutoff is 13Å and the only difference is that the distance in GNM is in the form of vectors, whereas the distance in ANM is in a scalar form and the product of these scalar values results in anisotropic fluctuations by picking up the second derivative of the potentials for the displacement along any axis in a 3D space. Thus, ANM provides dimensionality, and also gives rise

to excessively high fluctuations [23]. In spite of lower accuracy in local relative degrees of flexibility, ANM is superior for accessing directional mechanism of motion as it creates a Kirchhoff's matrix [24].

3.2.1 ***ProDy*** (Protein Dynamics and Sequence Analysis) is an open-source and free Python package [25]. It is used for protein structural dynamics analysis and is suitable for development of various applications and for the interactive sessions. ANM and GNM analyses were performed using ProDy by generating the normal modes in VMD and visualizing the results in VMD [26]. Specifically the following were analyzed-Theoretical and experimental B-factor cross correlation map for GNM (Figure 3)

3.2.2 Individual slow modes for first 10 slowest modes for GNM (Figure 3)

3.2.3 First 10 slowest ANM modes showing the open-closed conformation states of CDH (Figure 3)

3.2.4 ***Visual Molecular Dynamics*** **(VMD)** was used to visualize the slowest modes generated. Besides the short movies of 30 seconds duration were generated for each of the 10 slowest ANM modes, which depict the open-close conformational states of CDH.

3.3 PyMOL

PyMOL is a python enhanced, OpenGL based, open source model visualization tool used widely in structural biology. It excels at 3D visualization of proteins, small molecules, density, surfaces, and trajectories. It also includes molecular editing, ray tracing, and movies. [27]

PyMOL was used for visualization of LPMO and CDH molecules and for the purpose of image generation. The tool was also used for locating the conserved residues of LPMO and verifying as to whether they lie within the interacting surface, where CDH binds. Further superposition of the different conformations of the LPMO-CDH complexes was carried out with the help of 'align' command.

3.4 Molecular Docking (ClusPro, PatchDock and ROSIE)

Molecular Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex[28]. It can be used to perform virtual screening on large libraries of compounds, rank the results, and propose structural hypotheses of how the ligands inhibit the target, which is invaluable in lead optimization. [29]

ClusPro

ClusPro is a protein-protein docking software that utilizes six different energy functions, depending on the type of protein, to define centers of highly populated clusters of low-energy docked structures, selecting those with good electrostatic and desolvation free energies for further clustering. [30] The tool involves consideration of small-angle X-ray scattering (SAXS) data, and application of attraction or repulsion, accounting for pairwise distance restraints. [31,34]

PatchDock

PatchDock is a geometry-based molecular docking algorithm, which finds docking transformations that yield good molecular shape complementarities. Contrary to Cluspro, it performs a fast transformational search since it takes into account local feature matching instead of the six-dimensional transformation space. This induces wide interface areas and lesser steric clashes[35].

Two different docking tools were hereby implemented so to validate the results obtained from the one, with the other. Moreover, superposition was carried out for conformation models obtained from one tool with the other. Both the tools justify the implication of their utility towards study of protein-protein interaction.

Docking2 (Rosie)

ROSIE is a web front-end to the Rosetta 3.x software suite, a molecular modeling software package that provides experimentally tested and rapidly evolving tools for the 3D structure prediction and high-resolution design of proteins, nucleic acids, and a growing number of non-natural polymers. The Docking2 module of RosettaDock Server performs a local docking search. In other words, the algorithm will search a set of conformations near the given starting conformation for the optimal fit between the two

partners. [36-38] This refers to as refinement of the docked structures, which was performed on the 10 structures obtained after performing protein-protein docking by ClusPro and PatchDock.

Both ClusPro and PatchDock were used for generating the docked structures which show the binding between LPMO and CDH (Figure 3). As input, LPMO (pdb id: 4qi8) was selected as receptor and CDH (pdb id: 4qi7) was selected as ligand. From each docking, 10 structures were selected among which 3 from ClusPro and 1 from PatchDock were selected on the basis of binding of LPMO with the CYT domain of CDH. Presence of conserved residues were checked and further models were refined from Rosie Docking2. Superpositions were performed then, using PyMOL and finally these models each from ClusPro and PatchDock were found to be potential ones and compared with each other for being the most suitable conformations for the interaction between LPMO and CDH.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 CDH and LPMO - Structure and Mechanism

Structures of CDH (Figure 4.1) and LPMO (Figure 4.2) from *Neurospora crassa* were obtained from Protein Data bank (PDB). They were checked in CAZy for the corresponding family information.

CDH is a two-domain protein that consists of a FAD-binding dehydrogenase (DH) domain coupled to a heme-binding cytochrome (CYT) domain. Specifically, the N-terminal CYT domain is linked to the C-terminal DH domain by a 20 residue long flexible linker ranging from residues 210 to 230 (Figure 4.1). Within the CDH, the DH domain catalyzes its substrate cellobiose by oxidation reaction to convert it to cellobiono-1.5-lactone, during with there is a reduction event on FAD. Subsequently, the reduced FAD transfers electron to the heme in CYT domain. This inter-domain electron transfer (ET) is completed only when an external electron acceptor, in this case LPMO, accepts the electron and carries out its redox reaction. In order to carry out the ET reaction in coordination with LPMO it is obvious that CDH interacts with LPMO via a specific protein-protein interaction (PPI). It is hypothesized that the ET transfer between the DH domain and CYT domain is possible due to the open-close conformational changes in CDH [42][43]. Also, it is highly likely that the ET between CDH and LPMO occur when the heme group is closer to the Cu^{2+} ion in LPMO.

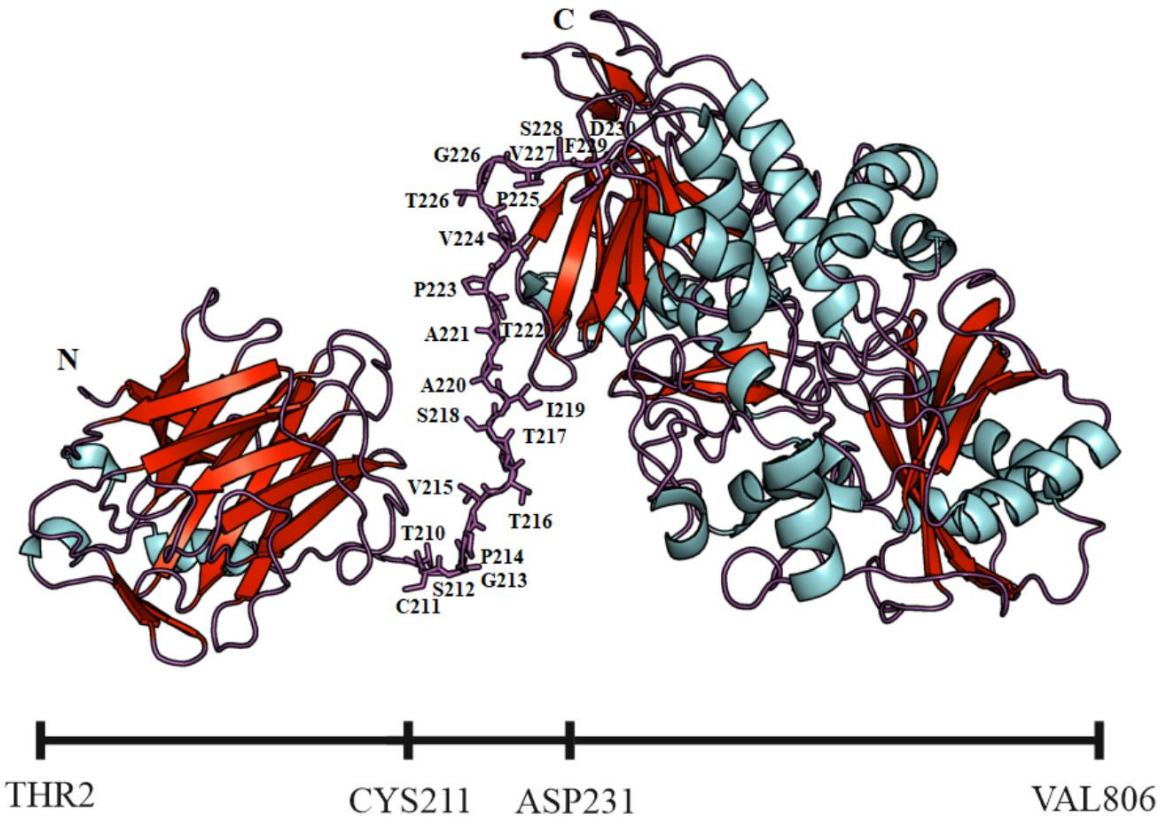

Figure 4.1: Structure of Cellobiose Dehydrogenase (pdb id: 4qi7). Cartoon representation of the structure generated using Pymol, where the helices are colored cyan, beta-strands are colored red, and loops are colored magenta; Residues depicting CYT and DH domains, and helix region. N and C terminals are labeled.

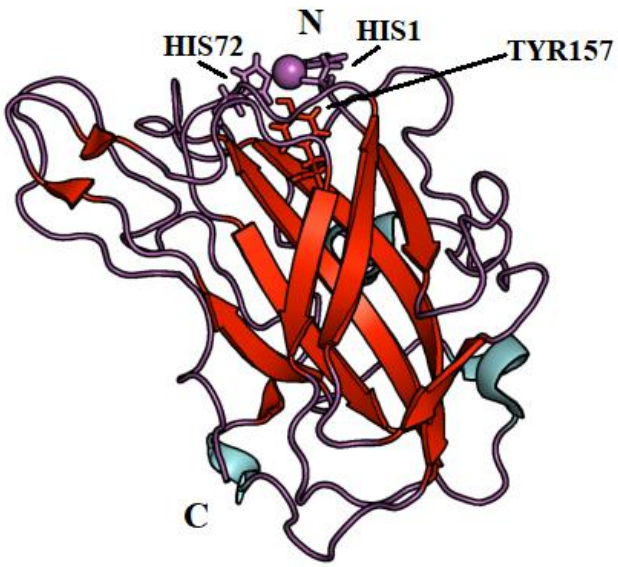

Figure 4.2: Structure of LPMO (pdb id: 4qi8). Cartoon representation of the structure generated using Pymol, where the helices are colored cyan, beta-strands are colored red, and loops are colored magenta; Residues depict the His-Tyr-His brace and the Cu ion centre is highlighted in purple. N and C terminals are labeled.

4.2 Normal Mode Analysis

The LPMO from *Neurospora crassa* belongs to the family AA9, as classified in CAZy. CDH on the other hand is classified into three families, AA3 1, AA8, and CBM1. The underlying principle of ENM is that the intrinsic dynamics of any protein is encoded within its structure. Thus, by calculating the normal modes of the protein, one can identify slow and fast modes, where the slow modes correlate to large structural changes and fast modes correlate to kinetically "hot residues". These residues are most likely to be catalytically important and thus directly involved in the protein's function. Here, using Gaussian Network Model (GNM) and Anisotropic Network Model (ANM), ENM models were generated to understand the open-close conformational changes in CDH. While, GNM provides the magnitude of fluctuations at the residue level (and is distributed in a Gaussian manner), ANM provides the directionality of the individual modes. Thus, GNM and ANM complement each other to provide collective fluctuations that are directional and are biologically functional in a protein.

4.2.1 B-factor cross correlation map of CDH

Debye-Waller temperature factors or B-factors are experimental values that quantify the degree of flexibility in a protein, where a residue with higher B-factor value is considered to be relatively more flexible than other residues with lower B-factor values. The B-factor is a representation of isotropic displacements of atoms in a molecule, involved during refinement. Refinement is performed in order to improve the agreement between the modeled structure and the reflections measured in the diffraction experiment. The isotropic displacement parameters are variables during a typical refinement and may therefore also contain contributions of apparent displacements resulting from the use of an inadequate model or from overlooked errors in the X-ray data. Displacement of atoms from their mean position in a crystal structure diminishes the scattered X-ray intensity. The displacement may be the result of temperature-dependent atomic vibrations or static disorder in a crystal lattice. [44]

The experimental B-factors and MSF are correlated; the experimental, shows higher fluctuations than the theoretical ones. Yet, the overall range remains the same, and the experimental values lie around the predicted, justifying the significance and accuracy of prediction. The correlation coefficient signifies the extent of similarity. It is considered to be good if the value is above a certain threshold (0.58) [45]. Here, the correlation coefficient of B-factors vs MSF is 0.64, indicating a higher accuracy of the ENM generated. (Figure 4.3)

Correlation refers, as to how the atomic motions of one residue in a molecule, effect the other residues. Accordingly, the graph is made taking residue numbers on both x and y axes. The MSF obtained corresponds to the respective motions of residues of CDH, where the trend of fluctuation clearly shows the boundary between the two domains, DH and CYT of CDH. The N-terminal residues (2 to 210) belong to CYT domain and the C-terminal (231 to 806) belong to DH domain, which is the relatively larger in size.

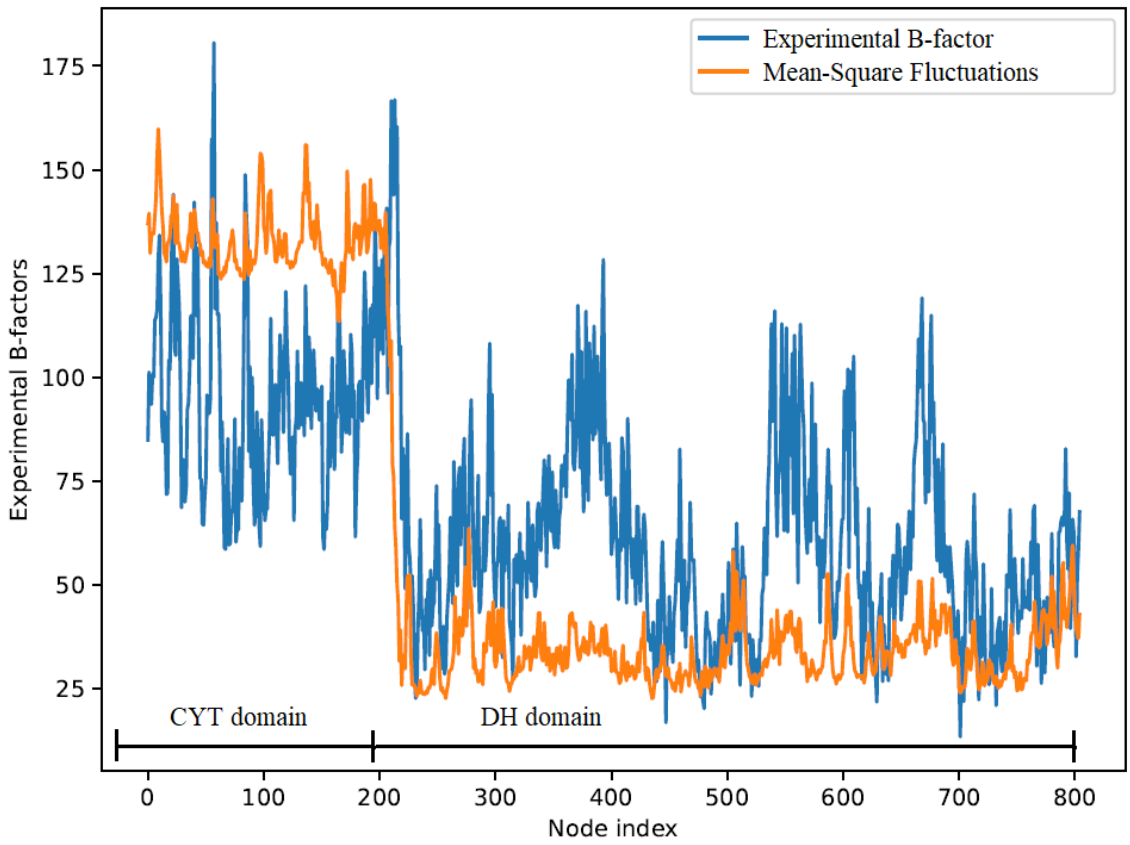

Figure 4.3: Theoretical and experimental B-factor cross correlation map for GNM. The experimental B-factors as mentioned in the pdb file and theoretical B-factors as depicted after running GNM for CDH.

4.2.2 Individual slow modes for first 10 GNM modes of CDH

GNM calculates mean-square fluctuations (MSF) for the given protein and it has been proved that the MSF obtained via a coarse-grained model is related to the B-factors and considering the crystal contacts, the MSF are highly significant [24]. The ProDy's Linux commands (or Python script) included a function 'showCrosscorr()' [Appendix 1], which depicted the cross-correlations between fluctuations in atomic positions, in the first 10 modes of GNM[39]. These 10 slowest modes generated are shown in Figure 4.4.

Visualizing individual GNM modes are valuable to identify the fluctuations in CDH. Specifically, the slowest modes correlate to large conformational changes. For example, the first slowest mode distinctly shows the domain boundary of CYT at 210^{th} residue, including the hinge region and DH domain, and the correlation map identifies the positively correlated (colored yellow) and negatively correlated (colored purple) residues of DH and CYT domain. A correlation of +1 signifies perfect positive correlation and -1 signifies perfect negative correlation. The positive correlation between the residues, suggest their synergistic motion in a coordinated way, i.e. motion of the specific fragment of the domain as a whole. Overall, some fragments move together and some move in an opposite way indicating that they are correlated or anticorrelated, respectively. As we begin visualizing the first mode and move on to consecutive modes, the perturbations seem to increase and vary significantly for each residue pair. Second and third slowest modes show a certain kind of diagonal pattern, which depicts the motion of the residues accordingly, together or in opposite direction. As per our hypothesis, the conserved residues shall be the positions with the most coordinated atomic motion in all the modes, thereby contributing towards regulation of smooth electron transfer between the two domains of CDH. This is referred to as Inter-domain electron transfer (IET).

The lighter regions on the maps indicate a positive correlation between the residues, suggesting their synergistic fluctuations. The darker regions on the maps on the other hand refer to the negative correlation between the corresponding residues, affecting their opposite fluctuations/motions with respect to each other. The positive-to-negative correlation of the structural dynamics between each residue pair is clearly visible for each of the modes.

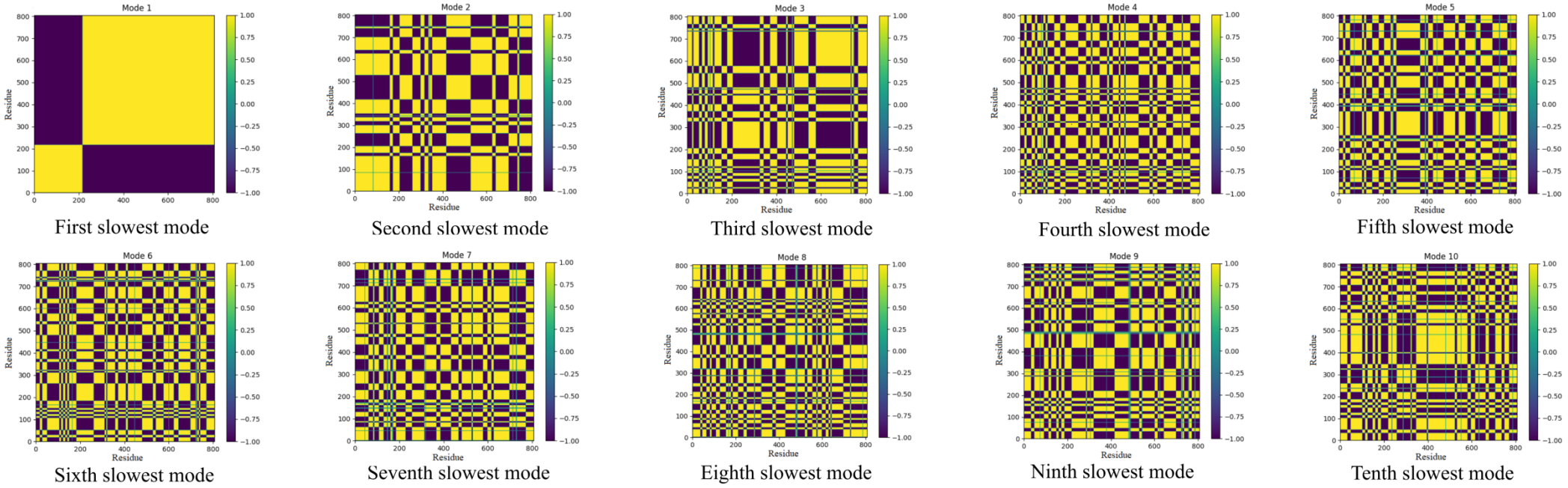

Figure 4.4 : Individual slowest modes for first 10 modes for GN

4.2.3 Identification of Open-Closed conformations of CDH

The interdomain electron transfer has been studied previously in *Myriococcum thermophilum* (MtCDH) whereby a closed state is observed in which the CYT domain is docked onto the DH domain in an arrangement so as to efficiently allow IET from FAD to haem b[40]. Haem b is a type of hemoprotein involved in electron transfer. It is attached to surrounding protein by a single coordination bond. In the case of *Neurospora crassa* (NcCDH), the CYT and DH domains present open states as well, apart from the closed one, with different conformations of the flexible linkers and different relative orientations of the domains [40]. The top 10 slowest modes obtained via ANM are shown in Figure 4.5. The results are tabulated in Table 4.1.

MODES	**CONFORMATIONAL STATE**
Slowest Mode 1	Closed
Slowest Mode 2	Closed
Slowest Mode 3	Closed
Slowest Mode 4	Closed
Slowest Mode 5	Unclear
Slowest Mode 6	Only certain residues in motion
Slowest Mode 7	Only certain residues in motion
Slowest Mode 8	Open
Slowest Mode 9	Only certain residues in motion
Slowest Mode 10	Open

Table 4.1: Slowest modes with their corresponding open or closed conformational states.

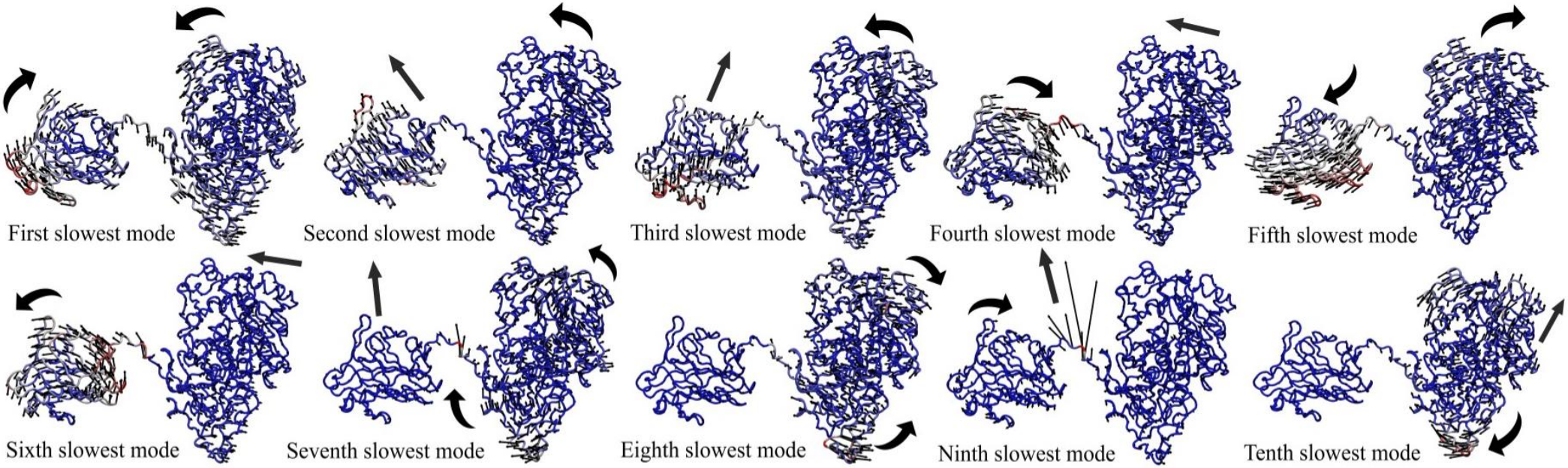

Figure 4.5 : First 10 slowest ANM modes showing the open-closed conformation states of CDH

The ANM modes, involve the movements of one domain as a whole, with respect to the other. These movements correspond to open and closed states of CDH, predicted accordingly.

Closed state refers to the nearby positioning of the domains, bending significantly inwards towards each other over the hinge region. Open state, in contrast, positions the domains bending away from each other. The to-and-fro movements for transition between the two states suggest that open and closed states facilitate electron transfer. The larger part of the CDH is the DH domain. The other one is CYT domain, to which LPMO shall bind for the electron transfer. In the figure, both the arrows when inwards, display the closed state and vice versa when open state is observed (Figure 8). Mode 1 has all the directions of both the domains, towards inside. This shows the closed position of the CDH. Also, the direction of the hinge region is synchronized with the inwards movement of domains, as it would certainly shift downwards to bring about that kind of motion. Modes 3 and 4 are typically closed on the same lines, with a little difference in the motion direction. Mode 2 has a slightly broader angle for of motion between the two domains. Nonetheless, this also seems a closed state. Modes 5 and 6 are of mixed dynamics with respect to a defined closed or open state, but the residues are definitely in a specific motion, although a closed state can be thought of. Modes 7 and 9 have a dominant motion of few residues of the hinge region, which has no practical significance with respect to its contribution in open or closed states. Modes 8 and 10 have the DH domain predominantly in motion. Both these modes can be inferred for open states with slight movements of the CYT domain.

The study in MtCDH [40] relates strongly to the results obtained for NcCDH. The movies generated using VMD, also indicate the same to-and-fro movements of the CYT and DH domains of CDH suggesting the open and closed states to facilitate electron transfer.

4.3 Molecular docking

In order to study the interaction between LPMO and CDH, molecular docking was carried out and protein-protein interaction maps were generated for all the models obtained after docking. The methodology followed and criteria for selection of models have been described in Figure 4.6.

The binding of LPMO with the CYT domain of CDH was observed using PyMOL. This is because the flow of electron transfer proceeds from DH domain to CYT domain within CDH and further from CYT domain to LPMO. The conformations which were bound more adhesively and closely to the CYT domain, were selected. These were:

Models selected from ClusPro: Model02, Model08, and Model09

Models selected from PatchDock: Res06

After this, the presence of conserved residues of LPMO [Li *et al.*], were observed in Pymol for positioning in the part interacting with CDH (Figure 4.7). Superposition carried out between each one of the models from ClusPro with the model from PatchDock gave rise to 3 total combinations. All these combinations had root mean square deviation (RMSD) values between 0.176 and 0.214 which indicates that the structures from ClusPro resemble to the structure obtained from PatchDock. This validates the accuracy and correlation of all the models.

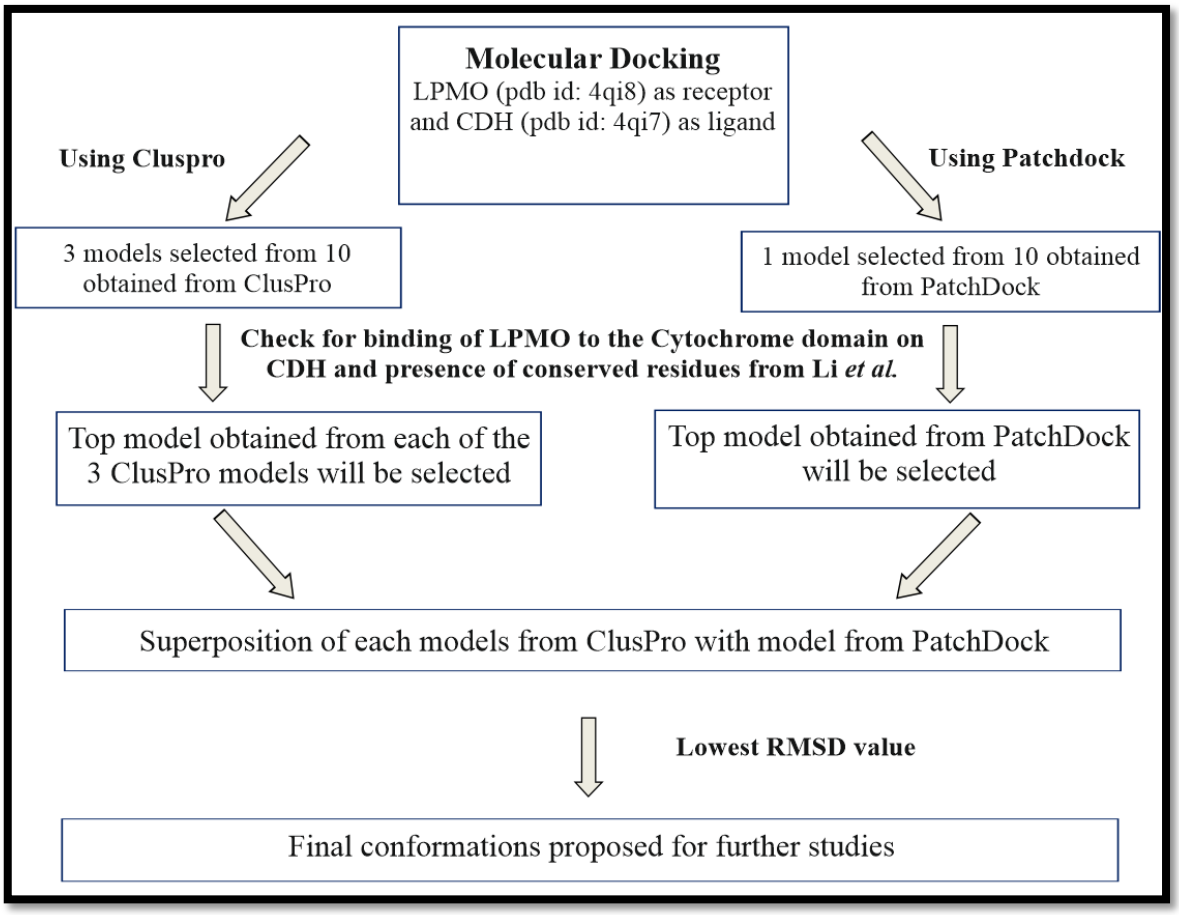

Figure 4.6 : Step-wise schematic representation for molecular docking

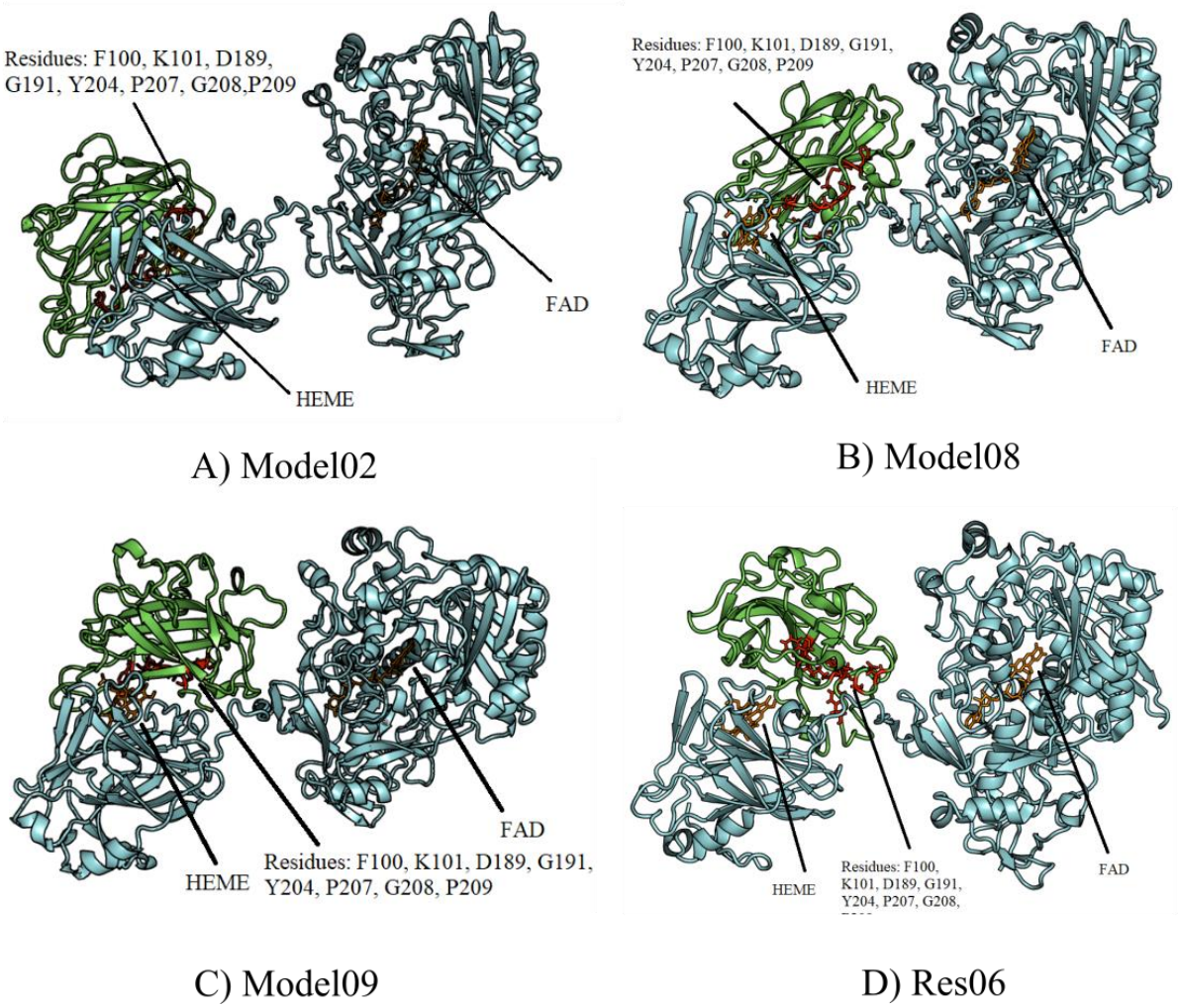

Figure 4.7: Cartoon representation of models selected after docking. 2A) Model02 B) Model08 C) Model09 D) Res06. The residues in blue belong to CDH, the ones in green are of LPMO and the ones in red are residue number 100, 101, 189, 191, 204, 207, 208, 209 of LPMO, as suggested by Li et al., meant to interact with CDH. Heme and FAD are labeled in red.

4.4 Protein-Protein interaction map

Protein-Protein interaction map is required so as to study the individual residue-residue interactions between LPMO and CDH, in order to cross check the importance of presence of conserved residues of LPMO, proposed by Li *et al.*, in the interaction map, and their involvement in the electron transfer mechanism. The docked files were entered into PDBsum and protein-protein interaction map was obtained for each of the 4 models.

As evident from the images of docked structures (Figure 4.7), all 4 models have LPMO binding significantly to the CYT domain of CDH. The heme group in CYT domain and FAD in DH domain are also shown in the figure. Heme group lies in close vicinity to LPMO, indicating its need in the electron transfer. The conserved residues as reported by Li *et al*., are F100, K101, D189, G191, Y204, P207, G208, and P209. Expectedly, all of these should be present in the interaction map obtained from PDBsum. However, only are present in the maps. Model02 from ClusPro has none of these residues interacting with any residues of CDH (Figure 4.8A). Although, I find that only the residues near these ones interact with CDH. Ile 102 and Tyr103 from LPMO interact with Tyr 99 and, Thr75 and Gly98 from CDH, respectively, through non-bonded contacts. Similarly, Lys205 interacts with Gln182 of CDH, and Ser210 interacts with Met180. Tyr99, as evident, seems to be a central residue for multiple interactions.

Model08 from ClusPro has Tyr204 from LPMO dominantly interacting with Gln175 from CDH (Figure 4.8B). Tyr204 is one of the residues found out to be conserved by Li *et al*. Gln175 is a residue which is one of the main centers of interaction from CDH, as it interacts with many residues of LPMO through different kinds of interactions. The map also indicates that the two groups of residues in CDH, one made by Phe159, Gly181, Met180, Asp177, Gln175 and other made by Lys27, Ala100, Pro102, His176, Tyr99, Pro225 are actively involved in interaction with LPMO. So, some of these residues must definitely be involved in electron transfer. It is clear that Tyr99 is the residue involved both in Model02 and Model08.

Similarly, Model09 also has Tyr99 interacting (Figure 4.8C). Like Model08, Tyr204 is present in the map and comparatively more number of scattered residues than other models, is found rather than a cleft-kind formation or cluster of interacting residues.

Res06 from PatchDock has the residue Asp189 which matches with the conserved residues reported by Li *et al.* This model also has Tyr99 from CDH quite dominant for interaction (Figure 4.8D). So, all the four models contain Tyr99 from CDH which seems to play an important role in the interaction. Also, Asp177 is common in three of the four models, with exception being in Model02. But then, Model02 has all neighboring residues of Asp177. So all these factors give us an insight that Tyr99 and Asp177 from CDH must be important and involved in the interaction with LPMO.

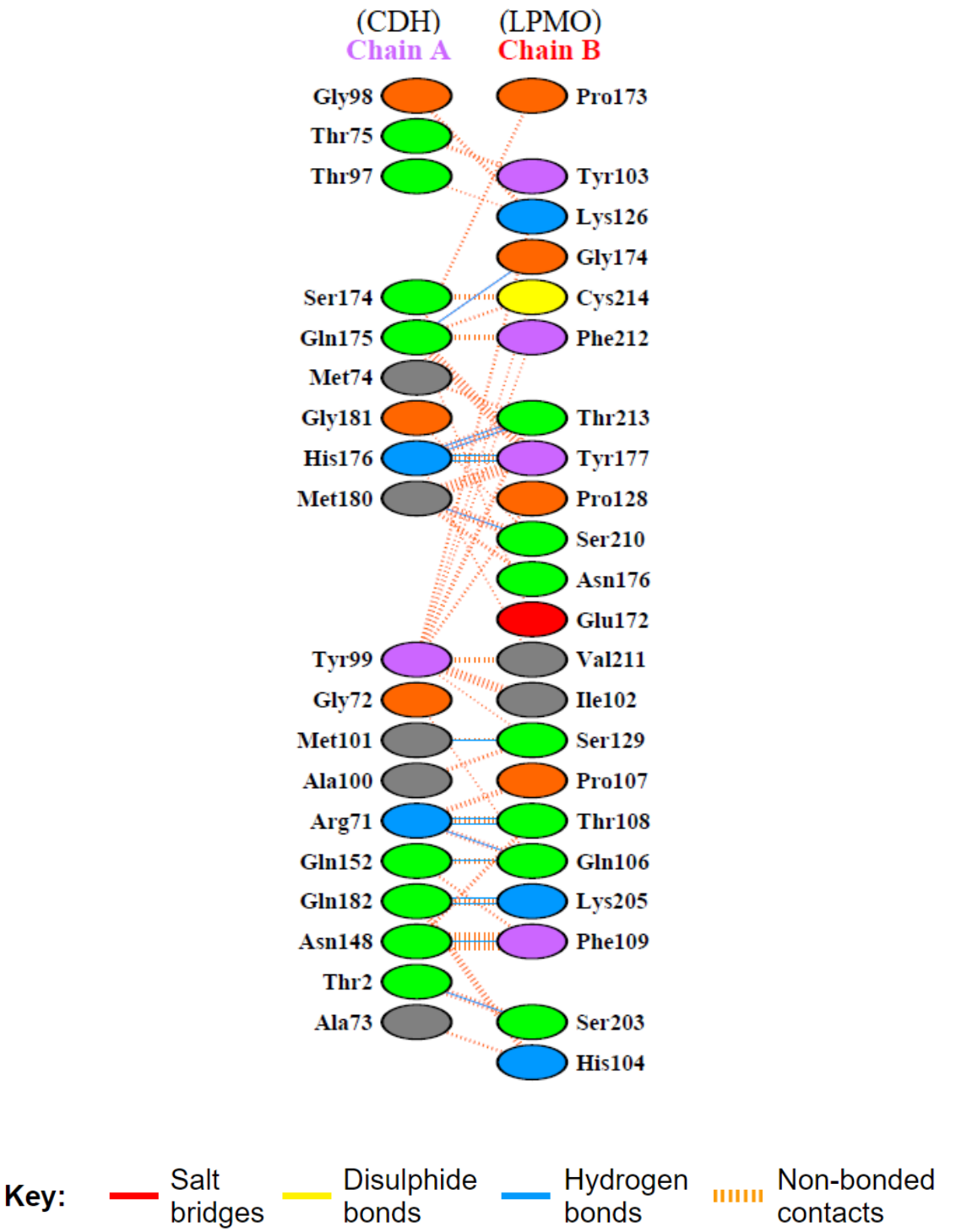

Figure 4.8 A): Protein-Protein interaction map for Model02 of ClusPro. The number of H-bond lines between any two residues indicates the number of potential hydrogen bonds between them. For non-bonded contacts, which can be plentiful, the width of the striped line is proportional to the number of atomic contacts. Chain A represents CDH and Chain B represents LPMO.

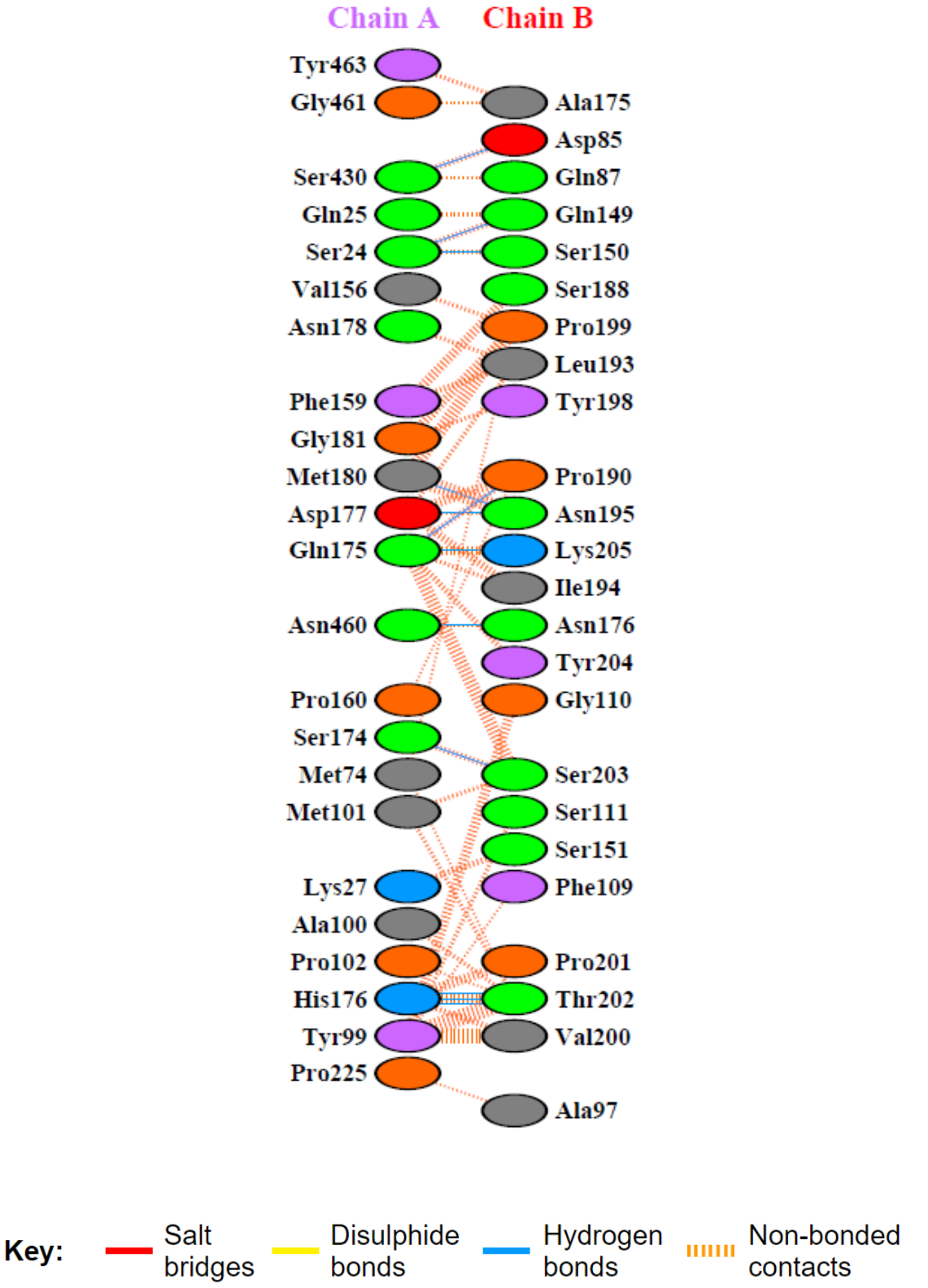

Figure 4.8 B): Protein-Protein interaction map for Model08 of ClusPro.

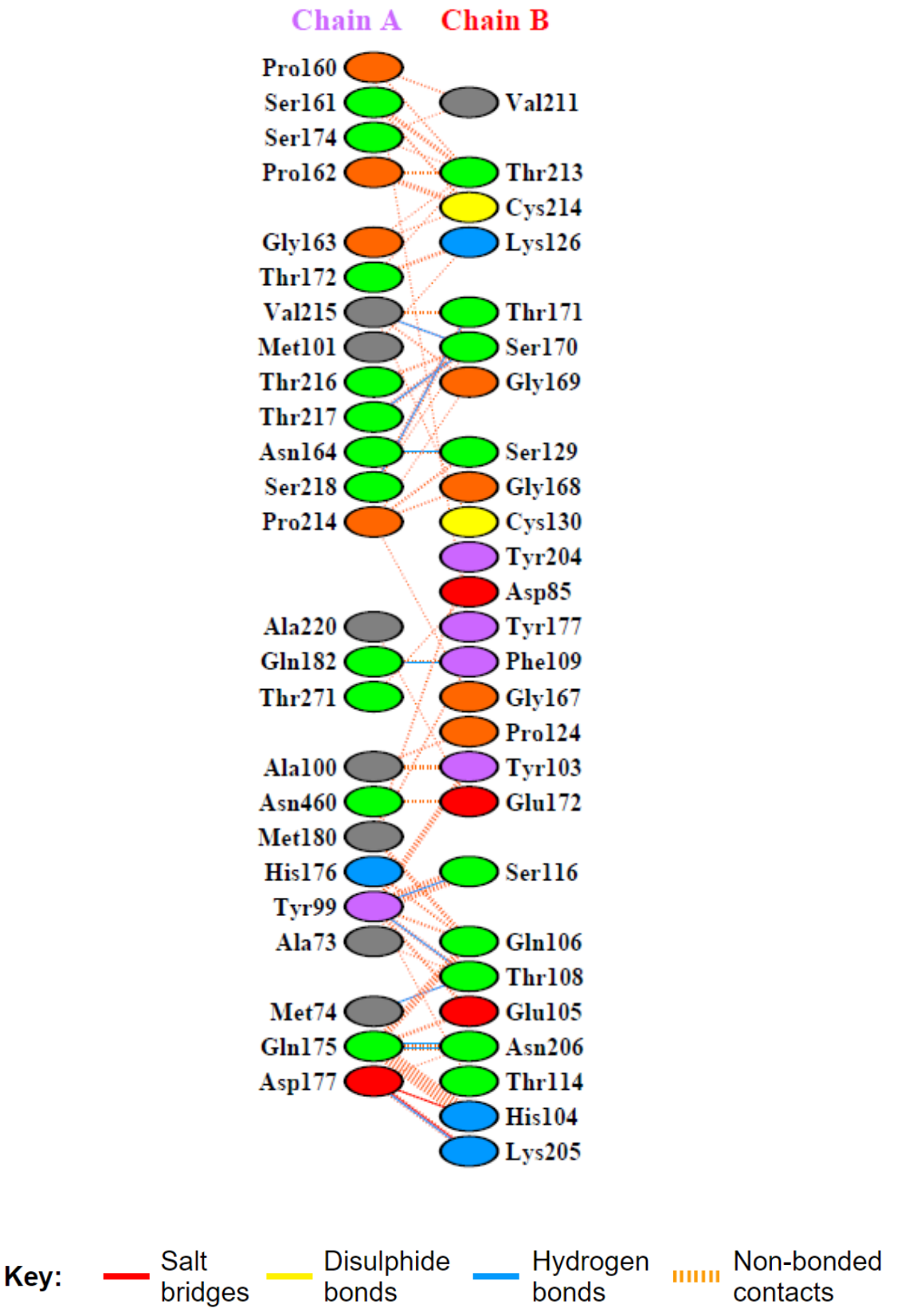

Figure 4.8 C): Protein-Protein interaction map for Model09 of ClusPro.

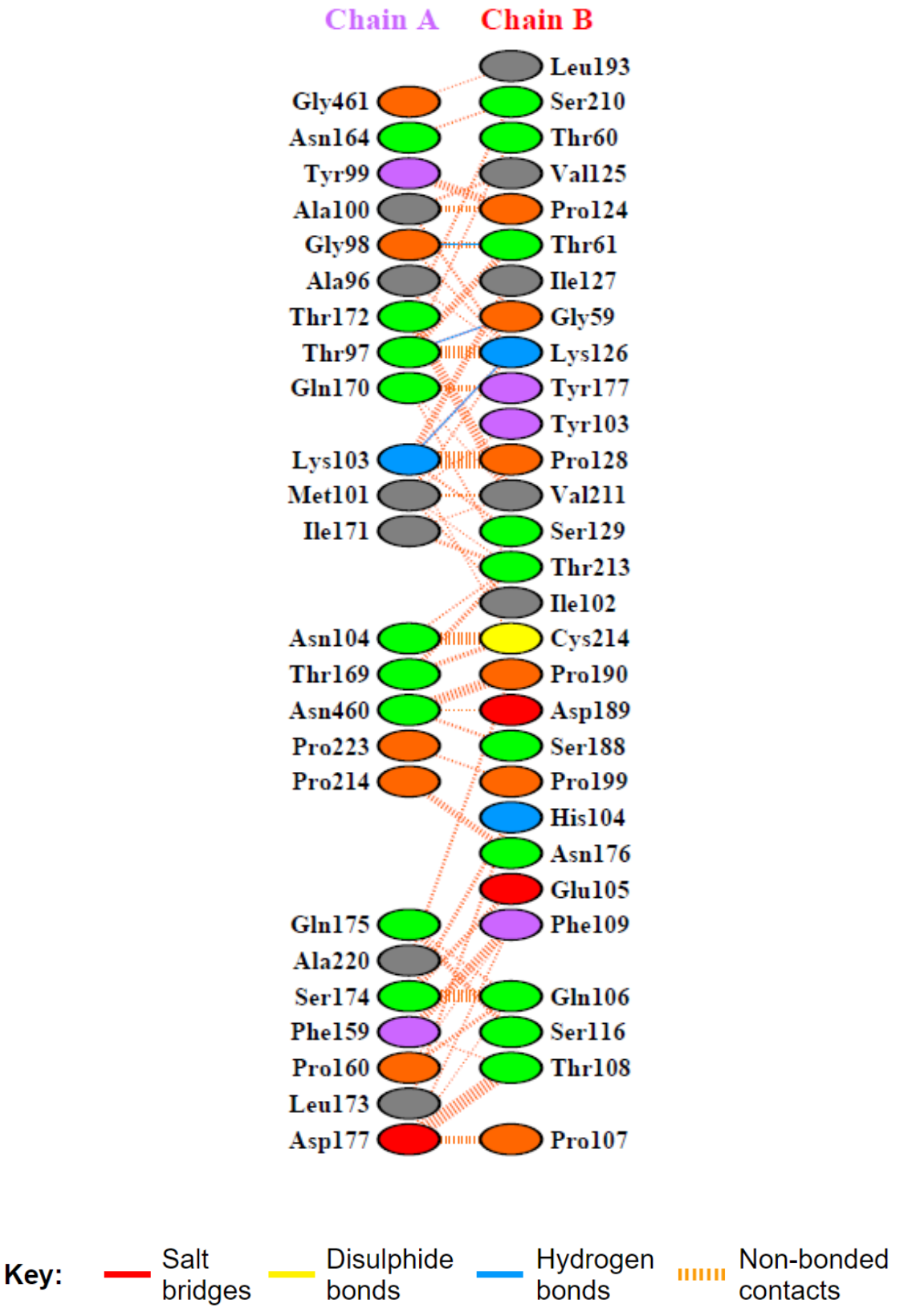

Figure 4.8 D): Protein-Protein interaction map for Res06 of PatchDock.

CHAPTER 5

CONCLUSION

Biofuels are a new source of energy from biomass. Microorganisms, specially fungi, degrade the plant cell wall through synergistic cocktails of enzymes whereby glucose is released in free form providing energy to the microbes. *Trichoderma reesei, Aspergillus niger, Neurospora crassa, Saccharomyces cerevisiae* and *Ruminococcus flavefaciens* are the most common microbes involved. Cellulose degradation happens by:

1) Endoglucanase
2) Cellobiohydrolase
3) β-glucosidase

LPMOs are the only enzymes to act on crystalline surfaces and break them into smaller substances. It can hereby be concluded that LPMO and CDH definitely hold a certain importance for their existing mechanism of action for interaction, which have been explored here.

The complete study revolves around LPMO and CDH and their structural and dynamical properties, the knowledge of which are required in order to understand the mechanism of their interaction, which is unexplored till date. For the same, my approach was to study the conformational changes the molecules undergo. The activation of LPMO is an important step in terms of cellulose degradation.

From this study, following conclusions can be drawn:

- Normal Mode Analysis indicates that the Mean square fluctuations are an indication of the motion of residues of CDH and that the ENM generated is accurate since the correlation is above the threshold of 0.58. The overall motion of CDH molecule as a whole has been observed by means of individual first 10 slowest modes whereby these indicate large conformational changes due to positive or negative correlation between the residues affecting their fluctuations. The ANM modes involve movement of CDH's domains towards or opposite to each other, in order to identify open or closed conformations. These to-and-fro movements are required to facilitate electron transfer.
- Molecular Docking results give us an insight about the interacting partners in LPMO-CDH complex. It is evident that LPMO interacts with CYT domain of

CDH. The individual protein-protein interaction maps for the four docked complexes indicate detailed information about the specific set of residues involved in the interaction. These residues must be important to facilitate electron transfer for the activation of LPMO.

APPENDIX 1

Python script to generate individual cross correlations for first 10 GNM modes

```
from prody import *
from matplotlib.pylab import *

ion()  # turn interactive mode on
pdb_input =  parsePDB('4qi7_atom.pdb')
Calphas = pdb_input.select('calpha') # selects only Cα atoms for building matrix
gnm = GNM('4qi7 GNM')
gnm.buildKirchhoff(calphas, cutoff=7, gamma=1) # build Kirchhoff matrix
gnm.getKirchhoff() # display Kirchhoff matrix
gnm.calcModes() # calculates normal modes
gnm.getEigvals().round(3) # display Eigenvalues matrix rounding off till 3 places of decimal
gnm.getEigvecs().round(3) # display Eigenvectors matrix rounding off till 3 places of decimal
gnm.getCovariance().round(2) # display Covariance matrix rounding off till 2 places of decimal
slowest_mode = gnm[0] # for first normal mode
slowest_mode.getEigval().round(3)
slowest_mode.getEigvec().round(3)
showCrossCorr(gnm) # displays the cross correlation map for first normal mode
slowest_mode = gnm[1] # for second normal mode
showCrossCorr(gnm) # displays the cross correlation map for second normal mode
slowest_mode = gnm[2] # for third normal mode
showCrossCorr(gnm) # displays the cross correlation map for third normal mode
```

```
slowest_mode = gnm[3] # for fourth normal mode
showCrossCorr(gnm)  # displays the cross correlation map for fourth normal mode
slowest_mode = gnm[4] # for fifth normal mode
showCrossCorr(gnm)  # displays the cross correlation map for fifth normal mode
slowest_mode = gnm[5] # for sixth normal mode
showCrossCorr(gnm)  # displays the cross correlation map for sixth normal mode
slowest_mode = gnm[6] # for seventh normal mode
showCrossCorr(gnm)  # displays the cross correlation map for seventh normal mode
slowest_mode = gnm[7] # for eighth normal mode
showCrossCorr(gnm)  # displays the cross correlation map for eighth normal mode
slowest_mode = gnm[8] # for ninth normal mode
showCrossCorr(gnm)  # displays the cross correlation map for ninth normal mode
slowest_mode = gnm[9] # for tenth normal mode
showCrossCorr(gnm)  # displays the cross correlation map for tenth normal mode
```

REFERENCES

1. G. Vaaje-Kolstad, B. Westereng, S. Horn, Z. Liu, H. Zhai, M. Sørlie and V. Eijsink, "An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides", *Science*, vol. 330, no. 6001, pp. 219-222, 2010.

2. 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.

3. 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.

4. K. Johansen, "Lytic Polysaccharide Monooxygenases: The Microbial Power Tool for Lignocellulose Degradation", *Trends in Plant Science*, vol. 21, no. 11, pp. 926-936, 2016.

5. A. Levasseur, E. Drula, V. Lombard, P. Coutinho and B. Henrissat, "Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes", *Biotechnology for Biofuels*, vol. 6, no. 1, p. 41, 2013.

6. S. Kim, J. Stahlberg, M. Sandgren, R. Paton and G. Beckham, "Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism", *Proceedings of the National Academy of Sciences*, vol. 111, no. 1, pp. 149-154, 2013.

7. C. Kjaergaard, M. Qayyum, S. Wong, F. Xu, G. Hemsworth, D. Walton, N. Young, G. Davies, P. Walton, K. Johansen, K. Hodgson, B. Hedman and E. Solomon, "Spectroscopic and computational insight into the activation of O2 by the mononuclear Cu center in polysaccharide monooxygenases", *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, pp. 8797-8802, 2014.

8. 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

FungusPhanerochaete chrysosporium", *Journal of Biological Chemistry*, vol. 288, no. 18, pp. 12828-12839, 2013.

9. V. Moses, Ö. Tastan Bishop and K. Lobb, "The evaluation and validation of copper (II) force field parameters of the Auxiliary Activity family 9 enzymes", *Chemical Physics Letters*, vol. 678, pp. 91-97, 2017.

10. E. Hedegård and U. Ryde, "Multiscale Modelling of Lytic Polysaccharide Monooxygenases", *ACS Omega*, vol. 2, no. 2, pp. 536-545, 2017.

11. E. Hedegård and U. Ryde, "Targeting the reactive intermediate in polysaccharide monooxygenases", *JBIC Journal of Biological Inorganic Chemistry*, vol. 22, no. 7, pp. 1029-1037, 2017.

12. R. Quinlan, M. Sweeney, L. Lo Leggio, H. Otten, J. Poulsen, K. Johansen, K. Krogh, C. Jorgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. Walter, P. Dupree, F. Xu, G. Davies and P. Walton, "Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components", *Proceedings of the National Academy of Sciences*, vol. 108, no. 37, pp. 15079-15084, 2011.

13. K. Sprenger, A. Choudhury, J. Kaar and J. Pfaendtner, "Lytic Polysaccharide Monooxygenases ScLPMO10B and ScLPMO10C Are Stable in Ionic Liquids As Determined by Molecular Simulations", *The Journal of Physical Chemistry B*, vol. 120, no. 16, pp. 3863-3872, 2016.

14. F. Aachmann, M. Sorlie, G. Skjak-Braek, V. Eijsink and G. Vaaje-Kolstad, "NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions", *Proceedings of the National Academy of Sciences*, vol. 109, no. 46, pp. 18779-18784, 2012.

15. J. Loose, Z. Forsberg, D. Kracher, S. Scheiblbrandner, R. Ludwig, V. Eijsink and G. Vaaje-Kolstad, "Activation of bacterial lytic polysaccharide monooxygenases with cellobiose dehydrogenase", *Protein Science*, vol. 25, no. 12, pp. 2175-2186, 2016.

16. G. Courtade, R. Wimmer, Å. Røhr, M. Preims, A. Felice, M. Dimarogona, G. Vaaje-Kolstad, M. Sørlie, M. Sandgren, R. Ludwig, V. Eijsink and F. Aachmann, "Interactions of a fungal lytic polysaccharide monooxygenase with β-glucan substrates and cellobiose dehydrogenase", *Proceedings of the National Academy of Sciences*, vol. 113, no. 21, pp. 5922-5927, 2016.

17. 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.

18. M. Chapman, "Normalizing Normal Mode Analysis", *Structure*, vol. 15, no. 2, pp. 135-136, 2007.

19. K. Hinsen, "Analysis of domain motions by approximate normal mode calculations", Proteins: Structure, Function, and Genetics, vol. 33, no. 3, pp. 417-429, 1998.

20. M. Tirion, "Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis", Physical Review Letters, vol. 77, no. 9, pp. 1905-1908, 1996.

21. Bahar, A. Atilgan and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential", Folding and Design, vol. 2, no. 3, pp. 173-181, 1997.

22. T. Haliloglu, I. Bahar and B. Erman, "Gaussian Dynamics of Folded Proteins", Physical Review Letters, vol. 79, no. 16, pp. 3090-3093, 1997.

23. A. Atilgan, S. Durell, R. Jernigan, M. Demirel, O. Keskin and I. Bahar, "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model", *Biophysical Journal*, vol. 80, no. 1, pp. 505-515, 2001.

24. S. Kundu, J. Melton, D. Sorensen and G. Phillips, "Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models", Biophysical Journal, vol. 83, no. 2, pp. 723-732, 2002.

25. A. Bakan, L. Meireles and I. Bahar, "ProDy: Protein Dynamics Inferred from Theory and Experiments", *Bioinformatics*, vol. 27, no. 11, pp. 1575-1577, 2011.

26. W. Humphrey, A. Dalke and K. Schulten, "VMD: Visual molecular dynamics", *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33-38, 1996.

27. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

28. Lengauer and M. Rarey, "Computational methods for biomolecular docking", *Current Opinion in Structural Biology*, vol. 6, no. 3, pp. 402-406, 1996.

29. Morris, G. and Lim-Wilby, M. (2008). Molecular Docking. *Methods in Molecular Biology*, pp.365-382.

30. Lengauer and M. Rarey, "Computational methods for biomolecular docking", *Current Opinion in Structural Biology*, vol. 6, no. 3, pp. 402-406, 1996.

31. D. Kozakov, D. Hall, B. Xia, K. Porter, D. Padhorny, C. Yueh, D. Beglov and S. Vajda, "The ClusPro web server for protein–protein docking", *Nature Protocols*, vol. 12, no. 2, pp. 255-278, 2017.

32. D. Kozakov, D. Beglov, T. Bohnuud, S. Mottarella, B. Xia, D. Hall and S. Vajda, "How good is automated protein docking?", *Proteins: Structure, Function, and Bioinformatics*, vol. 81, no. 12, pp. 2159-2166, 2013.

33. D. Kozakov, R. Brenke, S. Comeau and S. Vajda, "PIPER: An FFT-based protein docking program with pairwise potentials", *Proteins: Structure, Function, and Bioinformatics*, vol. 65, no. 2, pp. 392-406, 2006.

34. S. Comeau, D. Gatchell, S. Vajda and C. Camacho, "ClusPro: an automated docking and discrimination method for the prediction of protein complexes", *Bioinformatics*, vol. 20, no. 1, pp. 45-50, 2003.

35. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov and H. Wolfson, "PatchDock and SymmDock: servers for rigid and symmetric docking", *Nucleic Acids Research*, vol. 33, no., pp. W363-W367, 2005.

36. S. Lyskov and J. Gray, "The RosettaDock server for local protein-protein docking", *Nucleic Acids Research*, vol. 36, no., pp. W233-W238, 2008.

37. S. Chaudhury, M. Berrondo, B. Weitzner, P. Muthu, H. Bergman and J. Gray, "Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2", *PLoS ONE*, vol. 6, no. 8, p. e22477, 2011.

38. S. Lyskov, F. Chou, S. Conchúir, B. Der, K. Drew, D. Kuroda, J. Xu, B. Weitzner, P. Renfrew, P. Sripakdeevong, B. Borgo, J. Havranek, B. Kuhlman, T. Kortemme, R. Bonneau, J. Gray and R. Das, "Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE)", *PLoS ONE*, vol. 8, no. 5, p. e63906, 2013.

39. A. Bakan, L. Meireles and I. Bahar, "ProDy: Protein Dynamics Inferred from Theory and Experiments", *Bioinformatics*, vol. 27, no. 11, pp. 1575-1577, 2011.

40. T. Tan, D. Kracher, R. Gandini, C. Sygmund, R. Kittl, D. Haltrich, B. Hällberg, R. Ludwig and C. Divne, "Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation", *Nature Communications*, vol. 6, no. 1, 2015.

41. 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.

42. C. Phillips, W. Beeson, J. Cate and M. Marletta, "Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation byNeurospora crassa", *ACS Chemical Biology*, vol. 6, no. 12, pp. 1399-1406, 2011.

43. 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.

44. K. Trueblood, H. Bürgi, H. Burzlaff, J. Dunitz, C. Gramaccioli, H. Schulz, U. Shmueli and S. Abrahams, "Atomic Dispacement Parameter Nomenclature. Report of a Subcommittee on Atomic Displacement Parameter Nomenclature", *Acta Crystallographica Section A Foundations of Crystallography*, vol. 52, no. 5, pp. 770-781, 1996.

45. S. Kundu, J. Melton, D. Sorensen and G. Phillips, "Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models", *Biophysical Journal*, vol. 83, no. 2, pp. 723-732, 2002.

46. M. McCann, B. Wells and K. Roberts, "Complexity in the spatial localization and length distribution of plant cell-wall matrix polysaccharides", *Journal of Microscopy*, vol. 166, no. 1, pp. 123-136, 1992.

47. P. Harris, D. Welner, K. McFarland, E. Re, J. Navarro Poulsen, K. Brown, R. Salbo, H. Ding, E. Vlasenko, S. Merino, F. Xu, J. Cherry, S. Larsen and L. Lo Leggio, "Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family", *Biochemistry*, vol. 49, no. 15, pp. 3305-3316, 2010.