

# **Structure-function-dynamics of PanPL, a bacterial polysaccharide lyase from the PL-5 family**

*By*

**PRERANA DASH**

**LIFE11201604011**

**National Institute of Science Education and Research,  
Bhubaneswar**

*A thesis submitted to the*

*Board of Studies in Life Science*

*In partial fulfilment of requirements  
for the Degree of*

**DOCTOR OF PHILOSOPHY**

*of*

**HOMI BHABHA NATIONAL INSTITUTE**




**March, 2023**

# Homi Bhabha National Institute<sup>1</sup>

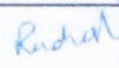
## Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Ms. Prerana Dash** entitled "**Structure-function-dynamics of PanPL, a bacterial polysaccharide lyase from PL-5 family**" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman - Name & Signature with date **Prof. A. Srinivasan**

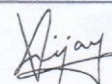
  
27.06.23

Guide / Convener - Name & Signature with date **Dr. Rudresh Acharya**

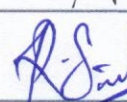
  
27.06.2023

Co-guide - Name & Signature with date (if any)

Examiner - Name & Signature with date **Prof. Vijay Thiruvengatam**

  
27.06.2023

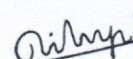
Member 1- Name & Signature with date **Dr. Ramanujam Srinivasan**

  
27/06/23

Member 2- Name & Signature with date **Dr. Tirumala Kumar Chowdary**

  
27.06.23

Member 3- Name & Signature with date **Dr. Dileep Vasudevan**

  
27/June/2023

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 27.06.2023

Place: Jatni

Signature

Co-guide (if any)

Signature

Guide

## **STATEMENT BY AUTHOR**

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.



**Prerana Dash**  
Name & Signature  
of the student

## **DECLARATION**

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

*Prerana Dash*

Prerana Dash  
Name & Signature  
of the student



## List of Publications arising from the thesis

### Publications:

1. # **P. Dash** and R. Acharya\*, Distinct Modes of Hidden Structural Dynamics in the Functioning of an Allosteric Polysaccharide Lyase, *ACS Cent. Sci.* **2022**, 8 (7), 933–947
2. # **P. Dash**, A. Mahapatra, R. Acharya\*, Structure-guided alteration of function and dynamics of PanPL, a polysaccharide lyase, (to be communicated)
3. # **P. Dash** and R. Acharya\*, Exploring polysaccharide lyases of PL-5 family through the lens of structure, function, and dynamics, *Acta Cryst.* **2021**, A77, C1121

#Pertaining to the thesis

### Conferences:

Poster presentations

International

1. **P. Dash** and R. Acharya\*, Exploring Polysaccharide lyases of PL-5 family through the lens of structure, function, and dynamics, 14-22 August 2021, XXV Congress & General Assembly of the International Union of Crystallography

National

1. **P. Dash**, Pranjal Mahanta, and R. Acharya\*, Structural and biochemical characterisation of bacterial polysaccharide lyases from PL5 family, 19-22 June 2019, 47<sup>th</sup> National seminar on Crystallography



Prerana Dash

Name & Signature

of the student

**Dedicated to**

**My mentor Dr. Rudresh Acharya.**

## ACKNOWLEDGEMENTS

I don't find words to express my gratitude towards my mentor, Dr. Rudresh Acharya. I remember, I joined my PhD program with zero work experience and lot of self-doubts. My mentor understood my background and assured me that I would get all the necessary training and support to carry out a well-defined doctoral work in his lab. His dedication to science, depth of knowledge, keenness to learn new things and ability to channelize pain into productivity during his toughest times without deviating from the path of truth inspires me. Professionally, he taught me how to ask scientific questions, design experiments and provided enough freedom and scopes to think independently on my PhD project. His timely guidance and wise advices made my PhD journey very smooth. I will always remember the days how my mentor relentlessly worked with me for my publications. Working with such a great scientist and good human being, helped me to develop patience, confidence and fearlessness to face any challenges in life. I consider myself among the luckiest persons who are blessed with great teachers (Guru) in their life. It won't be an exaggeration to admit that my mentor, Dr Rudresh Acharya is my role model whom I want to follow.

I would like to thank to all my past and present lab members for making the work environment enjoyable. I express my gratitude towards our collaborator Prof. Bryan Berger from University of Virginia for introducing the PL project to our lab. I am thankful to my senior Dr. Shubhant Pandey for his efforts in writing the grant on PL-5 enzymes with my mentor, which made my PhD work very well designed. I am grateful to Ramakrushna, Joydeep and Smruti for helping me in my initial days in my lab and course works. I will remain in debt to Venkatreddy Dadireddy from IISC who taught me all the molecular biology experiments and for collecting diffraction data at ESRF for my crystals. I had fun working with Pranjal, Nilima, Shreya Biswas, Lipsa, Sajith, Shanice, Vaishnavi, Vidhee, Ashwin, Gyanaranjan and Ranjita. I would love to thank Sandhya and Ankita for helping me in my project during their internships. My special thanks to Shreya Sharma, Avhra Biswas and Hariom for accompanying me in PL project. I should mention about the Saturday Lab meetings. The brain storming discussions and the critical scientific suggestions improved the basic understanding of structural biology and helped me in my doctoral work.

I would like to thank Dr Smitha Mohanlal, the women scientist and wife of Dr Rudresh Acharya; who always inspired me to work hard, taught me to maintain healthy work-life balance. I want to acknowledge Dr Nagendra Sharma's lab and Dr Sharnappa Nanbenna's Lab from SCS for providing help with chemicals generously. I should thank Dr. Himanshu Biswal, SCS for appreciating my work publicly to motivate NISER student community. I am thankful to Dr Dileep Vasudevan's lab from ILS especially to Ashish, Ajit, Manas and Ruchir for helping me whenever I needed. I am grateful to my doctoral committee members for their valuable suggestions.

I would like to thank my batchmates especially Aranya and Bratati. Aranya was always generous to give suggestions and consumables related to molecular biology. A million thanks to Laxmipriya, Ashutosh, Avraj, Soumya who were my emotional supports in NISER. I would like to thank Subhankar for motivating me all the times. I am thankful to Anup and Dhyanendra for collecting Mass spec data for my samples. I am grateful to Suryanarayan Sir who taught me Yoga and meditations during the pandemic time to maintain physical and mental health.

Finally, I am grateful to my family and relatives. My father, Mr Balaram Dash, is my strength and advised me to be sincere and disciplined throughout my PhD. He kept on the track of my every little progress and it was pleasurable to discuss my work with him. My mother, Mrs. Sarojini Devi, is a loving lady and advised me to take care of my physical, mental and emotional health during this strenuous life style of research. I am grateful to my brother Pratyush, who took care of parents behind me. Finally, I am grateful to Lord Jagannatha, the lord of the Universe to bless me with the best parents, the best mentor, beautiful friends and a magical PhD life.

# CONTENTS

<b>SUMMARY</b>	<b>VIII</b>
<b>LIST OF ABBREVIATIONS</b>	<b>X</b>
<b>LIST OF FIGURES</b>	<b>XI</b>
<b>LIST OF TABLES</b>	<b>XV</b>
<b>CHAPTER 1: Introduction</b>	<b>1</b>
<b>1.1. Enzymes, the biocatalysts</b>	<b>2</b>
<b>1.2. Enzyme structure: pre-organized active site and enzyme-ligand complex</b>	<b>3</b>
<b>1.3. Enzyme Dynamics</b>	<b>5</b>
<b>1.3.1. Methods to investigate dynamic properties</b>	<b>6</b>
<b>1.3.2. Dynamics and catalysis</b>	<b>7</b>
<b>1.4. Enzyme kinetics</b>	<b>8</b>
<b>1.4.1. Factors influencing enzyme catalysis</b>	<b>9</b>
<b>1.5. Allostery</b>	<b>10</b>
<b>1.6. Editing catalytic power by mutation</b>	<b>11</b>
<b>1.7. The Polysaccharide Lyase</b>	<b>12</b>
<b>1.8. The Anionic Polysaccharides</b>	<b>14</b>
<b>1.9. Significance of Polysaccharide lyase</b>	<b>16</b>
<b>1.10. The Beta elimination reaction mechanism</b>	<b>17</b>
<b>1.11. PanPL</b>	<b>19</b>
<b>1.12. Conclusion</b>	<b>21</b>



<b>CHAPTER 2: Materials and Methods</b>	<b>23</b>
<b>2.1. Introduction</b>	<b>24</b>
<b>2.2. Cloning and overexpression of PanPL</b>	<b>24</b>
<b>2.3. Extracellular secretion</b>	<b>25</b>
<b>2.4. Purification of PanPL</b>	<b>26</b>
<b>2.5. Thiobarbituric acid (TBA) assay</b>	<b>28</b>
<b>2.6. Enzyme kinetics</b>	<b>28</b>
<b>2.7. Product Analysis</b>	<b>29</b>
<b>2.8. Mutant protein production</b>	<b>30</b>
<b>2.9. Crystallization</b>	<b>33</b>
<b>2.9.1. Co-crystallization with substrates</b>	<b>34</b>
<b>2.10. X-ray Diffraction Data Collection and Structure Determination</b>	<b>37</b>
<b>2.11. Electrostatic Surface charge analysis</b>	<b>39</b>
<b>2.12. Rosetta flexible ligand docking</b>	<b>40</b>
<b>2.13. Ensemble refinement</b>	<b>40</b>
<b>2.14. Homology Modeling</b>	<b>41</b>
<b>2.15. Molecular Dynamics Simulations</b>	<b>42</b>
<b>2.16. Principal component analysis of MD trajectories</b>	<b>43</b>
<b>2.17. MMPBSA calculation</b>	<b>44</b>
<b>2.18. Electrostatic interaction network modelling</b>	<b>44</b>
<b>CHAPTER 3: PanPL, unique allosteric PL-5 enzyme</b>	<b>45</b>

<b>3.1. Introduction</b>	<b>46</b>
<b>3.2. Biochemical characterization reveals the allosteric nature of PanPL</b>	<b>46</b>
3.2.1. PanPL works optimally around neutral pH	46
3.2.2. PanPL activity fits non-linear kinetics with positive cooperativity	47
3.3. PanPL is endolytic lyase	50
<b>3.4. Structural characterization of PanPL</b>	<b>51</b>
3.4.1. The overall structure of PanPL forms a catalytic tunnel	53
3.4.2. Tunnel architecture	55
3.4.3. Location of Active site residues	56
3.4.4. Enzyme Substrate interaction: PanPL-tetra ManA complex crystal structure	59
3.4.5. Docking of PanPL with tetra-ManA	64
3.4.6. The role of active site residues in enzyme-substrate interaction	67
3.4.7. Crystal Packing Analysis	71
<b>3.5. The electrostatic surface charge distribution as a function of pH</b>	<b>76</b>
<b>3.6. Structural flexibility and the B-factor analyses as a function of pH</b>	<b>78</b>
<b>3.7. The Normal Mode Analysis</b>	<b>84</b>
<b>3.8. Ensemble refinement displaying the alternative conformations in crystal structures</b>	<b>85</b>
<b>3.9. The hidden open state of PanPL in Molecular Dynamics simulation</b>	<b>91</b>
3.9.1. The N-loop-lid dynamics and the H-bond lock	93
3.9.2. PCA clustering of MD trajectories	95
<b>3.10. Overall dynamics of PanPL</b>	<b>97</b>
<b>3.11. Allostery as a result of conformational flexibility and structural dynamics</b>	<b>99</b>
<b>3.12. Significance of N-loop-lid length in selecting open/close or only close state</b>	<b>100</b>
<b>3.13. Conclusion</b>	<b>105</b>

<b>CHAPTER 4: Structure-guided mutations that altered the dynamics and enzyme kinetics</b>	<b>109</b>
<b>4.1. Introduction</b>	<b>110</b>
<b>4.2. What changes by mutation?</b>	<b>110</b>
<b>4.3. Identification of flanking residues in the tunnel</b>	<b>111</b>
<b>4.3.1. Mutation at the entry site of tunnel: L222R</b>	<b>112</b>
<b>4.3.2. Mutation at the exit site of the tunnel: R48G</b>	<b>112</b>
<b>4.3.3. Mutation at the entry and exit site: L222R and R48G</b>	<b>113</b>
<b>4.4. Biochemical characterization of mutants</b>	<b>115</b>
<b>4.4.1. Biochemistry of L222R</b>	<b>116</b>
<b>4.4.2. Biochemistry of R48G</b>	<b>116</b>
<b>4.4.3. Biochemistry of L222R R48G</b>	<b>116</b>
<b>4.5. Crystal structures of PanPL mutants</b>	<b>120</b>
<b>4.6. Docking of tetra-ManA with PanPL mutants</b>	<b>123</b>
<b>4.7. Binding energy calculation for PanPL L222R mutant (MM-PBSA)</b>	<b>124</b>
<b>4.8. Dynamics of PanPL L222R</b>	<b>125</b>
<b>4.9. Electrostatic network</b>	<b>127</b>
<b>4.10. Electrostatic network of R48 residue</b>	<b>128</b>
<b>4.11. Dynamics of PanPL R48G mutant</b>	<b>129</b>
<b>4.12. PanPL R48G shifts the pKa of Y226</b>	<b>129</b>
<b>4.13. Dynamics of L222R R48G mutant</b>	<b>132</b>
<b>4.14. Discussion</b>	<b>134</b>
<b>4.14.1. Mutation modulates the biochemistry</b>	<b>135</b>
<b>4.14.2. Mutation alters the dynamics</b>	<b>136</b>

<b>4.15. Conclusion</b>	<b>138</b>
<b>CHAPTER 5: Future Directions</b>	<b>140</b>
<b>5.1. Introduction</b>	<b>141</b>
<b>5.2. PanPL as a model enzyme for research enthusiasts.</b>	<b>141</b>
<b>5.3. PanPL as a model enzyme for enzyme engineering</b>	<b>141</b>
<b>5.4. Medical application of PanPL</b>	<b>143</b>
<b>APPENDIX</b>	<b>145</b>
<b>REFERENCES</b>	<b>147</b>

## SUMMARY

Enzymes are biocatalysts that enhance the rate of chemical reactions, which are otherwise difficult to achieve ordinarily. The amino acid sequence of an enzyme encodes its three-dimensional structure as well as its intrinsic dynamics. In my Ph.D. thesis work, I have studied the structure-function-dynamics of PanPL, a bacterial polysaccharide lyase from the PL-5 family.

Polysaccharide Lyase (PL) catalyzes the depolymerization of anionic polysaccharides via a beta-elimination reaction mechanism. The present work was inspired by the unique biochemical property of a PL-5 family enzyme, the Smlt1473, that showed a pH-regulated multi-substrate specificity. We hypothesized that other PL-5 enzymes with similar amino acid distribution patterns around the catalytic site, like in Smlt1473, might possess a pH-regulated substrate specificity. We selected eight different PL-5 enzymes initially to investigate our hypothesis. After a rigorous process of cloning and purification of eight proteins, we stumbled upon the PanPL from *Pandoraea apista* that was found amenable for biochemical assay and crystallization.

The biochemical characterization suggests PanPL be alginate specific with an endolytic mode of action. We observed a unique trend of positive cooperativity in enzyme kinetic studies. To provide the structural basis for the functioning, we have determined the crystal structures of PanPL in apo forms across the pH spectrum and substrate-bound forms. The PanPL folds into a characteristic  $(\alpha/\alpha)_5$  fold forming a closed tunnel with an N-terminal loop lid (N-loop-lid) arched over the active site in all of its crystal structures. Our study suggests that pH modulates



the electrostatic surface charge distribution and N-loop-lid fluctuations to govern the enzyme's functionality. Our B-factor analysis shows huge fluctuations at N-loop-lid in co-crystallized and substrate-bound crystal structures. This result inspired us to perform a Normal mode analysis that supports the functional dynamics of N-loop-lid.

Nevertheless, our 1 $\mu$ s molecular dynamics (MD) simulations on both apo and substrate-bound crystal structures of PanPL revealed the presence of a hidden but transiently feasible open state in the solution phase. We performed an ensemble refinement (ER) to display the hidden conformational substates in the static crystal structures. The ER analyses on the apo structures revealed a large number of conformational substates for N-loop-lid as a function of pH.

Further, ER could model a huge backbone and side chain fluctuations at N-loop-lid in co-crystallized and substrate-bound structures without dismantling the closed state. Thus, ER reveals that the substrate enhances the N-loop-lid fluctuations when maintaining the close state. While the loop flexibility and transiently accessible open state are intrinsic to PanPL. In addition, our work on PL-5 enzyme structures suggests that the insertion near the N-loop-lid stabilizes the low-energy closed state causing a shift in paradigm for the mechanism: to adopt the closed state with functional fluctuations at N-loop-lid over an open-close state transition for functioning. This work demonstrates the role of dynamics; both intrinsic and substrate-induced conformational dynamics are vital for enzyme functioning and allostery.

The study further motivated us to decode the role of the flanking residues around the catalytic sites. We introduced structure-guided mutations around the active site to modify the inter-residue electrostatic networks, resulting in altered dynamics and catalytic activity. Overall, our study has provided the visualization of the enzyme's functionality through the lens of structural dynamics, and will guide the field of enzyme engineering to make versatile and efficient enzymes.

## LIST OF ABBREVIATIONS

<b>PL-5</b>	<b>P</b> olysaccharide <b>L</b> yases family <b>5</b>
<b>ManA</b>	<b>M</b> annuronic <b>A</b> cid
<b>GlcA</b>	<b>G</b> lucuronic <b>A</b> cid
<b>HA</b>	<b>H</b> yaluronic <b>A</b> cid
<b>ER</b>	<b>E</b> nsemble <b>R</b> efinement
<b>MD</b>	<b>M</b> olecular <b>D</b> ynamics
<b>r. m. s. d.</b>	root <b>m</b> ean square <b>d</b> eviation
<b>r. m. s. f.</b>	root <b>m</b> ean square <b>f</b> luctuation
<b>NMR</b>	<b>N</b> uclear <b>m</b> agnetic <b>r</b> esonance
<b>DHFR</b>	<b>D</b> ihydrofolate <b>r</b> eductase
<b>IPTG</b>	<b>I</b> sopropyl $\beta$ -D-1- <b>t</b> hiogalactopyranoside
<b>LB</b>	<b>L</b> uria <b>B</b> ertani
<b>MMPBSA</b>	<b>M</b> olecular <b>m</b> echanics <b>P</b> oisson– <b>B</b> oltzmann surface <b>a</b> rea

# LIST OF FIGURES

## CHAPTER 1

- Figure 1.1. Schematic representation of energetics of an enzyme catalyzed reaction vs uncatalyzed reaction
- Figure 1.2. Schematic representation of an enzyme kinetic plot
- Figure 1. 3. Schematic representation of positive cooperativity of an allosteric enzyme.
- Figure 1. 4. The glycosidic bond cleavage reaction
- Figure 1. 5. The fold diversity among PL families
- Figure 1. 6. The structure of anionic polysaccharides
- Figure 1.7. Schematic representation of  $\beta$  elimination reaction
- Figure 1.8. Schematic representation of *syn* and *anti*  $\beta$  elimination reaction

## CHAPTER 2

- Figure 2.1 workflow for the objective-I.
- Figure 2.2. The over-expression of PanPL.
- Figure 2.3. The extracellular lyase activity of PanPL
- Figure 2.4. Purification of PanPL
- Figure 2.5. workflow for the objective 2, 3, and 4
- Figure 2.6. Crystal images of PanPL across the pH spectrum
- Figure 2.7. The crystal images for the mutants of PanPL
- Figure 2.8. The overall workflow for structure determination of PanPL

## CHAPTER 3

- Figure 3.1. TBA assay for PanPL across the pH spectrum
- Figure 3.2. The optimum pH and temperature of PanPL
- Figure 3.3. Enzyme kinetic assay for PanPL
- Figure 3.4. Product analysis by anion exchange chromatography
- Figure 3.5. Mass Spectrometric analysis of cleaved products
- Figure 3.6. Primary structure, topology, and overall structure of PanPL
- Figure 3.7. Tunnel architecture: entry site, constriction region, exit site
- Figure 3.8. Amino acid distribution and active site location in the tunnel
- Figure 3.9. The conformations of active site residues and R48 across the pH
- Figure 3.10. The substrate in the tunnel view
- Figure 3.11. Enzyme-Substrate interactions with tunnel residues PanPL H172A and tetra ManA complex Crystal structure
- Figure 3.12. The mosaic floor formed by aromatic residues
- Figure 3.13. Polder map for the ligand
- Figure 3.14. Docking of PanPL and tetra-ManA
- Figure 3.15. The enzyme substrate interactions in docked structure
- Figure 3.16. The active site residues of PanPL
- Figure 3.17. Enzyme-substrate interactions at active sites in PL-5 proteins
- Figure 3.18. Schematic representation of the geometry of a hydrogen bond
- Figure 3.19. Electrostatic surface charge visualization as a function of pH

- Figure 3.20. Average B factor and B factor putty representation of PanPL across the pH spectrum
- Figure 3.21. Normalized B-factor fluctuations across the pH spectrum for PanPL
- Figure 3.22. The enzyme activity, the electrostatic surface charge and the B factor putty representation of PanPL as a function of pH
- Figure 3.23. B factor putty representation and the Normalised B factor plots for mutant structures
- Figure 3.24. Normal mode analysis of PanPL at all pH and substrate bound structure
- Figure 3.25. Ensemble refinement of PanPL crystal structures across pH spectrum
- Figure 3.26. Ensemble refinement of PanPL inactive mutant crystal structures in bound and apo forms
- Figure 3.27. The quantitative analysis of N-loop-lid fluctuation
- Figure 3.28. Analysis of the Molecular dynamics simulation trajectories of apo and tetra-ManA bound crystal structures
- Figure 3.29. The dynamics of H- bond lock the N-loop-lid
- Figure 3.30. Principal component analysis on Molecular dynamics simulation trajectories (100ns) of apo and tetra-ManA bound structures of PanPL
- Figure 3.31. Dynamic behavior of N-loop-lid of PanPL.
- Figure 3.32. The open-close state transition in *Sphingomonas sp.* Alginate lyase AIII
- Figure 3.33. Sequence Structure comparison PL-5 family enzymes
- Figure 3.34. shows the structural analysis of RpPL
- Figure 3.35. shows the structural analysis of BcPL



- Figure 3.36. The TBA assay for BcPL and RpPL
- Figure 3.37. Summary of PanPL dynamics.

## CHAPTER 4

- Figure 4.1 Multiple sequence analysis among PL5 family.
- Figure 4. 2. The surface representation of PanPL L222R mutation
- Figure 4.3. Hydrogen bonds between the conserved Arg and the N-loop-lid of PL-5 enzymes.
- Figure 4. 4. The Arginine at the exit site of PL-5 enzymes
- Figure 4.5. Enzyme kinetic plots for PanPL mutants.
- Figure 4.6. The specific activities of PanPL WT with mutants for different substrates.
- Figure 4.7. The crystal structure of PanPL mutants.
- Figure 4.8. Interaction between PanPL L222R mutant with tetra manA in stereo view
- Figure 4.9. Binding energy decomposition analysis
- Figure 4.10. The dynamics of PanPL L222R mutation
- Figure 4.11. The coulombic interaction network for R222 residue
- Figure 4.12. Electrostatic interaction network of R48
- Figure 4.13. The r. m. s. d. of the PanPL R48G mutant
- Figure 4.14. The electrostatic interaction network of Y226
- Figure 4.15. The r. m. s. d. of the PanPL L222R R48G

# LIST OF TABLES

## CHAPTER 1

- Table 1.1. The PL-5 enzymes under the Smlt1473-UNLIKE group
- Table 1.2. The PL-5 enzymes under the Smlt1473-LIKE group

## CHAPTER 2

- Table 2.1. The list of primer sequence used to create point mutation
- Table 2.2. Reagents added in the PCR mix
- Table 2.3. The thermocycler protocols
- Table 2.4. Reagents added for phosphorylation and ligation
- Table 2.5. The crystallization conditions for PanPL and its mutants

## CHAPTER 3

- Table 3.1. The enzyme kinetic parameters for PanPL for its substrates
- Table 3.2. Crystallographic data collection and refinement statistics of PanPL apo structures across the pH spectrum
- Table 3.3. The list of interactions found in enzyme-substrate (PanPL and tetra-ManA) complex crystal structure
- Table 3.4. Crystallographic data collection and refinement statistics of PanPL active site mutants in apo and substrate bound forms
- Table 3.5. The list of interactions seen in PanPL and tetra-ManA docked structure
- Table. 3.6. The active site residues and their putative role in catalysis

- Table 3.7. Hydrogen bonds observed at N-loop-lid (42-52: YYTDAARSVID) between the symmetry related molecules in crystal packing
- Table 3.8. Solvent interactions with polar atoms of N-loop lid (42-52: YYTDAARSVID)
  - 3.8.1. PanPL\_pH3.5: 7WXJ
  - 3.8.2. PanPL\_pH4.5: 7WXX
  - 3.8.3. PanPL\_pH5.5: 7WXL
  - 3.8.4. PanPL\_pH6.5: 7WXM
  - 3.8.5. PanPL\_H172A+tetra-ManA: 7WXP
  - 3.8.6. PanPL\_H17A: 7WXQ
  - 3.8.7. PanPL\_N171L\_apo: 7XTE
  - 3.8.8. PanPL\_Y226F\_apo: 7XTF
- Table 3.9. ER statistics of crystal structures

## CHAPTER 4

- Table 4.1. The Enzyme kinetic parameters for PanPL mutants
- Table 4.2. Crystallographic data collection and refinement statistics for mutants
- Table 4.3. PanPL L222R and tetra ManA docked structure interactions
- Table 4.4. list of residues interacting with R222 by coulombic interaction
- Table 4.5. list of residues interacting with R48 by coulombic interaction
- Table 4.6. list of residues interacting with Tyr 226 by coulombic interaction
- Table 4.7. The list of newly formed and lost interactions due to mutation in PanPL L222R R48G structure
- Table 4.8. Summary of the PanPL mutants



# **CHAPTER 1: Introduction**

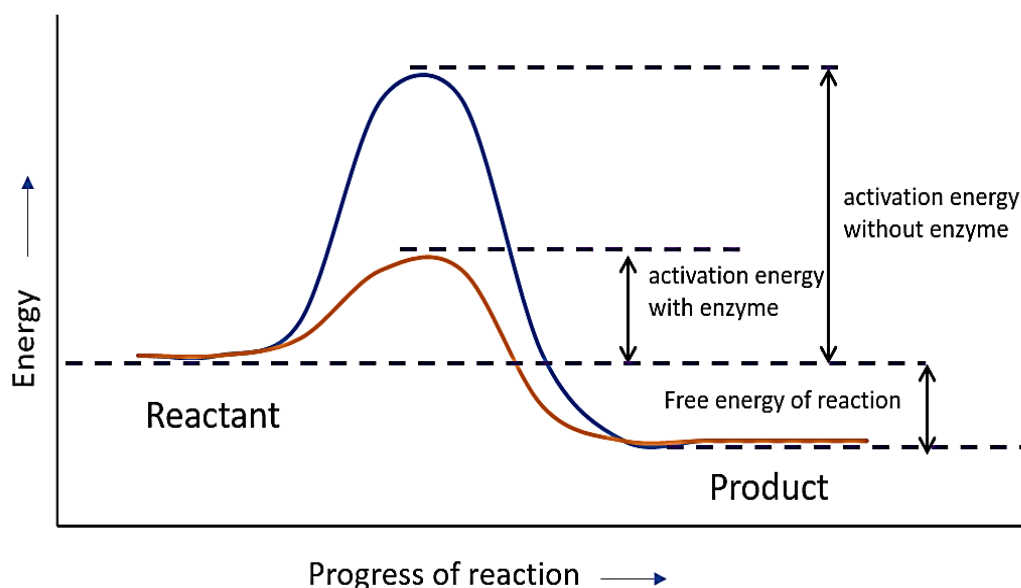
## **1.1. Enzymes, the biocatalysts**

Chemical reactions are the driving force to create new compounds, and can be assisted by catalysts to accelerate the process. Enzymes are the biocatalysts with unbeatable rate enhancing capacities. Nature uses enzymes as work horses to achieve difficult chemical reactions in a simplistic and energy efficient manner. Enzymes are usually proteins except some ribozymes, which are RNA molecules. The primary structure of the enzyme provides the amino acid sequence. The protein molecules get folded into various folds composed of secondary structure elements. The sequence determines the structure of the protein. The enzymes create unique shapes or binding pockets for their substrates known as active site or catalytic site. The enzyme interacts with its substrates, stabilizes the transition state and assists the chemical reaction by lowering the activation energy to make products without being consumed in the process [Figure 1.1]. Based on the functions they are associated; the enzymes are categorized into different classes such as oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases and translocases (1). The world of biological chemistry has been dominated by enzymes. Almost all the biological processes are catalyzed by enzymes efficiently and flawlessly. The precision of the enzyme catalysis is commendable and the speed is dramatic. Life without enzymes is almost unimaginable. Here, evolution plays a significant role in creating these molecular machines with diverse functionalities. The continuous evolutionary selection pressure keeps on adding new enzymes by editing the old ones.

The diversity in the enzyme world fascinates researchers to understand the structure, function and dynamics of these tiny molecules in laboratories. Enzymes are prevalent in all the strata of life. From viruses to higher eukaryotes, all have their own enzymes encoded in their genetic material. The medical industries target the enzymes, which are crucial for the survival of pathogens during their life cycle. Inhibitors are designed to stop their



functions to serve the mankind. The fundamental function of enzymes as catalysts are further extended to carry out several non-biological chemistries. Inspired by nature, the enzymes are now utilized in chemical industries for their enormous catalytic properties to carry out challenging chemical reactions. In addition to this, they also come under the green and clean energy without disposing any waste to nature.



**Figure 1.1. Schematic representation of energetics of an enzyme catalyzed reaction vs uncatalyzed reaction,** Enzyme lowers the activation energy of the chemical reaction, in return speed up the reaction.

## **1.2. Enzyme structure: pre-organized active site and enzyme-ligand complex**

Enzymes are known for their precision in accommodating their specific substrates in the catalytic pockets based on the shape complementarity (2). Some of the enzymes undergo local conformational changes around the binding site to accommodate their substrates (3). This supports the lock and key model and the induced fit model for enzyme substrate

interaction(2). The spatial arrangement of the amino acid side chains in the catalytic sites provides the microenvironment to facilitate the chemical reaction. To appreciate the visualization of amino acid arrangements in the catalytic sites, one needs the crystal structures of the enzymes. Several studies have pointed out that the geometry of the active site residues is preorganized in the catalytic pocket(4). Further, the catalytic geometry has been found to be conserved across a specific enzyme family (5). Due to the shape complementarity property some of the shape mimicking alternative substrates also can bind to the pocket in order to alter the chemistry. Inhibitors are such molecules which are found/designed to bind at the active site and inactivate the enzyme. Based on this knowledge, the bioinformatics studies help to predict the putative ligands and their binding sites in protein sequences and structures.

The substrate bound crystal structures depict the visualization of substrates at the catalytic site. The proper orientation and positioning of substrates in the catalytic pocket is crucial for the catalysis. The preorganized active site residues facilitated the stabilization of the transition state for the chemical reaction. This lowers the activation energy resulting in enhancement of the rate of the reaction. However, it should be noted that the substrate bound crystal structure represents the pre-transition state. The substrate is held in the catalytic pocket by several interactions like hydrogen bonds (polar), hydrophobic (non-polar), van der Waals or electrostatic interactions based on the nature of the substrate and its amino acid interacting counterparts. The interactions of the substrate with the active site residues are observed to be evolutionary conserved across the enzyme family (5). Therefore, the active site mutants of the enzymes are designed intending to change the side chain functional groups which disrupt the crucial enzyme-substrate interactions. This results in loss or reduction or increasing/enhancing in enzyme activity. As the native structures are maintained in crystal lattice, the enzymes retain their catalytic activity in the

crystals. The substrate specificity and the conserved enzyme-substrate interactions enable the enzymes to perform catalysis even in the crystal lattice to form products. However, it is difficult to capture and visualize such rapid process of catalysis by X-ray crystallography. Hence, it is a general practice to trap the substrate in the catalytic site of the enzyme in its inactive form to visualize the pre-transition state; the enzyme-substrate interactions. To attain the inactive forms of the enzymes, active site mutants are generated or the conditions that favor the inactive state of the intact-enzyme are maintained. Thus, we can perturb the external factors like pH, salt concentrations of the solution that influences the enzyme activity and crystallize the enzyme in that condition. The substrate bound crystal structures helps to create a list of interaction counterparts and interaction types in the enzyme-substrate complex. In absence of crystal structures, the docking studies can provide the list of the predicted interactions between the enzyme and the substrate.

### **1.3. Enzyme Dynamics**

Enzymes are inherently fluctuating molecules. They are not static or rigid in the solution. The rhythm of vibrations or the correlated motions creates a breathing pattern for the enzymes. The enzymes are tiny, fluffy molecular machines with distinct structural flexibilities. The flexible parts of the enzymes show different mechanisms of dynamics based on their sequence, structures and the process of catalysis. The fate of the enzyme dynamics is encoded in their amino acid sequences and largely influenced by external perturbations. The enzyme dynamics epitomizes diverse molecular motions ranging from amino acid side chain conformational change to domain movement within protein structures. A few large multi domain oligomeric proteins show sequential dynamic mechanism involving domain movements to carry out their functions(6). However, small monomeric enzymes with a single domain show various mode of dynamics involving loop motions, open to close state transition, and side chain conformational change etc. The

timescale of these dynamics ranges from femtosecond to millisecond(7) The dynamics of the proteins set them into different energy states in the energy landscapes. The discrete energy difference allows the enzymes to access multiple inter-convertible alternative states in their crystal structures. However, the narrow energy difference between the states shows no noticeable structural changes in the crystal structures. The dynamics of the enzymes generate the conformational heterogeneity in the solution and maintains an equilibrium among the heterogenous population. The external factors can cause the shift in the equilibrium of the population distribution by tuning the enzyme dynamics.

### **1.3.1. Methods to investigate dynamic properties**

There are different methods to investigate the dynamic behavior of the enzymes. The global motions are detected by methods like FRET (fluorescent resonance energy transfer)(8) and AFM (Atomic force microscopy)(9,10) etc. The most popular method to investigate internal motions at atomic level in solution is the Nuclear magnetic resonance (NMR) spectroscopy(11,12). These methods provide the information about the time scale of the internal dynamics experimentally.

The crystal structure provides the average electron density which is like a static snapshot. However, the crystal structure carries the footprint of the protein dynamics in the diffraction pattern. The flexible part of the protein reflected as disordered regions in the crystal structure. The B-factor recorded during the diffraction data collection provides an idea about the structural flexibility of the enzymes(13). The B-factor has its limitations as it gets affected by the resolution of the data and the crystal packing. Recently developed Ensemble Refinement (ER) uses the time-averaged refinement, and molecular dynamic simulations for sampling local molecular motions to generate an ensemble of structures using X-ray diffraction data(14). ER is considered to be a powerful tool which reveals the alternative conformations hidden in the crystal structures. A few case studies of ER

analysis on the crystal structures shows agreement to the solution NMR results (14). The normal mode analysis on the crystal structures also provides the information about functionally relevant motions(15).

The molecular dynamics simulation is considered as an extremely useful computational tool to understand the dynamic behaviour of the enzyme (16). This computational study complementary to other experimental methods can sample a wide range of conformational states for the protein in the energy landscape. Unlike other experimental methods, MD simulation can access different transiently accessible hidden conformers along its trajectory. Thus, MD simulation has become an irreplaceable method in enzyme dynamics, and provide essential dynamics (17).

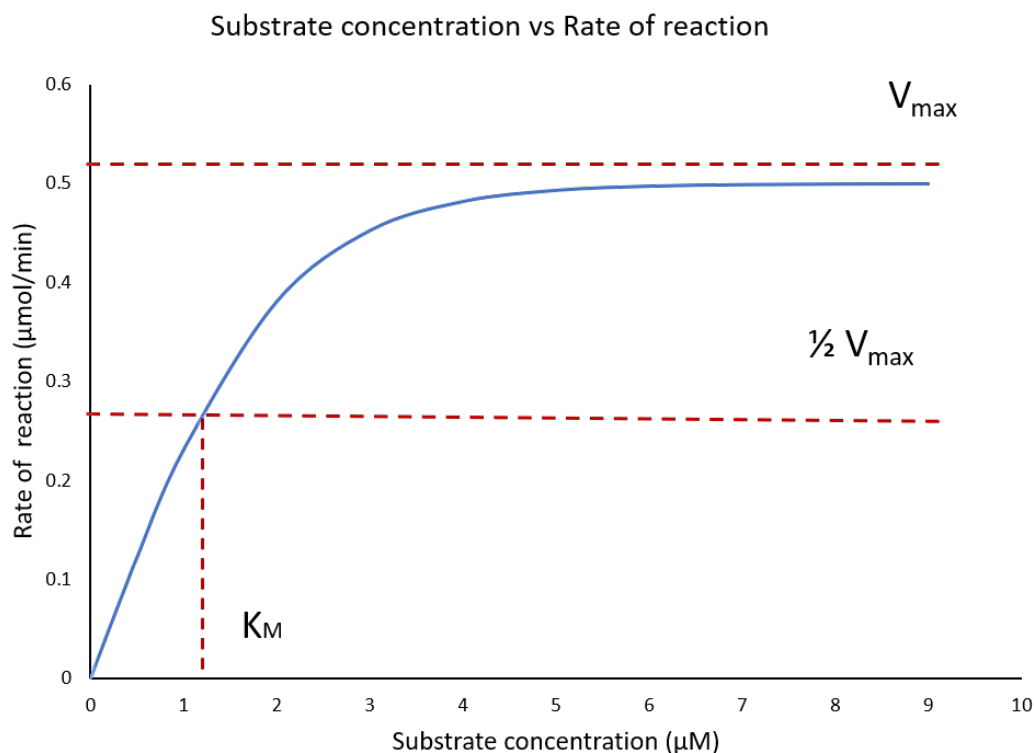
### **1.3.2. Dynamics and catalysis**

The dynamics property and the catalytic ability are intrinsic to enzymes(18). However, the direct correlation between the enzyme dynamics and catalysis has been poorly established. The enzymes along with their preorganized active site are granted a little amount of structural flexibility and are expected to undergo dynamics to aid the substrate acquisition and processing step. An NMR study demonstrates that the enzyme cyclophilin A shows conformational fluctuation at the active site during catalysis, and the timescale of the dynamics coincide with the substrate turnover(19). Another study on DHFR dihydrofolate reductase shows a dynamic knockout mutant N23pp/S148A ecDHFR diminish the enzyme activity which explain the importance of dynamics for the enzyme function(20). The vibrational spectroscopy study on the Lactate dehydrogenase- pyruvate complex demonstrates the important dynamics of the enzyme- substrate (Michaelis) complex (18). The mode and the timescale of the enzyme dynamics can vary during different steps of the catalysis i.e., substrate binding, rearrangement of enzyme-substrate complex, product formation and release. However, the significance of enzyme dynamics in catalysis is

largely context dependent and need extensive study to delineate the correlation. The dynamics might enhance, impede or might be ineffective on the enzyme catalysis. Several studies on the evolutionary aspect of enzyme dynamics have been performed which suggest that evolution pressure might play a role in alteration in structural dynamics (21,22). A few evolutionary significant substitution, insertion or deletion in protein sequence has been observed to alter the mechanism of structural dynamics(22,23). The studies on *E.coli* Dihydrofolate reductase (ecDHFR) and human Dihydrofolate reductase (hDHFR) shows that a single residue insertion in hDHFR at Met20 loop alters the dynamic mechanism considerably(23).

#### **1.4. Enzyme kinetics**

Enzymes as catalysts enhance the rate of the chemical reactions. Enzyme kinetics involves the study of rate of the reactions in variable conditions. The factors controlling the rate of the enzyme catalysis are pH, temperature, salt concentration and the substrate concentrations. In order to perform enzyme kinetics, the optimum conditions for the catalysis should be characterized. Generally, the rates of the reactions as a function of substrate concentration are plotted. This popular plot provides the information about the binding affinity of substrates, maximum rate of the enzyme catalyzed reactions, and catalytic efficiency etc. The plot showing a hyperbolic curve [Figure 1.2] follows the classical Michaelis Menten equation, whereas the sigmoid curve reflects allosteric behavior. The  $V_{\max}$  is the maximum velocity achieved. The Michaelis constant  $K_M$ , represents the substrate concentration at which the half of the maximum rate is achieved. The lower values of  $K_M$  signify better binding affinity. The turnover number  $K_{\text{cat}}$ , signifies the number of substrate molecules transformed into products. The ratio  $K_{\text{cat}}/K_M$  signifies the catalytic efficiency of the enzyme.



**Figure 1.2. Schematic representation of an enzyme kinetic plot**, it is a hyperbolic curve following typical Michaelis Menten behavior.

#### 1.4.1. Factors influencing enzyme catalysis

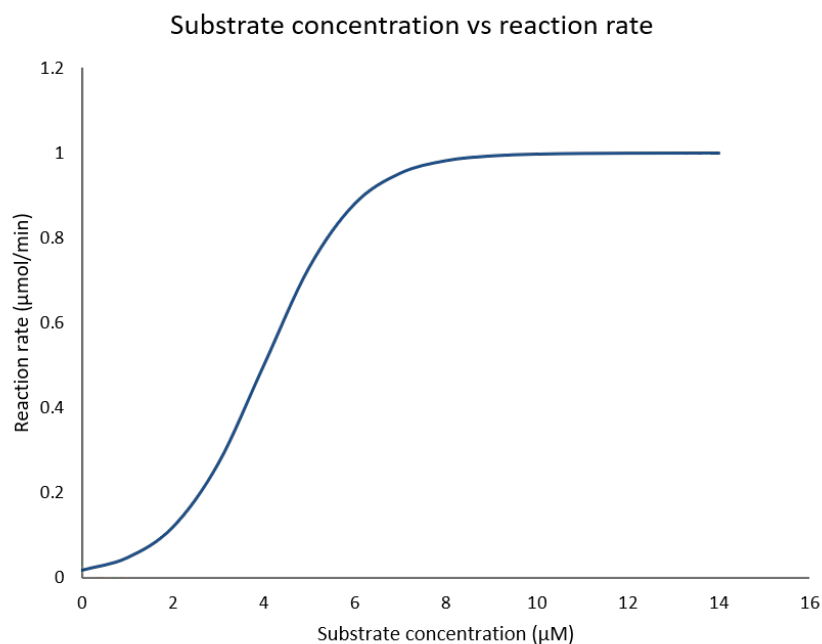
The catalytic power of the enzymes can be influenced by several external factors. The temperature of the chemical reaction regulates the thermal vibration of the molecules in the reaction mixture. At the optimum temperature of the reaction, the catalysis attains its maximum, beyond that the enzyme gets inactivated. The enzymes are called as psychrophilic (low temperature  $\sim -20$  to  $20^\circ\text{C}$ ), mesophilic (medium temperature  $\sim 20$  to  $45^\circ\text{C}$ ) and thermophilic (high temperature  $\sim 41$  to  $122^\circ\text{C}$ ) based on their optimum temperature of functioning and the source organisms (24,25). The factor pH sets the residues in the protein at different ionization states and regulates the surface charge density of the enzyme. The enzyme catalysis occurs within a pH range and reaches maximum activity at the optimum pH. The pH of the solution also influences the enzyme dynamics,

which has been demonstrated in this thesis work. The salt concentration of the solution controls the enzyme solubility in the reaction mixture and might influence the catalysis.

### **1.5. Allostery**

Allostery is a specific behavior observed in enzyme kinetics(26). Allosteric enzymes don't follow the classical Michaelis Menten kinetics. The kinetic plot shows a sigmoidal curve reflecting either a positive or negative cooperativity [Figure 1.3]. The allosteric enzymes show a time lag to achieve the maximum velocity at initial substrate concentrations. There are different allosteric enzymes functions in different ways. The general notion about an allosteric enzyme is that there should be multiple binding site and an effector molecule. The binding of effector molecule at the allosteric site further regulates the rate of reaction by enhancing or diminishing it. However, there are homotropic allosteric enzymes with a single binding site where substrate acts as an effector or allosteric regulator. Allosteric enzymes are known to undergo conformational change during catalysis. The conformational change might vary in time scale (ms-ns), i.e., loop opening/closing to side chain conformational change(27). A few enzymes undergo subtle fluctuations which are structurally unnoticeable upon substrate binding. The allosteric enzymes are suggested to have ensemble of structures with a large number of hidden conformational substates. The substrate acts as a perturbation and cause a population shift skewing the equilibration among the ensemble structures which causes allostery during catalysis (27). The time lag corresponds to the conformational rearrangements that the enzyme has to undergo for achieving saturation and maximum velocity.





**Figure 1.3. Schematic representation of positive cooperativity of an allosteric enzyme.**

The sigmoidal curve signifies that the enzyme kinetic doesn't follow the typical Michaelis-Menten behavior, rather it reflects allostery.

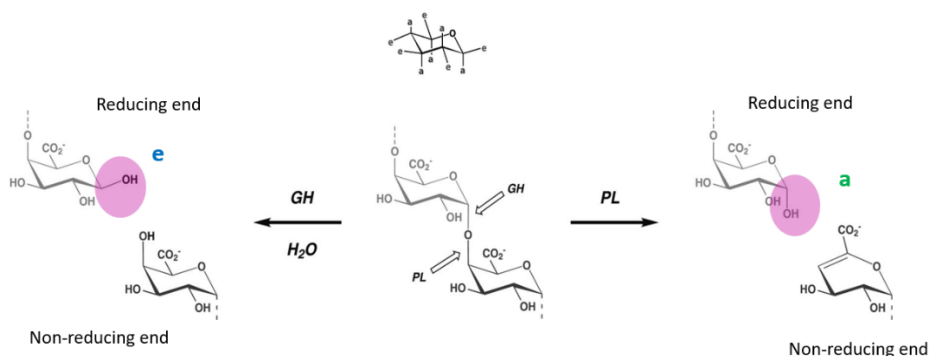
### 1.6. Editing catalytic power by mutation

Enzymes are protein molecules, sequentially placed amino acid entangled to make a molecular machine. Each amino acid is encoded by its specific nucleotide triplet. By editing the nucleotide sequence, we can edit the amino acid sequence in the protein. The active site residues along with the flanking residues near the catalytic pocket hold the catalytic power of the enzyme. The catalytic power of the enzyme can be modified by editing these amino acid sequence by mutation i.e., substitution, insertion and deletion. Mutations change the microenvironment of the catalytic region by disrupting inter-residue interaction network and or stabilizing or destabilizing effect on substrate binding. It can alter the charge, hydrophobicity as well as the dynamics of the enzyme molecule. We have discussed previously how insertion of single amino acid in a Met 20 loop changes the mechanism of dynamics for hDHFR (23). However, nature uses mutation as a tool to create

new enzymes with new functions during the process of evolution. It keeps on editing the nucleotide sequence, selects productive alteration and discards unfit alterations based on the demand of the environment. Inspired by nature, a new laboratory method has been emerged to mimic the natural evolution process. This method is called directed evolution. The background idea of the directed evolution is to use the promiscuous behavior of the enzyme, modify it by incorporating random mutations, then train it in a new environment by creating a selection pressure and check the fitness of the enzyme to take up new substrates from environment to perform new functions. Many eminent research groups have trained enzymes to catalyze new abiological reactions, by directed evolution (28–30).

### **1.7. The Polysaccharide Lyase**

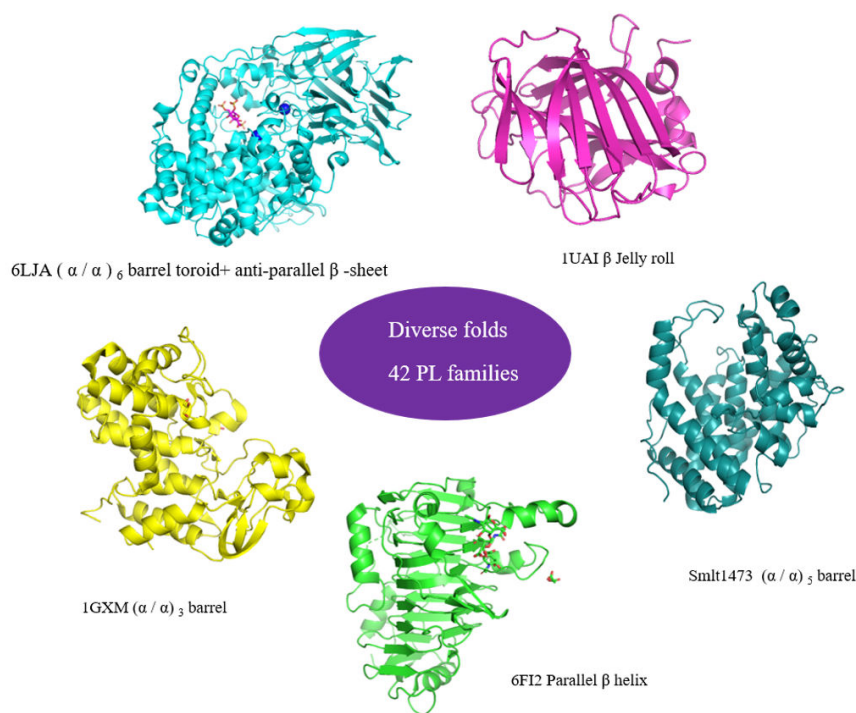
This thesis work is dedicated to delineate the structure-function-dynamics of a novel enzyme PanPL, a polysaccharide lyase from PL-5 family. PanPL is a polysaccharide lyase from *Pandoraea apista*. The polysaccharide lyases (PLs) are the enzymes which catalyze the non-hydrolytic cleavage of negatively charged polysaccharides [Figure 1.4] (31). These enzymes are prevalent in all kingdoms of life. These carbohydrate-active enzymes (CAZymes; [www.https://cazy.org](https://cazy.org)), are categorized into 42 families based on the sequence similarities and the structural folds [Figure 1.5]. The PLs shows a large fold diversity suggesting that PLs have been invented multiple times during course of evolution with a diverse structural scaffold. The PLs utilizes a beta-elimination reaction mechanism for the depolymerization of the anionic polysaccharides.



**Figure 1.4. The glycosidic bond cleavage reaction;** Glycoside Hydrolase (GH) and PL both catalyse the glycosidic bond cleavage. The GH and PL used here to degrade the pectin make different products with varying stereochemistry at the reducing ends. GHs cleave the glycosidic bond (C-1':O-4) by the addition of water, maintaining the 4-OH group at the new non-reducing chain end. PLs, in contrast, generate a hexeneuronic acid moiety at the new non-reducing end by eliminative cleavage of the O-4:C-4 bond. **image source,** (31).

The major source for PLs include bacteria, fungi and lower eukaryotes. Bacteria like *Pseudomonas*, *Stenotrophomonas*, *Azotobacter*, *Burkholderia*, *Pandoraea* etc are seen producing alginate specific PLs(32–35). Some molluscs from *Littorina* genus produce alginate lyases(36). The hyaluronate specific lyases include pathogenic bacterial genera like *Propionibacterium*, *Staphylococcus*, *Clostridium*, *Streptococcus* etc as their source organisms(37). Pectin lyases have filamentous fungi and plant sources like *Arabidopsis thaliana*(38,39). Bacteria like *Proteus vulgaris*, *Arthrobacter aurescens*, *Bacteroides thetaiotamicron*, and *Flavobacterium heparinum* produce chondroitin lyases(40–42). The *Flavobacterium heparinum* is a soil isolated bacterium that utilizes heparin as the sole carbon and nitrogen source by producing a heparin lyase(43). The Smlt1473 from a gram-negative, nosocomial bacteria *Stenotrophomonas maltophilia*, is a unique PL-5 family

polysaccharide lyase showing pH regulated multi-substrate specificity and cleaves diverse polysaccharides like hyaluronic acid, poly glucuronic acid and poly mannuronic acid at pH 5 (acidic), pH 7 (neutral) and pH 9 (basic) respectively(44).

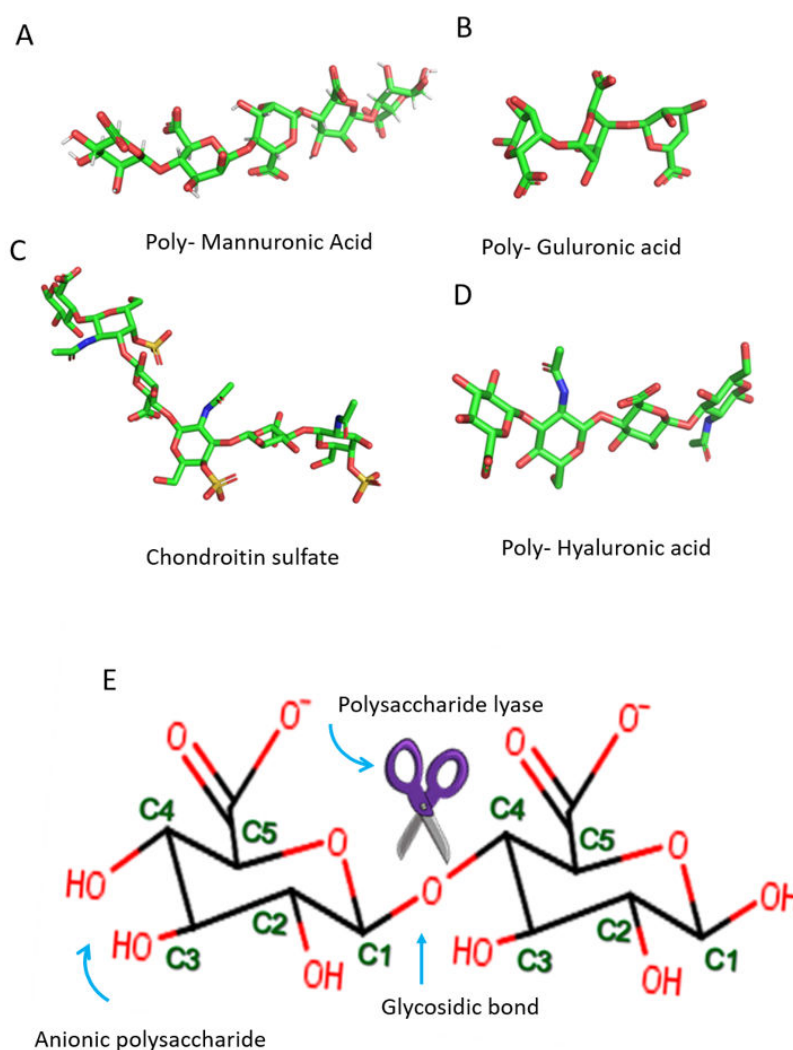


**Figure 1.5.** The fold diversity among PL families. The figure shows some of the folds from PL families from CAZY database. There are 42 families with different folds.

### 1.8. The Anionic Polysaccharides

Polysaccharides are important molecules of life and have tremendous applications in the pharmaceutical, food, cosmetic industries, and energy sectors. Anionic polysaccharides are negatively charged as they have -COOH group at C5. The subunits form a 1,4 glycosidic bond between C1 and C4. They are classified into three groups based on the sugar compositions at [+1] subsite of PLs (3). The Group-I represents polysaccharides composed of Galacturonic acid (GalA) like pectates, pectins, and rhamnogalacturonans. Pectins are present in the cell wall of

higher plants. The Group-II represents polysaccharides composed of Glucuronic (GlcA) or Iduronic acid (IdoA) e.g., hyaluronans, chondroitins, chondroitin sulfates, heparin (anti-coagulant) etc. Hyaluronic acid is present in the extracellular matrix of many higher organisms. The Group-III represents Mannuronic (ManA) or Guluronic (GulA) containing alginates. Many seaweeds, marine algae and bacteria produce alginate.



**Figure 1. 6. The structure of anionic polysaccharides.** The three-dimensional structures of different anionic polysaccharides are shown (A, B, C and D). The schematic representation shows the glycosidic bond between two monomeric sugar unit(€).

### 1.9. Significance of Polysaccharide lyase

The PLs are considered as useful enzymes in chemical industries as they can tailor the long chain polysaccharides into short oligomers. Synthesizing short oligomers chemically is a daunting task and not economical. In nature, PLs are utilized by their source organisms for biofilm production, seed germination, arranging food or carbon source by degrading anionic polysaccharides available in their natural environment. PLs are important in the lifetime of their source organism, loss of their function might risk the survival of the organisms they belong to. This property of PLs renders PL as valuable biotechnological editing tools.

Besides the industrial and medical applications, the PLs captivate the researchers by their molecular attributes like pH regulation, substrate specificity, conformational dynamics, mode of actions etc. Recently, a unique PL-5 enzyme Smlt1473, was reportedly to possess a pH regulated multi-substrate specificity. This enzyme cleaves Hyaluronic acid (HA) at pH 5, poly-glucuronic acid (GlcUA) at pH 7 and poly- Mannuronic acid (ManA) at pH 9. The polysaccharide lyase named vAL-1 from PL-14 family shows pH dependent alteration in mode of action. It switches from endolytic to exolytic with a shift in pH from pH 7 to pH 10(45). The *Sphingomonas sp* alginate lyase A-III from PL-5 family shows dynamical state transition between open and close conformations(46). The crystal structure of apo alginate lyase A-III form an open cleft like structure and switches to closed tunnel like structure upon substrate binding. The Normal Mode Analysis of a PL-12 Heparin lyase BtHepIII discovers the dynamical open-close movement around the en'yme's catalytic site(47). In PL-18 alginate lyase Aly-SJ02 the substrate entry is facilitated by an open-close gating function of lid loops (Ala 208-Gly217; Ala260-Thr265) obtained by their side-chain conformational changes(48). The catalytic activity for Aly-SJ02 was abolished by an N214C/T236C mutant which restricted the open-close dynamics by a disulphide bond formation.

### 1.10. The Beta elimination reaction mechanism

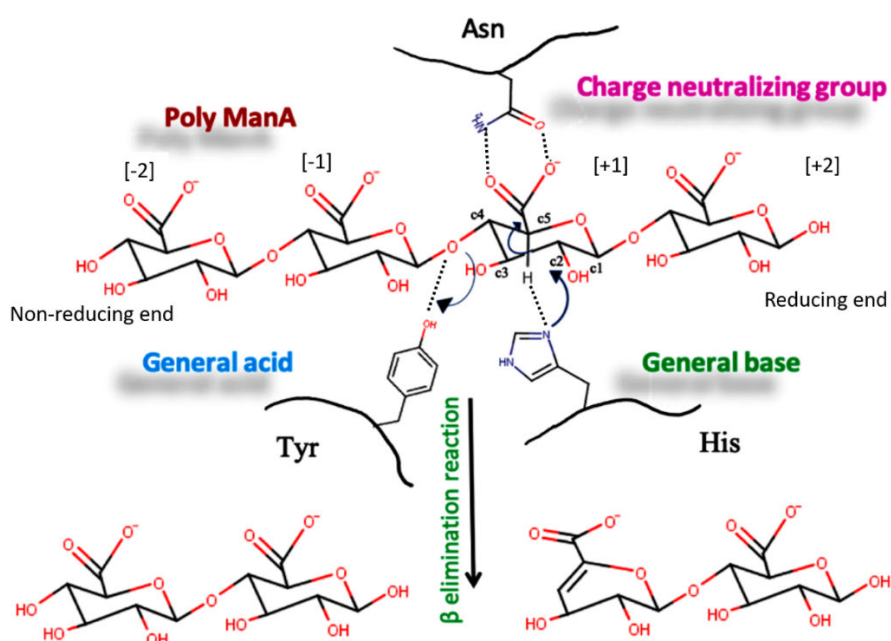
This is an eliminative cleavage reaction mechanism employed by all PLs for a nonhydrolytic cleavage of 1,4- glycosidic bond of the anionic polysaccharides. There are two types of products formed due to the lyase activity. The saturated cleaved product is formed at reducing end and the unsaturated cleaved product is formed at non reducing end [Figure 1.7]. The unsaturated product is named as 4-deoxy-erythro-hex-4-enpyranosyluronic acid and it absorbs UV at 235nm. This property helps in biochemical characterization of the enzymatic lyase activity.

There are three major steps in  $\beta$  elimination reactions [Figure 1.7] listed below:

1. Negative charge neutralization of C5 carboxylate group of +1 sugar ring. This step reduces the pKa of C5 proton and facilitate its abstraction.
2. Removal of C5 proton by a base and enolate anion intermediate formation.
3. Between C4 and C5 double bond formation and cleavage of C4-O-C1 glycosidic bond via electron transfer.

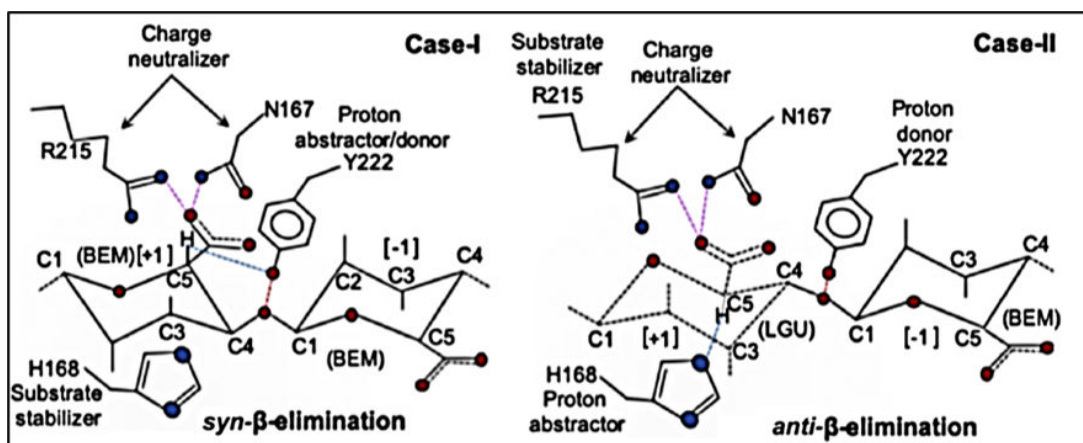
There are two types of degradation system are observed among PLs; (i) Lys/Arg system and (ii) His/Tyr system(3). The Lys/Arg system utilizes a metal assisted charge neutralization where positively charged metal ions act as charge neutralizer. Lys or Arg acts as a base in this system. The pectate lyases utilises H<sub>2</sub>O as a general acid. The residue acting as a general acid in this system is not conserved. All pectates and some chondroitin lyases use this degradation system. In the His/Tyr system, the PLs utilize Asn or Gln residue as charge neutralizer. The role of general acid and base is determined by the positioning of His and Tyr residues in substrate bound structure. The geometric conformation and accessibility of C5 proton to the His or Tyr in the catalytic side in the

enzyme-substrate complex chooses the general acid. If the proton protrudes itself towards the His, then His acts as a general base and Tyr acts as a general acid. Whereas, if the proton is in the proximity of Tyr, then Tyr acts as both general acid and base. When the C5 proton and C4 bridging oxygen are on the same side the configuration is called *syn*-configuration, and if they are on opposite side then they form the *anti*-configuration. Based on this geometric arrangement, the residues for catalytic base and acid are chosen and the beta elimination reaction mechanism is denoted as *syn*- $\beta$  elimination or *anti*- $\beta$  elimination [Figure 1.8]. The multi-specific PLs can perform both *syn* and *anti*- $\beta$  elimination based on the substrate it acts on(49).



**Figure 1.7. Schematic representation of  $\beta$  elimination reaction;** In this case Asparagine acting as the charge neutralizer, Histidine as a general base and Tyrosine as a general acid. Two types of products are forming. The unsaturated product has a double bond between C4 and C5.





**Figure 1.8. Schematic representation of *syn* and *anti*  $\beta$  elimination reaction.** The figure shows the active site residue interactions with the sugar subunits at [+1] and [-1] subunits in *syn* and *anti*  $\beta$  elimination reactions. In *syn*  $\beta$  elimination reaction the Tyr222 residue of Smlt1473 acts as both proton abstractor and donor, whereas in *anti*  $\beta$  elimination reaction the Tyr222 residue acts as proton donor and His168 acts as proton abstractor. (Image source(50)).

### 1.11. PanPL

The biological significance, structural diversity, multi substrate specificity and dynamic properties of polysaccharide lyases motivated this thesis work to understand the structure-function-dynamics of PLs from PL-5 family. Our work was inspired by Smlt1473, a PL-5 family enzyme that shows a unique pH-directed multi-substrate specificity. The Smlt1473 catalyzes the depolymerization of poly Hyaluronic acid (poly-HA) at pH 5, poly glucuronic acid (poly-GlcA) at pH 7, and poly Mannuronic acid (poly-ManA) at pH 9. We hypothesized that other PL-5 enzymes with similar amino acid distribution patterns like in Smlt1473 around their catalytic site, might exhibit a pH-regulated multi substrate specificity. Initially, we selected eight different PL-5 enzymes to examine our hypothesis. Based on the amino acid distribution pattern, we divided these eight PL-5 enzymes into two groups, i.e., Smlt1473-like and Smlt1473-unlike [Table 1.1 and 1.2]. After a strenuous

process of cloning and purification of eight proteins, we stumbled upon PanPL from *Pandoraea apista*, which was found amenable for biochemical assay and crystallization. The source organism, *Pandoraea apista* is a pathogenic gram-negative bacterium first isolated from the sputum of a cystic fibrosis patient.

Hypothesis: Smlt1473-like sequence pattern might enable the enzymes to possess the pH directed functional regulations among PL-5 family.

We coined the following objectives:

1. To investigate whether selected PL of the PL-5 family possess **pH directed catalytic activity** on various substrates
2. To determine the **structure** of the selected PL in apo and substrate bound form by X ray crystallography.
3. To investigate the **role of catalytic residues** in lyase activity by both biochemical and crystallographic studies.
4. To establish the **effect of active site flanking residues** on substrate specificity and enzyme activity.

**Table 1.1. The PL-5 enzymes under the Smlt1473-UNLIKE group**

Sr. no	Gene Id	Smlt1473-UNLIKE Smlt1473 flanking residues around conserved Trp								Sequence identity w.r.t. Smlt1473
	Smlt1473	115 Y	162 K	163 R	171 W	218 R	221 H	225 Y	312 R	
1	<b>AGK12841.1</b> (AvPL)	S	L	K	W	R	A	Y	R	32%
2	<b>CAA11481.1</b> (AcPL)	S	L	K	W	R	A	Y	R	32%
3	<b>ABP60714.1</b> (EsPL)	S	L	E	W	R	A	Y	R	29%
4	<b>AGI80338.1</b> (PaPL)	S	L	K	W	R	A	Y	R	29%
CONSENSUS		<b>S</b>	<b>L</b>	<b>X</b>	<b>W</b>	<b>R</b>	<b>A</b>	<b>Y</b>	<b>R</b>	

**Table 1.2. The PL-5 enzymes under the Smlt1473-LIKE group**

Sr. no	Gene Id	Smlt1473-LIKE Smlt1473 flanking residues around conserved Trp								Sequence identity w.r.t. Smlt1473
	Smlt1473	115 Y	162 K	163 R	171 W	218 R	221 H	225 Y	312 R	
1	<b>AIO37757.1</b> (BcPL)	F	N	R	W	R	H	F	R	34%
2	<b>ACS64514.1</b> (RpPL)	Y	N	H	W	R	H	F	R	35%
3	<b>AEU36899.1</b> (GmPL)	Y	T	K	W	K	H	Y	Y	36%
4	<b>AJE99968.1</b> (PanPL)	Y	K	H	W	L	H	Y	R	60%
CONSENSUS		<b>Y</b>	<b>X</b>	<b>X</b>	<b>W</b>	<b>X</b>	<b>H</b>	<b>Y</b>	<b>X</b>	

## 1.12. Conclusion

Even though our hypothesis found to be negative, this thesis work gave us an opportunity to explore holistically the structure-function-dynamics, quest for a novel polysaccharide

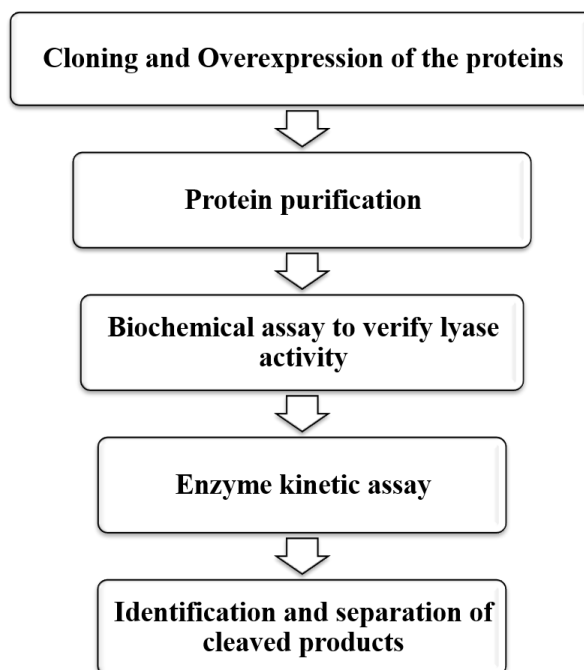
lyase, and propose mechanistic model for the functioning of novel polysaccharide lyase PanPL. The PanPL, localizes at periplasmic space and performs an extracellular lyase activity. The biochemistry of PanPL characterized it to be an alginate specific PL with an activity range within pH 5.5 to pH 7.5. Although PanPL shares 60% sequence identity with Smlt1473, it didn't show pH regulated multi substrate specificity. The enzyme kinetic studies revealed PanPL to be an allosteric enzyme. The product analysis characterized PanPL as an endolytic enzyme. We have determined the crystal structures of PanPL across the pH spectrum from acidic pH 3.5 to alkaline pH 8.5. The substrate bound crystal structure was obtained by co-crystallizing tetra-Mannuronic acid with active site mutant of PanPL H172A. In this thesis, we have performed several analyses on the crystal structures to characterize the dynamical behavior of PanPL. We tried to explain how pH influences enzyme activity, surface charge and dynamics. Further, we discussed the effect of substrate acquisition on the dynamics. In addition, we have generated various mutations to understand the role of the active site flanking residues in enzyme catalysis. The thesis work will delineate the overall structure-function-dynamics of PanPL. Further, our work demonstrates the distinct modes of dynamics of PanPL (both intrinsic and substrate-induced) which are vital for enzyme functioning and allostery, along with the structure guided mutational studies to alter the enzyme catalysis and dynamics.



## **CHAPTER 2: Materials and Methods**

## 2.1. Introduction

This chapter provides the experimental and computational methods employed to achieve the four objectives. This chapter enlists the materials has been used in the research. The detailed description of the methods utilized in this research work has been provided. For each objective, workflow has been drawn [Figure 2.1, 2.8].



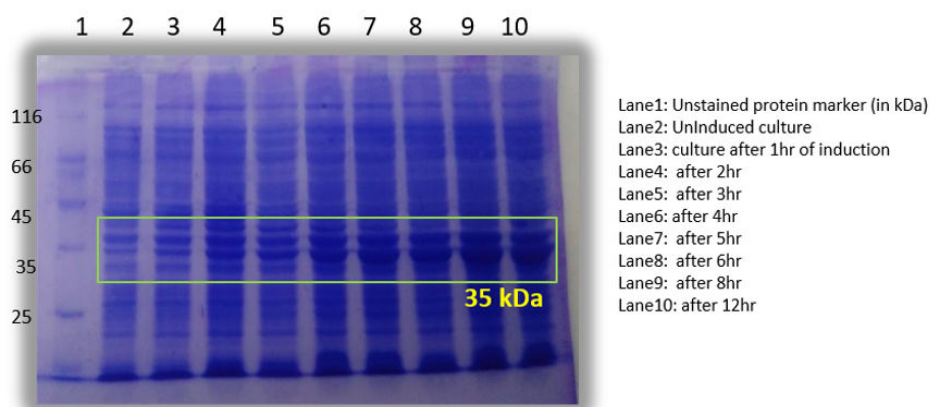
**Figure 2.1 workflow for the objective 1 [section 1.11 and page 20].**

## 2.2. Cloning and overexpression of PanPL

We cloned the PanPL gene into pET28a (+) vector between NdeI and XhoI restriction sites to utilize the (His)<sub>6</sub> tag protein purification process. The plasmid containing the PanPL gene was transformed into *E. Coli*. Lemo21(DE3) cells. The transformed cells were spread on a pre-warmed antibiotic agar plate with two antibiotics i.e., Kanamycin and chloramphenicol. The pET28a (+) vector has Kanamycin resistance and the Lemo21 (DE3) cells have chloramphenicol resistance.

The overexpression of PanPL in Lemo21(DE3) cells was optimized [Figure 2.2] in a small scale initially and then performed in a large scale. The next day of the transformation, a single colony was inoculated into a 10ml of LB broth with the above-mentioned antibiotics. The cells were allowed to grow overnight at 37° C for primary culture.

Following next morning, the primary culture was inoculated in 1L LB broth with the antibiotics for a secondary culture and after three hours the culture was induced with 1.5mM of IPTG at OD 0.6. The induced secondary culture was grown for 12 hours at 18° C in the shaker incubator. We noticed that during the protein expression, the N-terminal signal peptide and the (His)<sub>6</sub> tag was getting cleaved making the protein tag-less with a molecular weight of 34.9 kDa.



**Figure 2.2. The over-expression of PanPL.** The gel image shows increasing in protein expression as a function of time in cell.

### 2.3. Extracellular secretion

The occurrence of N-terminal signal peptide in the primary structure indicates the secretory nature of PanPL. The plate assay was performed to confirm PanPL secretion into the periplasmic space. We prepared the assay plate (44) by adding 1mg of alginate to LB agar.

The induced bacteria culture (12hr of induction) was placed on the agar-alginate plate and left at room temperature for 24 hours. The plate was then flooded with a 10% acetyl butyl solution, and the clear sections was observed which confirms the extracellular lyase activity [Figure 2.3].



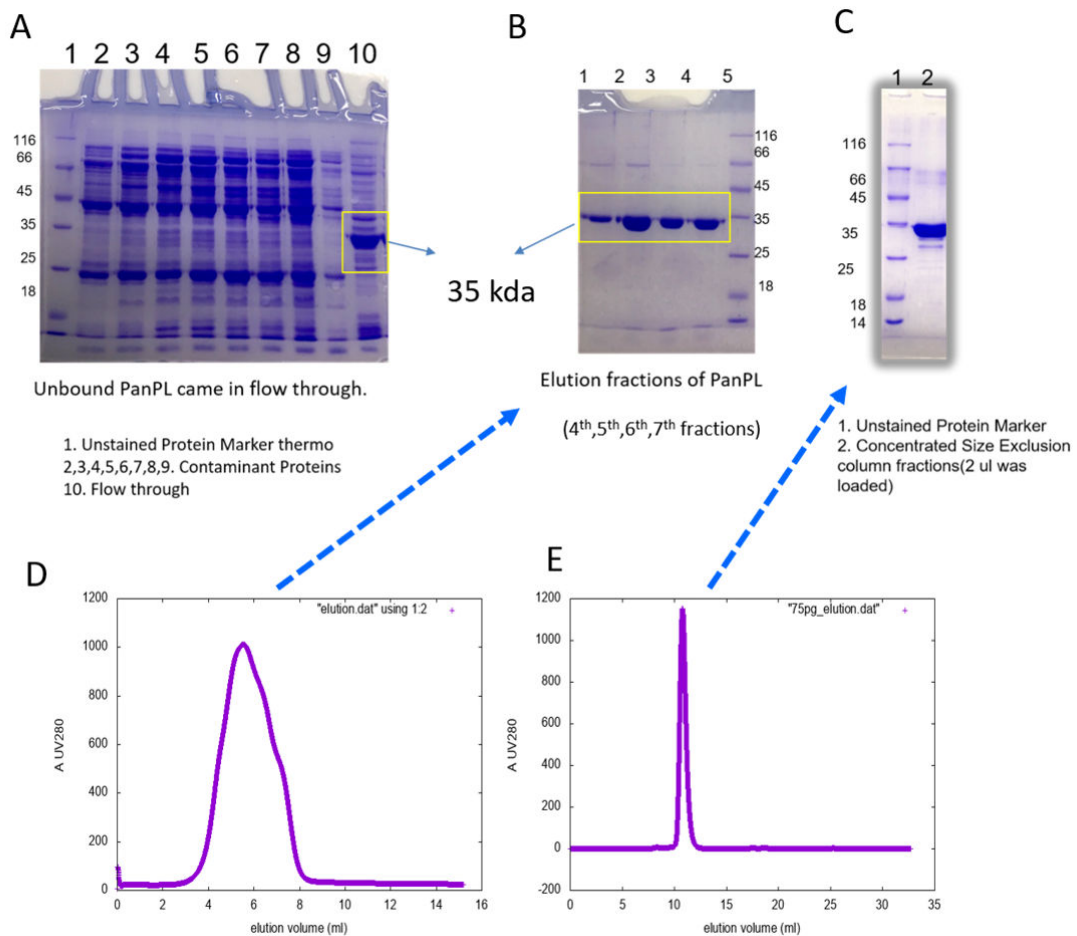
**Figure 2.3. The extracellular lyase activity of PanPL.** The plate assay of 1. uninduced culture shows no clear sections, hence no lyase activity. 2. The induced culture shows clear section or degradation in agar-alginate plate, signifies the extracellular lyase activity.

## 2.4. Purification of PanPL

We performed anion exchange chromatography for the purification of tag-less PanPL using the RESOURCE-Q column (GE healthcare). To keep the protein negatively charged, the purification buffers were chosen to be at pH-9.0 (above the pI, i.e., pH 7.2). The lysis buffer contained 50mM Tris pH-9.0, 150mM NaCl, 0.02% Triton X-100, and 5mM beta-mercaptoethanol. The cells were suspended in lysis buffer containing PMSF and lysed for 15 minutes at 45% amplitude using a tip-sonicator with an ON (5s) and OFF (10s) pulse rate. The lysate was centrifuged for 1 hour at 18000 rpm. To keep the salt concentration at 30mM NaCl, the supernatant was diluted with 50mM Tris pH-9.0 buffer. For the following purification steps, the buffers A: 50mM Tris pH 9, 30mM NaCl, 5mM beta-



mercaptoethanol and B: 50mM Tris pH 9, 1M NaCl, 5mM beta-mercaptoethanol were prepared. The column-equilibration was done using buffer A prior to sample loading. The gradient elution was performed and the PanPL eluted at 100mM NaCl [Figure 2.4 B, D]. To further purify, size exclusion chromatography using Superdex 75 10/300 pg was performed [Figure 2.4 C, E].



**Figure 2.4. Purification of PanPL;** The upper panel shows the gel image for elution fractions of PanPL after A. first round ion exchange chromatography, B. second round ion exchange chromatography and C. size exclusion chromatography. The lower panel shows the chromatogram after D. ion exchange chromatography, E. size exclusion chromatography.

### **2.5. Thiobarbituric acid (TBA) assay**

To confirm PanPL lyase activity, the standard colorimetric-based TBA assay was used (51). We used three solutions for the TBA assay: a periodate solution, an arsenite solution, and a TBA solution. 20µg of protein was added to 500µg/ml substrates dissolved in 200µl buffers with pH range from 4.0 to 9.0 in steps of 1.0 unit. For each pH the reactions time was set to be 10 mins. Each reaction received 50ul of periodate solution and was incubated for 20 minutes. In the presence of periodate, the unsaturated cleaved product formed during the reaction results in a pre-chromogen. To destroy the extra periodate in the reaction, 200µl of arsenite was added to the reaction mixture. The reaction mixture was heated close to 100°C (after adding 500µl of TBA solution). The pre-chromogen reacts with TBA to give pink colour chromogen. This pink colour formation indicates lyase activity. We then dissolved the chromogen in cyclohexanone, and measured absorption at 550nm to determine the amount of chromogen formed (hence, the unsaturated product formed). The lyase activity of PanPL mutants was tested using cell lysate taking PanPL wild type as a positive control. TBA assays for RpPL and BcPL were performed using cell lysate and purified enzyme, respectively.

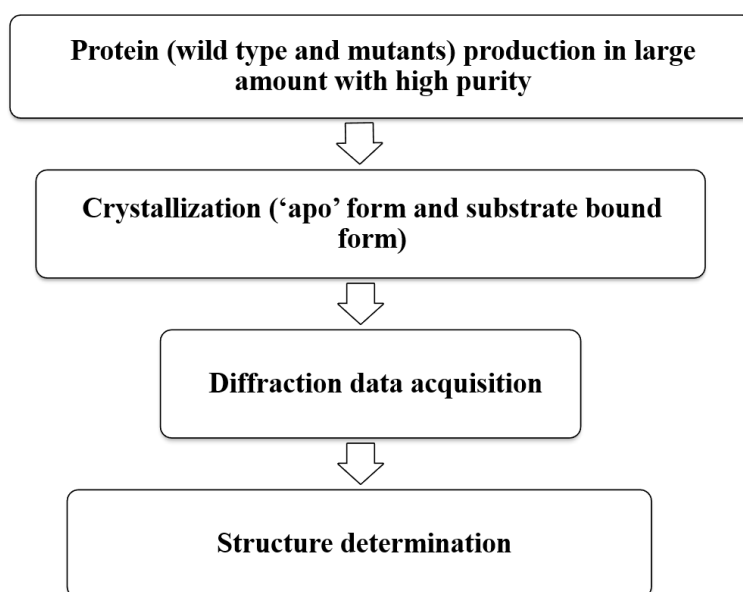
### **2.6. Enzyme kinetics**

The TBA assay was used to scan for the optimal pH of PanPL lyase activity for different substrates. Further, we performed the enzyme kinetic assay to determine the optimal pH of activity accurately for the given substrates. We incubated 50µg of protein and 100µM of substrates in buffers with pH 4.0 to 9.0 in 1 unit steps for a pH scan in enzyme kinetics. Enzyme activity at 235nm was measured on an "Eppendorf Bio-spectrometer Kinetic" with a 10mm quartz cuvette for each pH(52). The optimal pH was discovered to be 7.0. The enzyme activity at 25°C was measured in µmol/min/ml. All enzyme kinetic measurements

were performed on 50 $\mu$ g of protein sample for the substrates Alginate and Polymannuronic acid with substrate concentration ranging from 20 $\mu$ M to 1mM. During the kinetic, 500 $\mu$ l of reaction mixture was used. The OriginLab software was used to plot the kinetic curves by using nonlinear curve fitting to determine the kinetic constants.

## 2.7. Product Analysis

Alginate is a polysaccharide composed of negatively charged monosaccharide units. Hence, we analyzed the products of the enzymatic lysis using the principle; its size proportional to charge. We incubated 5mg of substrates (alginate/poly-ManA) with PanPL overnight in phosphate buffer pH 7. In the next day, using a 10kDa cutoff Amicon-Ultra centricon tube, the enzyme was separated from the reaction mixture. The absorbance at 235nm was used to detect the presence of cleaved products. The end products of the enzymatic degradation were separated using anion exchange chromatography on a Hi-trap HP (Cytiva) column. The peak fractions were collected and lyophilized. The mass of the products was analyzed by mass spectrometry.



**Figure 2.5. workflow for the objectives 2, 3, and 4 [Section 1.11 page 20].**

## 2.8. Mutant protein production

We performed site-directed mutagenesis (SDM) to create the active site mutants N171L, H172A, and Y226F, as well as R219L mutant. A few non-active site mutations i.e., L222R, R48G, L222R\_R48G were also created by using SDM. PCR was carried out with mutant primers [Table 2.1, 2.2, 2.3] and NEB Q5 polymerase. A QIAquick PCR purification kit was used to purify the PCR products. The T4 PNK (NEB) enzyme was used to phosphorylate the PCR product ends [Table 2.4], which were then ligated using Quick ligase (NEB). We used the Dpn1 (NEB) enzyme to degrade the product's parental strands. The *dpn1* digested product (10ul) was transformed into DH5 alpha cells. A single colony was inoculated in 10ml LB with 0.1% Kanamycin and incubated overnight to grow the primary culture. Qiagen MiniPrep kit was used to isolate mutant plasmids from cells. Sanger sequencing was used to confirm the correctness of the mutations. The mutant plasmids were transformed into *E. Coli*. Lemo21(DE3) cells. At various IPTG concentrations, the mutant protein expression was screened for optimization. To purify the mutant proteins, we used PanPL wild-type purification protocols.

**Table 2.1. The list of primer sequence used to create point mutation****FP: forward primer, RP: reverse primer**

Mutation	Primer
N171L-FP	CTGCACTACTATTGGACCGGTGTGGG
N171L-RP	GTTACGTTTGTGCTTCGGGTTATCCC
H172A-FP	GCGTACTATTGGACCGGTGTGGGTATCATGG
H172A-RP	GTTGTTACGTTTGTGCTTCGGGTTATCCC
Y226F-FP	TTTCACGACTATGCGACCGCGCC
Y226F-RP	GTGCAGCGCCAGACGCTTACGCGC
R219L-FP	CTGAAGCGTCTGGCGCTGCAC
R219L-RP	CGCCATTTCCATCGGCAGACGACC
L222R-FP	CGTGCGCTGCACTACCACGACTATGC
L222R-RP	ACGCTTACGCGCCATTTCATCG
R48G-FP	GACGCGGCGGGCAGCGTTATTGATCCGCGTCTGAA
R48G-RP	GGTATAGTAGCCAATCGCACGGATGTCCGGGCTACC

The following steps (1-6) are utilized for site directed mutagenesis.

#### 1. Mix gently and prepare the PCR mix.

**Table 2.2. Reagents added in the PCR mix**

Reagents	20 µl reaction
Q5 Buffer	4 µl
10mM dNTPs	0.4 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
Template DNA (100 ng/µl)	0.5 µl
Q5 High-fidelity DNA Polymerase	0.3 µl
Nuclease-free water	12.8 µl

## 2. Thermocycling Conditions for a Routine PCR

**Table 2.3. The thermocycler protocols.**

STEP	TEMPERATURE	TIME
Initial Denaturation	98°C	1 min
30 Cycles	98°C	15 sec
	50–72°C*	1 min 30 sec
	72°C	2 min
Final Extension	72°C	3 min
Hold	4°C	

## 3. Phosphorylation of ends and ligation

**Table 2.4. Reagents added for phosphorylation and ligation**

Reagents	Volume
Quick Ligase Buffer	20µl
PCR Product	10µl
20mM ATP	2 µl
Quick Ligase	0.7µl
T4 PNK	0.7µl
Nuclease-free Water	3.5µl

**4. DpNI digestion**

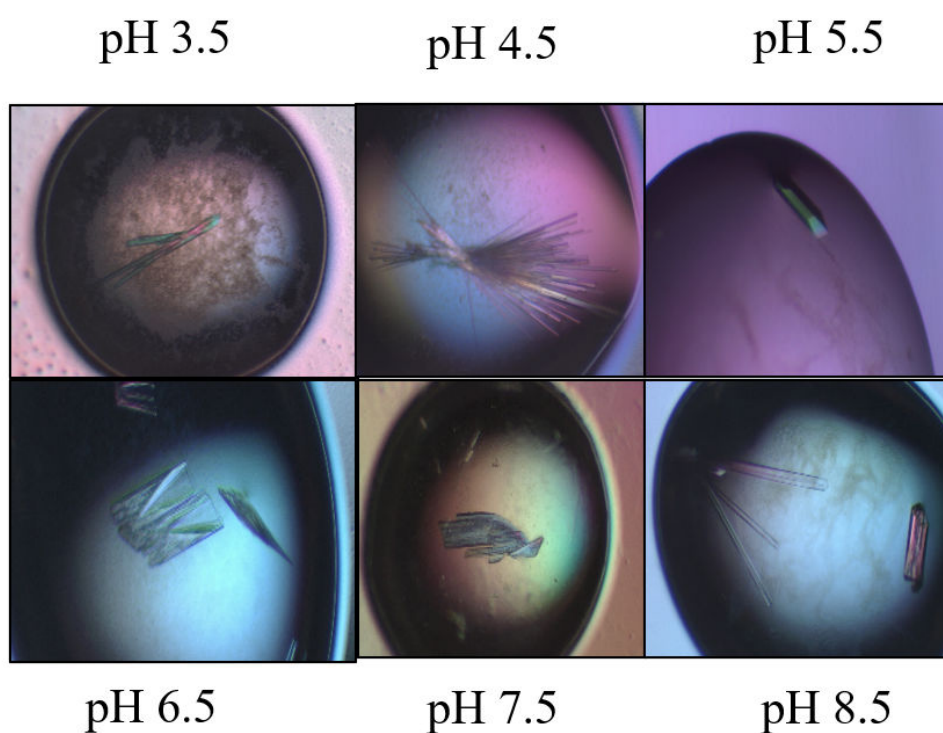
**5. Transformation of plasmid into DH5 $\alpha$**

**6. Confirm the mutation by sanger sequencing**

**2.9. Crystallization**

We used the purified protein in 50mM Tris-HCl, pH-9.0, 200mM NaCl for the initial crystal screening of wild-type PanPL. Initially we screened for crystal growth in commercially obtained crystallization conditions (Hampton research Index Screen). The single crystals were found grown in 0.1 M Bis-Tris pH 5.5 and 25% PEG 3350 crystallization condition. We then optimized crystallization for various pH. The protein samples were prepared at different pH values (50mM Citric Acid for pH 3.5, 50mM Acetate buffer for pH 4.5, pH 5.5; 50mM HEPES buffer for pH 6.5; 50mM sodium cacodylate buffer for pH 7.5; 50mM Tris buffer for pH8.5) with 200mM NaCl using a buffer exchange protocol. To grow crystals at a specific pH, both the protein sample buffer and the crystallization condition were maintained at the same pH. Crystals were formed under crystallization conditions Index 40-45 (pH 3.5 to pH 8.5) of commercially available Index screen [Figure 2.6, Table 2.5]. We utilized both hanging drop and sitting drop methods to crystallize the protein.

The active site N171L mutant was crystallized at 0.1M citric acid pH 3.5, 25% PEG 3350. H172A and Y226F were crystallized at 0.1mM Bis-Tris pH 5.5, 25% PEG 3350. PanPL L222R was crystallized at various pH, where as we could obtain PanPL R48G crystal at pH 5.5 and PanPL L222R R48G crystals at pH 4.5 [Figure 2.7, Table 2.5].

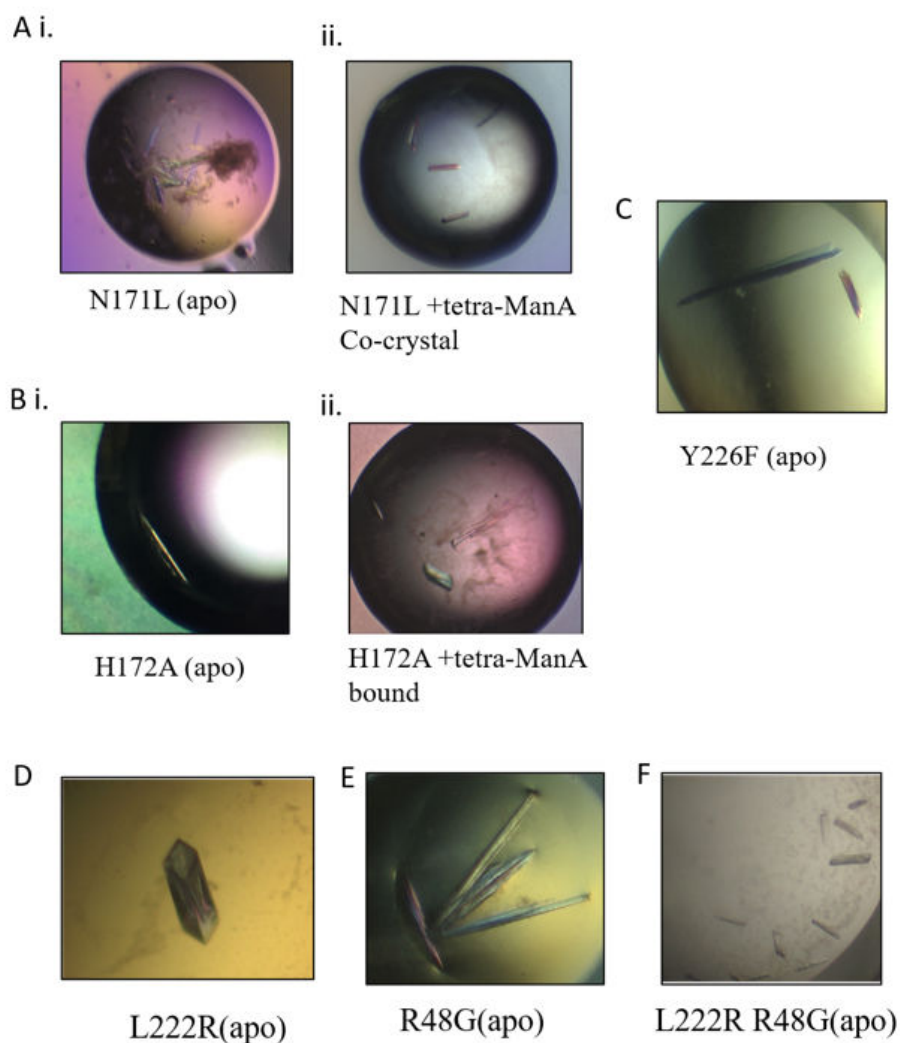


**Figure 2.6. Crystal images of PanPL wildtype across the pH spectrum**

### **2.9.1. Co-crystallization with substrates**

The mutant proteins N171L and H172A were incubated with tetra-ManA for 30 minutes before being lyophilized. The lyophilized samples were reconstituted with deionized water before crystallisation with 0.1M Citric acid pH3.5 and 25% PEG 3350. The diffraction data for N171L and H172A co-crystals [Figure 2.7, Table 2.5] were collected with soaking and without soaking.





**Figure 2.7. The crystal images for the mutants of PanPL;** A. PanPL N171L crystals in i. apo and ii. Co-crystallised forms. B. PanPL H172A crystals in i. apo form and ii. Tetra-ManA bound forms. C. PanPL Y226F crystals in apo form. A, B, C are the crystal pictures of active mutations. D. PanPL L222R crystal in apo form, E. PanPL R48G in apo form, F. PanPL L222R R48G in apo form. D, E, F show the crystals for the flanking residues around the active site.

**Table 2.5. The crystallization conditions for PanPL and its mutants**

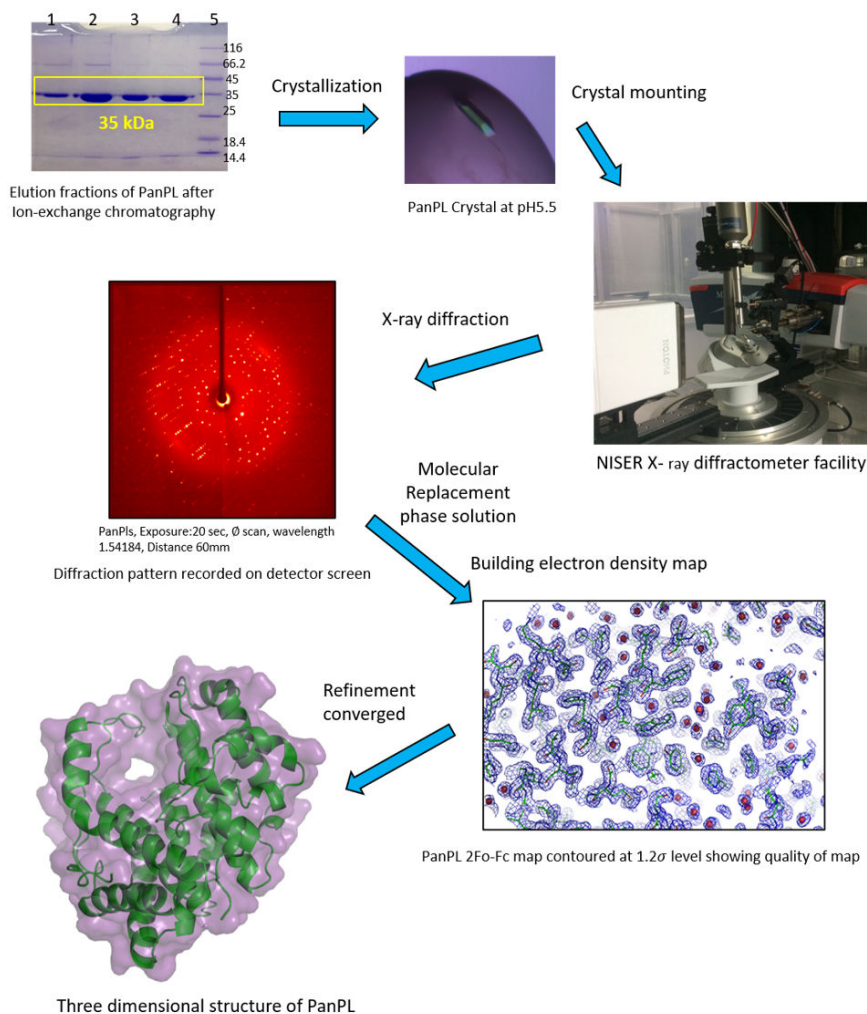
PanPL variant	Apo/Substrate bound	Crystallization conditions
PanPL pH 3.5	Apo	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350
PanPL pH 4.5	Apo	0.1 M Sodium acetate trihydrate pH 4.5, 25% w/v Polyethylene glycol 3,350
PanPL pH 5.5	Apo	0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350
PanPL pH 6.5	Apo	0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350
PanPL pH 7.5	Apo	0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350
PanPL pH 8.5	Apo	0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350
PanPL N171L	Apo	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350
PanPL N171L	Co-crystallized, tetra-ManA, no substrate density	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350
PanPL H172A	Apo	0.1 M Sodium acetate trihydrate pH 5.5, 25% w/v Polyethylene glycol 3,350
PanPL H172A	tetra-ManA, Substrate bound	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350
PanPL Y226F	Apo	0.1 M Sodium acetate trihydrate pH 5.5, 25% w/v Polyethylene glycol 3,350
PanPL L222R	Apo	0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350
PanPL R48G	Apo	0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350
PanPL L222R R48G	Apo	0.1 M Sodium acetate trihydrate pH 4.5, 25% w/v Polyethylene glycol 3,350

## 2.10. X-ray Diffraction Data Collection and Structure

### Determination

The very first crystal for PanPL was obtained in the crystallization condition at pH 5.5. We used 20% ethylene glycol as a cryo-protectant before mounting the crystal on a home source diffractometer (Bruker) outfitted with a rotation anode Cu-K $\alpha$  radiation source, LN2 cryo-stream at 100K, and a photon100 detector. The cell parameter was determined using the unit cell determining strategy implemented in BRUKER's PORTEUM software. The phi-scan was used to collect the entire dataset, and diffraction data sets were recorded to a maximum resolution of 1.89Å. PROTEUM software was used to integrate and scale the raw data sets. The statistics for data collection are shown in Table 3.2 Chapter 3. Since the sequence similarity with Smlt1473 was 60%, we solved the phase problem (53) using the molecular replacement (MR) method of PHASER(54) using the poly-ala model of Smlt1473 as the probe. The refinement was carried out using the refinement protocol implemented in Phenix (55), and all of the side chains were modelled in the electron density map. Then the crystallographic water molecules were added satisfying the geometric parameters for hydrogen bond formation with polar atoms using Fo-Fc maps contoured at 3.0 sigma level. Coot(56), the visualization program was used to build the model and trace the side chain and solvent molecules in the electron density map. The refinement was performed until it converged. The statistics for refinement are shown in Table 3.2 Chapter 3. This structure was used as a probe for solving phase problem by MR for the diffraction data sets of other pH variants and mutant crystals. Similarly, diffraction data sets were collected on the home source for PanPL pH 7.5 crystal. Again, for mutants H172A, N171L home source was used to collect data in both apo and co-crystallized with tetra-ManA forms. Ethylene glycol 20% was used as a cryoprotectant before flash freezing the crystals. The statistics for data collection are shown in Table 3.4, Chapter 3. To solve these

structures, we used the molecular replacement method implemented in phaser. The auto-build protocol i.e., *phenix.autobuild* (57) is implemented for model building, and the structures were refined using *phenix.refine* (58). We did not find the density for the substrate in the N171L co-crystallized structure. However, for the H172A co-crystallized structure, the substrate was modelled in the Fo-Fc map at 3-sigma level, and only two sugar units were resolved in the map. The diffraction data sets for the crystals grown in conditions with pH values of 3.5, 4.5, 5.5, 6.5, and 8.5 and Y226F mutant were collected at the ESRF ID29 beamline. The statistics for data collection are shown in Table 3.2 and 3.4 Chapter 3. The diffraction data for L222R, R48G and L222R R48G mutants were collected at the ESRF ID30-A beamline. We built the model with *phenix.autobuild* (57) and refined it with *phenix.refine* (58). We used a Fo-Fc map contoured at 3.0 sigma to model the water and solvent molecules. Table 4.2 Chapter 4 displays the refinement statistics for each structure. The structure validation is an important step before deposition to RSCB-PDB database. The space group validation was performed on the structures using Zanuda(59). The sequences of crystal structures of PanPL-WT and mutants were thoroughly checked before PDB deposition. The electron densities of the side chains of amino acids, water molecules, ligands and ions were traced carefully. The geometry and rotamer corrections were performed. The clashes within the molecule and with symmetry related molecules if any were resolved. The structures were validated using RAMPAGE(60).



**Figure 2.8. The overall workflow for structure determination of PanPL.** The purified enzyme has been crystallized and the first diffraction data has been collected at NISER X-ray diffractometer facility. The diffraction data then processed and structure has been solved by molecular replacement method. Following the iterative model building, and when iterative refinement converged, the structure was used for analysis and deposition.

### 2.11. Electrostatic Surface charge analysis

The *pqr* models were generated using PDB2PQR to create an electrostatic surface charge map(61). Using .pdb files as input, and the PARSE force field the electrostatic calculations were performed. The PROPKA was used to calculate the pKa at their respective pH(62).

The outputs were used by the Pymol APBS plugin to generate and visualize the electrostatic potential surface map (63,64). The electrostatic potential range was set (+5000 Ke/T to -5000 Ke/T) with 0.05 grid spacing.

### **2.12. Rosetta flexible ligand docking**

Rosetta flexible protocol (65) was used for ligand docking, which includes low resolution docking followed by high-resolution docking (65). The ligand molecule was translated around a binding hotspot as a rigid body and then rotation and translation were performed on it. This step is called the low-resolution docking. The side chains were repacked in high-resolution docking. The scoring was performed in *ligand\_soft\_rep* energy terms followed by the complex minimization in *hard\_rep* energy terms. When applied to different conformational states of ligand, the above protocol results in ligand being sampled with different conformations and orientations. The side chains surrounding the ligands were repacked and relaxed alongside the backbone sampling the protein conformational flexibility. 5000 docking models were generated in total. To find the best docking pose, the models were ranked based on the total energy score. Then the top 20% of the models from the previous sorting were selected to screen the highest interfacial binding energy. The model with the lowest interface energy and reasonable ligand interactions with active site residues corresponds to the most likely ligand configurations.

### **2.13. Ensemble refinement**

The Ensemble refinement (ER) was performed using *Phenix* software package (14,66). All alternative conformations from the crystal structures were removed, and occupancies in the structures for the ensemble refinement were set to 1.0. We performed the ER for each pH structure at various settings of the ER parameters  $p_{tls}$ ,  $T_{bath}$ ,  $T_x$ , and  $nmodels$  to determine their optimal values. The program allowed the resolution-specific parameters

T<sub>x</sub> and nmodels to be configured, while T<sub>bath</sub> was set at 10K, 5K, and 2K values, and the p<sub>tls</sub> was set from 0.6 to 1.0 in 0.1 steps for each T<sub>bath</sub> value. The protein monomer was defined as a single TLS group in each ER. The ER parameters are taken as optimal that result in a model with a low R<sub>free</sub>. To test the reproducibility, we repeated ER four more times independently using the optimized parameters. The ER statistics and associated parameters are shown in Table 3.9 Chapter 3. We adapted the ER to gain atomic level dynamics insight as function of pH, since we had structures deduced across the pH spectrum. This aided us in analysing the structure-function-dynamics relationship.

## 2.14. Homology Modeling

To generate the homology structural models for BcPL and RpPL, we used the comparative modelling method implemented in RosettaCM(67). We used four structures as templates for the closed states: 4OZV, 4OZW, Smlt1473 (7FHZ), and PanPL (7WXM), and one for the open state: 1QAZ. Before modelling, we deleted the predicted signal peptide region of RpPL and BcPL. For the BcPL and RpPL sequences, we first used Rosetta-server to generate sequence-based 3-mer and 9-mer fragments. The sequence alignment was created between the template structures (both in open and close state) and the target sequences of BcPL and RpPL. The individual sequence alignment and corresponding template structure were used to build threaded models of the target sequences using Rosetta's partial threading protocol (*partial\_thread*). We used all three stages of RosettaCM: stage-1 generates the full-length model using sequence-based fragments (3-mer and 9-mer) and template-derived segments. In stage-2, the local structure and loop closure of the generated full-length model were optimized. Stage-3 involved performing all-atom refinement (*fastrelax*) on the stage-2 model, including backbone and sidechain refinement. Appropriate weights were assigned to each stage. Total four templates for closed state modelling were chosen, which we treated as multiple templates in RosettaCM to generate hybridized structures. The

RosettaCM was set to generate 100 models. As the best homology structure, we chose the model with the lowest rosetta score. We also used alpha-fold suite (68) to generate the BcPL and RpPL structures.

## **2.15. Molecular Dynamics Simulations**

A Classical molecular dynamic simulation(69) was performed on the PanPL apo crystal structure (pH 6.5). Firstly, the protein was kept in a cubic box, then solvated it with addition of water molecules. A periodic boundary was establish keeping the protein at the centre of the box. Gromacs version 2019.227 (69) , CHARMM 36 force field (70)and the TIP3P water model (71)were used for simulation. The particle mesh Ewald (PME) method and LINCS constraint algorithm were used to evaluate long-range electrostatic interactions. The Lennard Jones potential cut-off was set to 1.2 nm. Initially, the energy minimization step was used to relax the system to ensure no presence of abnormal geometries and steric clashes. The goal was to achieve the lowest possible potential energy with a maximum target force of no more than  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . The steepest descent method was used for energy minimization. The 4ns of equilibration was performed followed by 100ns of production. Throughout the equilibration and production stages, a 2fs time step was used. After analyzing the 100ns production outcomes, we extended the production to one micro second.

Simulation restraints were applied to all protein atoms during equilibration, but they were removed during the production stage. The Nose Hoover thermostat was used in the production stage to maintain constant temperature of 303.15 K and the Parrinello-Rahman barostat with isotropic coupling was used to control the pressure. A multistep leapfrog integrator was used to integrate the equation of motions. The above method was used for substrate-bound structures (X-ray and docked). In case of the crystal structures, we



introduced a computational mutation from Alanine to Histidine at 172 sites because two H-bond interactions were observed in our docked structure. The CHARMM 36 force field was used to model the protein, and the CHARMM General Force Field was used to parameterize the substrate Di/Tetra-ManA (CGenFF, version 3.0.1) (72).

The trajectories of PanPL simulations were processed using `trjconv` command to correct the jumps or breaks that protein might experience across the periodic boundary during the production step. All of the analyses were performed on the corrected trajectory, which had the protein in the centre of the box. After aligning with the crystal structure, r.m.s.d. (root mean square deviation) was calculated for each snapshot at 100 ps intervals of the 100ns trajectory. Similarly, for the backbone atoms, r.m.s.f. (root mean square fluctuations) per residue was calculated by taking the initial structure as reference.

### **2.16. Principal component analysis of MD trajectories**

The principal component (PC) analysis of  $C\alpha$  atoms coordinates were performed aligning the  $C\alpha$  atoms time evolved coordinates to the  $C\alpha$  atoms of the starting structure(73). Only first few PCs were used to analyze the essential dynamics. The first Principal component PC1 had the highest proportion of variance, followed by the second principal component PC2. PC1 highlight the major conformational states or dynamics. Hierarchical clustering in PC space was used to further decode the dynamical states. This resulted in three distinct subgroups [Figure 3.31, Chapter 3]. The average structures of the subgroups represented three distinct conformational states; one represents the closed state, while the other two represent the open state. To identify the protein regions as major contributors of concerted motions, individual residue contributions to the principal components were assessed. The trajectories of first two PCs were extracted to analyze the concerted motions. The PC1 reveals N-loop-lid motions for closed to open state transition. The PC2 shows a twisting

motion or sliding back and forth along the tunnel. The entire PC analysis was performed using bio3D package(15).

### **2.17. MMPBSA calculation**

We performed the MMPBSA calculation for L222R and tetra-ManA docked complex by using gmx\_MMPBSA tool (74). The gmx\_MMPBSA tool has been developed to perform AMBER's MMPBSA calculation on GROMACS trajectories. The binding free energy decomposition calculation for the protein-ligand complex has been performed to generate a residue wise binding energy profile. The .xtc file from GROMACS output has been used as input. The GROMACS topology file and index file has used as input. The ligand parameter file has been generated by using Antechamber ligand parameterization. The final output files for end state energy calculation and decomposition analysis have been generated.

### **2.18. Electrostatic interaction network modelling**

The PROPKA31 tool was used to perform pKa prediction (62,75). It predicts the pKa of the ionizable groups in protein. This analysis generates the information about the solvent accessibility, and the columbic interaction counterparts for charged residues with their contribution to influence the pKa. Based on this information we modelled the columbic interaction network by using Pymol as the visualization tool(64).



# **CHAPTER 3: PanPL, unique allosteric PL-5 enzyme**

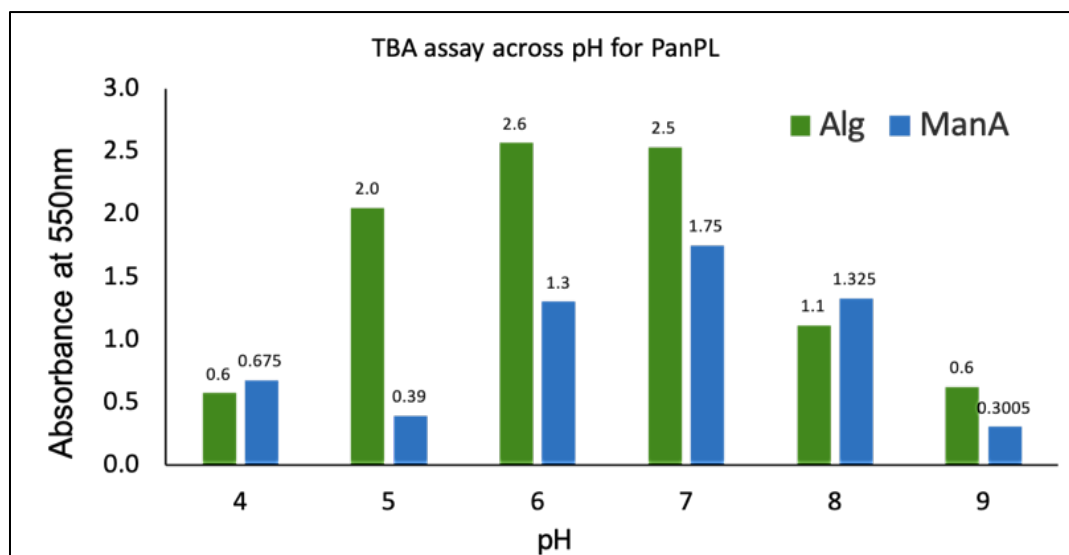
### **3.1. Introduction**

This chapter provides the structural-dynamics basis for the functioning of the PanPL while achieving the first three objectives defined for the thesis work. We found that PanPL is a novel PL-5 enzyme with unique biochemical properties. We have described the substrate specificity, enzyme kinetic behavior, and product analysis of PanPL in this chapter. The structural characterization for PanPL in its apo forms across the pH and substrate bound form has been performed and discussed in detail. The extensive studies for the dynamics of PanPL by the B-factor analysis, Normal mode analysis, Ensemble refinement analysis and Molecular dynamics simulation have been performed. The studies reveal distinct modes of hidden dynamics of PanPL.

### **3.2. Biochemical characterization reveals the allosteric nature of PanPL**

#### **3.2.1. PanPL works optimally around neutral pH**

The TBA assay (51) is the colorimetric functional assay for the polysaccharide lyase reaction. The unsaturated end products formed during the depolymerization forms a pink chromogen at the end of the assay, which confirms the lyase activity. We performed the TBA assay for PanPL with two substrates alginate and poly-ManA at different pH [Figure 3.1]. The TBA assay suggests the optimum pH for PanPL for both of its substrates to be pH 7.



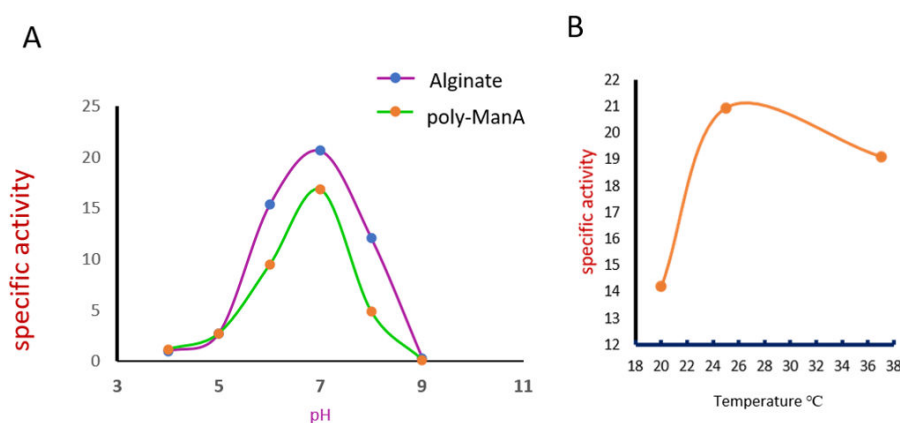
**Figure 3.1. TBA assay for PanPL across the pH spectrum.** The plot shows the absorbance at 550nm for TBA assay product. Both alginate (green) and poly-ManA (blue) was used for the assay. The number above the bar is the absorbance value.

### 3.2.2. PanPL activity fits non-linear kinetics with positive cooperativity

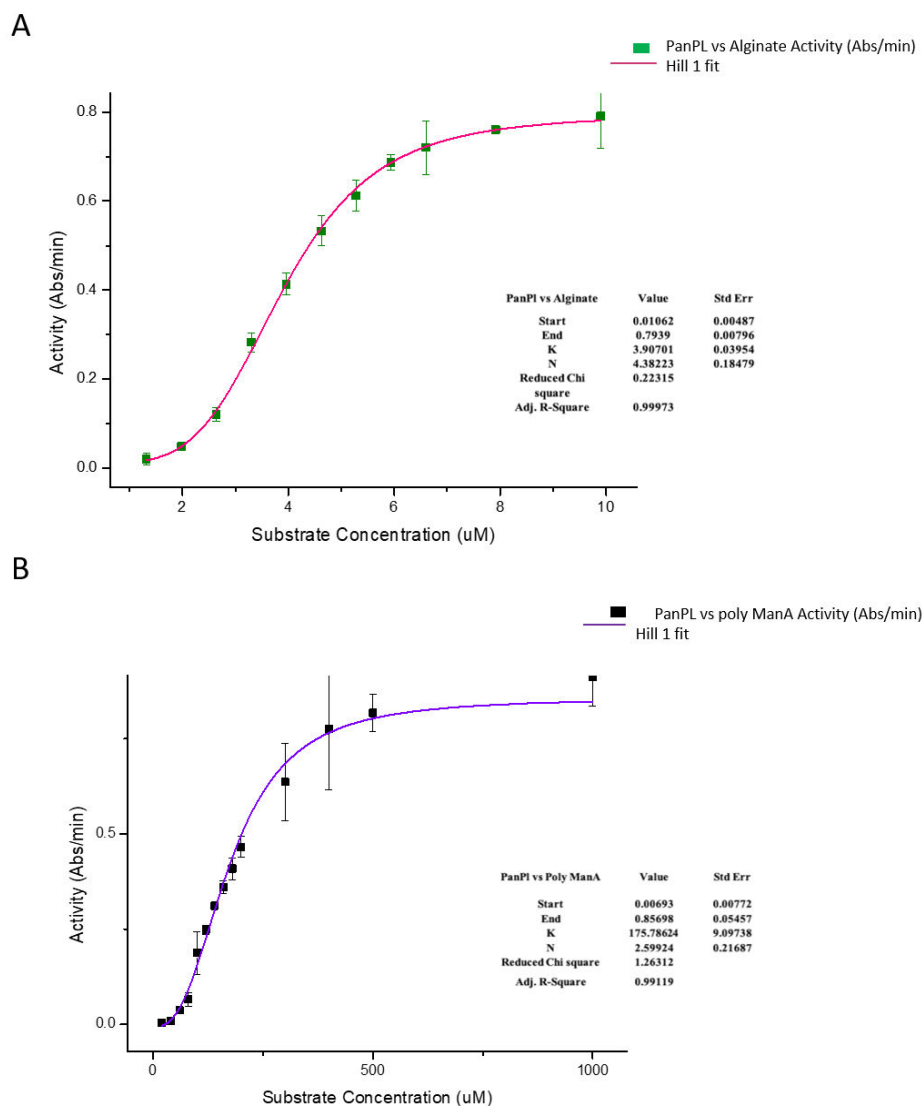
The unsaturated product formed in enzymatic degradation of polysaccharide via  $\beta$ -elimination reaction mechanism absorbs UV at 235nm. We recorded the 235nm UV absorbance to detect the products for both the alginate and poly-ManA substrates at different temperatures and across the pH spectrum [Figure 3.2]. Our results suggest that the optimum temperature to be 25°C and the optimum pH to be pH 7. Later, we measured 235nm UV absorbance at 25°C and pH 7 for a range of substrate concentrations and recorded the rates of unsaturated product formation. Enzyme activity at various substrate concentration was plotted using OriginLab software. While fitting the curve, we observed that the kinetic plots deviate from the traditional Michaelis-Menten kinetic behaviour. The enzyme activities increase slowly at low substrate concentrations, with rise in substrate concentration, the activities increase further and the curves assumed a sigmoidal shape [Figure 3.3]. We attempted to fit the curves by nonlinear curve fitting using the Hill's

equation by varying the Hill's coefficient ( $n$ ) [Figure 3.3]. For alginate and poly-ManA, the Hill's coefficient ( $n$ ) values were found to be 4.4 and 2.6, respectively ( $n > 1$ ). Here the enzyme kinetic patterns reflect the allosteric mechanism and show positive cooperativity. The allosteric behaviour observed in PanPL has never been reported in other PL-5 family enzymes, making it a distinct feature of PanPL (35).

Further, we used the data to calculate the enzyme kinetic parameters. One unit of activity was defined as an increase in absorbance of 0.1 unit per minute at 235nm (U). The rate of product formation was calculated using the product's molar extinction coefficient ( $6150 \text{ M}^{-1} \text{ cm}^{-1}$ ) (76). We discovered that the  $K_{\text{cat}}$  of the enzyme activities for alginate and poly-ManA are  $0.766 \text{ S}^{-1}$  and  $0.825 \text{ S}^{-1}$ , respectively. Furthermore,  $K_{\text{M}}$  values for alginate and poly-ManA substrates were calculated to be  $0.0039 \text{ mM}$  and  $0.175 \text{ mM}$ , respectively. The specific activities for poly-ManA and alginate were found to be  $8.56 \text{ mol min}^{-1} \text{ mg}^{-1}$  and  $7.94 \text{ mol min}^{-1} \text{ mg}^{-1}$  respectively. We used the alginate substrate, which has a degree of polymerization (dp) of 700, while the poly-ManA substrate has a dp of 25.



**Figure 3.2. The optimum pH and temperature of PanPL.** A. The optimum pH for both alginate (Alg) and poly-ManA substrate found to be pH 7. B. The optimum temperature has characterized to be 25°C.



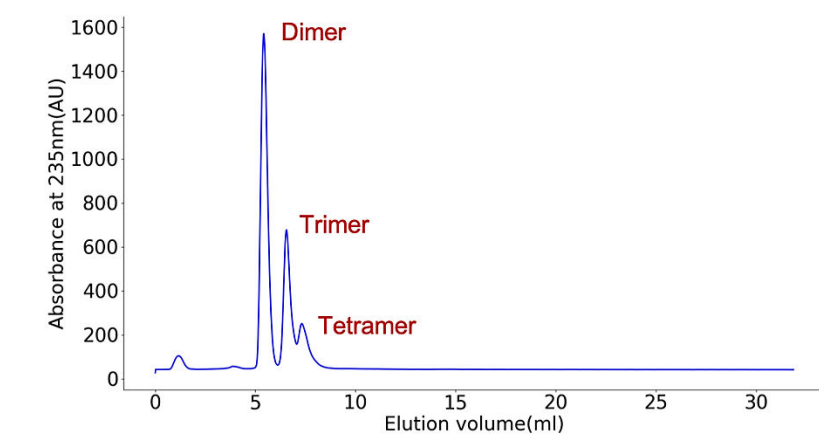
**Figure 3.3. Enzyme kinetic assay for PanPL.** The enzyme activity vs substrate concentration plot from enzyme kinetic assay for PanPL and its substrates; alginate and poly ManA. The sigmoid plot shows positive cooperativity for alginate and poly-ManA enzyme with standard deviations at each point. The equation used for the sigmoidal curve fitting was  $y = START + (END - START) \frac{x^n}{k^n + x^n}$ , which is a modified version of Hill equation, where  $n$  is hill's coefficient and  $k$  is  $K_M$ .

**Table 3.1. The enzyme kinetic parameters for PanPL for its substrates**

Substrates (dp)	Optimum pH	Enzyme concentration ( $\mu\text{g/ml}$ )	Hill's coefficient n	$K_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{M}}$ (mM)	$K_{\text{cat}}/K_{\text{M}}$ ( $\text{mM}^{-1}\text{s}^{-1}$ )
Poly-ManA (~25)	7	100	2.7	0.825	0.175	4.7
Sodium Alginate (~700)	7	100	4.4	0.766	0.0039	196

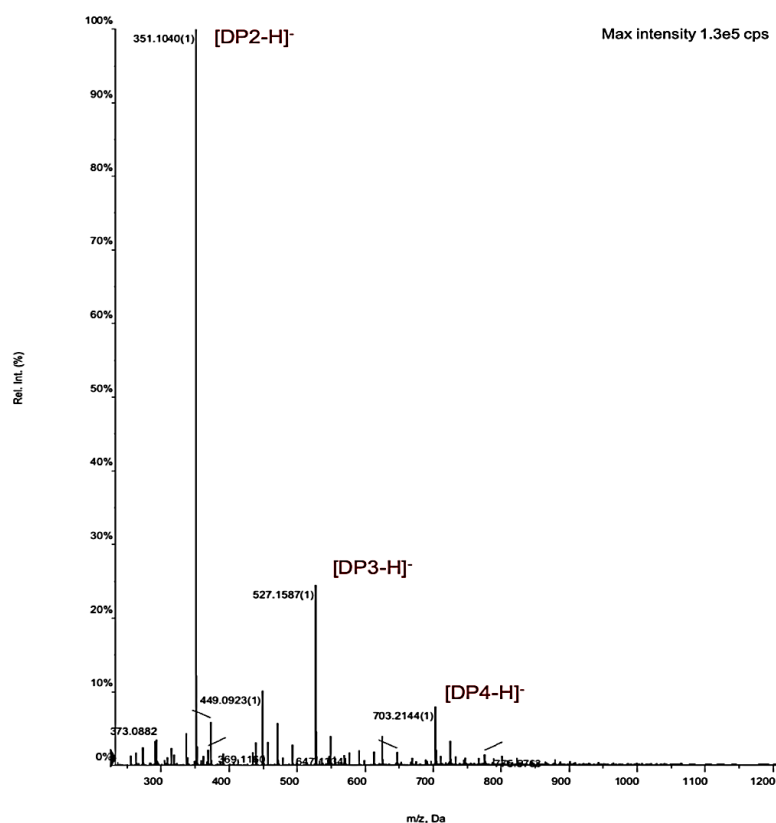
### 3.3. PanPL is endolytic lyase

The end products of the enzymatic degradation of polysaccharide substrates were separated by an anion exchange chromatography by monitoring UV absorbance at 235nm [Figure 3.4]. The charges on the oligomeric products are proportional to their length as each sugar unit carries unit negative charge on it. Both anion exchange chromatography [Figure 3.4] and mass spectrometry [Figure 3.5] confirm that the dimeric product is the most abundant among the unsaturated products. Thus, PanPL is characterized to be an endolytic alginate lyase.



**Figure 3.4. Product analysis by anion exchange chromatography.** The chromatogram shows alginate digestion product after 24 hours incubation with PanPL. Based on the charge on the oligomeric products, the dimeric unit eluted first as the major product followed by trimeric and tetrameric minor products.





**Figure 3.5. Mass Spectrometric analysis of cleaved products.** The mass spectrometric analysis of the enzymatic degradation products shows the dimeric unit as major peak (highest abundance). Trimer is the second major peak. The tetramer is the minor product.

### 3.4. Structural characterization of PanPL

The structural characterization of PanPL has been performed using X-ray crystallography. The crystal structures of PanPL have been solved at different pH in apo form [Table 3.2]. The substrate bound structure has been obtained by co-crystallization of substrate with inactive mutants of PanPL [Table 3.4].

Table 3.2. Crystallographic data collection and refinement statistics of PanPL apo structures across the pH

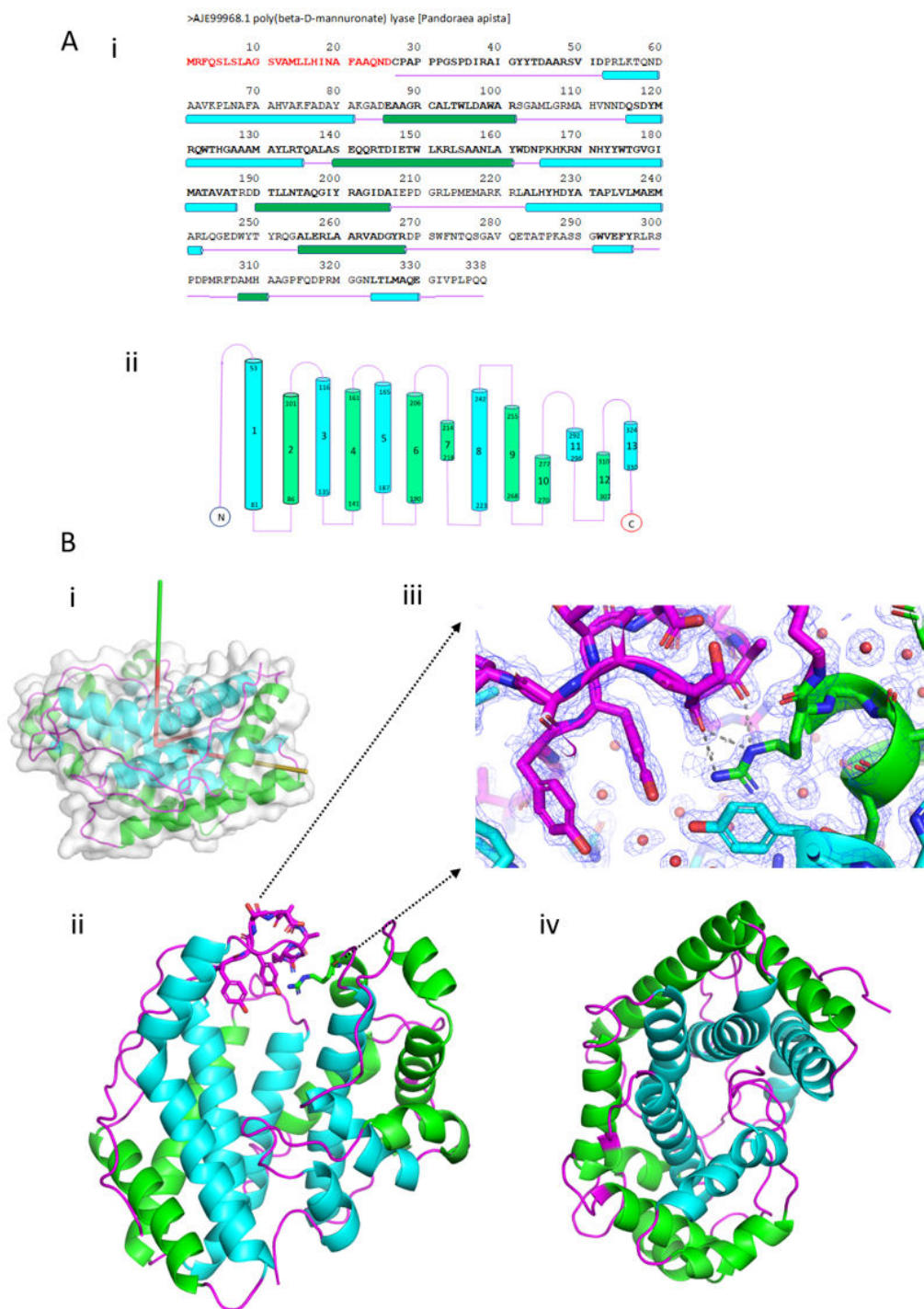
Crystal Name	PanPL_pH3.5	PanPL_pH4.5	PanPL_pH5.5	PanPL_pH6.5	PanPL_pH7.5	PanPL_pH8.5
PDB ID	7WXI	7WXX	7WXL	7WXM	7WYN	7WYO
Crystallization condition	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350	0.1 M Sodium acetate trihydrate pH 4.5, 25% w/v Polyethylene glycol 3,350	0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350	0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350	0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350
<b>Data collection statistics</b>						
Beamline	ID29	ID29	ID29	ID29	NISER Home source, BRUKER-PROTEUM	ID29
Wave length (Å)	1.07234	1.07234	1.07234	1.07234	1.54178	1.07234
Detector	Pilatus	Pilatus	Pilatus	Pilatus	Photon 100	Pilatus
Processing software	XDS/autoproc	XDS/autoproc	XDS/autoproc	XDS/autoproc	Protein 3	XDS/autoproc
Cell Parameter	c=64.59(Å) β=91.1(°)	c=98.89(Å)	c=91.16(Å)	c=91.37(Å)	a=36.03(Å), b=84.62(Å), c=91.30(Å)	c=90.82(Å)
Space group	C2	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution range (Å)	64.58-2.10(2.14-2.10)*	54.56-1.68(1.70-1.68)*	61.89-1.45(1.47-1.45)*	62.09-1.24(1.26-1.24)*	62.06-2.14(2.24-2.14)*	62.25-1.96(1.99-1.96)*
Rmerge	0.076(0.146)	0.104(0.823)	0.123(0.944)	0.054(0.658)	0.185(0.488)	0.102(0.810)
Unique reflections	15375(408)*	34187(1525)*	50270(2460)*	78113(2704)*	15538(1605)*	20836(1047)*
I/σ(I)	16.7(6.4)*	17.2(2.1)*	15.2(3.3)*	21.9(2.1)*	10.54(2.51)*	19.2(3.3)*
Completeness (%)	89.2(48.9)*	97.5(88.0)*	100(100)*	96.1(67.8)*	96.6(80.3)*	100(100)*
Redundancy	5.1(2.9)*	11.2(5.9)*	12.1(10.3)*	11.3(4.9)*	10.37(2.19)*	12.8(12.6)*
CC(1/2)	0.996(0.975)*	0.998(0.683)*	0.998(0.822)*	1.0(0.759)*	0.961(0.650)*	0.998(0.878)*
<b>Data Refinement Statistics</b>						
Refinement	phenix	phenix	phenix	phenix	phenix	phenix
Rwork/Rfree	0.1893/0.2297	0.2122/0.2554	0.1680/0.1903	0.1697/0.1860	0.2449/0.2854	0.1959/0.2325
ΔR=[Rwork-Rfree]	0.0404	0.0432	0.0223	0.0163	0.0405	0.0366
Matthews Coefficient (Å <sup>3</sup> /I)	2.11	2.13	1.99	2	1.99	2
Number of molecules in asym	1	1	1	1	1	1
Solvent content (%)	41.8	42.3	38.1	38.5	38.4	38.5
<b>Number of atoms</b>						
Proteins	2431	2405	2413	2442	2409	2407
Water and other (ligands or ions)	186	278	322	276	5	121
<b>Overall B-factor (Å<sup>2</sup>)</b>						
Proteins	16.0	17.6	11.0	15.0	11.3	24.9
Ligands or ions	19.4	25.4	19.9	22.5	3.0	26.9
<b>RMSD from ideal values</b>						
rms bond length (Å)	0.002	0.006	0.006	0.009	0.004	0.003
rms bond angle (°)	0.49	0.85	0.81	1.06	0.85	0.66
<b>Ramachandran plot statistics</b>						
Favored (%)	96.8	96.8	97.4	97.4	96.4	97.1
Allowed (%)	3.2	3.2	2.6	2.6	3.6	2.9
Outlier (%)	0	0	0	0	0	0

R<sub>work</sub>=Σ||F<sub>o</sub>|-|F<sub>c</sub>||/Σ|F<sub>o</sub>| R<sub>free</sub> is the R<sub>work</sub> value for 5% of the reflections excluded from the refinement.R<sub>merge</sub>=Σ|I-<I>|/ΣI

\* Values in parentheses are for the highest resolution shell

### 3.4.1. The overall structure of PanPL forms a catalytic tunnel

We determined the crystal structures of PanPL across the pH spectrum to provide a structural basis for the functioning. PanPL folds with six  $\alpha$ -helices making the inner core, five  $\alpha$ -helices forming the outer surface, and connecting loops resulting in a  $(\alpha/\alpha)_5$  incomplete toroid fold. The topology diagram is shown in Figure 3.6 Aii. the inner helices are labelled  $H_{i1}$  to  $H_{i6}$ , the outer helices are labelled  $H_{o1}$  to  $H_{o5}$ , and the loops are labelled  $L_1$  to  $L_{15}$  [Table A1, Appendix section]. The N-terminal segment forms a loop ( $L_1$ ), and a portion of it partially wraps the cavity formed by the inner helices, forming a lid (aa 42-52). We denoted it as N-loop-lid [Figure Biii]. The longest helix i.e., the first inner helix ( $H_{i1}$ ) bends with angle of  $\sim 27^\circ$  at Val63, is followed by the N-loop-lid. The first outer helix  $H_{o1}$  orienting across both the inner helices  $H_{i1}$  and  $H_{i2}$  defines an antiparallel three-helix bundle. This configuration of three-helix bundle consisting of two consecutive inner helices and an outer helix, continues across the toroid. Inner helices  $H_{i1}$  to  $H_{i4}$  form an approximately parallel helical bundle ( $H_{ib}$ ), while  $H_{i5}$  and  $H_{i6}$  are short segments that pack at the  $H_{ib}$ 's N-terminus. The inner helices define the core of the toroid. The major loop segments protrude and pack on the surface, allowing the helices to pack optimally. A few loop sections insert into the cavities to fill in the gaps, and a few interact with one another. The N-loop-lid has hydrogen bond interactions with the Loop (aa 218-223) between R219 sidechain atoms (NE, NH2) and R48 and A47 backbone carbonyl oxygen atoms.  $L_{14}$  folds and fits snugly into the cavity formed by the inner helices, and packing is favored by hydrophobic and a few hydrogen bonds. The  $L_{15}$  plugs into the cavity formed at the C-terminal ends of the inner helical bundle and stabilizes the hydrophobic interactions and a hydrogen bond [Figure Bii]. The interface between inner helices  $H_{i1}$ – $H_{i6}$  and outer helices  $H_{o1}$ – $H_{o5}$  is dominated by hydrophobic interactions with a few electrostatic interactions. Figure 3.7A depicts the surface representation, which clearly shows the tunnel formation.

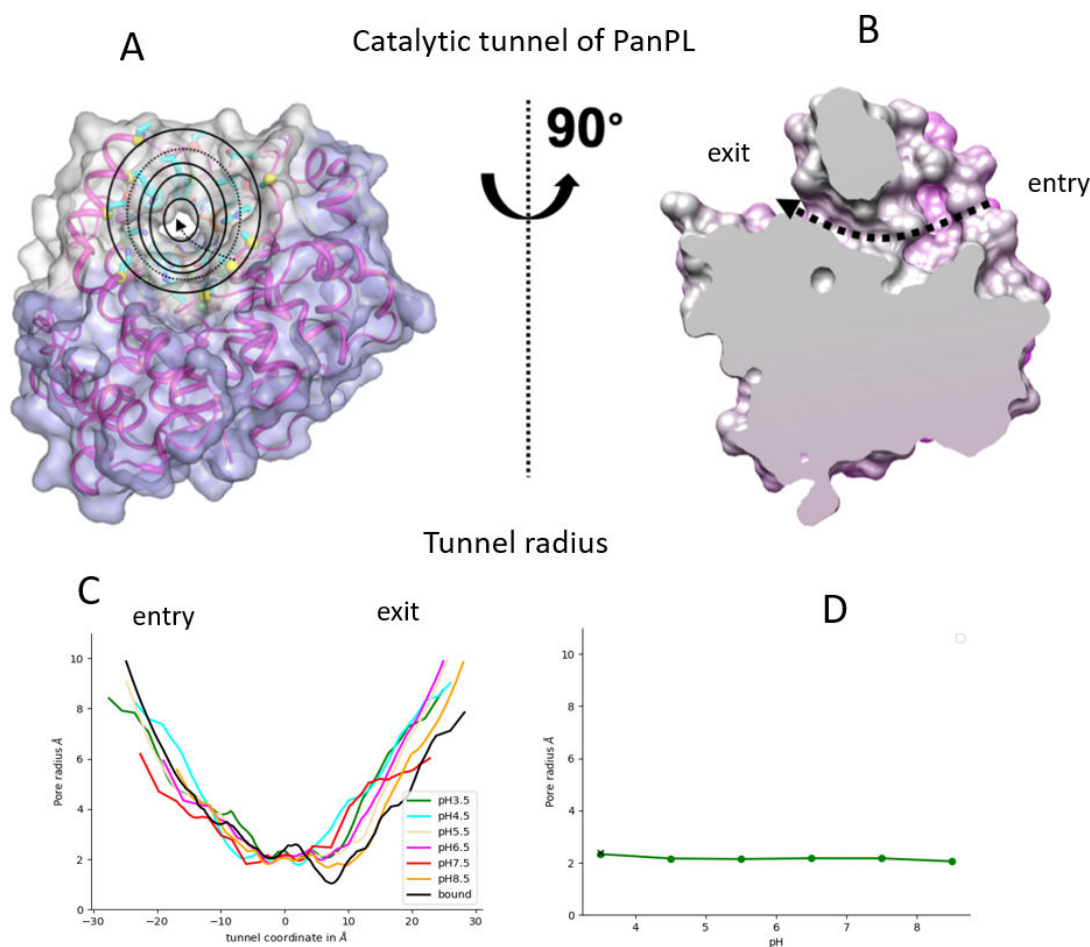


**Figure 3.6. Primary structure, topology, and overall structure of PanPL.** Ai. The sequence of the protein has been overlaid with the secondary structural elements. Aii. The topology of the PanPL shows the sequential arrangement of helices; H<sub>11</sub>-H<sub>16</sub> and H<sub>01</sub>-H<sub>05</sub> connected through loops; L1 to L15. The helical and loop segment length are also

indicated. Bi. The three-dimensional structure of PanPL, when aligned with its principal axis along the Z-axis represents the toroidal shape. Bi, Bii, Biv. The overall fold appears to be  $(\alpha/\alpha)_5$  toroid. The inner helices are colored cyan, outer helices green, and inter-connecting loops magenta. Biii. The loop segment represented in stick mode is the N-lid-loop; the inset shows the N-lid-loop locked in position by the side-chain to main-chain hydrogen bonds.

### 3.4.2. Tunnel architecture

The catalytic tunnel is defined by the N-terminal lid, the segment [aa 53-67] of H<sub>i1</sub>, a part [114-115] of L<sub>2</sub> and [aa 116-123] of H<sub>i2</sub>, a part of [aa 168-170] of L<sub>3</sub> and [aa 171-175] of H<sub>i3</sub>, a segment [aa 215-219] that folds into a short  $\alpha$ -helix, a turn [aa 219-22], a segment [aa 225-229] of H<sub>i4</sub>, a part of [aa 282-289] L<sub>12</sub> and [aa 317-319] of L<sub>14</sub> [Table A1, Appendix section]. The tunnel is represented by five concentric shells [Figure 3.7.A]. The tunnel is wider at the first shell and narrows as it approaches the fifth shell, with a constriction ( $\sim 2.0$  Å) at the fourth shell [Figure 3.7.C]. The constriction is maintained across the pH spectrum [Figure 3.7.D]. The residues are located in each of the tunnel's shell points into the interior [Figure 3.8], and their physiochemical properties contribute to the tunnel's surface texture. Aromatic residues, such as W175, forms the floor, creating a mosaic texture. The entry site is continuous with the first and second shell, while the exit site is defined by the fifth shell and beyond. The active site residues are found in the third and fourth shells. The tunnel's wall defines the outline, and the residues protruding from the outline into the interior of the tunnel determine the shape and size, and aid the substrate recognition process.

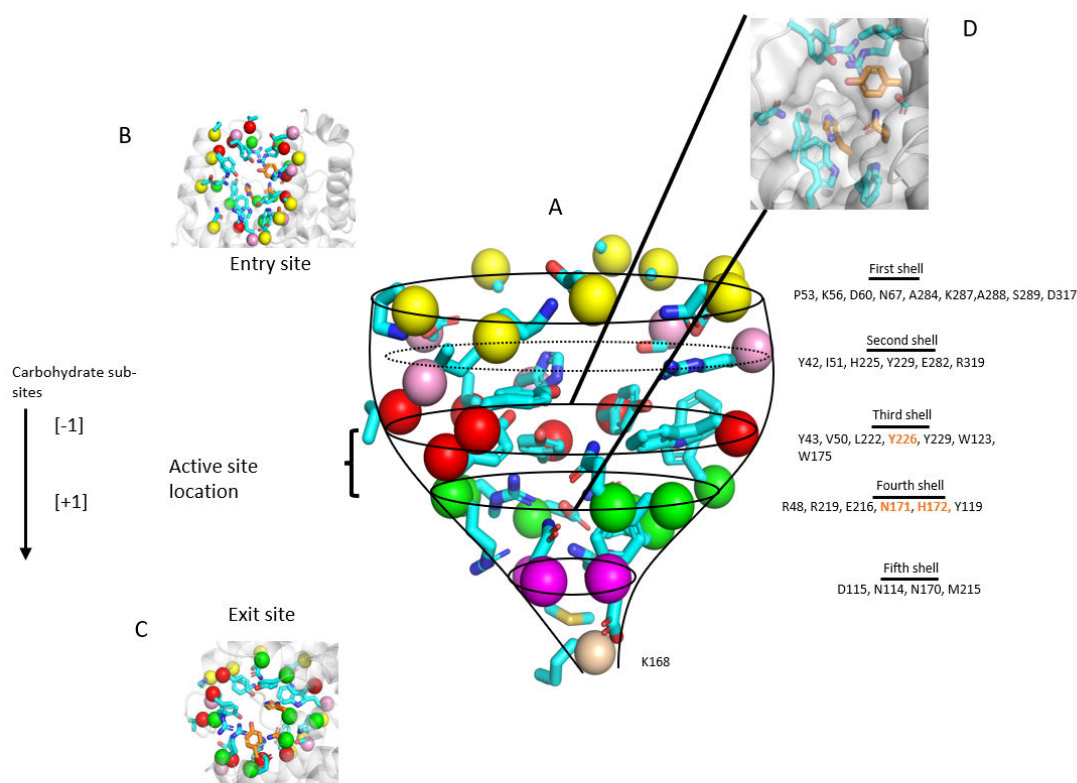


**Figure 3.7. Tunnel architecture: entry site, constriction region, exit site.** A. Tunnel formed can be visualized as a passage defined by concentric shells with decreasing radius, the arrow points from the entry site to exit site. B. The tunnel's side view defines a passage with a wide opening at the entry site following the constriction region and exit site. C. The tunnel radius for all pH crystal structures has been plotted from entry to exit site. D. The pore radius at the constriction remains constant as a function of pH.

### 3.4.3. Location of Active site residues

According to the tunnel architecture, the active site residues N171, H172, and Y226 are located in the fourth and third shells, respectively [Figure 3.8A, D]. N171, H172 are located near the C-terminus of H<sub>i3</sub>, and Y226 is located near the N-terminus of H<sub>i4</sub>. Tunnel

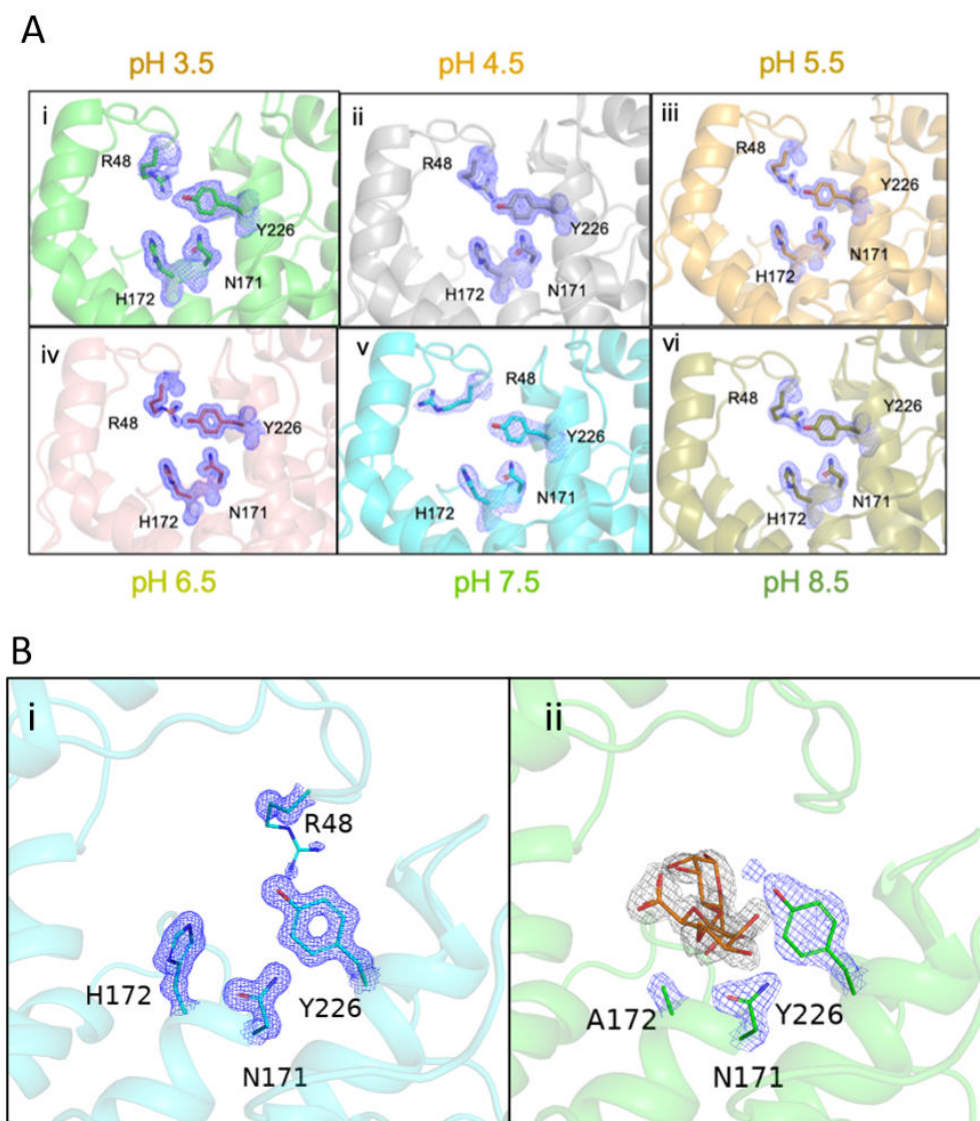
constriction is observed at the active site residue constellation  $\sim 2\text{\AA}$ . Given that activity varied as a function of pH, it was interesting to see whether the architecture of the tunnel varied as a function of pH. However, the tunnel architecture and constriction at active site remained constant across the pH spectrum in all crystal structures. Except for R48 on the exit side, the side chain conformations of flanking residues remain invariant across the pH spectrum [Figure 3.9]. This motivated us to conduct detailed structural analyses.



**Figure 3.8. Amino acid distribution and active site location in the tunnel.** A. The catalytic tunnel resembles the wine glass and can be constructed using the concentric shells with decreasing radius: The first shell to the fifth shell. Each shell is considered when we trace C $\alpha$  atoms on the circumference. B. The first shell has the largest radius and favours the polar residues, while the second shell circumference with the polar and aromatic residues. D. It's the third and fourth shell active site residues (Y226, N171, and H172) reside. C. The fifth shell is favored by hydrophilic residues, which lie closer to the exit site.



The active site residues are represented in stick mode with orange color. The carbohydrate sub-sites [-1] and [+1] are also depicted and correspond to the third and fourth shell.



**Figure 3.9. The conformations of active site residues and R48 across the pH. A.** The active site residues and the R48 residue has been shown in stick model with the electron density map at 1 sigma level. The active site residues don't show any structural change as a function of pH. The R48 residue at the exit site shows rotameric transition as a function of pH. **B.** The active site residues at PanPL pH 6.5 structure (i) and Tetra- ManA PanPL H172



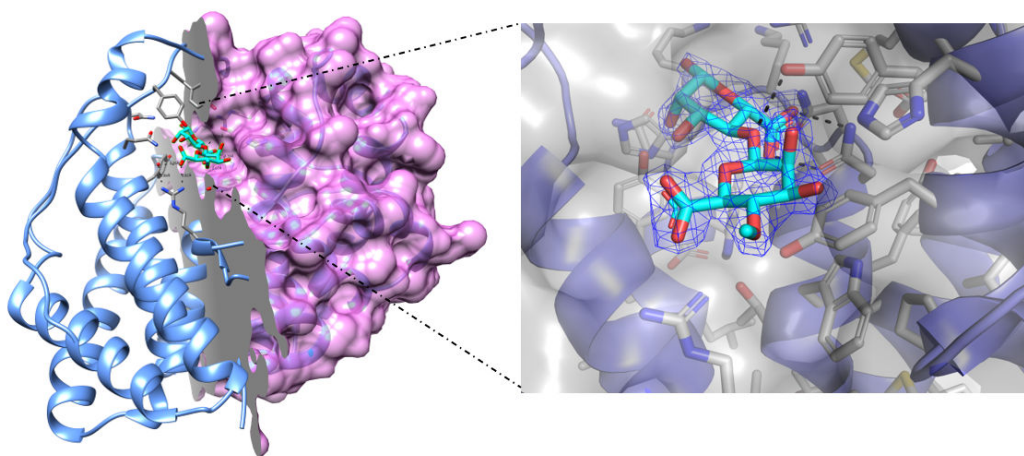
mutant structure (ii) has been shown at 1.5 sigma level show no structural changes. However, the electron density for R48 for substrate bound structure (ii) was diffused.

#### 3.4.4. Enzyme Substrate interaction: PanPL-tetra ManA complex crystal structure

After determining the apo structure details, we proceeded to characterize the substrate-bound structure with the intention to visualize how the substrate fits into the narrow active site region. Further this would describe the possible side-chain configurations at the tunnel, as well as their interactions with active site residues and the rest of the tunnel residues. We created active site mutants (Y226F, H172A, and N171L) to co-crystallize the substrate (tetra-ManA) bound protein and discovered that the H172A mutant could trap the substrate in the crystal structure [Figure 3.10].

The crystal structure of the PanPL\_H172A and tetra-ManA complex was determined at 2.2Å resolution. The statistics for data collection and refinement are shown in Table 3.4. We could only resolve two sugar units in the electron density map ( $F_o - F_c$ ) while building the model [Figure 3.13]. Our structural analyses revealed several intriguing features. To begin with, the tunnel constriction [Figure 3.7C], and the active site residues conformations remain invariant in all apo structures and substrate bound structure [Figure 3.9 A, B] This crystal structure of PanPL in the substrate bound state catalogues the amino acid residues interacting with the substrate in the catalytic tunnel [Table 3.3]. Although the H-bond interactions dominate, the distribution of aromatic residues in the tunnel also helps in optimal substrate positioning, binding, and processing [Figure 3.12]. Proper substrate orientation and binding are critical for catalysis. The sugar units at [+1] and [-1] subsites are held together by a dense network of H-bonds [Figure 3.11]. The first sugar subunit of the substrate is located at the [+1] subsite in the region between 3<sup>rd</sup> and 4<sup>th</sup> shells and interacts with the catalytic residue N171 as well as other residues such as R219 and Q116. Another catalytic residue Y226 residing in 3<sup>rd</sup> shell interacts with the O4 atom of the

glycosidic bond as well as with the O8 atom of sugar unit at the [-1] subsite via the H-Bond [Figure 3.11]. Y42 in the second shell interacts with the sugar subunits at [+1] and [-1] subsites. The second sugar unit at the [-1] subsite has more interactions with second shell residues such as H225 and R319. We couldn't model the side chain because we saw a diffused electron density for R48 residue in the catalytic tunnel. The average electron density was found to be diffused at the N-loop-lid in substrate bound structures. Furthermore, we observed a partial electron density for the 3<sup>rd</sup> sugar unit at [-2] subsite [Figure 3.13]. There was no electron density observed for the 4<sup>th</sup> sugar unit at [-3] subsite. Therefore, we were unable to build the model for the last two sugar units of tetra mannuronic acid. The fewer interactions with the tunnel residues are suggested to compromise their steady positioning in the tunnel.

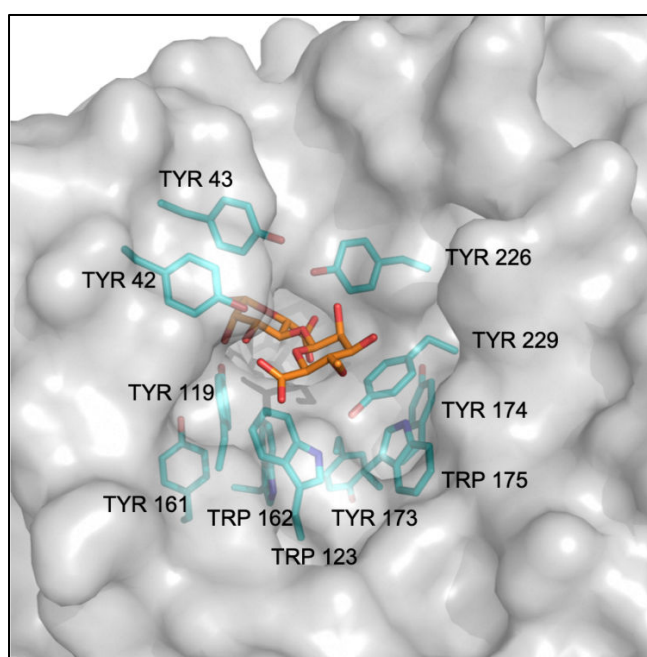


**Figure 3.10. The substrate in the tunnel view.** The substrate-bound structure shows the substrate being snugly fit into the tunnel cavity. The  $2F_o - F_c$  electron density map quality for the substrate contoured at 1.2 sigma level is also shown.



**Table 3.3. The list of interactions found in enzyme-substrate (PanPL H172A and tetra-ManA) complex crystal structure**

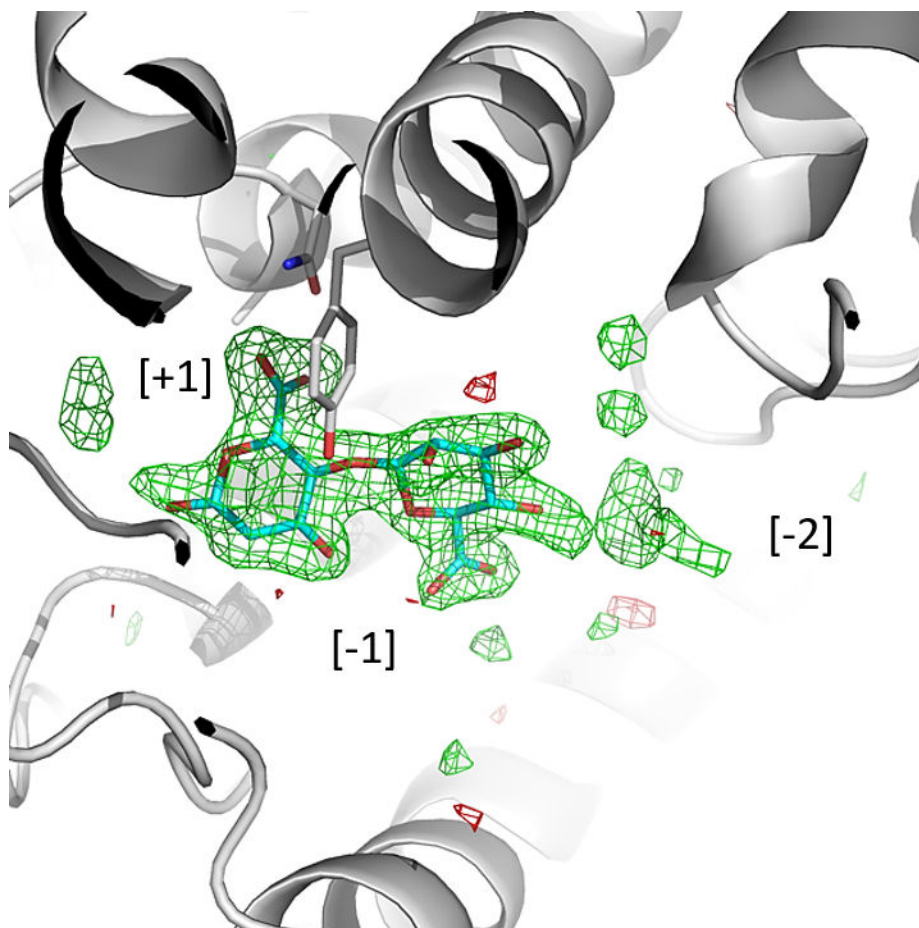
ATOM:Residue	ATOM: Substrate	Subsite	Interaction Type	Distance (Å)
ND2:Asn171	O6A:BEM1	[+1]	H-BOND	2.8
NH1:Arg219	O6A:BEM1	[+1]	H-BOND	3.1
OE1:Gln116	O2:BEM1	[+1]	H-BOND	2.8
NE2:Gln116	O3:BEM1	[+1]	H-BOND	2.9
NE2:Gln116	O2:BEM1	[+1]	H-BOND	3.2
OH:Tyr42	O3:BEM1	[+1]	H-BOND	3.1
OH:Tyr226	O4:BEM1	[+1]	H-BOND	2.8
O:HOH298	O1:BEM1	[+1]	H-BOND	3.1
O:HOH298	O5:BEM1	[+1]	H-BOND	2.7
O:HOH298	O6A:BEM1	[+1]	H-BOND	3.6
OH:Tyr226	O2:BEM2	[-1]	H-BOND	3.0
NE2:His225	O2:BEM2	[-1]	H-BOND	2.9
NE2:His225	O3:BEM2	[-1]	H-BOND	3.2
OH:Tyr42	O6B:BEM2	[-1]	H-BOND	2.9
NH1:Arg319	O6A:BEM2	[-1]	H-BOND	2.9
NH2:Arg319	O6A:BEM2	[-1]	H-BOND	3.5
O:HOH282	O6B:BEM2	[-1]	H-BOND	3.5
O:HOH220	O3:BEM2	[-1]	H-BOND	3.6



**Figure 3.12. The mosaic floor formed by aromatic residues.** The catalytic tunnel has aromatic residues on its floor. The aromatic residues and the substrate (ManA) are shown in the catalytic tunnel.

Table 3.4. Crystallographic data collection and refinement statistics of PanPL active site mutants in apo and substrate bound forms

Crystal Name PDB ID	PanPL_N171L(tetra-ManA co-crystallisation) 7WXR	PanPL_N171L(apo) 7XTE	PanPL_Y226F(apo) 7XTF	PanPL_H172A(apo) 7WXQ	PanPL_H172A + tetra-ManA 7WXP
Crystallization condition	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350	0.1 M Sodium acetate trihydrate pH 5.5, 25% w/v Polyethylene glycol 3,350	0.1 M Sodium acetate trihydrate pH 5.5, 25% w/v Polyethylene glycol 3,350	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350
<b>Data collection statistics</b>					
Beamline	NISER Home source, BRUKER-PROTEUM	NISER Home source, BRUKER-PROTEUM	ESRF	NISER Home source, BRUKER-PROTEUM	NISER Home source, BRUKER-PROTEUM
Wave length (Å)	1.54178	1.54178	0.96777	1.54178	1.54178
Detector	Photon 100	Photon 100	Eiger X 4M	Photon 100	Photon 100
Processing software	Proteum 3	Proteum 3	XDS/autoproc	Proteum 3	Proteum 3
Cell Parameter	a=46.95(Å), b=52.86(Å), c=119.27(Å)	β=90.6(°)	c=91.57(Å)	a=35.92(Å), b=84.25(Å), c=91.78(Å)	a=46.65(Å), b=53.01(Å), c=121.41(Å)
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C2	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution range (Å)	59.64-1.88(1.98-1.88)*	49.86-2.30(2.40-2.30)*	45.78-1.73(1.76-1.73)*	62.07-2.03(2.13-2.03)*	60.71-2.20(2.30-2.2)*
Rmerge	0.080(0.448)*	0.152(0.350)*	0.066(0.068)*	0.096(0.278)*	0.058(0.180)*
Unique reflections	24617(3190)*	13303(1572)*	30829(1551)*	18270(2227)*	15844(1839)*
I/σ(I)	27.9(4.6)*	8.1(2.0)*	14.0(2.2)*	18.6(4.4)*	26.84(5.61)*
Completeness (%)	98.9(92.2)*	99.8(99.9)*	89.8(96.4)*	97.8(92.7)*	99.2(94.6)*
Redundancy	14.42(4.28)*	4.96(2.18)*	4.3(4.6)*	13.25(5.06)*	9.86(2.55)*
CC(1/2)	0.998(0.727)*	0.982(0.747)*	0.997(0.793)*	0.998(0.946)*	0.998(0.948)*
<b>Refinement Statistics</b>					
Refinement	phenix	phenix	phenix	phenix	phenix
Rwork/Rfree	0.1798/0.2258	0.2425/0.2808	0.1797/0.2221	0.1917/0.2396	0.1695/0.2171
ΔR=[Rwork-Rfree]	0.046	0.0438	0.0424	0.0485	0.0476
Matthews Coefficient	2.12	2.32	1.94	1.99	2.15
Number of molecules in asymm	1	1	1	1	1
Solvent content (%)	42.0	47.0	36.2	38.2	42.9
<b>Number of atoms</b>					
Proteins	2415	2403	2410	2411	2394
Water and other (ligands or ligands or ions)	268	71	262	276	221
<b>Overall B-factor (Å<sup>2</sup>)</b>					
Proteins	14.1	14.7	21.4	13.3	13.4
ligands or ions	19.2	12.7	29.6	15.0	19.7
Water			30.8	18.0	13.9
<b>RMSD from ideal values</b>					
rms bond length (Å)	0.006	0.003	0.009	0.004	0.009
rms bond angle (°)	0.78	0.66	0.89	0.59	0.91
<b>Ramachandran</b>					
Favored (%)	96.8	96.8	97.4	97.4	97.1
Allowed (%)	3.2	3.2	2.6	2.6	2.9
outlier (%)	0	0	0	0	0
R <sub>work</sub> =Σ  F <sub>o</sub>  - F <sub>c</sub>   /Σ F <sub>o</sub>        R <sub>free</sub> is the R <sub>work</sub> value for 5% of the reflections excluded from the refinement.					
R <sub>merge</sub> =Σ  I-⟨I⟩ /ΣI					
*Values in parentheses are for the highest resolution shell					



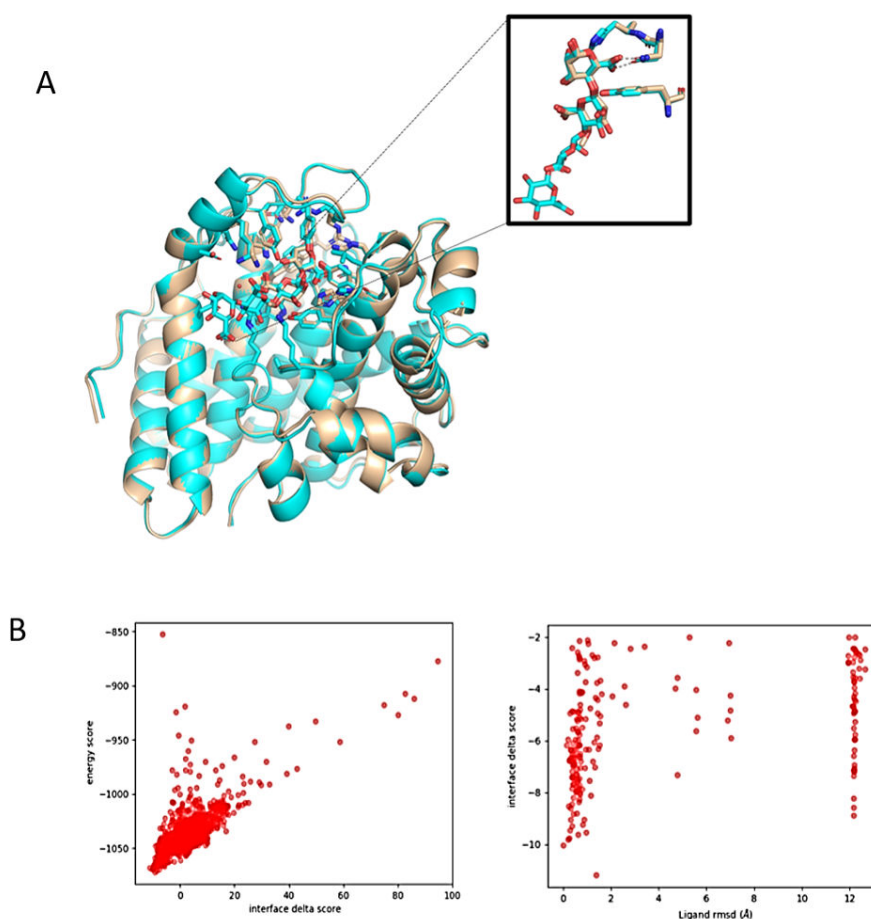
**Figure 3.13. Polder map for the ligand.** Figure shows the  $F_0-F_c$  polder map contoured at 3.0 sigma level which clearly indicates the trailing partial density for the next sugar atoms i.e., at [-2] sub-site. [+1] site is at the reducing end.

#### 3.4.5. Docking of PanPL with tetra-ManA

We used the Rosetta flexible docking protocol to perform docking of tetra-ManA with PanPL. For the docking studies, we used a pH 6.5 apo structure. We created 5000 docking models and sorted them by total energy (rosetta energy units) and chosen top 20% structures to sort the structure by interface energy [Figure 3.14]. We compared the top docked structure with the substrate orientation in the substrate-bound crystal structure and finalized the docked structure to validate the docking results. The docked structure revealed that K56 interacts with the third sugar unit at [-2] subsite, whereas no interaction was found with the 4<sup>th</sup> sugar unit at [-3] subsite [Figure 3.15, Table 3.5]. The lesser interactions at [-



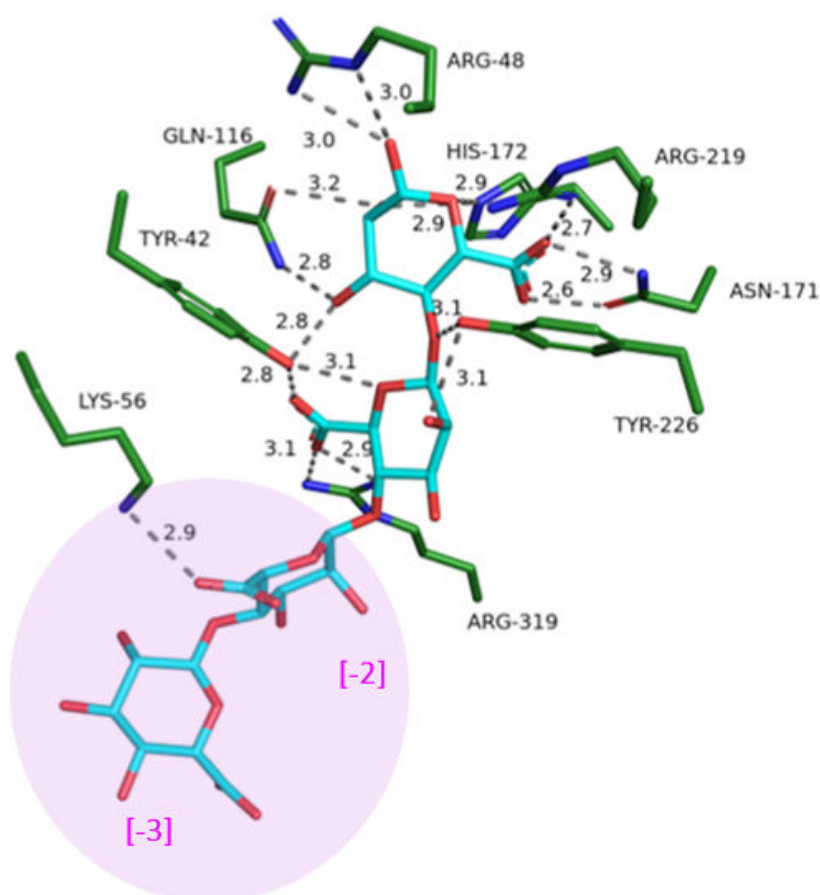
2] and [-3] subsites account for higher flexibility for 3<sup>rd</sup> and 4<sup>th</sup> sugar units resulting in diffused electron densities in the crystal structure. Similar observation was reported in the GH7 CBH US analysis where the first two subunits of the sugar chain encountered a greater number of H-bond interactions with lesser interactions at the above-mentioned ends(77).



**Figure 3.14. Docking of PanPL and tetra-ManA** The docking of PanPL and tetra ManA was performed using Rosetta flexible docking. The figure A shows the structural agreement between the PanPL H172A tetra-ManA bound crystal structure (wheat) and the docked structure (cyan). The two sugar units at [+1] and [-1] subsites are in same pose both in crystal structure and the docked structure. The plots B shows docking statistics, interface

delta score vs energy score and ligand r.m.s.d. (Å) vs interface delta scores respectively.

The interactions between tetra-ManA and active site residues in tunnel are shown.



**Figure 3.15. The enzyme substrate interactions in docked structure.** The polar interactions with the tunnel residues with the tetra ManA are shown in dashes labelled with distance in Å. The third sugar unit at [-2] subsites show an interaction with Lys 56. The fourth sugar unit has no interaction.



**Table 3.5. The list of interactions seen in PanPL and tetra-ManA docked structure**

ATOM:Residue	ATOM: Substrate	Subsite	Interaction Type	Distance (Å)
ND2:Asn171	O6A:BEM1	[+1]	H-BOND	2.9
NH1:Arg219	O6A:BEM1	[+1]	H-BOND	2.7
NE:Arg48	O1:BEM1	[+1]	H-BOND	3.0
NH2:Arg48	O1:BEM1	[+1]	H-BOND	3.0
OE1:Gln116	O2:BEM1	[+1]	H-BOND	3.2
NE2:Gln116	O3:BEM1	[+1]	H-BOND	2.8
NE2:Gln116	O2:BEM1	[+1]	H-BOND	3.6
OH:Tyr42	O3:BEM1	[+1]	H-BOND	2.8
OH:Tyr226	O4:BEM1	[+1]	H-BOND	3.1
OH:Tyr226	O2:BEM2	[-1]	H-BOND	3.1
NE2:His225	O2:BEM2	[-1]	H-BOND	2.9
NE2:His225	O3:BEM2	[-1]	H-BOND	3.1
OH:Tyr42	O5:BEM2	[-1]	H-BOND	3.1
OH:Tyr42	O6B:BEM2	[-1]	H-BOND	2.8
NH1:Arg319	O6A:BEM2	[-1]	H-BOND	2.9
NH2:Arg319	O6A:BEM2	[-1]	H-BOND	3.1
NZ:Lys56	O6A:BEM3	[-2]	H-BOND	2.9
NE2:His172	O5:BEM1	[+1]	H-BOND	2.9
NE2:His172	O2:BEM1	[+1]	H-BOND	2.9

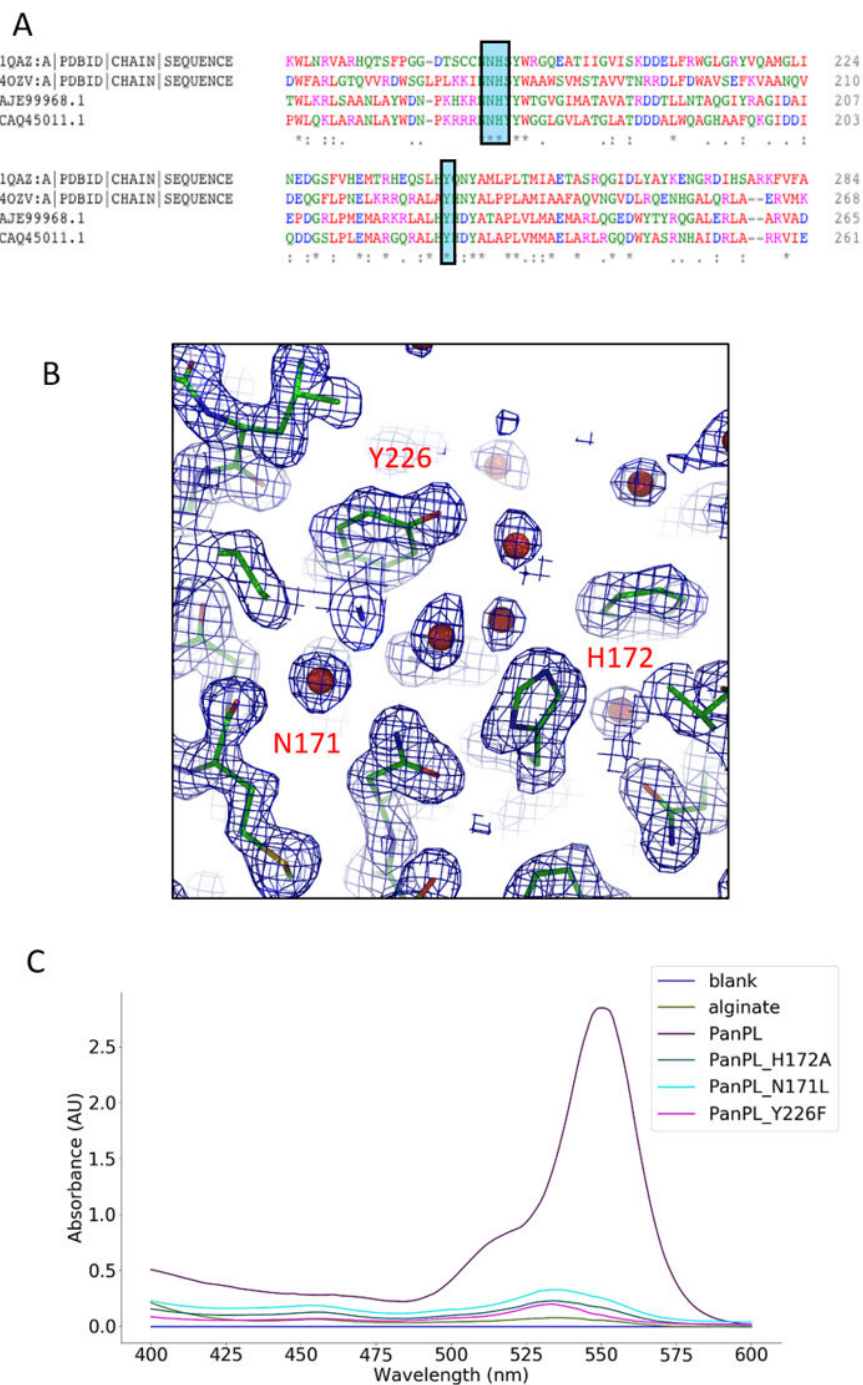
### 3.4.6. The role of active site residues in enzyme-substrate interaction

The characteristic fold of PanPL maintained across pH spectrum indicates a robust pH stable fold. The closed tunnel accommodates the active site of PanPL. Across the pH spectrum, the active site geometry remains unchanged [Figure 3.9]. To determine the substrate binding mode, we identified the conserved active site residues based on sequence analysis [Figure 3.16A] and created active site mutants N171L, Y226F and H172A. The loss of lyase activity for the mutants were recorded by the TBA assay [Figure 3.16C]. The inactive mutants were used for co-crystallization with the tetra-ManA substrate. The co-crystallization of tetra-ManA with N171L couldn't trap the substrate; however, we were able to capture the ligand at the active site for H172A mutant [Table 3.6]. The N171 side chain polar atom form hydrogen bond interaction with the carboxylate group of the sugar unit at [+1] subsite in the crystal structure of H172A and tetra-ManA complex. This indicates that N171L mutation with its hydrophobic side chain atoms couldn't provide the

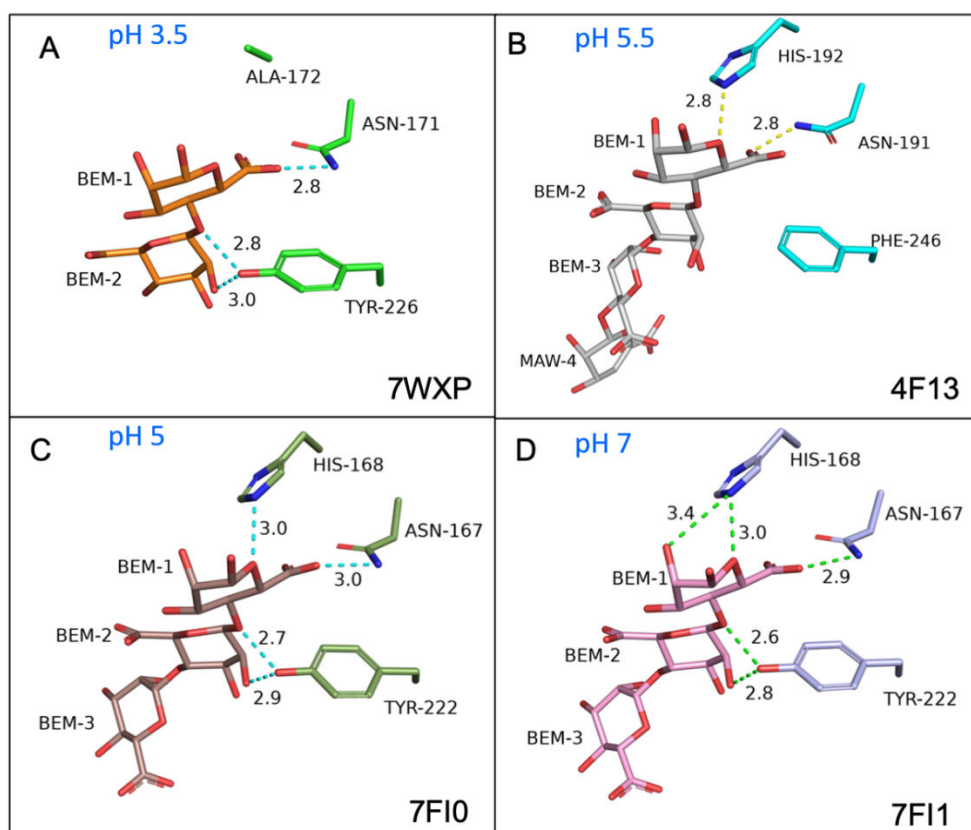
critical contact for substrate stabilization and positioning at the active site. Our understanding suggests that the active site residue N171 acts as the charge neutralizer and is important in substrate binding. It supposedly establishes the initial contact for the first sugar unit at the [+1] subsite. The substrate-bound H172A mutant crystal structure displays the binding pose of sugar subunits as well as the other interacting amino acids required to precisely orient the substrate. Despite the fact that the H172A being an inactive mutation, the role of H172 as a catalytic base cannot be established where the C5-H is inaccessible to H172. Here the substrate-bound crystal structure suggests a *syn*  $\beta$ -elimination reaction mechanism as the catalytic mechanism. Our docking and biochemical studies suggests the H172 residue to be a substrate stabilizer based on its interaction with the substrate. The crystal structure of the Smlt1473-tetraManA complex (7FI0) supports the similar kind of observation(49). We suggest that the H171 residue might act as a catalytic base in case of the guluronic acid, a C5 epimer of ManA, by an anti  $\beta$ -elimination reaction mechanism. As the mechanism favours *syn*  $\beta$ -elimination, the Y226 residue acts as both proton abstractor and proton donor for ManA substrate and the catalytically inactive Y226F mutant support this. Further, our analysis on PL-5 substrate bound structures suggest that the enzyme-substrate interaction pattern at the catalytic site remains conserved in PL-5 family at different pH [Figure 3.17].

**Table. 3.6. The active site residues and their putative role in catalysis**

Residues	Role in catalysis	Mutation	Inference from TBA assay of mutants
Asparagine N171	Charge neutralizer, Critical for substrate binding	N171L	Loss of functionality
Histidine H172	Substrate Stabilizer	H172A	Loss of functionality
Tyrosine Y226	Proton donor/ Proton Acceptor	Y226F	Loss of functionality



**Figure 3.16. The active site residues of PanPL.** A. The multiple sequence alignment shows the conserved catalytic residues. B. The electron density  $2F_o - F_c$  map contoured at 1.2 sigma level shows the active site of the PanPL in apo form. The water molecules are sitting nicely at the active site in the catalytic tunnel. C. The TBA assay shows loss of activity for active site mutants (C).



**Figure 3.17. Enzyme-substrate interactions at active sites in PL-5 proteins**

Figure shows the substrate interactions with the active site residues for different PL-5 enzymes at different crystallization conditions. Irrespective of pH the enzyme-substrate interactions are found to be conserved among PL-5 enzymes. The enzyme interacts with the reducing end of the substrate. 7WXP; PanPL active site residues with dimeric sugar unit at pH3.5 (tetra ManA was used for co-crystallization). 4F13; Alginate lyase A-III from *Sphingomonas sp.* complexed with tetra saccharide at pH 5.5 (Alginate lyase digested product was used for co-crystallization). 7FI0; Smlt1473 active site residues and three sugar unit interaction at pH 5 (tetra mannuronic acid was used for co-crystallization) 7FI1; Smlt1473 active site residues and three sugar unit interaction at pH 7 (tetra mannuronic acid was used for co-crystallization). Hydrogen-bond interactions are shown with dashed line labelled with distance in Å units.

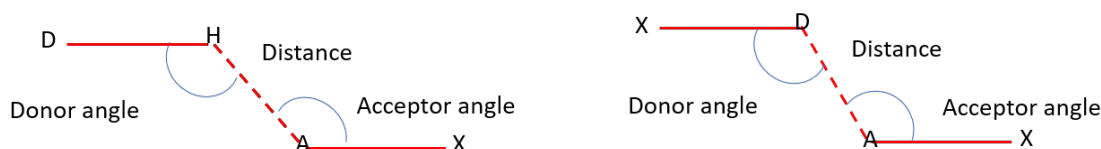
### 3.4.7. Crystal Packing Analysis

We observed only closed state in all crystal structures. It was necessary to check the effect of crystal packing on N-loop-lid flexibility. The crystal packing analysis was performed to learn more about the interactions between symmetry-related molecules on all the structures, especially on the N-loop-lid. We discovered that H-bond crystal contacts between symmetry related molecules at the end of N-terminal or C-terminal of the N-loop-lid in all crystal structures [Table 3.7]. For the H172A bound structure and the N171L and Y226F apo structures no crystal contacts were observed. Our analysis showed that most of the loop is solvent exposed in the lattice [Table 3.8]. This implies that the loop was not found to be restricted regardless of cell parameters or space group.

Geometric Parameters used to find the hydrogen bond(78):

$$D \cdots A < 3.5(\text{\AA}), D-H \cdots A > 90(^{\circ}), \text{ and } D \cdots A-X > 90(^{\circ}).$$

Where D is the donor atom, A is the Acceptor atom and H is the Hydrogen atom.



**Figure 3.18. Schematic representation of the geometry of a hydrogen bond.** On the left is the definition of geometry when proton positions are defined; on the right when they are not(78).

**Table 3.7. Hydrogen bonds observed at N-loop-lid (42-52: YYTDAARSVID) between the symmetry related molecules in crystal packing**

Donor (D) ATOM: RESI	Acceptor(A) ATOM: RESI	D ... A (Å)	X-D... A (°)	D... A-X (°)	SYMMETRY	pH	PanPL	PDB CODE
NZ: Lys166	O: Ile51	3.18	154.3	152.1	-X+1/2, Y+1/2, -Z	3.5	WT	7WXJ
OG1:Thr285	O: Thr44	2.83	112.4	135.3	X+1/2, -Y+1/2, -Z	4.5	WT	7WXK
NH2:Arg188	O: Thr44	3.51	111.8	110.0	-X, Y+1/2, -Z+1/2	8.5	WT	7WYO
NH2:Arg188	OD2:Asp52	3.16	132.7	119.2	-X, Y+1/2, -Z+1/2	5.5	WT	7WXL
NH2:Arg188	OD2:Asp52	3.00	142.0	137.3	-X, Y+1/2, -Z+1/2	6.5	WT	7WXM
NH2:Arg188	OD2:Asp52	3.54	127.6	107.3	-X, Y+1/2, -Z+1/2	7.5	WT	7WXN
NH2:Arg108	O: Ile51	3.44	114.4	139.8	-X, Y+1/2, -Z+1/2	3.5	N171L+4ManA*	7WXR
NH2:Arg188	OD2:Asp52	2.80	140.6	135.8	-X, Y+1/2, -Z+1/2	5.5	H172A	7WXQ
						3.5	N171L	7XTE
						5.5	Y226F	7XTF

N171L+4ManA\*: Co-crystallized with tetra-ManA but no trace of substrate in the electron density map.

**Table 3.8. Solvent interactions with polar atoms of N-loop lid (42-52: YYTDAARSVID)**

### 3.8.1. PanPL\_pH3.5: 7WXJ

Donor (D) ATOM: RESIDUE	Acceptor(A) ATOM: RESIDUE	D ... A (Å)	H ... A (Å)	D-H... A (°)	SYMMETRY
N: Tyr42	O: HOH201	3.2	2.4	157	X, Y, Z
N: Ala46	O: HOH170	2.9	2.1	157	X, Y, Z
N: Ala47	O: HOH133	2.9	2.1	152	X, Y, Z
O: HOH282	OH: Tyr42	3.0			X, Y, Z
O: HOH60	OD2:Asp52	3.1			-X+1/2, Y+1/2, -Z
O: HOH199	OD1: Asp52	2.8			-X+1/2, Y+1/2, -Z
O: HOH199	O: Gly313	3.1			X+1/2, Y+1/2, -Z
O: HOH60	O: Ala312	2.8			X+1/2, Y+1/2, -Z

## 3.8.2. PanPL\_pH4.5: 7WXX

Donor (D) ATOM: RESIDUE	Acceptor(A) ATOM: RESIDUE	D ... A (Å)	H ... A (Å)	D-H ... A (°)	SYMMETRY
N: Tyr42	O: HOH111	3.0	2.4	155	X, Y, Z
N: Tyr43	O: HOH335	3.0	2.1	165	X, Y, Z
N: Ala47	O: HOH301	3.0	2.2	140	X, Y, Z
NH2:Arg48	O: HOH128	3.1	2.2	168	X, Y, Z
NZ: Lys287	O: HOH197	3.0	2.3	145	O2
O: HOH197	O: Thr44	3.2			X, Y, Z
O: HOH302	OH: Tyr42	3.0			X, Y, Z
O: HOH366	O: Tyr43	3.2			X, Y, Z
O: HOH302	OH: Tyr43	3.2			X, Y, Z
O: HOH156	OG1: Thr44	2.8			X, Y, Z
O: HOH395	OG1: Thr44	2.5			X, Y, Z
O: HOH231	O: Ser49	2.8			X, Y, Z
O: HOH77	O: Ser49	2.9			X, Y, Z
O: HOH160	OD1: Asp52	2.8			X, Y, Z
O: HOH333	OD2: Asp52	2.5			X, Y, Z
O: HOH402	OD2: Asp45	2.6			X, Y, Z

## 3.8.3. PanPL\_pH5.5: 7WXL

Donor (D) ATOM: RESIDUE	Acceptor(A) ATOM: RESIDUE	D ... A (Å)	H ... A (Å)	D-H ... A (°)	SYMMETRY
N: Tyr42	O: HOH62	3.1	2.3	158	X, Y, Z
N: Ala46	O: HOH265	2.9	2.1	166	X, Y, Z
N: Ala47	O: HOH310	3.0	2.2	156	X, Y, Z
N: Ile51	O: HOH277	2.9	2.1	179	X, Y, Z
NH1:Arg48	O: HOH181	3.1	2.3	162	X, Y, Z
NH1:Arg48	O: HOH490	3.2	2.4	160	X, Y, Z
NH2:Arg48	O: HOH490	3.5	2.7	146	X, Y, Z
O: HOH240	O: Tyr43	3.2			X, Y, Z
O: HOH393	O: Tyr43	2.6			X, Y, Z
O: HOH240	O: Asp45	2.8			X, Y, Z
O: HOH304	OD2:Asp45	2.5			X, Y, Z
O: HOH310	OD2:Asp45	3.3			X, Y, Z
O: HOH187	O: Ser49	2.7			X, Y, Z
O: HOH130	OG: Ser49	2.8			X, Y, Z
O: HOH214	O: Ile51	2.8			X, Y, Z
O: HOH369	OD1: Asp52	2.7			X, Y, Z

## 3.8.4. PanPL\_pH6.5: 7WXM

Donor (D) ATOM: RESIDUE	Acceptor(A) ATOM: RESIDUE	D... A (Å)	H... A (Å)	D-H... A (°)	SYMMETRY
N: Tyr42	O: HOH92	3.1	2.3	157	X, Y, Z
N: Ala46	O: HOH239	2.9	2.1	169	X, Y, Z
NH2: Arg48	O: HOH140	3.0	2.1	176	X, Y, Z
N: Ile51	O: HOH184	3.0	2.1	177	X, Y, Z
O: HOH150	O: Ser49	2.8			X, Y, Z
O: HOH112	OG: Ser49	2.7			X, Y, Z
O: HOH309	O: Ile51	2.9			X, Y, Z

PanPL\_pH7.5: 7WXN: No H bond contacts found

PanPL\_pH8.5: 7WYO: No H bond contacts found

## 3.8.5. PanPL\_H172A+tetra-ManA: 7WXP

Donor (D) ATOM: RES	Acceptor(A) ATOM: RESI	D... A (Å)	H... A (Å)	D-H... A (°)	SYMMETRY
N: Ala46	O: HOH278	3.1	2.2	162	X, Y, Z

## 3.8.6. PanPL\_H17A: 7WXQ



Donor (D) ATOM: RESIDUE	Acceptor(A) ATOM: RESIDUE	D... A (Å)	H... A (Å)	D-H... A (°)	SYMMETRY
N: Tyr42	O: HOH67	3.0	2.2	154	X, Y, Z
N: Tyr43	O: HOH279	3.2	2.4	161	X, Y, Z
N: Ala46	O: HOH170	2.8	2.0	159	X, Y, Z
N: Ala47	O: HOH156	3.0	2.2	152	X, Y, Z
N: Ile51	O: HOH190	3.1	2.2	170	X, Y, Z
NH2:Arg48	O: HOH35	2.8	1.9	167	X, Y, Z
O: HOH124	OH: Tyr42	2.7			X, Y, Z
O: HOH287	OH: Tyr42	2.8			X, Y, Z
O: HOH140	O: Tyr43	3.3			X, Y, Z
O: HOH196	O: Thr44	2.9			X, Y, Z
O: HOH140	O: Asp45	2.8			X, Y, Z
O: HOH156	OD2: Asp45	3.5			X, Y, Z
O: HOH210	O: Ser49	2.7			X, Y, Z
O: HOH152	OG: Ser49	2.5			X, Y, Z
O: HOH159	O: Ile51	3.0			X, Y, Z
O: HOH198	O: Ile51	2.7			X, Y, Z
O: HOH134	OD1: Asp52	3.5			X, Y, Z
O: HOH351	O: Ala46	2.9			-X, Y+1/2, -Z+1/2
O: HOH365	O: Ala46	3.7			-X, Y+1/2, -Z+1/2
O: HOH185	OD2: Asp52	3.5			-X, Y+1/2, -Z+1/2

PanPL\_N171L (tetra-ManA Co-crystallized):7WXR: No H bond contacts found

### 3.8.7. PanPL\_N171L\_apo: 7XTE

Donor (D) ATOM: RESIDUE	Acceptor(A) ATOM: RESIDUE	D... A (Å)	H... A (Å)	D-H... A (°)	SYMMETRY
O: HOH67	N: Tyr43	3.0			X, Y, Z
O: HOH101	OD2: Asp52	3.1			X, Y, Z

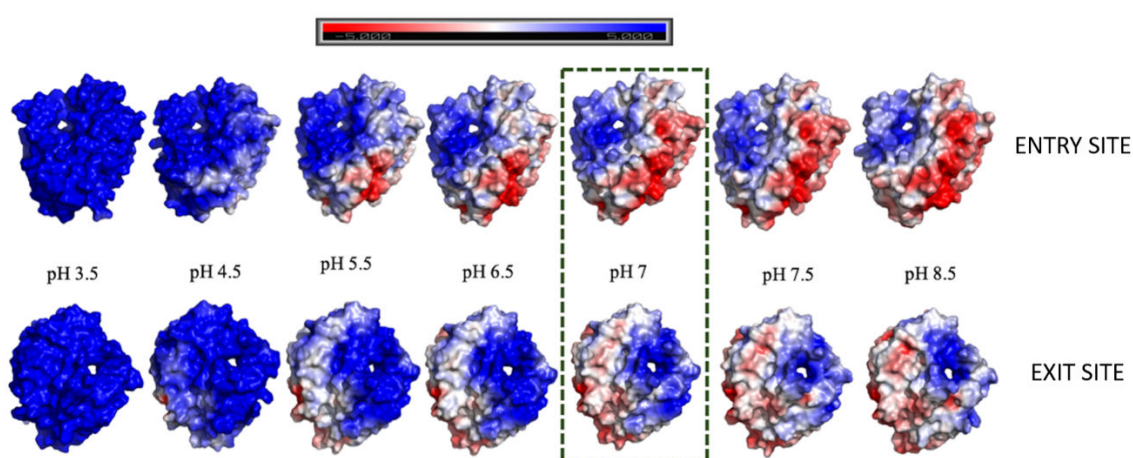
## 3.8.8. PanPL\_Y226F\_apo: 7XTF

Donor (D) ATOM: RES	Acceptor(A) ATOM: RESI	D... A (Å)	H... A (Å)	D-H... A (°)	SYMMETRY
N: Tyr42	O: HOH43	3.1	2.3	157	X, Y, Z
N: Ala46	O: HOH218	2.9	2.1	166	X, Y, Z
N: Ala47	O: HOH253	3.3	2.5	155	X, Y, Z
N: Ile51	O: HOH112	2.8	2.0	171	X, Y, Z
O: HOH246	OH: Tyr42	3.2			X, Y, Z
O: HOH246	OH: Tyr43	2.9			X, Y, Z
O: HOH281	OH: Tyr43	2.7			X, Y, Z
O: HOH189	O: Asp45	2.9			X, Y, Z
O: HOH253	OD2: Asp45	3.3			X, Y, Z
O: HOH335	OD2: Asp45	3.2			X, Y, Z
O: HOH85	O: Ser49	2.8			X, Y, Z
O: HOH23	OG: Ser49	2.8			X, Y, Z
O: HOH273	OD1: Asp52	2.8			X, Y, Z
O: HOH320	OD2: Asp52	3.2			X, Y, Z

### 3.5. The electrostatic surface charge distribution as a function of pH

We calculated the electrostatic surface charge distribution for PanPL apo structures across the pH spectrum, as the pH changes the ionization states [Figure 3.19]. The electrostatic models revealed that the surface charge was more electropositive at lower pH (pH 3.5 and 4.5) and the positive surface charge faded away at higher pH (pH 5.5 to 8.5), where electronegative surface charge increase [Figure 3.19]. To allow enzyme catalysis to occur, the anionic substrate must enter the tunnel precisely. The electrostatic surface charge helps to avoid nonspecific substrate binding in the catalytic activity range (pH 5.5 to 7.5). At pH 6.5, the electropositive surface charge becomes more confined around the catalytic tunnel shown as a confined blue patch. It helps the enzyme in attracting the anionic substrate optimally. Electrostatic interactions between the positively charged tunnel and the negatively charged polysaccharide substrate can result in the catalysis's substrate

acquisition step. Furthermore, the mosaic floor of aromatic residues [Figure 3.12] in the tunnel aids in the proper orientation of the sugar rings of the substrates. This is consistent with the findings of a study of Umbrella Sampling (US) simulations of Glycoside Hydrolase 7 cellobiohydrolase, which found that the strong electrostatic interaction between the substrate with tunnel polar residues is the driving force for cellulose chain processivity, where conserved aromatic residues in the tunnel also assist substrate processivity (77)



**Figure 3.19. Electrostatic surface charge visualization as a function of pH**

The electrostatic map of charged molecular surface of PanPL at different pH values (PDB2PQR generated by PROPKA (62,75) and visualized using APBS plug-in of pymol). At pH 3.5 and 4.5 the electropositive charge extended though out the surface, which might lead to nonspecific binding of anionic substrate. With increase in pH the electropositive blue patch on the molecular surface getting confined around the catalytic tunnel. At the optimum pH 7 the electrostatic map was calculated by using pH 6.5 structure. The electropositive patch guides the negatively charged substrate to bind specifically into the tunnel. At pH 7.5 and 8.5 the electropositive charge gets weaker to hold the substrate optimally, hence enzyme activity decreases.

### 3.6. Structural flexibility and the B-factor analyses as a function of pH

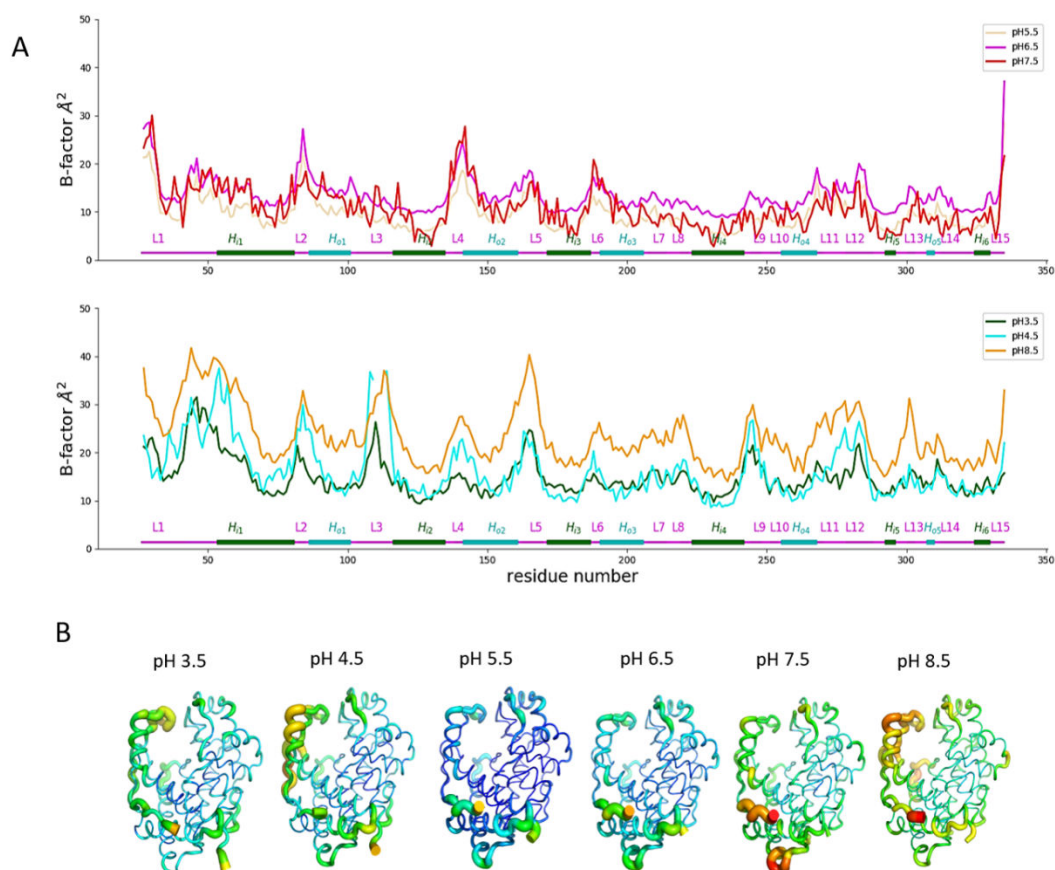
The average B-factor for the main chain atoms per residue was calculated for each apo structure, as the main chain atoms are responsible for structural flexibility. Since we observed the flexible regions with higher B-factors than the core of the protein, we identified the peaks as the flexible segments in the proteins. We sorted the B-factor plot into groups to see if there was a pattern among plots within a group [Figure 3.20A]. The average B-factor plots for the structures at pH 5.5, 6.5, and 7.5 had a similar trend falling into group one, while pH 3.5, 4.5, and 8.5 fell into group two with a similar B-factor pattern [Figure 3.20B]. The plot pattern in group one is similar but elevated to a different degree, with pH 7.5 having a higher elevation than pH 6.5 and pH 5.5, thus the elevation follows pH 7.5 > pH 6.5 > pH 5.5. Similarly, plots for structures at pH 3.5, pH 4.5, and pH 8.5 in group two show a similar trend with elevations following pH 8.5 > pH 4.5 > pH 3.5. A careful examination shows that L<sub>1</sub>-H<sub>i1</sub>, L<sub>3</sub>, L<sub>5</sub>, and L<sub>9</sub> peaks are especially prominent in the group two plots compared to group 1. Similarly, in group 1 plots, the peaks at L<sub>4</sub>, L<sub>6</sub>, and L<sub>13</sub> are more elevated than in group 2 plots. To make the raw average B-factor values independent of resolution effects and crystal packing, we calculated normalised B-factor values. Figure 3.21. depicts a plot of the normalised B-factor calculated for main-chain atoms per residue versus the number of residues. To identify the peaks corresponding to the flexible protein segments, we used a normalised B-factor threshold of 1.0 units in the plot. The plots confirm the previous categorization and reveal several interesting features. In comparison to group 2, the N-loop-lid and a portion of the first inner helix (H<sub>i1</sub>) fluctuate less, while L<sub>2</sub> and L<sub>4</sub> fluctuate more in group 1. Furthermore, in group 2, the N-loop-lid and a portion of the first inner helix, as well as segments in L<sub>3</sub> and L<sub>5</sub>, are more fluctuating.

Because the molecules in group 1 maintain higher activities (with optimal activity at pH 7.0), the pattern and strength of fluctuations observed in group 1 suggest the molecular dynamics that drive the molecule to be in functional states. Similarly, the signature and strength of fluctuations observed in group 2 are driving the molecules for the dynamics to reduce activity. Referring back to our earlier discussion of tunnel architecture, the strength of the fluctuations at the arc and left-wall of the tunnel (N-loop-lid and first inner helix ( $H_{i1}$ ), respectively), distal to the active site ( $L_2$ ,  $L_4$ ), third shell, and fourth shell of the tunnel ( $L_3$  and  $L_5$  respectively) are regulated by pH change, and thus determine the dynamics for the functional state. Our analysis suggests that the pH tunes the dynamics in the structures and sets them into different energy states. As a result, while some structures on the continuum makes the enzyme functionally active, others cause it to be non-functional/reduced activity. Furthermore, the fluctuations at the N-loop-lid and a portion of the first inner helix ( $H_{i1}$ ) are in sync with those at  $L_3$ , but anti-sync with those at  $L_2$  and  $L_4$ .

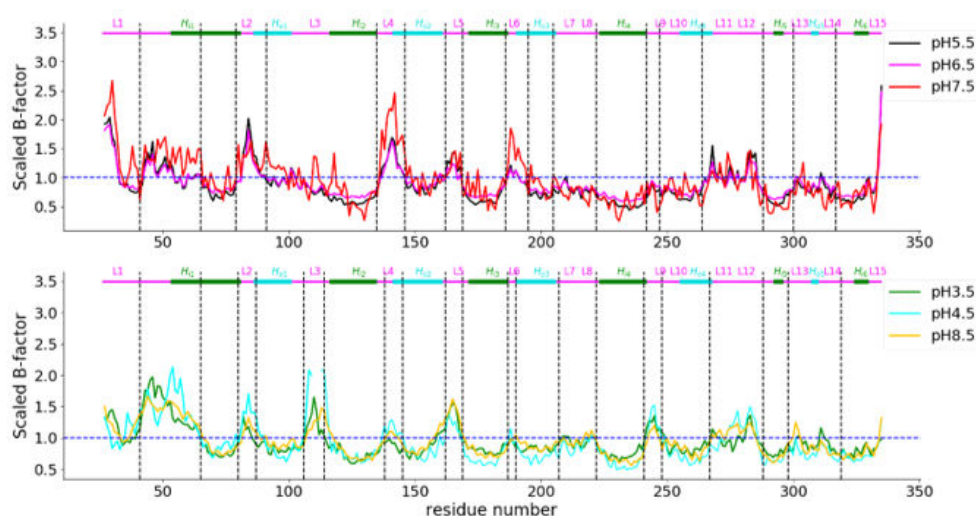
We examined the B-factor putty representation of all crystal structures in apo form at different pH levels and found distinct trends of fluctuations at the N-loop-lid as a function of pH [Figure 3.20 B]. The N-loop-lid vibrates more at the extreme ends (pH 3.5, 4.5 and 8.5) than its pH range of activity of the pH spectrum (5.5 - 7). It implies that neither too much nor too little vibration at the N-loop-lid aids in catalysis. Rather, an optimal degree of dynamics at pH 6.5 (close to optimal pH) ensures maximum catalytic activity.

The B-factor analysis of the active site mutant structures both in their apo, co-crystallized and substrate bound forms have been calculated. The residue wise normalized B-factor plots have been shown [Figure 3.23B] The plots show the mutant structures follow the similar trends of fluctuations as a function of pH with respect to PanPL wildtype structures. However, huge elevations in the N-loop-lid regions of the plot have been noticed in

H172A+tetra-ManA bound form and N171L+tetra-ManA co-crystallized structures as compared to the mutant apo structures. The B-factor putty representations help to visualize the huge vibration at the N-loop-lid of above two structures [Figure 3.23A]. Our study suggests the substrate induced the N-loop-lid vibration in the crystal structures.

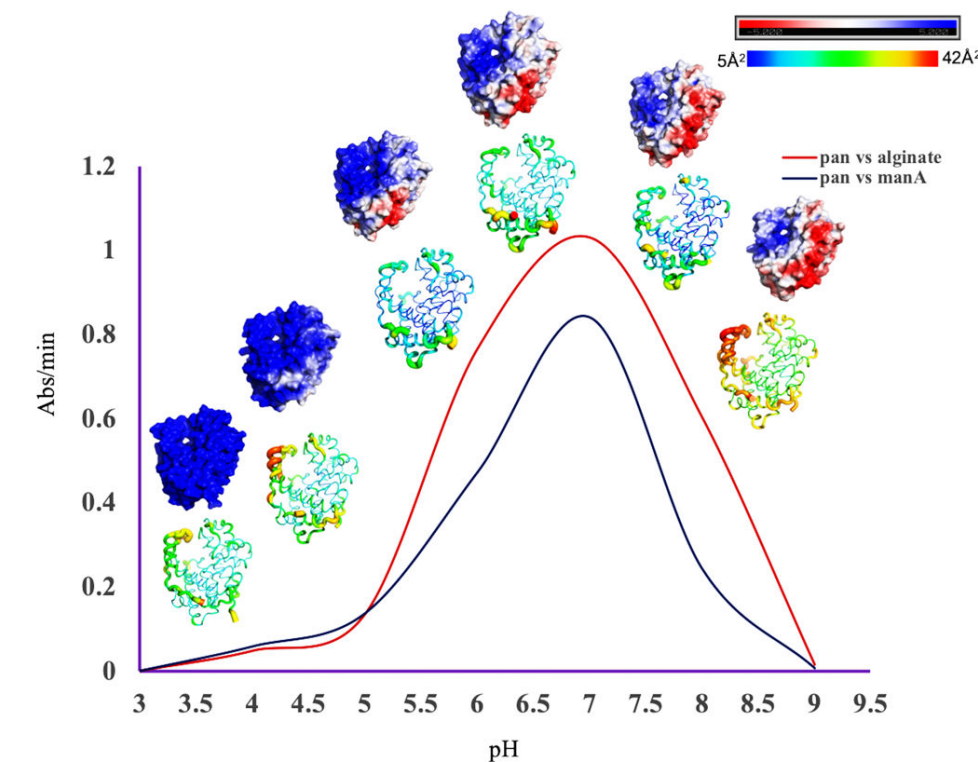


**Figure 3.20. Average B-factor fluctuations and B factor putty representation of PanPL across the pH spectrum.** A. The average B-factor fluctuation per residue for PanPL structures at different pH. The plots for different pH are grouped into two groups based on the fluctuation pattern observed at different regions of PanPL structure. The secondary structural components are represented on the bottom of each group. B. The B-factor putty representation of PanPL as a function of pH.



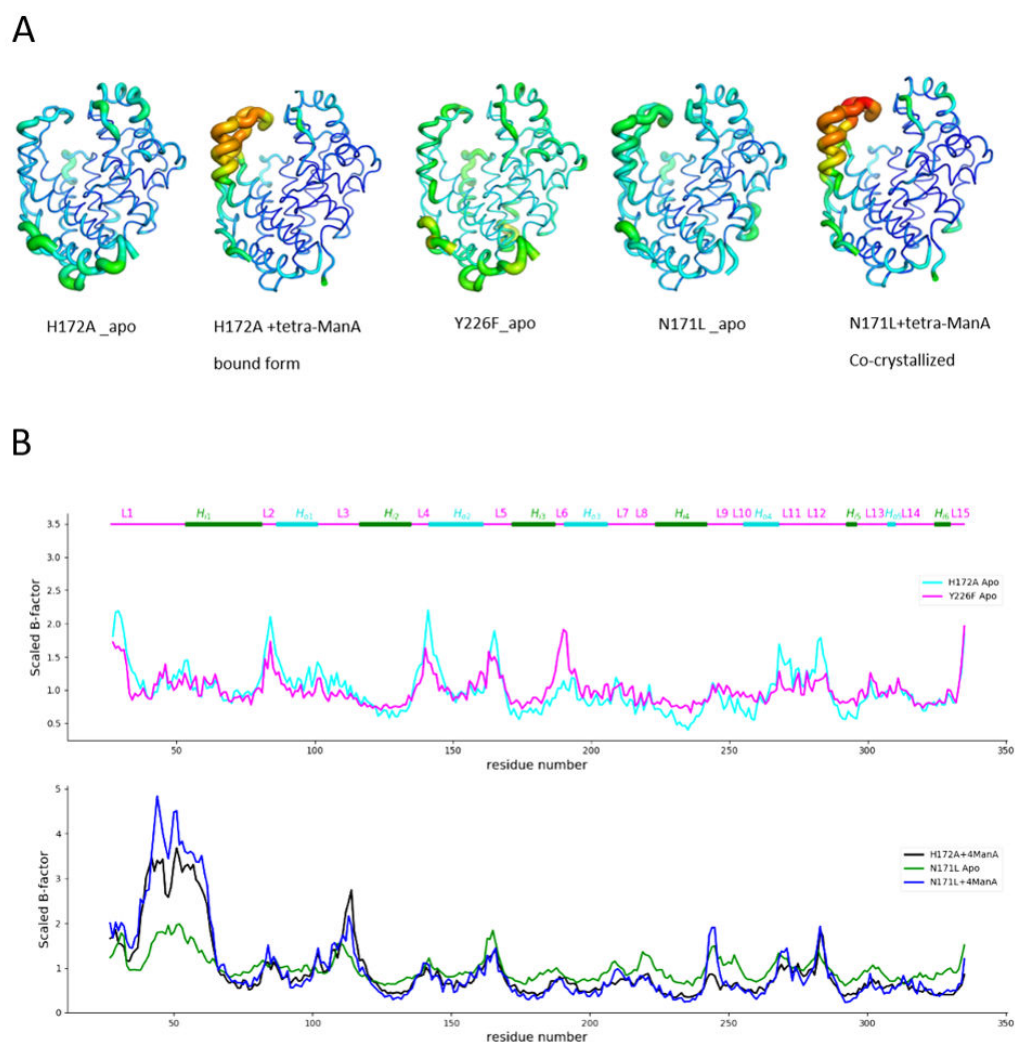
**Figure 3.21. Normalized B-factor fluctuations across the pH spectrum for PanPL.**

Normalized B-factor fluctuation per residue for PanPL structures at different pH. The plots for different pH are grouped into two groups based on the fluctuation pattern observed at different regions of PanPL structure. The secondary structural components are represented on the top of each group. The vertical dotted bars are used to show the vibrational hotspots distinctly. The fluctuation pattern for pH 3.5, 4.5 and 8.5 structures resemble and represent the low to no enzyme activity group. The pH 5.5, 6.5 and 7.5 structures show similar the fluctuation pattern and represent the range of enzyme activity group.



**Figure 3.22. The enzyme activity, the electrostatic surface charge and the B factor putty representation of PanPL as a function of pH.** The plot represents the enzyme activity as a function of pH, corresponding B-factor putty representation and electrostatic surface charge distribution of PanPL structures determined across the pH spectrum (3.5 to 8.5 in steps of 1.0 units). The PanPL shows higher enzyme activity in the pH range of 5.5 to 7.5, with optimal pH-7.0. The N-terminal lid loop vibrations at pH 3.5, 4.5, 8.5 are relatively higher in contrast to pH 5.5, 6.5, and 7.5. pH 5.5 has the least vibration across the pH spectrum. The comparison of fluctuations among different pH structures provides a basis of pH optimum of enzyme activity around pH 6.5~7.0. The electrostatic surface charge distribution shows at the range of pH 5.5 to 7.5 the blue patches (positive charge) around the catalytic tunnel are well defined, which might help anionic substrate to bind into the tunnel. However, at pH 3.5 and 4.5 the whole surface becomes positively charged and at pH 8.5 the negative charge dominates the surface, which might affect the substrate binding in the tunnel.

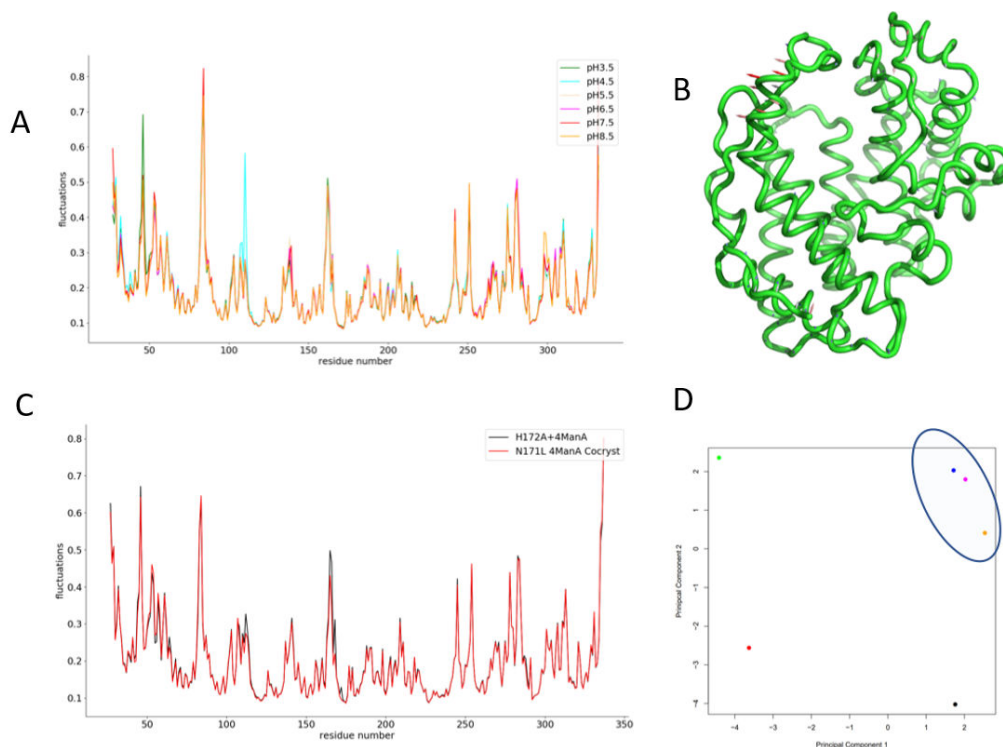




**Figure 3.23. B factor putty representation and the Normalised B factor plots for mutant structures.** Figure A shows the B -factor putty representation of PanPL active site mutants. The H172A+tetra-ManA bound form and N171L+tetra-ManA co-crystallized structures shows huge vibration at N-loop-lid. B. The residue wise normalized B-factor plot also shows enhanced fluctuation at N-loop-lid region in presence of substrate in contrast to apo structures.

### 3.7. The Normal Mode Analysis

To gain insight into the functional dynamics, we used normal mode analysis. Normal modes provide internal molecular motions indicating the flexibility and rigidity in proteins as well as correlated motions. We conducted an all-atom normal mode analysis using the Cartesian coordinate system. We used the ensemble normal mode analysis because we had structures across the pH spectrum. Figure 3.24A, C depicts the normal mode analysis fluctuations agreeing to the B-factor plot, which depicts the flexible parts of protein structures. This clearly demonstrates the N-loop-lid's flexibility. We generated the trajectories for mode 7, the first non-trivial mode, to gain more insight into the nature of motions. Figure 3.24B depicts the vector field representation of these trajectories, which depicts normal mode 7, denoting the possible correlated functional motions. We used principal component analysis (PCA) to decode the relationship between different structures because the structures were available across the pH spectrum. We discovered that pH 5.5, 6.5, and 7.5 structures are confined in the PC-space, whereas pH 3.5, 4.5, and 8.5 structures are distributed distantly throughout the PC-space [Figure 3.24 D]. This implies that conformations closer to optimal activity may be constrained in a small space, whereas other conformations are spread widely.



**Figure 3.24. Normal mode analysis of PanPL at all pH and substrate bound structure**

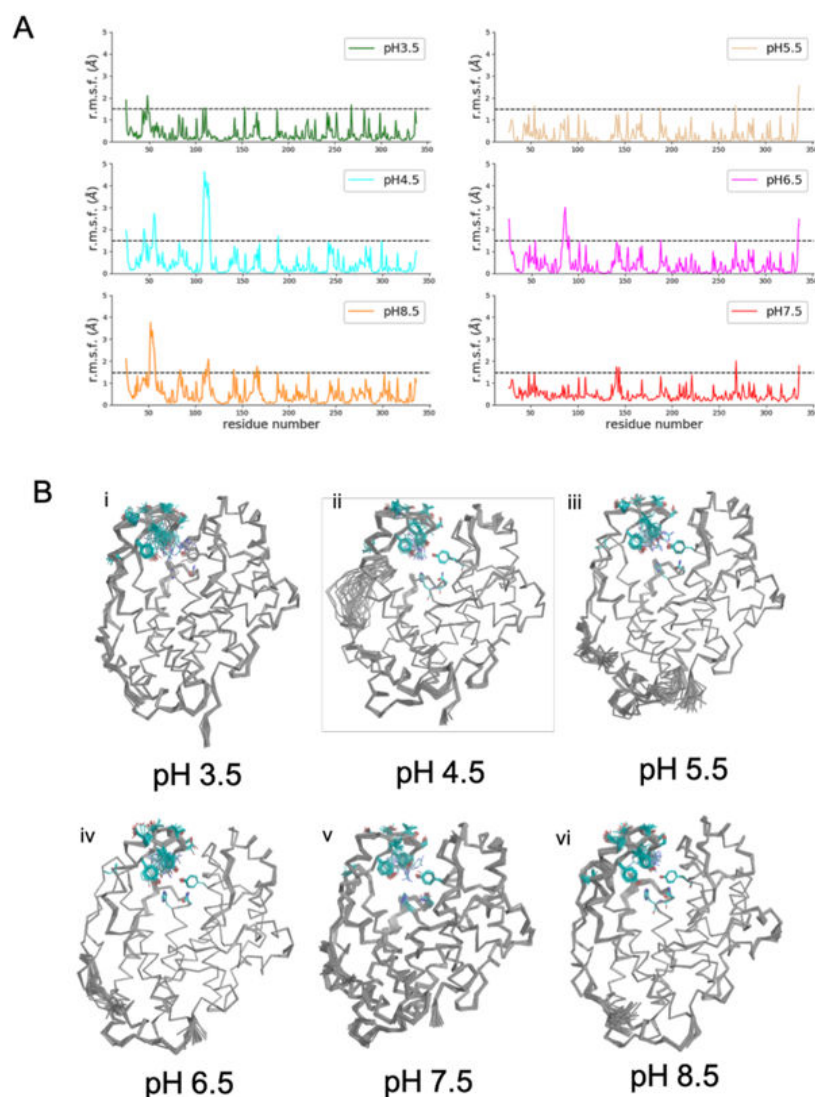
A. The residue wise fluctuations for all pH structures are shown as a result of all atom ensemble normal mode analysis. B. The vector representation of directionality of motion is shown. C. The r. m. s. f. for co-crystallized structures after NMA. D. The PC space for all pH structures from Normal Mode Analysis is shown. The functionally active structures are grouped together and other structures are spread widely (color codes in D, pH 3.5, 4.5, 8.5: red, green, black; pH 5.5, 6.5, 7.5: blue, magenta, orange).

### 3.8. Ensemble refinement displaying the alternative conformations in crystal structures

The crystal structures like snapshots reveal little about substrate processivity in the narrow catalytic tunnel. We expect some local fluctuations here to give the enzyme the flexibility to process the huge polysaccharide chains without jeopardising enzyme stability. With an

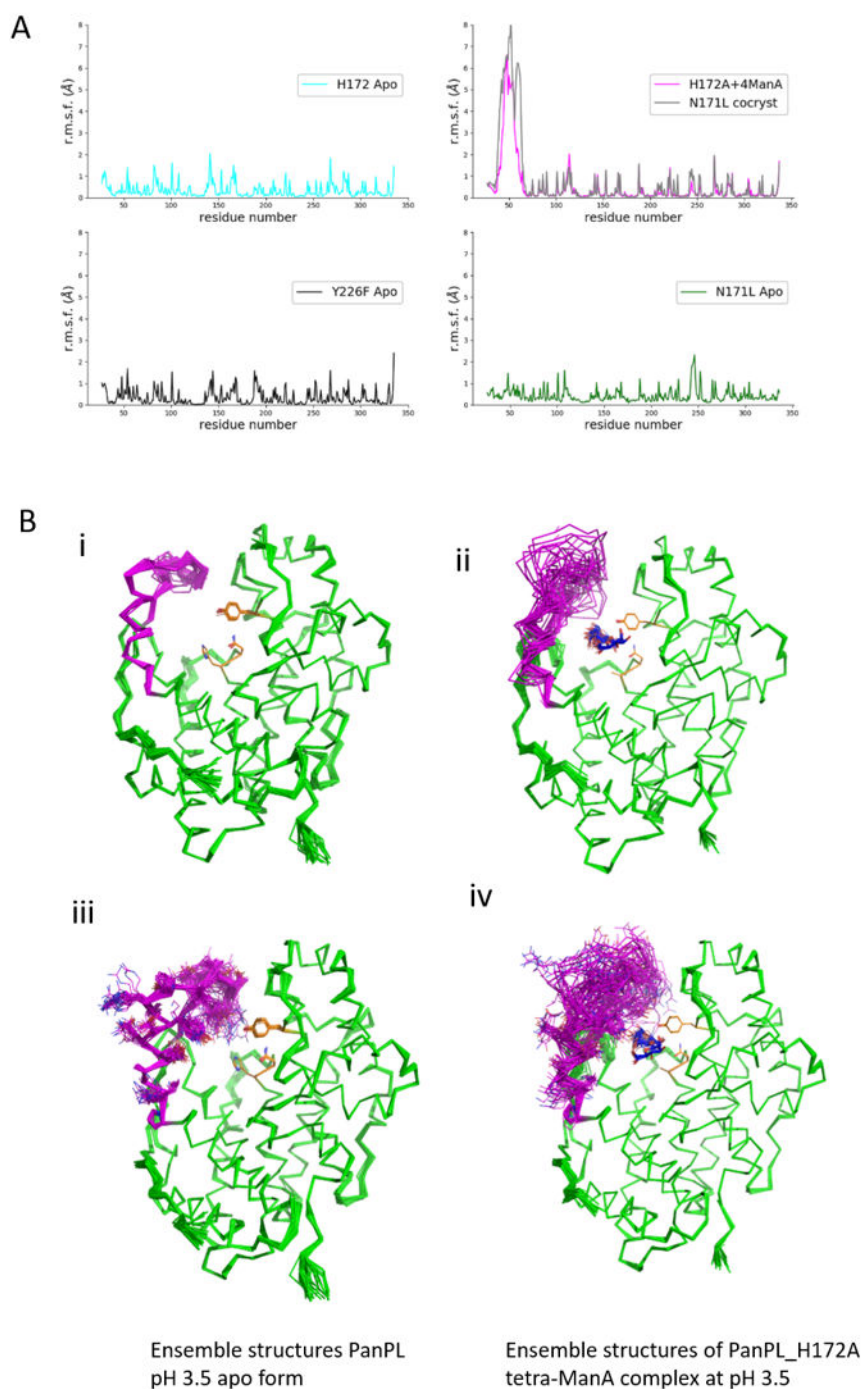
effort to reveal the fluctuations hidden as alternative conformers in the static crystal structures, we used ensemble refinement (ER). To generate an ensemble of structures, ER employs time-averaged refinement and molecular dynamic simulations to sample local molecular motions. We performed ER on each structure to generate the ensemble of structures and determined optimal empirical refinement parameters ( $t_x$ ,  $p_{tls}$ ,  $T_{bath}$ ). Structure ensembles fit X-ray data better than single structures, as evidenced by lower  $R_{free}$  values in comparison to single structures. ER generated a number of structures for each structure [Table 3.9], but we took an equal number of structures in ensembles of each structure for analysis. Figure 3.25] depicts the apo structure ensembles. Our analysis indicates the presence of well-ordered residues in the protein core and loop regions with flexible residues. ER modelled a large number of alternative conformations for the N-loop-lid region which denotes the higher loop flexibility of structures at pH 3.5, 4.5, and 8.5, compared to the structures at pH 5.5, 6.5, and 7.5 [Figure 3.27A]. This implies that PanPL in its activity range, adopts a less dynamic state. We performed ER on all the mutant and co-crystallised structures. We noticed the co-crystallised structures i.e.; N171L and H172A\_tetraManA bound structures shows huge fluctuation at the N-loop-lid [Figure 3.27B]. Not only substrate acquisition even the presence of the substrate itself increase the loop dynamics, as evidenced by the ensemble refinement. The N-loop-lid undergoes local backbone dynamics as well as side chain fluctuations induced by substrate. The alternative conformers trapped during crystal growth phase are now displayed from the crystal structures with the help of ER. We could model huge number of alternative conformers for the N-loop-lid residues representing substrate induced fluctuations. Such kind of loop dynamics upon substrate binding in ABL kinase were recorded in solution NMR, which was revealed during the case study of ER on the ABL kinase crystal structure(14). The loop dynamics of PanPL revealed by ER analysis suggest the occurrence of the substrate

induced fluctuations in solution. Our overall ensemble refinement analysis suggests that fluctuations play important role in enzyme function.



**Figure 3.25. Ensemble refinement of PanPL crystal structures across pH spectrum**

A. The residue wise fluctuations (r. m. s. f.) were calculated for ER ensembles and are plotted in different colors to represent apo structures of PanPL from pH 3.5 to pH 8.5. B (i-vi) Ensemble refinement of PanPL shows the flexibility of the core region is less than the loop regions. ER shows larger number of alternative conformations in N-loop-lid region. The side chain conformational flexibility of the N-loop-lid is shown in cyan color sticks for different pH along with active site residues.



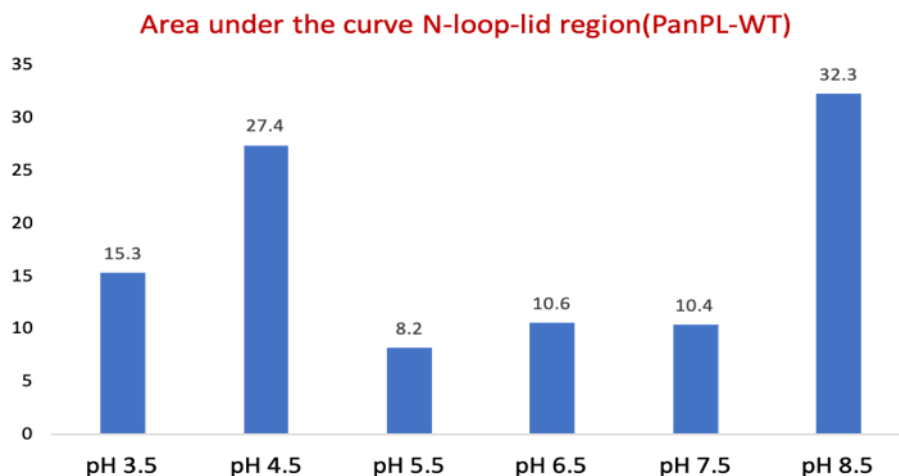
**Figure 3.26. Ensemble refinement of PanPL inactive mutant crystal structures in bound and apo forms.** A. The r. m. s. f. plot for the Ensemble Refinement analysis shows that the apo structures active site mutant H172A, Y226F, and N171L show less dynamics at N-loop-lid region in comparison to H172A tetra-ManA bound and N171L co-crystallized structures. This indicates, substrate induces the enhanced fluctuations at the loop. B. The ensemble

structures obtained during the ER analysis. In B i and ii only backbones of N-loop-lid are shown, whereas side chains are shown in B iii and iv. It showed a large number of alternative conformations at N-loop-lid of the substrate bound structure compared to the apo structure.

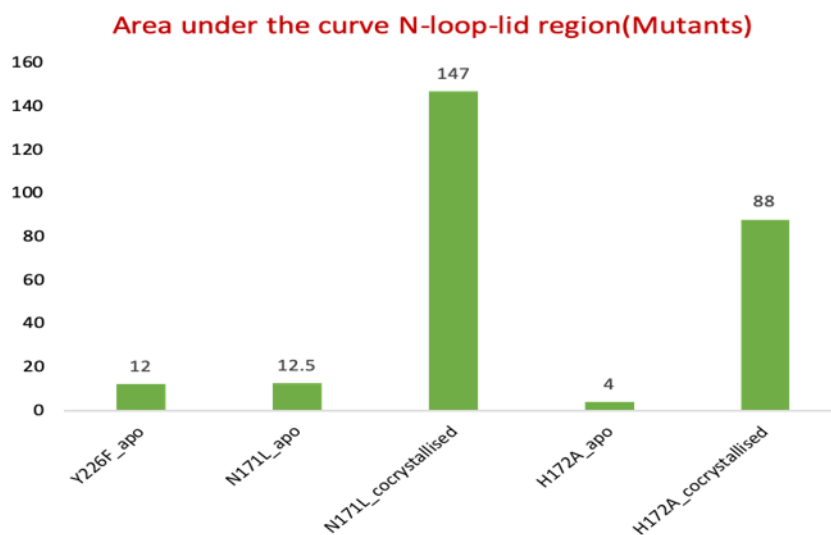
**Table 3.9. ER statistics of crystal structures**

PDB ID	phenix.refine		Ensemble Refinement								
	Rwork	Rfree	resolution (Å)	$\tau$ (ps)	No. of Models	ptls	Rwork	Rfree	Tbath (K)	$\Delta$ Rwork	$\Delta$ Rfree
7WUJ	0.1893	0.2297	2.1	0.5	36	0.9	0.1587	0.2144	5.0	0.0306	0.0153
7WVK	0.2122	0.2554	1.68	0.9	56	0.9	0.1823	0.2251	10.0	0.0299	0.0303
7WVL	0.168	0.1903	1.45	1.0	72	0.9	0.1399	0.1680	5.0	0.0281	0.0223
7WVM	0.1697	0.186	1.23	1.3	75	1.0	0.1432	0.1625	2.0	0.0265	0.0235
7WVN	0.2449	0.2854	2.4	0.5	28	0.6	0.198	0.2736	5.0	0.0469	0.0118
7WVO	0.1959	0.2325	1.96	0.5	32	0.8	0.1501	0.1989	10.0	0.0458	0.0336
7WVP	0.1695	0.2171	2.20	0.4	27	1	0.1489	0.2029	10.0	0.0206	0.0142
7WVQ	0.1911	0.2396	2.03	0.5	63	0.9	0.1548	0.2191	10.0	0.0363	0.0205
7WVR	0.1798	0.2258	1.88	0.6	50	0.9	0.1557	0.2057	5.0	0.0241	0.0201
7XTE	0.2425	0.2808	2.30	0.4	27	0.8	0.2156	0.2755	5.0	0.0269	0.0053
7XTF	0.1797	0.2221	1.73	0.7	35	0.9	0.1701	0.2159	2.0	0.0096	0.0062

A



B

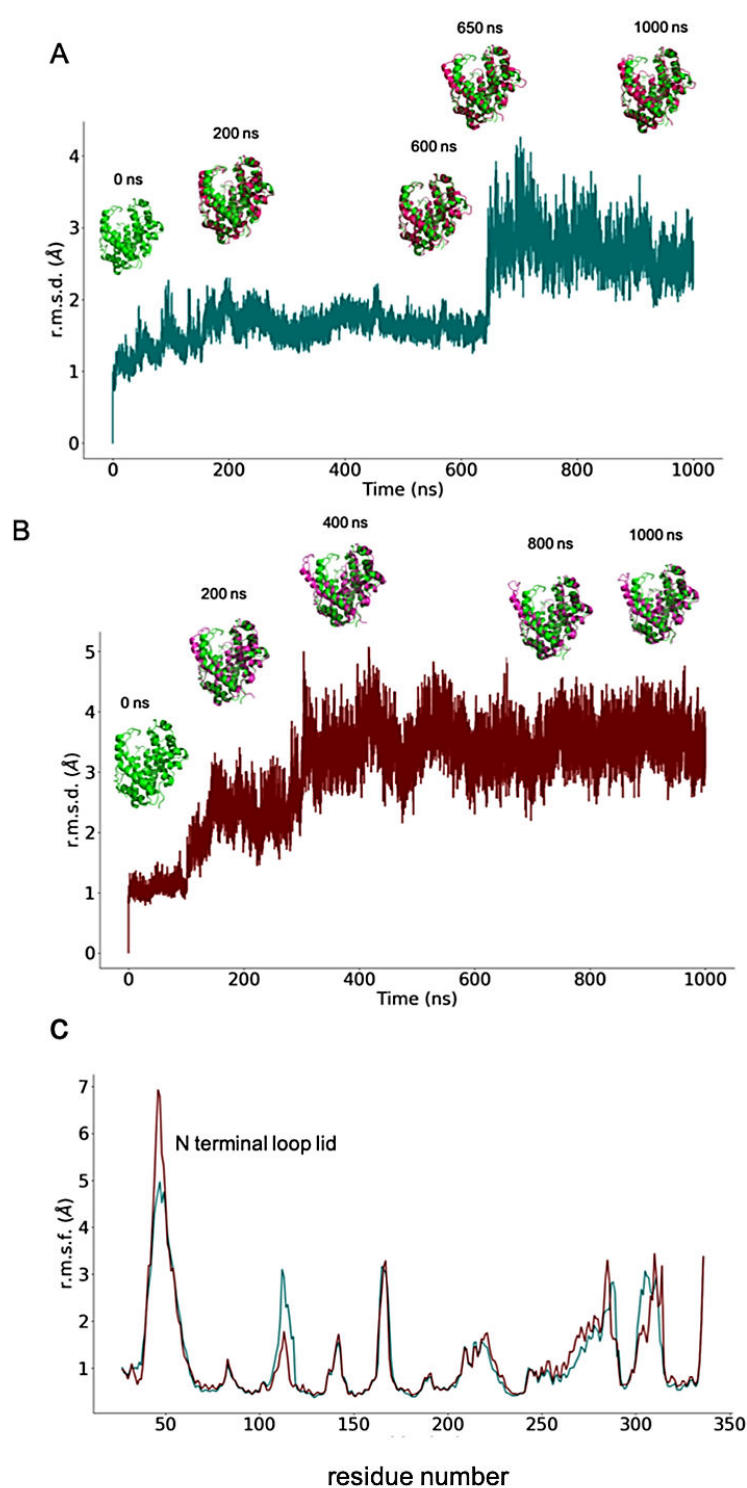


**Figure 3.27. The quantitative analysis of N-loop-lid fluctuation.** A. The plot shows the area under the curve of r. m. s. f. of N-loop-lid region of PanPL as a function of pH. At pH 3.5, 4.5 and 8.5 the fluctuations are more compared to those at pH 5.5, 6.5 and 7.5, which suggests that N-loop-lid of PanPL in its functionally active state prefers lesser fluctuations (pH 5.5, 6.5 and 7.5) than its functionally inactive states (pH 3.5, 4.5 and 8.5). B. The plot shows the area under the curve for the N-loop-lid fluctuations for the mutant structures. This shows enhanced fluctuations for H172A tetra-ManA bound and N171L co-crystallized structures.



### **3.9. The hidden open state of PanPL in Molecular Dynamics simulation**

The B-factor analysis and normal mode analysis on the apo crystal structures established the functional relevance of the N-loop-lid dynamics [Figure 3.21, 3.22, 3.24]. In support to the previous data, the substrate-bound structure exhibits a significant vibration at the N-loop-lid [Figure 3.23, 3.26]. All of these findings prompted a more in-depth investigation of the enzyme's loop dynamics. To observe the dynamic property of PanPL, we performed molecular dynamic simulation studies on the pH 6.5 apo structure and substrate-bound crystal structure. To match the order of time scale required for loop motion, we observed the time evolution of atomic coordinates for 1 micro second. The apo structure r.m.s.d. plot shows a transition in deviation around 650ns. Similarly, the r.m.s.d. plot for a substrate-bound structure shows a transition around 200ns. We compared the structure at various time points to the initial structure for both apo and substrate bound structures. It was fascinating to realize that these transition time points represent a conformational shift from a close tunnel state to an open cleft state [Figure 3.28A, B]. We also analyzed the dynamics of the H-bond interactions that help in maintaining the closed tunnel structure [Figure 3.29].

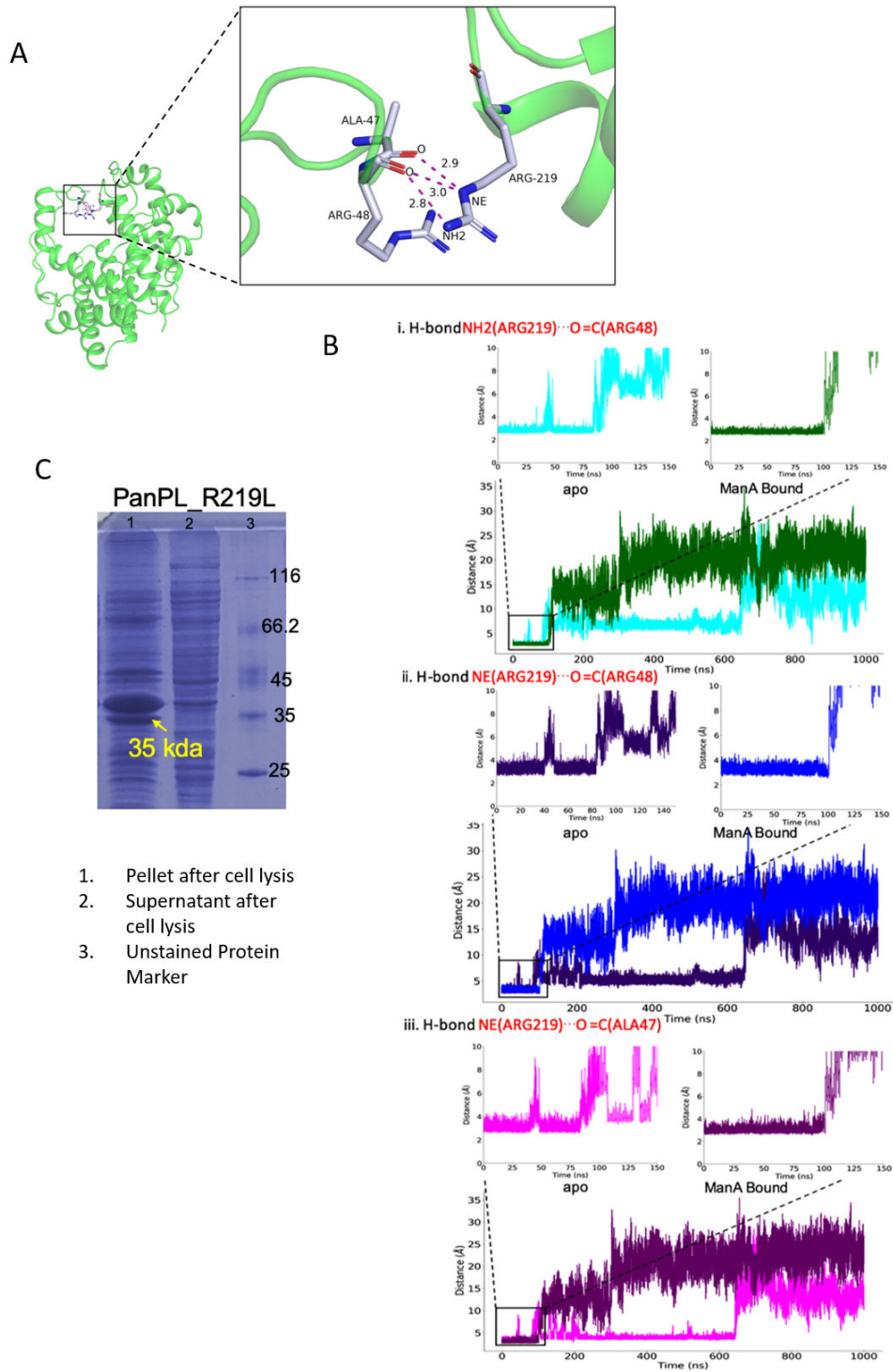


**Figure 3.28. Analysis of the Molecular dynamics simulation trajectories of apo and tetra-ManA bound crystal structures.** Figure shows r. m. s. d. (Å) as a function of time (ns). (A) Shows the r. m. s. d. plot for the apo structure trajectory with snapshots of the coordinates at different time points are shown. (green color: initial structure, and magenta

color: The jump at 650ns corresponds to the N-loop-lid opening. and the transition from close tunnel to open cleft. (B) shows the r. m. s. d. of substrate bound crystal structure during 1 $\mu$ s simulation with snapshots of coordinates at different time points. C. shows the residue wise atomic fluctuations during the length of trajectory, suggesting the residues range ~30-60 has large movements during the course of simulation.

### **3.9.1. The N-loop-lid dynamics and the H-bond lock**

There are three hydrogen bond interactions which favour the tunnel formation between the R48 and A47 main chain carbonyl group of the N-loop-lid (aa 42-52) and the side chain of the R219 residue loop (aa 218-223) [Figure 3.29A]. These types of hydrogen bonds are reported to preserve the fold's architecture (79). We examined the trajectories of 1 $\mu$ s Molecular dynamics simulation data to determine the importance of H-bonds. Our analysis shows that the H-bonds are broken during the simulation in both the case of the apo and substrate bound structures, resulting in the opening of the N-loop-lid [Figure 3.29 B]. The N-loop-lid moves away from the tunnel, hence the distance increases in the plot [Figure 3.29B] allowing the overall scaffold to access the open state. The r.m.s.f. plot depicts fluctuations in the N-loop-lid region of both the apo and substrate bound structures [Figure 3.28C]. Furthermore, we created the R219L mutation to test the significance of H-bonds in maintaining the close state, which resulted in a loss of lyase activity as confirmed by the TBA assay [Figure 3.29C].

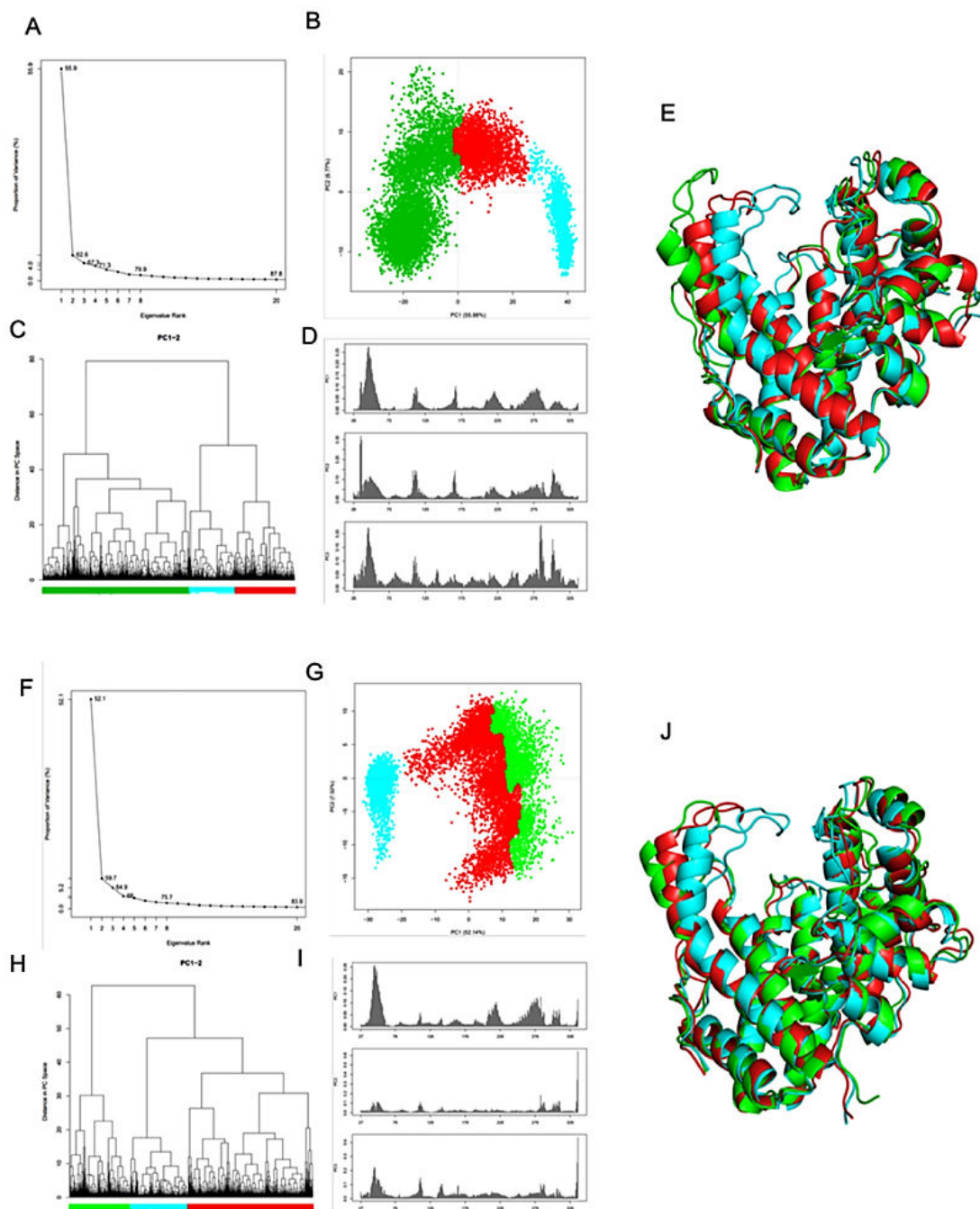


**Figure 3.29. The dynamics of H- bond lock the N-loop-lid.** A.The main chain of R48 and A47 residues with the side chain of R219 residue are important to maintain the closed tunnel architecture. Bi, ii and iii show time evolution the H-bond length between

NH2(R219) and C=O(R48), NE(R219) and C=O(R48), NE(R219) and C=O(A47) in both apo and substrate bound structures. In case of apo structure trajectory, the H-bond interactions are lost permanently beyond 650ns (cyan, indigo, magenta). In case of substrate bound structure trajectory the H-bond interactions are lost around 100ns (green, blue, purple). C. The gel image shows that the R219L mutant is not stable in solution and remain in pellet fraction after lysis.

### **3.9.2. PCA clustering of MD trajectories**

We used Principal component (PC) clustering of MD trajectories of both apo and substrate bound structures to delineate the conformations spanned in MD trajectories. The PC analysis segregates the essential dynamics as well as samples of the major conformational shifts that occurred during the MD simulation. Three distinct states were observed upon hierarchical clustering. The average structures for each group were examined and compared to the starting structures. During simulation in both apo and substrate bound structures, presence of three distinct states were observed, namely a closed state, an intermediate open state, and a widely open state. This result demonstrates the inherent flexibility of the enzyme's N-loop-lid. [Figure 3.30].



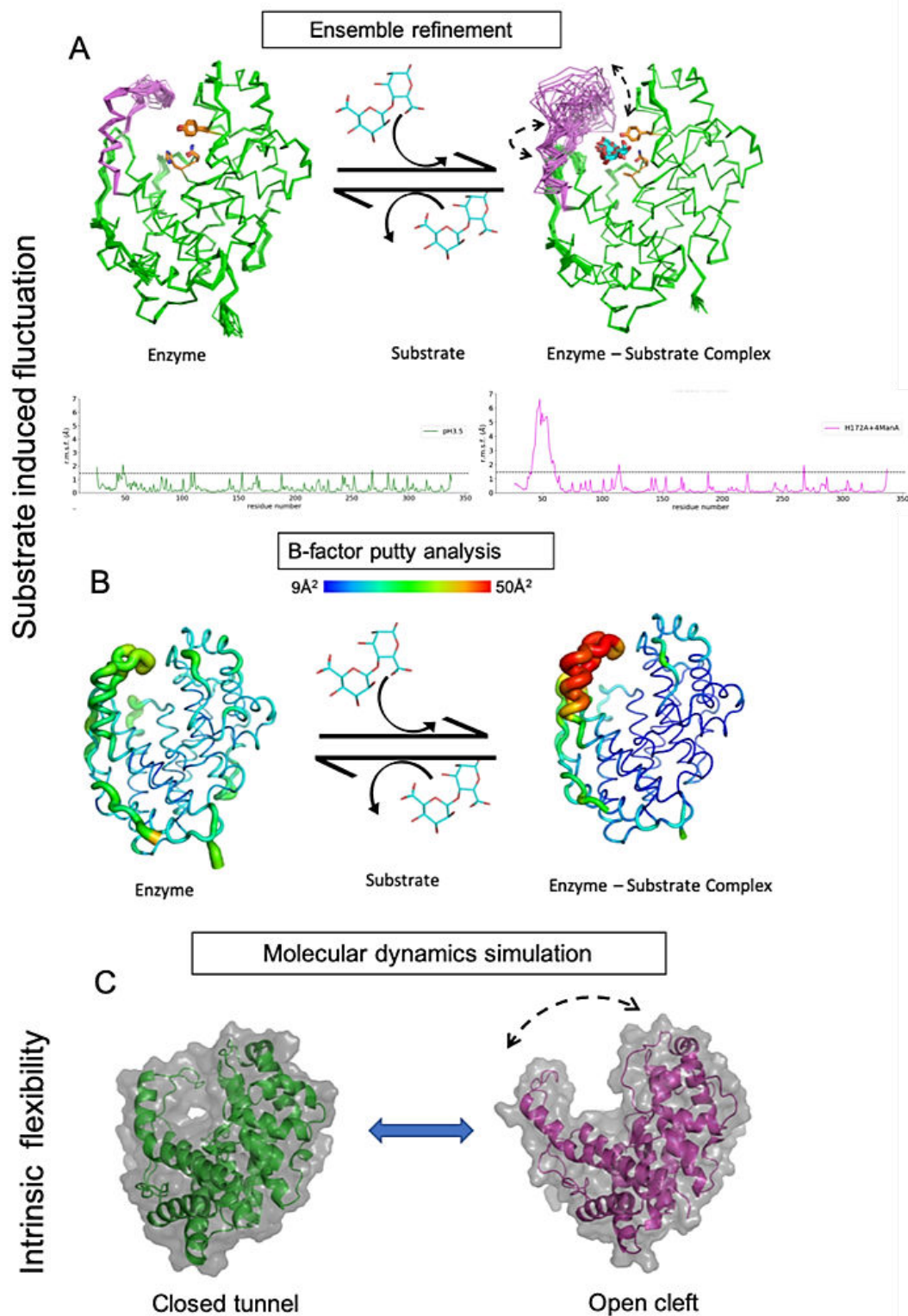
**Figure 3.30. Principal component analysis on Molecular dynamics simulation trajectories (100ns) of apo and tetra-ManA bound structures of PanPL.**

Panels A-E are for the apo structure. A. Scree plot depicts first three components contribution amounts to ~67.3% of variance of which 55.9% is PC1 component, 6.7% PC2 component and 4.7% PC3 component. B score plot in PC1-PC2 space, colored based on the groups

recognized in hc, cyan colored points represent closed state conformers, while red and green colored points belong to intermediate open and widely open state. C hierarchical clustering (hc) in PC space into three main groups, backbone conformations in each group are similar. D contribution of each residue to the principal components (PC1, PC2, PC3). E shows the average structure of three groups; closed state(cyan), intermediate open state (red) and widely open state (green). Panels F-J are for the tetra-ManA bound crystal structure. E. PC1 has dominating 52.1% variance while PC2 and PC3 have 7.6% and 5.2 % variance respectively. F. score plot in PC1-PC2 space showing the clustering of data colored based on states. (Cyan closed state, Red intermediate open state, Green widely open state). G. hc in PC space shows two main groups similar to apo structure. H. residue level contribution towards the first three principal components.

### 3.10. Overall dynamics of PanPL

The B-factor analysis, ER analysis and the normal mode analysis on the PanPL apo crystal structures congruently suggests the pH tunes the N-loop-lid dynamics and influences the enzyme activity. In addition to that, the B-factor analysis and the ER analysis on the co-crystallized mutant structures show that the presence of the substrate in the crystallization solutions induce tremendous dynamics in the N-loop-lid without disintegrating the catalytic tunnel. However, the MD simulation studies on the PanPL apo and substrate bound structures witness the loop dynamic, where the N-loop-lid moves away to form an open cleft like structure in the MD trajectories. The overall dynamics of PanPL have been shown in Figure 3.31. We denoted these distinct modes of dynamics of PanPL as i. substrate induced enhanced fluctuations, where N-loop-lid fluctuates upon encountering the substrate without dismantling the close state and ii. inherent flexibility, where the N-loop-lid opens up to access the open state.



**Figure 3.31. Dynamic behaviour of N-loop-lid of PanPL.** A and B shows the substrate induced fluctuation. A. The ensembles of structures (ribbon) for apo and substrate bound state generated by ensemble refinement. Substrate binding enhances the fluctuation at N-



loop-lid. The plots show the r.m.s.f. per residue. B. The B-factor putty representation of PanPL apo and substrate bound structure. Substrate bound structure shows a high atomic fluctuation at the N-terminal loop lid, reflecting high fluctuation. C. shows the intrinsic dynamic. The MD simulation analysis shows closed tunnel (0 ns) to open forming a cleft like structure (1  $\mu$ s) during simulation.

### **3.11. Allostery as a result of conformational flexibility and structural dynamics**

In addition to NMA and MD simulation, ER of crystal structures from all pH displays the hidden alternative substates in the average crystal structures. These ensemble structures fall within a narrow range of energy variation. External perturbations such as pH change, substrate binding, mutations can influence the population distributions of conformers. The N-loop-lid denotes the vibrational hotspot around the catalytic tunnel, according to our ensemble refinement study of substrate bound structure. The presence of the substrate in solution brings about local lid loop dynamics and corresponding conformational substates for the loop residues. This substrate-induced loop flexibility with loop residue side chain conformational heterogeneity, can be able to shed light on PanPL's allosteric property. Along with the conformational heterogeneity, the hidden open-close state transition recorded during MD simulation may also assist the positive cooperativity observed in the biochemical characterization. This unique positive cooperativity has never been reported previously in other PL-5 family enzymes. This unique property of PanPL adds it to the list of allosteric enzymes which show dynamics driven allostery with no observable structural changes. Our thorough studies of PanPL structure-function and dynamics assisted us in explaining the enzyme's behaviour comprehensively. Our findings suggest that all PL-5 family enzymes should include dynamics in addition to the structure-function aspects.

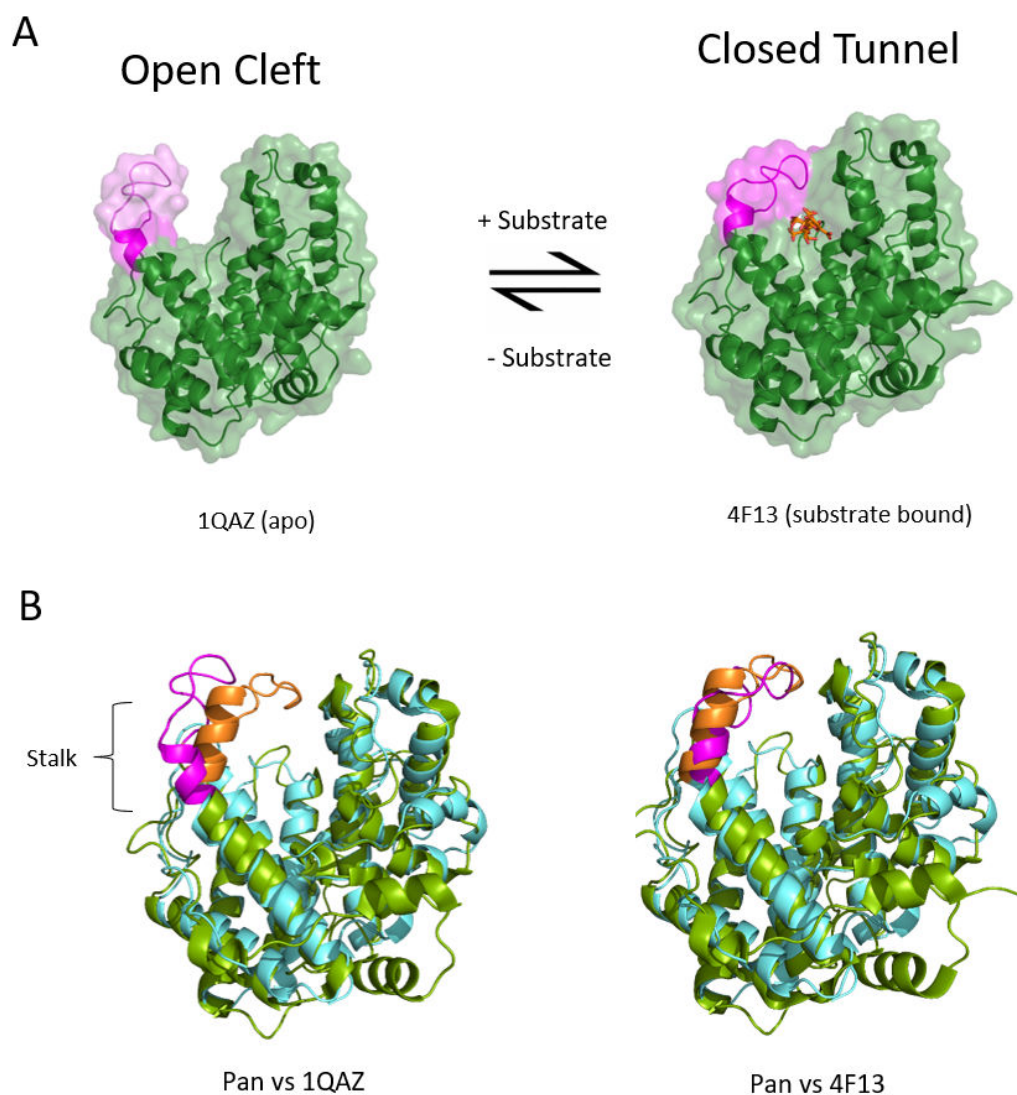
### **3.12. Significance of N-loop-lid length in selecting open/close or only close state**

We compared the crystal structures of PL-5 enzymes from PDB to make a list of the open cleft-like and closed tunnel-like architecture. An open cleft-like configuration was observed for the apo crystal structure (1QAZ) of alginate lyase AIII from *Sphingomonas sp.* However, the enzyme adopts an induced fit mechanism to form a closed tunnel like configuration upon substrate binding (46). The tunnel-like closed state provides a catalytically competent microenvironment seen in the substrate bound crystal structure (4F13) [Figure 3.32A]. Whereas the closed tunnel conformation was preferred by Smlt1473 (49), alginate lyase from *P. aeruginosa* (4OZV) and PanPL in all of their apo and substrate bound forms. The sequence alignment analysis [Figure 3.33] reveals that Alginate lyase AIII has a four-residue deletion at the N-terminal long helix region, which reduce the length of the bent portion of the first inner helix i.e., the stalk of the N-loop-lid. This shortening of stalk results in a loose the H-bond lock required to maintain the tunnel-like structure and give an access to the open cleft [Figure 3.32B]. Upon substrate binding a conformational transition to catalytically competent tunnel-like state has been witnessed (46). Other PL-5 enzyme structures possess an extended N-terminal helix to allow the loop lid to form only tunnels in their crystal structures (5,49). In this case, the length of the lid loop is critical in determining the mechanism of dynamics (i) an open and closed conformation, or (ii) a predominant catalytically competent closed state with functional fluctuations around the active site.

Our ER analysis on the co-crystallised structures ensure the occurrence of conformational fluctuation at the N-loop-lid both in the side chain and the back bone, without disturbing the catalytically competent closed state [Figure 3.26]. However, the MD analysis allows

the access to the hidden open state transiently. These findings indicate that the PanPL N-loop-lid is flexible intrinsically and susceptible to alternative dynamics (an open/close state transition and a close state with functional fluctuation). The extended stalk of the N-loop-lid favours the close state with functional fluctuation in case of PanPL over the open/close state transition. Another study on *E.coli*. Dihydrofolate reductase (ecDHFR) and human Dihydrofolate reductase (hDHFR) shows a similar kind of divergent dynamics (Bhabha et al., 2013). A single residue insertion in the Met20 loop of hDHFR (8 residues) significantly alters the dynamics compared to the Met20 loop of ecDHFR (7 residues). The ecDHFR has two different loop conformations, both closed and occluded states, which follows the closed to occluded state transition mechanism. Whereas the hDHFR Met20 loop conformation has only a closed state which undergoes fast fluctuations to facilitate ligand binding and product expulsion (23).

According to our sequence alignment studies [Figure 3.33], we chose two PL-5 enzymes from *Burkholderia cenocepacia* (BcPL) and *Ralstonia picketti* (RpPL) with 10 residue deletions at the stalk region of the N-terminal helix. We created homology models by taking both open and closed states as template using the Rosseta Comparative Modelling module. The closed state models of RpPL and BcPL, the tunnels were found to be constricted which cannot accommodate any substrates [Figure 3.34, 3.35]. Furthermore, the open state models form cleft-like structure which has never been observed to bind substrates. By using Alphafold, we modelled RpPL and BcPL. The tunnel formation in Alphafold models lack the hydrogen bonds and fails to maintain the tunnel architecture. Furthermore, in both RpPL and BcPL, the putative catalytic residue Histidine is replaced by Leucine. To test their activities, we overexpressed RpPL and BcPL. The enzymes did not confirm lyase activity in the TBA assay, as anticipated [Figure 3.36].

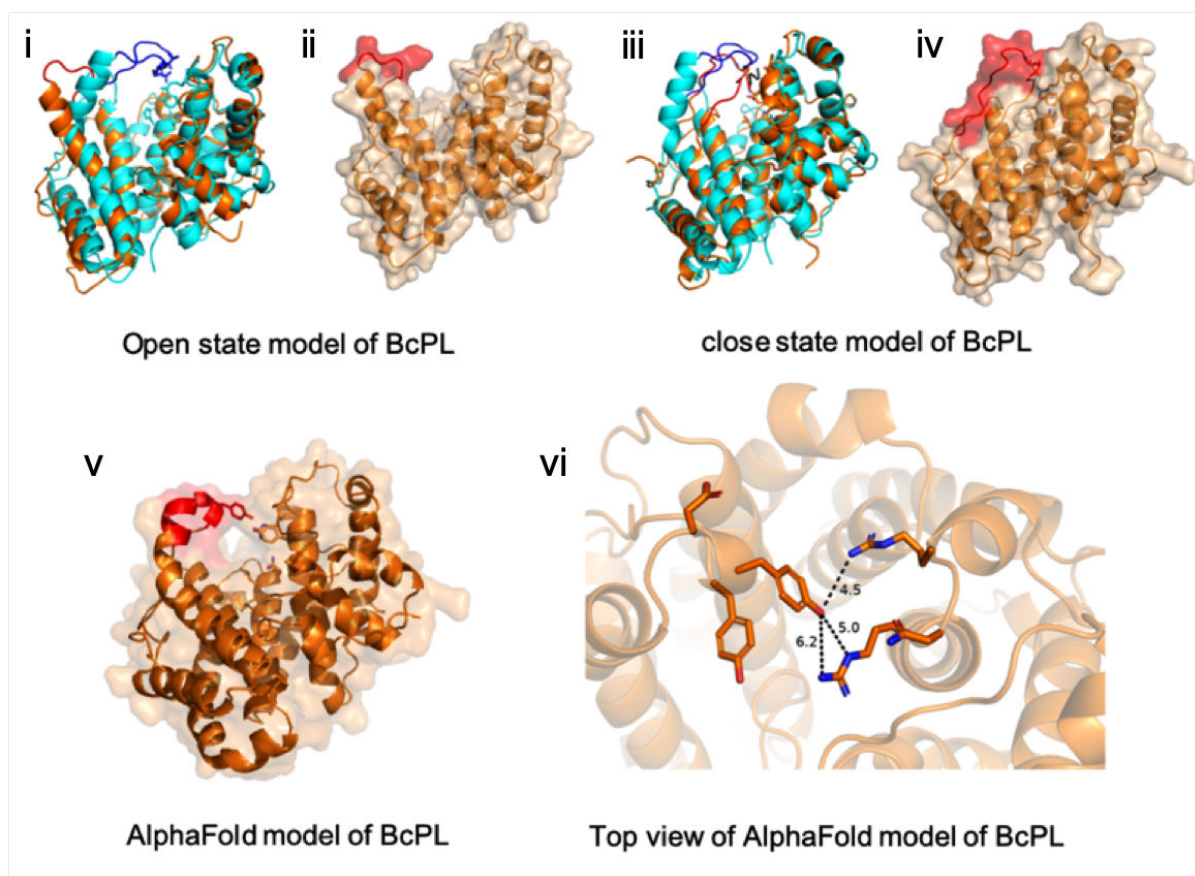


**Figure 3.32. The open-close state transition in *Sphingomonas sp.* Alginate lyase AIII.**

**A.** The crystal structure of this PL-5 enzyme forms a open cleft like structure in apo form(1QAZ). Upon substrate binding the open state transforms to a closed tunnel like structure (4F13). B shows the 1QAZ structure (green, left) and 4F13 structure (green, right) superimposed with PanPL structure(cyan). The N-loop-lid of 1QAZ and 4F13 are shown in magenta. The N-loop-lid of PanPL is shown in orange.

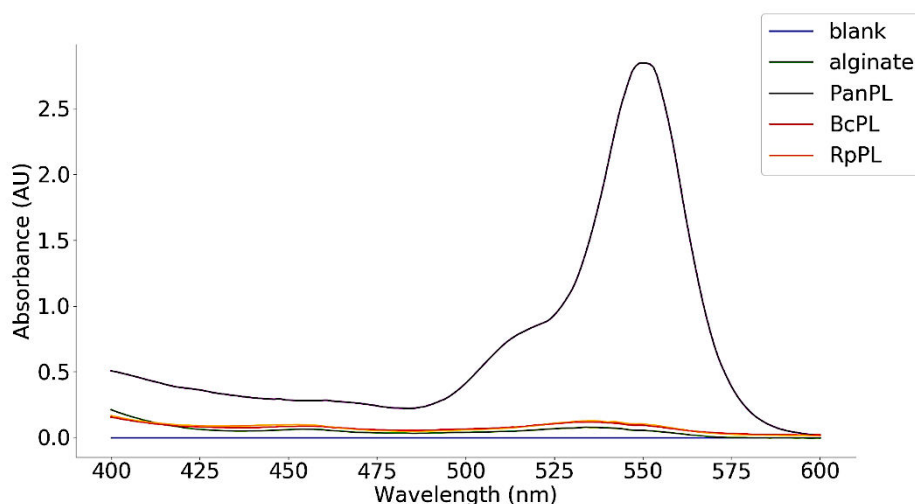


no tunnel formation. v shows the AlphaFold Model of RpPL(wheat) shows a tunnel formation. vi shows the top view of RpPL alpha fold model shows no H-bond formation between the loops of the tunnel ceiling.



**Figure 3.35. shows the structural analysis of BcPL.** i. shows the super imposed structure of PanPL(cyan) and BcPL(orange) Rosetta homology model, BcPL has a short N-terminal helix compared to PanPL. ii. shows the open cleft state of BcPL. iii. shows the closed state Rosetta model of BcPL superimposed with PanPL structure. iv. shows no tunnel formation for BcPL closed state. v. The AlphaFold Model of BcPL shows a tunnel formation. vi. The top view of BcPL alpha fold model shows no H-bond formation between the loops of the tunnel ceiling.





**Figure 3.36. The TBA assay for BcPL and RpPL.** This plot shows the RpPL and BcPL has no enzyme activity, i.e., the absorbance spectrum shows negligible at 550 nm for RpPL and BcPL compared to PanPL.

### 3.13. Conclusion

In this chapter our study on the novel PL-5 enzyme PanPL covers the biochemical and structural characterization along with exploring the dynamical behaviour of the enzyme, which provides a holistic idea about the enzyme's function and mechanism. During protein expression the signal peptide cleavage has been witnessed and the extracellular lyase activity has been confirmed by the plate assay [Figure 2.3, Chapter 2]. The tag-less protein has been purified by ion exchange chromatography [Figure 2.4, Chapter 2]. The purified protein has been used for the biochemical and the structural characterizations. The enzyme was stable across the pH spectrum. The TBA assay confirmed the lyase activity of PanPL towards alginate and poly-ManA [Figure 3.1]. The enzyme activity range falls between pH 5.5 to 7. The enzyme kinetic assays have been performed at pH 7 (optimum pH) to determine the enzyme kinetic parameters of PanPL with its substrates. Interestingly, a trend of positive cooperativity or allostery has been observed in PanPL's biochemistry, which

has never been reported in PL-5 enzymes [Figure 3.3]. The protein was crystallized at different pH (pH 3.5-8.5) and the crystal structures of PanPL across the pH spectrum have been solved [Table 3.2]. Further active site mutations were generated and co-crystallised with tetra-ManA to trap the substrate in the catalytic tunnel of PanPL. The apo structures of PanPL across the pH and the mutant structures in their apo and co-crystallised forms have been deposited to RCSB-PDB [Table 3.2, 3.4]. All of the crystal structures of PanPL form a tunnel like close state and maintain the characteristic  $(\alpha/\alpha)_5$  fold of PL-5 family [Figure 3.6]. The structures remain invariant across the pH with little rotameric transitions of amino acids except R48 [Figure 3.9]. However, the pH has been observed to influence the structural flexibility, suggested by B-factor analysis and the electrostatic surface charge distribution, calculated by PDB2PQR [Figure 3.22]. Further, our B-factor analysis and the Ensemble refinement analysis on the apo structures show that in the pH range of enzyme activity i.e., at pH 5.5, 6.5, 7.5, the fluctuations at the N-loop-lid of PanPL have been observed to be less compared to the pH 3.5, 4.5 and 8.5, where the enzyme shows reduced activity [Figure 3.21, 3.25]. In case of N171L co-crystallised and H172A tetra-ManA bound structures, the B-factor analysis show huge vibration at the N-loop-lid [Figure 3.23] and the ER analysis reveals large number of hidden alternative conformations at N-loop-lid [Figure 3.26]. Both of the analyses suggest the substrate induced fluctuation at the N-loop-lid, where the presence of substrate in the crystallization solution causes huge fluctuations at the N-loop-lid, which supports the allosteric behaviour of PanPL. However, these fluctuations maintain the catalytically competent closed state of PanPL. The MD simulation studies on apo and substrate bound structures reveal the N-loop-lid accessing an open cleft like state which has never been observed in the crystal structures [Figure 3.28]. This suggests that the N-loop-lid of PanPL has an intrinsic flexibility to obtain a hidden open state transiently. Our study suggests the closed state to be the major and open

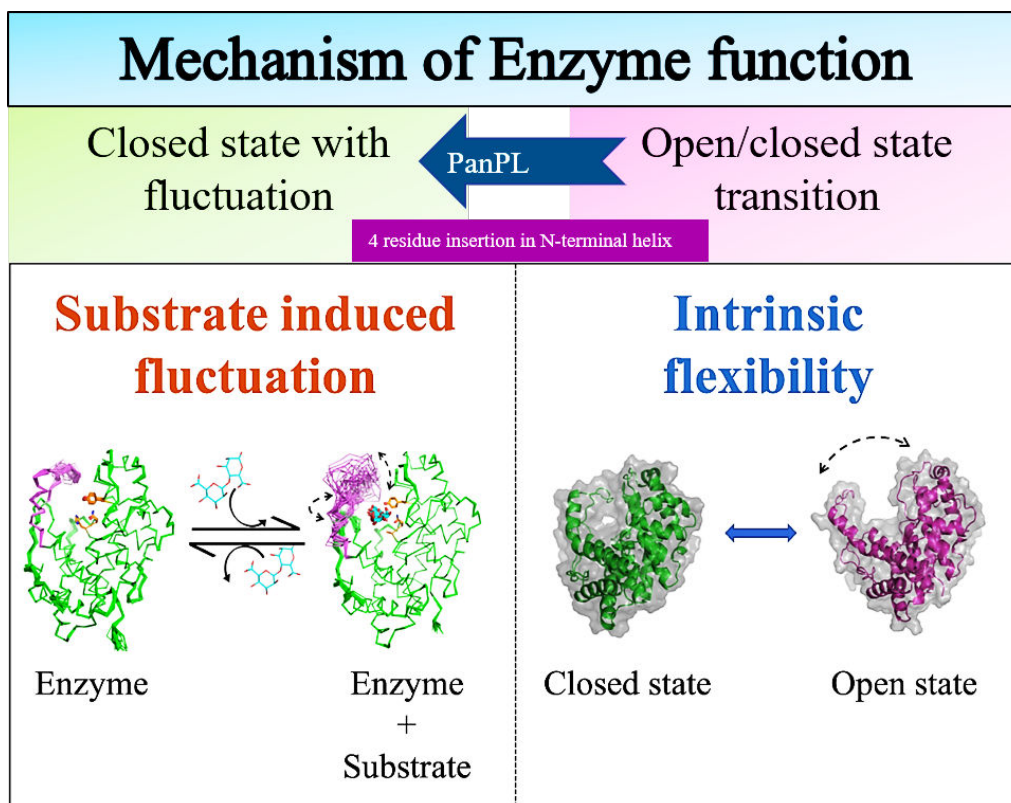


state to be the minor conformer in the energy landscape. From the sequence analysis studies we found that there is a 4 residues insertion at the N-terminal helix region which helps the N-loop-lid to access the close state as a major conformer [Figure 3.33]. The insertion helps PanPL to select the mechanism of substrate induced fluctuation at N-loop-lid maintaining the closed state over the open to close state transition mechanism [Figure 3.37].

The key findings of our work on PanPL are listed below.

- PanPL is an alginate specific, endolytic, homotropic, allosteric enzyme with single ligand binding site.
- The close state is catalytically competent and more stable.
- The substrate induced fluctuations at N-loop-lid allows the loop residues to access large numbers of Conformational Substates. This brings about the allostery in biochemistry.
- PanPL has an intrinsic property to access the open state.
- PanPL selects close state with fluctuation for its functioning.
- Insertion at N-terminal helix helps to select the mechanism.

The biochemical and structural characterization of two other PL-5 enzymes; i.e., RpPL and BcPL have been incorporated in this chapter.



**Figure 3.37. Summary of PanPL mechanism.** The PanPL shows two distinct modes of dynamics. The substrate induced fluctuations at N-loop-lid has been shown by the ensemble structures modelled after ER analysis (Left). The PanPL can access a hidden open state along with the closed state, which have been observed during MD trajectories (Right). Although PanPL has an intrinsic flexibility to undergo open/close state transition, it chooses the mechanism of undergoing fluctuations in its closed state upon substrate encounter. The 4 residues insertion at the N-terminal helix helps to select the mechanism.



# **CHAPTER 4: Structure- guided mutations that altered the dynamics and enzyme kinetics**

## 4.1. Introduction

Enzymes are nature's wonderful gadgets to aid difficult chemical reactions. However, nature constantly keeps on editing their nucleotide sequences by evolution which in turn update, upgrade or alter the enzyme function. Inspired by evolution, enzyme have been engineered to attain their best version by carving their catalytic site architecture by introducing mutations. The directed evolution that mimics the process of natural evolution in laboratory has shown its exemplary contribution in making new enzymes editing old enzymes to give new functions. The structural characterization of enzymes in apo and substrate bound forms provide the details of flanking amino acid orientations around the active sites. A careful analysis of the enzyme-substrate interaction patterns can provide a guided insight to select the non-catalytic residues in the catalytic region, which can enhance or abate enzyme activity. These flanking non-catalytic residues are important as they maintain the web of electrostatic interactions, H-bonds and hydrophobic interactions. Introducing mutations can rewire the pre-organized interaction networks. These modifications can alter enzyme-substrate affinity, specificity, loop dynamics, most importantly the catalytic efficiency. Here, we report modulating the enzyme efficiency of PanPL, a polysaccharide lyase from PL5 family by introducing mutations around the catalytic tunnel. We studied the impact of a few selected mutations on biochemical properties, allosteric behaviour, structural and dynamics aspects and compare and contrast with wildtype PanPL.

## 4.2. What changes by mutation?

The side chain orientations of amino acids in a protein maintain a specific local microenvironment. The geometry and the inter-residue interactions among the side chains create distinct network patterns for the polar interactions, electrostatic interactions and

hydrophobic interactions. However, introducing a point mutation at a specific site of the protein changes the local geometry as well as the interaction patterns. The mutation can establish new interactions or can disrupt the previous network based on their side chain orientation, size, charge or hydrophobicity etc. Some active site mutations are introduced to the enzymes for inactivating the catalysis. The loss of function is achieved by altering the functional groups involved in catalysis. Some mutations are selected in order to reshape the active site pockets of the enzymes. The size of the active site pocket can be controlled by mutations with variable side chain length. Similarly introducing a charged residue mutation alters the local electrostatic environment of the enzyme. Mutations involving hydrophobic residues might alter the hydrophobic interaction pattern locally. Although the point mutation changes the local environment of the specific site, it is capable of altering the global dynamics of the enzyme in some cases. The mutational study also includes insertion and deletion in the nucleotide sequence. A certain length of nucleotide sequence encoding amino acids can be inserted to elongate the specific region of protein. Similarly, the deletion of certain length in the protein chain can also be achieved to shorten it. Generally, loop regions are targeted for insertion and deletion. Mutations are considered as a powerful tool to alter the enzyme's structure, function and dynamics. The above discussion highlights the significance of mutation in enzyme engineering.

### **4.3. Identification of flanking residues in the tunnel**

The amino acid composition of the catalytic tunnel of PanPL was categorised into three sites; entry site, exit site and tunnel interior [Figure 4.2]. The crystal structures of PanPL showed the conserved active site residues maintain the PL-5 family specific conserved geometry deep inside the tunnel. However, overall composition of flanking amino acids around the catalytic tunnel varies for different enzymes of PL-5. We analysed the amino acid distributions of available PL-5 crystal structures. We observed mostly positively

charged amino acids near the entry site and inside the tunnel in the structures. This peculiar distribution of positive amino acids facilitates negatively charged substrate acquisition by electrostatic attraction.

#### 4.3.1. Mutation at the entry site of tunnel: L222R

The multiple sequence alignment [Figure 4.1] and the structural comparison among the PL-5 enzymes show that PanPL has lost a conserved Arginine residue at the entry site of the tunnel. In other PL-5 structures (4OZV, 7FHZ) the conserved Arginine residue forms H-bonds with the N-loop-lid main chain carbonyl group [Figure 4.3]. These extra H-bonds are considered to add to the stability of the closed state, and can contribute to maintain the architecture of the fold [Figure 4.2]. However, PanPL lacks these H-bonds as there is a Leucine at the entry site in place of Arginine. We selected the L222 residue and mutate it to Arginine to understand functional the role of the Arginine in other PL-5 enzymes.

EsPL	ABP60714.1	IDDNGQLPNEQKRGARAVAYHNYALQPLVMIAFKAQNVDTQENHGALTRLASYVLR	237
PaPL	AGI80338.1	VDEQGFLPNELKRRQRALAYHNYALPPLAMIAAFAQVNGVDLRQENHGALQRLAERVMKG	296
AcPL	CAA11481.1	VDPRGFLANELKRRQRALAYHNYSLPPLMMIAAFAQANGVDLRGDNDGALGRLAGNVLAG	292
AvPL	AGK12841.1	VDSNGFLPNELKRRQRALAYHNYSLPPLMMVAAFALANGVDLRGDNDGALGRLAGNVLAG	293
BcPL	AIO37757.1	IGPDGALAREVKRGNRALHYHTFALLPLVFAAELVQRRHIDLYRENDGAIGRLANLVIDA	249
RpPL	ACS64514.1	IGPDGALAREVKRGNRALHYHTFALLPLVFAAELVQRRHIDLYRENNHAIGRLANLVIDA	249
PanPL	AJE99968.1	IEPDGRLPMEMARKKLLALHYHDYATAPLVLMAEMARLQGEDWYTYRQGALERLAARVADG	266
Smlt1473	CAQ45011.1	IQDDGSLPLEMARQQALHYHDYALAPLVMMALARLQGDWYASRNHAIDRLARRVIEG	262

**Figure 4.1 Multiple sequence analysis among PL5 family.** The sequence alignment shows the conserved Arginine at the entry site of the catalytic tunnel in PL-5 enzymes. In case of PanPL, there is Leucine residue in place of Arg.

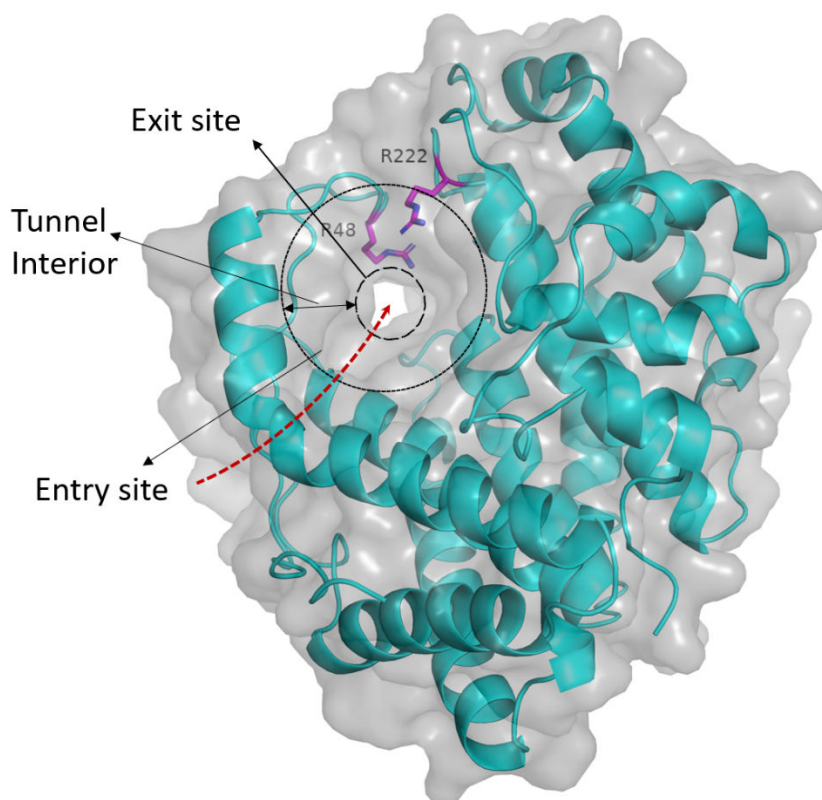
#### 4.3.2. Mutation at the exit site of the tunnel: R48G

The crystal structures of PanPL wild type shows that the R48 residue near the exit site of the tunnel attains various alternative conformers [Figure 3.9, Chapter 3]. We compared the exit site of 4OZV(AlgL), 7FHZ (Smlt1473), 7WXM (PanPL, pH 6.5) structures of PL-5

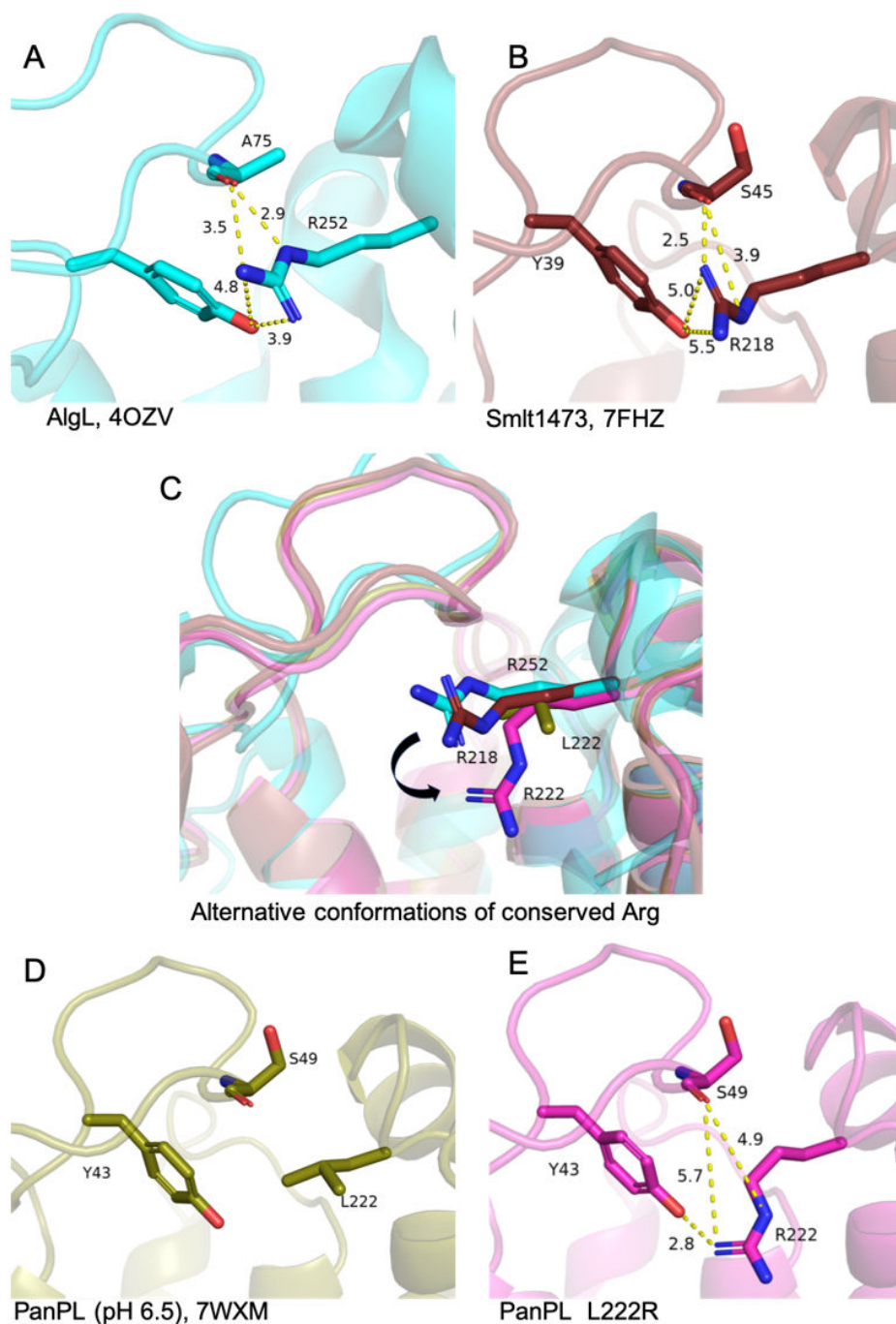
enzymes [Figure 4.4]. We observed alternative conformations for the Arg residue at the exit site in 4OZV and 7WXM structures. The Smlt1473 has a glycine residue in place of Arginine. We selected R48 of PanPL to create a mutation R48G to delineate its functional role.

#### 4.3.3. Mutation at the entry and exit site: L222R and R48G

Based on the structural comparison and sequence analysis with Smlt1473 we created the double mutant L222R R48G to understand the effect of these mutations in substrate binding, processivity and catalysis.



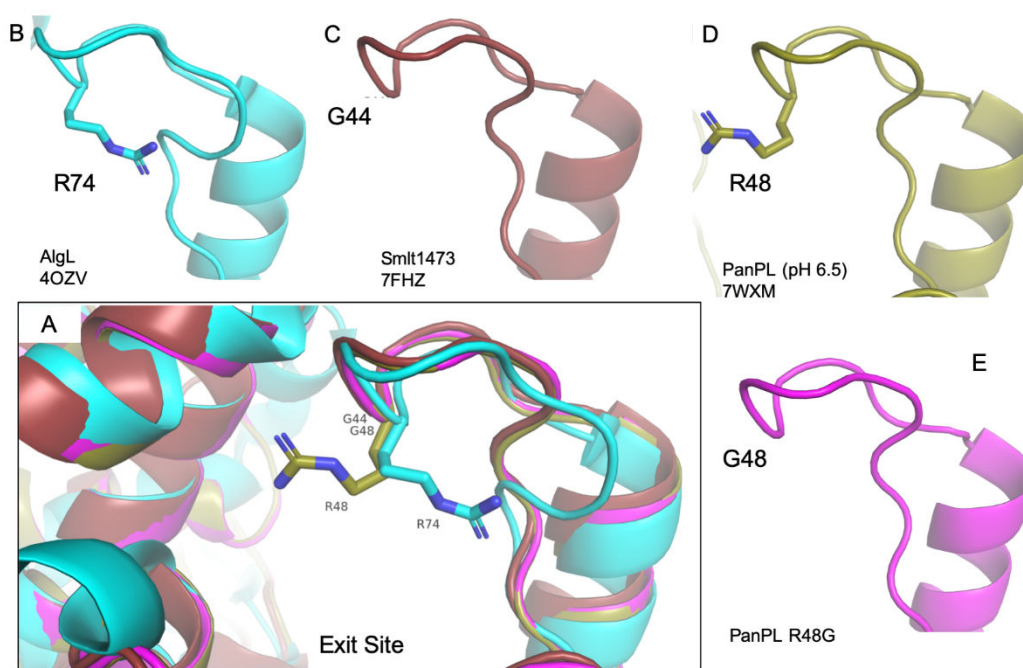
**Figure 4. 2. The surface representation of PanPL L222R mutation.** The figure shows the entry site, exit site and the interior part of the tunnel in surface and cartoon view. The R48 and R222 residues are shown in stick at the exit and entry site respectively.



**Figure 4.3. Hydrogen bonds between the conserved Arg and the N-loop-lid of PL-5 enzymes.** A. There two H bonds formed between R252 and A75 in 4OZV (AlgL) structure (cyan) ; B. In case of 7FHZ, there is one H bond between R218 and S45 (ruby). C. There are four crystal structures 4OZV(cyan), 7FHZ(ruby), 7WXM(smudge) and PanPL L222R(pink) are superimposed. The figure shows the alternative conformation of R222 of



PanPL L222R mutation in comparison to R252(4OZV) and R218(7FHZ). PanPL wildtype has L222. D. There is no H bond forms with L222 in case of 7WXM, PanPL wildtype structure at pH 6.5; E. There is no H bond formation between R222 and S49, but one H bond is formed Y43 and R222 in case of PanPL L222R mutant structure.



**Figure 4.4. The Arginine at the exit site of PL-5 enzymes;** A. There are four crystal structures i.e., 4OZV(AlgL, cyan), 7FHZ(Smlt1473, ruby), 7WXM(PanPL at pH 6.5, smudge) and PanPL L222R(pink) are superimposed. The figure shows alternative conformation of R48 (PanPL) and R74(AlgL), in addition to G44 (Smlt1473) and G48 (PanPL R48G) are shown at the N-loop-lid. B, C, D, E shows the individual N-loop-lid for 4OZV, 7FHZ, 7WXM and PanPL R48G respectively.

#### 4.4. Biochemical characterization of mutants

The mutations were generated using mutagenic primers [Table 2.1, Chapter 2] and the mutant proteins were purified by anionic exchange chromatography. The optimal pH of activity for the mutants were characterised by TBA assay for the substrates, alginate and

poly-ManA followed by enzyme kinetic measurement. The specific activities for the mutants were compared with wildtype enzyme [Figure 4.6].

#### **4.4.1. Biochemistry of L222R**

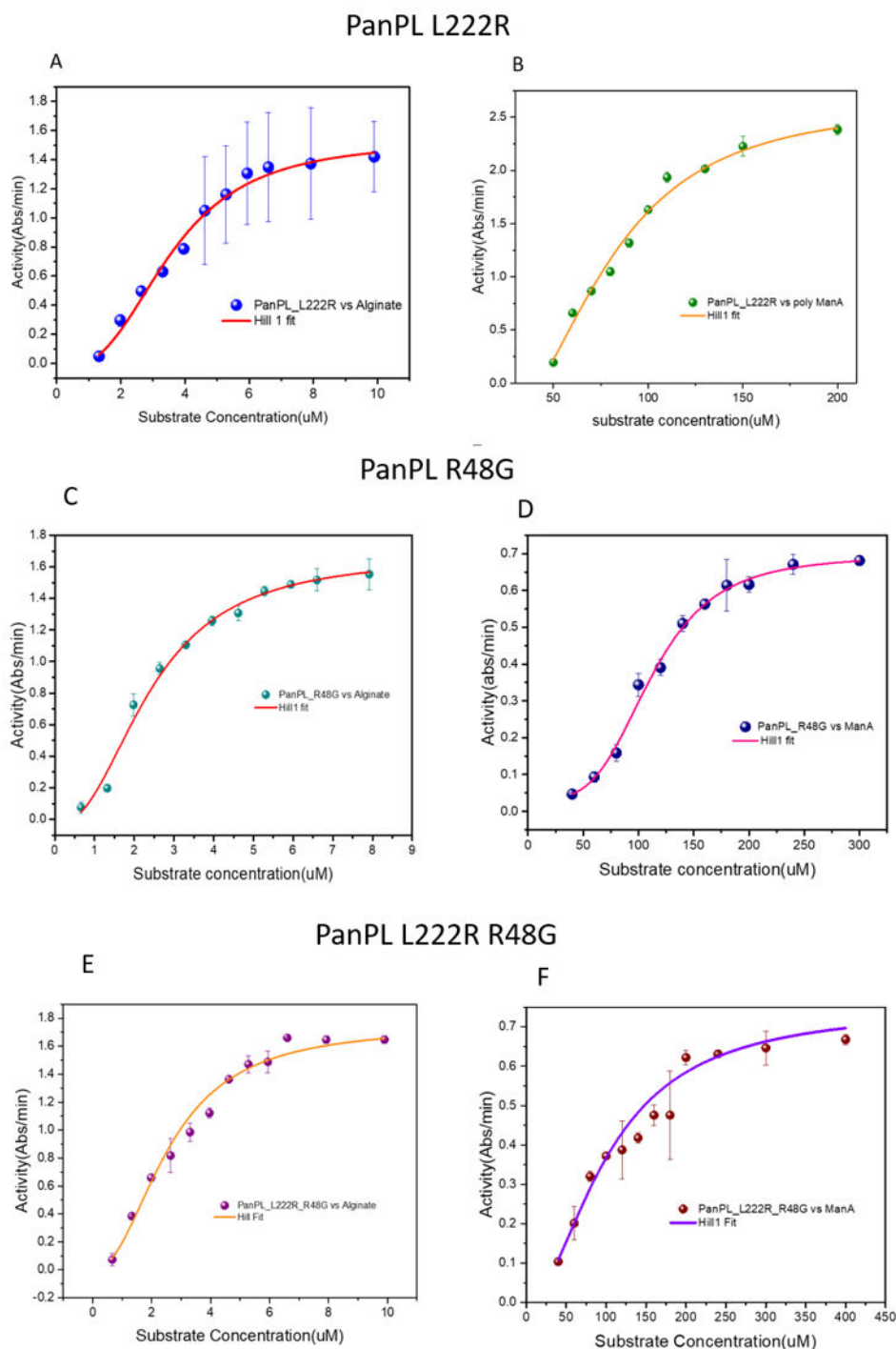
The optimum pH for PanPL L222R was found to be pH 7 for the substrates Alginate and poly-ManA. The enzyme kinetics for alginate and poly-ManA substrate has been performed for the mutant [Figure 4.5 A, B]. The L222R mutation shows an enhanced the overall enzyme activity for both of the substrates compared to wildtype PanPL. Specially for the poly-ManA the catalytic efficiency  $K_{cat}/K_M$  has been increased 14-fold. The specific activity for the poly-ManA substrate for PanPL L222R has been increased [Figure 4.6]. The Michaelis constant,  $K_M$  of the mutant is less compared to the wildtype which indicates that PanPL L222R is a better binder [Table 4.1].

#### **4.4.2. Biochemistry of R48G**

The R48G mutant shows optimal activity at pH 7 for alginate, while the optimum pH shifted to 8.5 in case of poly-ManA [Table 4.1]. The enzyme kinetic plot has been shown [Figure 4.5 C, D] for both the substrates. The overall activity compared to wildtype has not changed much.

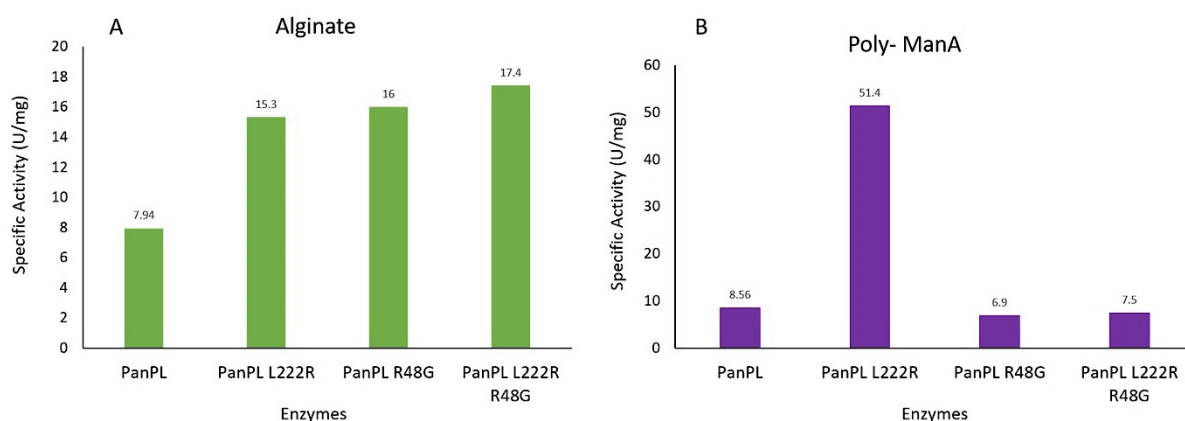
#### **4.4.3. Biochemistry of L222R R48G**

In case of double mutant L222R R48G the optimum pH was found to be pH 7. The enzyme kinetic assay was performed for both of the substrates, alginate and poly-ManA [Figure 4.5 E, F] and the allostery was diminished [Table 4.1] although the overall activity did not change much.



**Figure 4.5. Enzyme kinetic plots for PanPL mutants.** A. PanPL L222R vs Alginate and B. PanPL L222R vs Poly- ManA. The assay was performed at 25°C with pH 7. For alginate substrate, 100µg/ml enzyme concentration was used, whereas for Poly-ManA substrate, 50µg/ml enzyme concentration was used to record the enzyme activity. C. PanPL R48G vs

Alginate and D. PanPL R48G vs poly-ManA. The assay was performed at 25°C with pH 7 for alginate substrate, whereas for poly-ManA substrate the assay was kept at pH 8.5, 100µg/ml enzyme concentration was used to record the enzyme activity. E. PanPL vs Alginate and B. PanPL vs poly- ManA. The assay was performed at 25°C with pH 7. For each substrate, 100µg/ml enzyme concentration was used to record the enzyme activity. The equation used for the sigmoidal curve fitting was  $y = START + (END - START) \frac{x^n}{k^n + x^n}$ , which is a modified version of Hill equation.



**Figure 4.6. The specific activities of PanPL WT with mutants for different substrates.**

A. The plot compares the specific activities for alginate, the mutations enhanced specific activity compared to PanPL-WT. B. The plot compares the specific activities for poly-ManA substrate. The PanPL L222R mutation enhanced the specific activity towards poly-ManA.

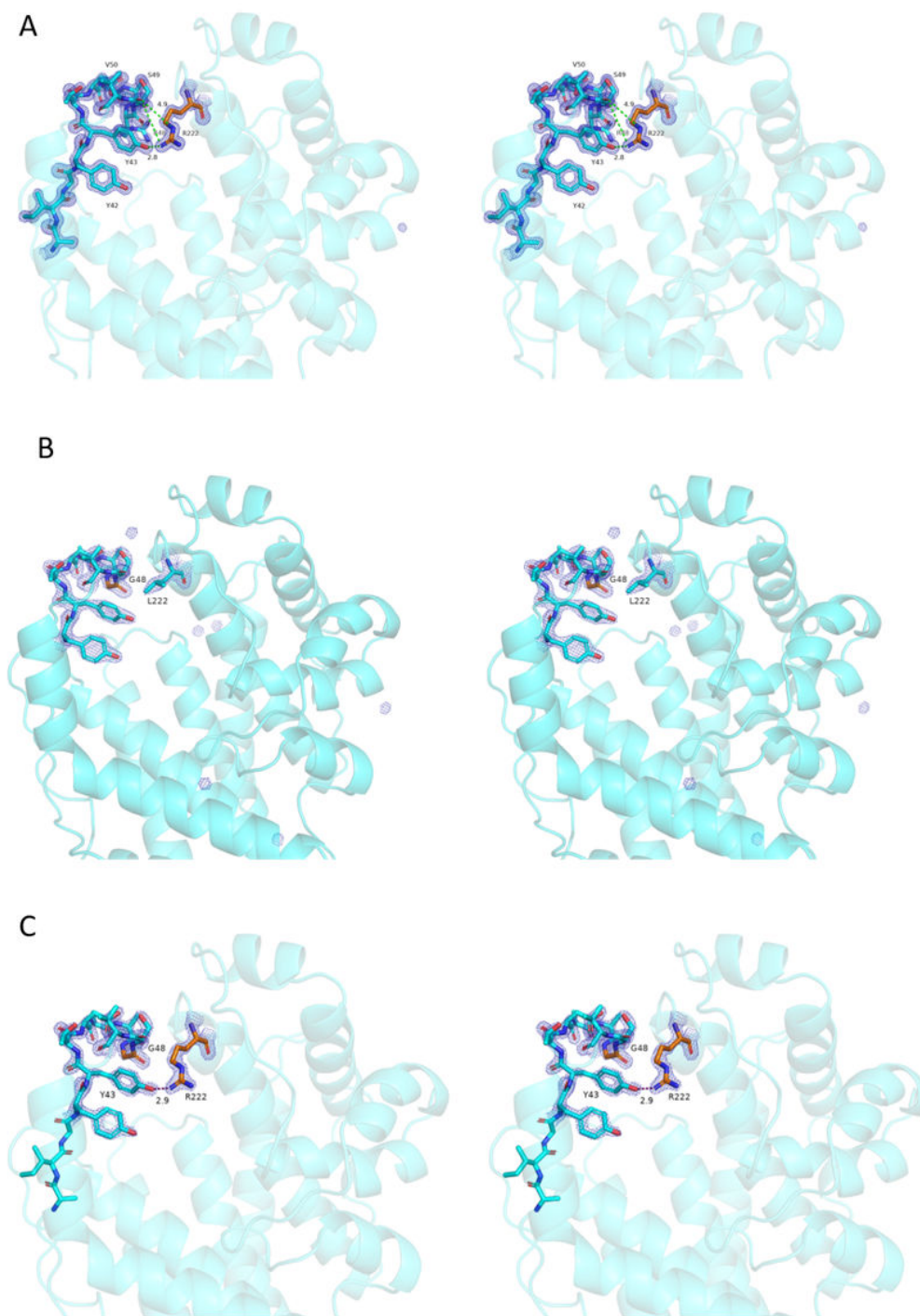
Table 4.1. The Enzyme kinetic parameters for PanPL mutants

Substrates	Enzymes	Optimum pH	Enzyme concentration (μg/ml)	Hill's co-efficient n	$K_{cat}$ (s <sup>-1</sup> )	$K_M$ (mM)	$K_{cat}/K_M$ (mM <sup>-1</sup> s <sup>-1</sup> )
Alginate	PanPL WT	7	100	4.4	0.766	0.0039	196
	PanPL L222R	7	100	2.8	1.47	0.0035	420
	PanPL R48G	7	100	2.32	1.6	0.002	666
	PanPL L222R R48G	7	100	2.17	1.6	0.0025	640
Poly-ManA	PanPL WT	7	100	2.7	0.825	0.175	4.7
	PanPL L222R	7	50	2.65	4.9	0.073	67.12
	PanPL R48G	8.5	100	3.83	0.66	0.111	6
	PanPL L222R R48G	7	100	1.87	0.72	0.102	7

## 4.5. Crystal structures of PanPL mutants

The crystals of the PanPL mutants were attempted to grow at different pH values. However, we could obtain crystals in a few pH values of solution for the mutants [Table 4.2]. The diffraction data sets were collected at ESRF, ID30A-1 beamline. [Table 4.2]. We utilized Molecular Replacement (MR) method to solve the phase problem using wildtype PanPL structure as search model. The model building and refinement iteratively were performed using *phenix* (55) software, and electron density map was traced in visualization program *Coot* (56).

We analysed the side chain conformation of the R222 residue. Based on the sequence and structural information we expected H-bond formation between the R222 side chain with the N-loop-lid main chain, which have been conserved among PL-5 structures [Figure 4.3]. The crystal structure of PanPL L222R and PanPL L222R R48G showed no H-bond formation between Ser 49 main chain and L222 side chain, whereas a new H-bond has been established between L222 and Y43 residues [Figure 4.7].



**Figure 4.7. The crystal structure of PanPL mutants.** A. PanPL L222R, B. PanPL R48G and C. PanPL L222R, R48G in stereo views shows the electron density map of the mutant residues with their side chain in stick models. The side chain atoms of the neighbouring residues are shown in stick. B, C. G48 has no side chain atoms. The side chain conformation of Arg222. There is no H-bonds with the N-loop-lid main chain.

Table 4.2. Crystallographic data collection and refinement statistics for mutants

Crystallographic data collection and refinement statistics			
Crystal Name	PanPL_L222R	PanPL_R48G	PanPL_L222R_R48G
Crystallization condition	0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350	0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350	0.1 M Sodium acetate trihydrate pH 4.5, 25% w/v Polyethylene glycol 3,350
<b>Data collection statistics</b>			
Beamline	ID30A-1	ID30A-1	ID30A-1
Wave length (Å)	0.96777	0.96777	0.96777
Detector	Eiger X 4M	Eiger X 4M	Eiger X 4M
Processing software	XDS/autoproc	XDS/autoproc	XDS/autoproc
Cell Parameter	a=35.96(Å), b=84.13(Å), c=91.12(Å)	a=36.14(Å), b=84.96(Å), c=91.48(Å)	a=35.93(Å), b=84.46(Å), c=90.42(Å)
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution range (Å)	42.06-1.35(1.37-1.35)*	62.25-1.99(2.02-1.99)*	45.21-1.77(1.8-1.77)*
Rmerge	0.128(0.878)*	0.096(0.743)*	0.170(1.408)*
Unique reflections	61622(2993)*	20008(968)*	27212(1341)*
I/σ(I)	8.9(2.6)*	16.8(2.2)*	13.4(2.2)*
Completeness (%)	99.9(99.1)*	99.9(98.6)*	98.6(97.2)*
Redundancy	8.9(8.6)*	12.5(7.1)*	20.1(20.2)*
CC(1/2)	0.995(0.818)*	0.999(0.756)*	0.999(0.849)*
<b>Data Refinement Statistics</b>			
Refinement	phenix	phenix	phenix
Rwork/Rfree	0.1835/0.2057	0.1760/0.2224	0.1794/0.2243
ΔR= Rwork-Rfree	0.0222	0.0464	0.0449
Matthews Coefficient	1.97	2.01	1.96
Number of molecules in asymmetric unit	1	1	1
Solvent content (%)	37.57	38.73	37.28
<b>Number of atoms</b>			
Proteins	2466	2409	2432
Water and other (ligands or ions)	338	155	255
<b>Overall B-factor (Å<sup>2</sup>)</b>			
Proteins	14.088	28.573	21.241
ligands or ions		41.2	
Water	19.2	34	29.4
<b>RMSD from ideal values</b>			
rms bond length (Å)	0.006	0.008	0.007
rms bond angle (°)	0.901	0.895	0.896
<b>Ramachandran</b>			
Favored (%)	97.4	96.13	97.42
Allowed (%)	2.6	3.87	2.58
outlier (%)	0	0	0
R <sub>work</sub> = $\sum   F_o - F_c  /\sum  F_o $			
R <sub>merge</sub> = $\sum   F_o - F_c  /\sum  F_o $			
*Values in parentheses are for the highest resolution shell			

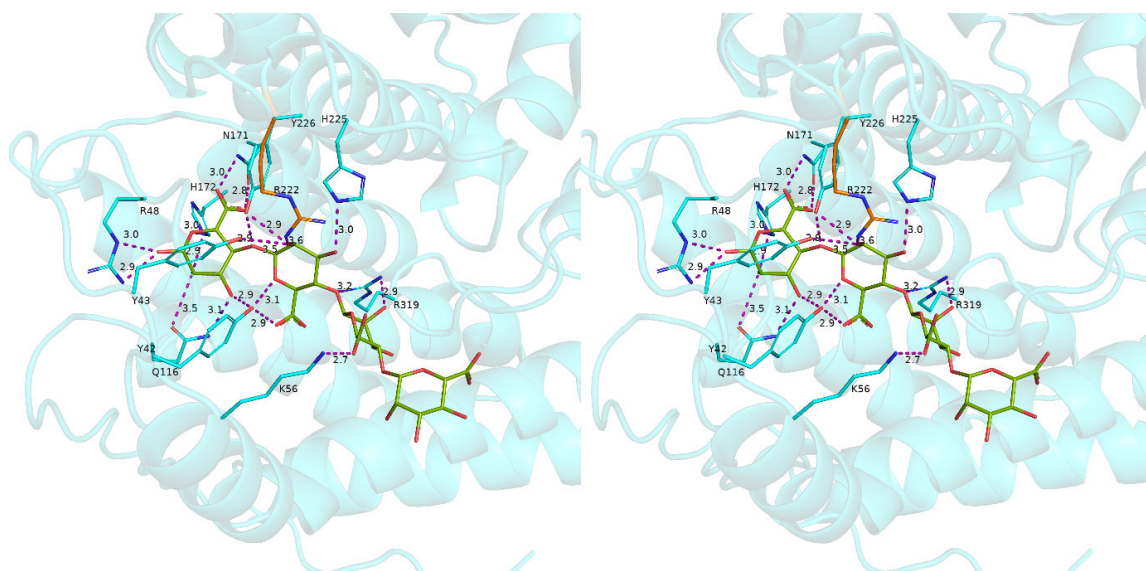
R<sub>free</sub> is the R<sub>work</sub> value for 5% of the reflections excluded from the refinement.

\*Values in parentheses are for the highest resolution shell



#### 4.6. Docking of tetra-ManA with PanPL mutants

We observed the enzyme kinetic parameters of the L222R mutant, which shows L222R is a better substrate binder. In order to visualize the interaction of R222 residue with the substrate we performed the docking using the protocol implemented in Rosetta Flexible docking. The final docked structure PanPL L222R-tetra-ManA complex was sorted in comparison with the crystal structure of wildtype PanPL-tetra ManA complex. We analysed the polar interaction pattern of the enzyme-substrate complex. Table 4.3 and Figure 4.8 show possible interactions between substrate and protein. R222 found to have one polar interaction with the second sugar ring at [-1] subsite.



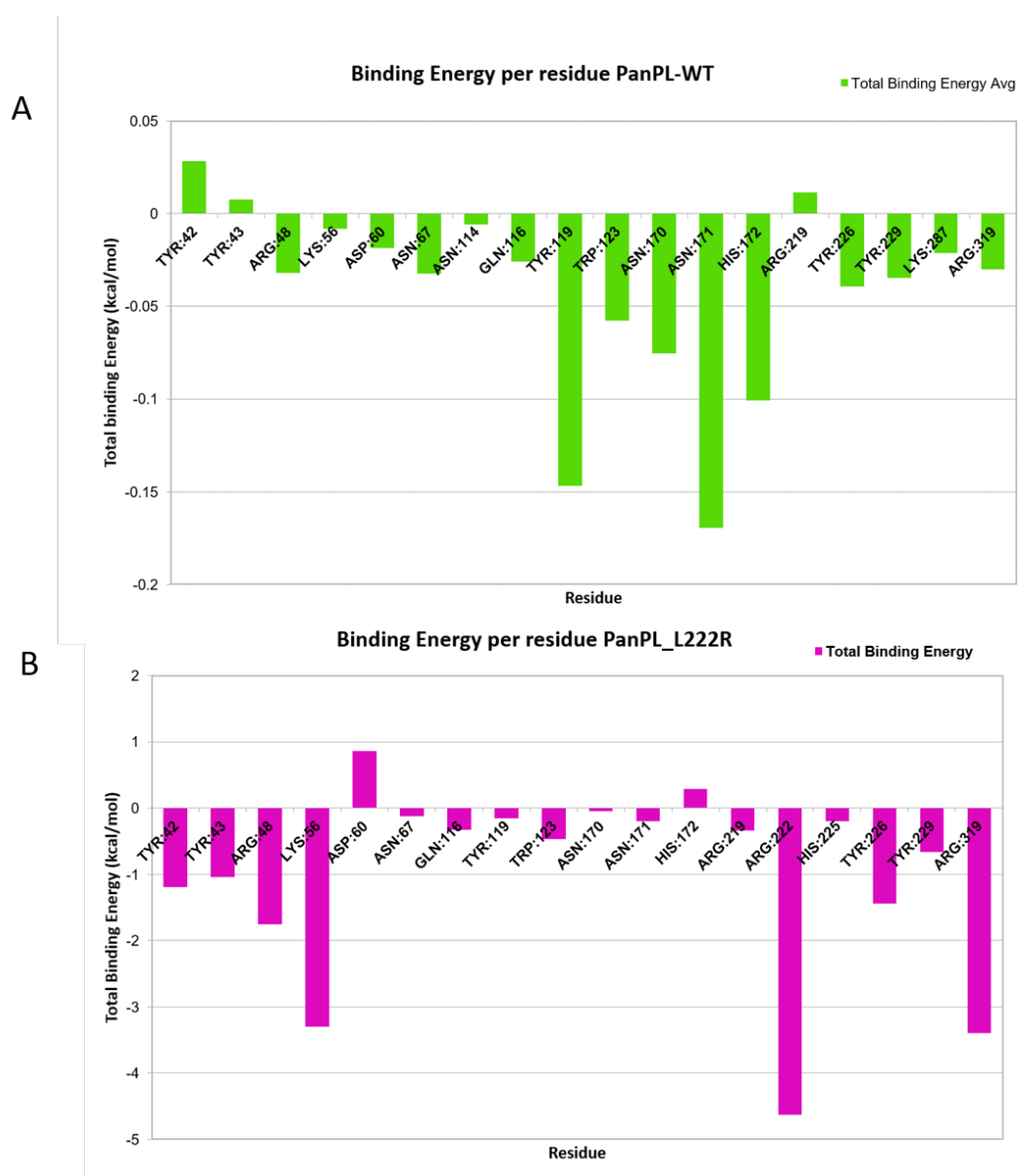
**Figure 4.8. Interaction between PanPL L222R mutant with tetra-ManA in stereo view.** The substrate interacting residues in cyan sticks, the R222 residue in orange stick and the tetra-manA in green sticks are shown. The polar interactions are shown with dashed lines and labelled with distance values in Å. R222 has one interaction with the second sugar subunit at [-1] subsite.

Table 4.3. PanPL L222R and tetra-ManA docked structure interactions.

ATOM:Residue	ATOM: Substrate	Subsite	Interaction Type	Distance (Å)
ND2:Asn171	O7:BEM1	[+1]	H-BOND	3.0
OD1:Asn171	O6:BEM1	[+1]	H-BOND	2.8
NE2:His172	O5:BEM1	[+1]	H-BOND	3.0
NE2:His172	O2:BEM1	[+1]	H-BOND	2.9
NE:Arg48	O1:BEM1	[+1]	H-BOND	3.0
NH2:Arg48	O1:BEM1	[+1]	H-BOND	2.9
OE1:Gln116	O2:BEM1	[+1]	H-BOND	3.5
NE2:Gln116	O3:BEM1	[+1]	H-BOND	3.1
OH:Tyr42	O3:BEM1	[+1]	H-BOND	2.9
OH:Tyr226	O4:BEM1	[+1]	H-BOND	3.1
OH:Tyr226	O2:BEM2	[-1]	H-BOND	3.1
NE2:His225	O3:BEM2	[-1]	H-BOND	2.9
OH:Tyr42	O5:BEM2	[-1]	H-BOND	3.1
OH:Tyr42	O6:BEM2	[-1]	H-BOND	2.9
NH1:Arg319	O4:BEM2	[-1]	H-BOND	3.2
NH2:Arg319	O2:BEM3	[-1]	H-BOND	3.1
NZ:Lys56	O6:BEM3	[-2]	H-BOND	2.7

#### 4.7. Binding energy calculation for PanPL L222R mutant (MM-PBSA)

We performed an MM-PBSA calculation for the tetra-ManA docked structure of PanPL WT and PanPL\_L222R mutant in order to analyse the residue wise contribution to the substrate binding. We used the Gromacs MD simulation trajectories and used Amber MM-PBSA calculation on it utilizing gmx MM-PBSA module. The decomposition analysis shows the residue wise contribution to substrate binding [Figure 4.9]. The result suggests that the contribution of L222R to the substrate binding is significant.

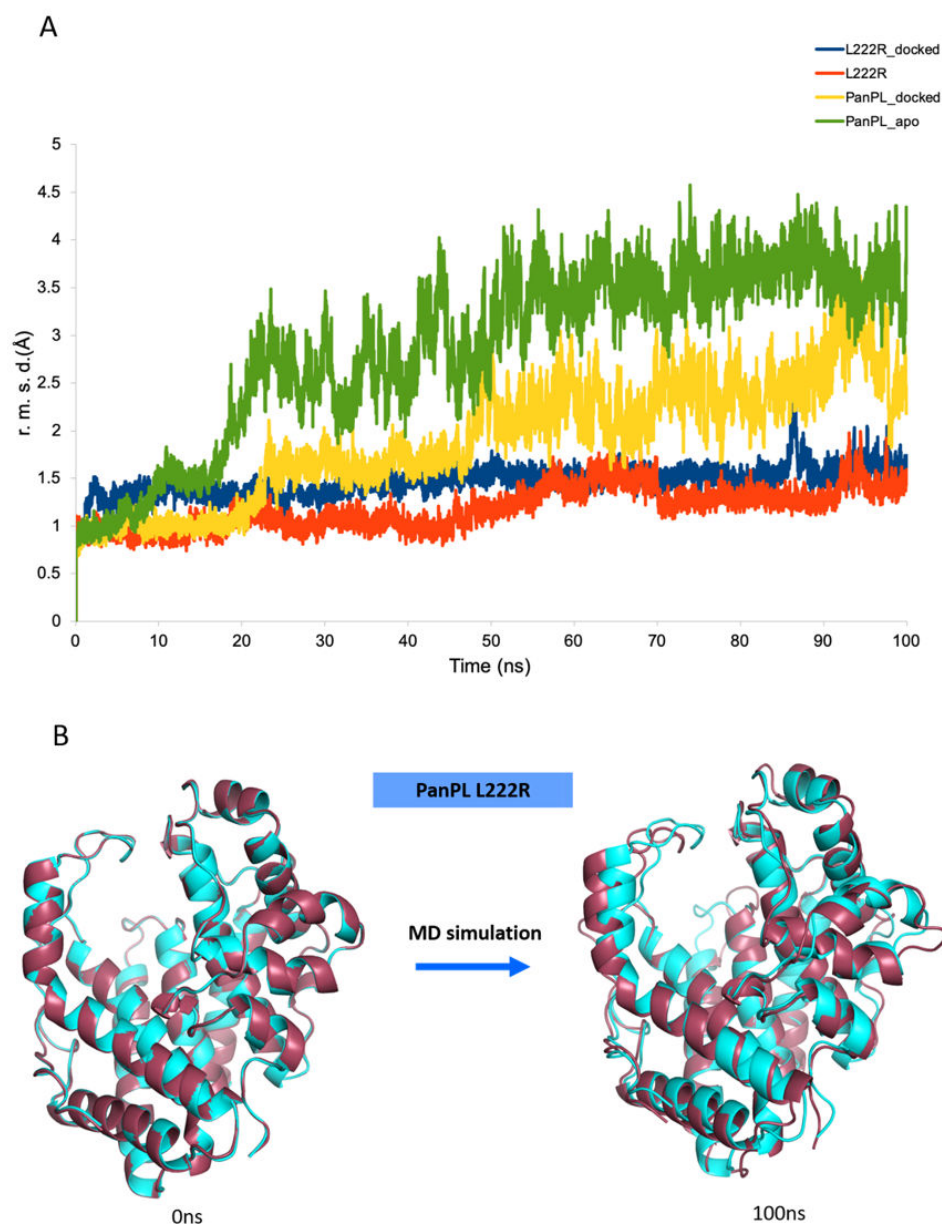


**Figure 4.9. Binding energy decomposition analysis for PanPL L222R mutant:** A. The residue wise contribution to the binding energy for the PanPL-WT- tetra ManA docked structure. B. PanPL L222R-tetra ManA docked structure. The result shows that L222R mutation contributes significantly the substrate binding.

#### 4.8. Dynamics of PanPL L222R

We performed the molecular dynamics study for 100ns on the PanPL L222R mutant crystal structure in apo form and the tetra-ManA docked structure. The evolution of the trajectories in both of the simulations shows no N-loop-lid opening to access the open state. The

catalytically competent closed structure is maintained throughout the simulation. We compared the r. m. s. d. of the mutant with the wildtype apo structure [Figure 4.10]. The analysis shows no jump in the r. m. s. d. corresponding to N-loop-lid opening which is previously observed in the wildtype simulations.

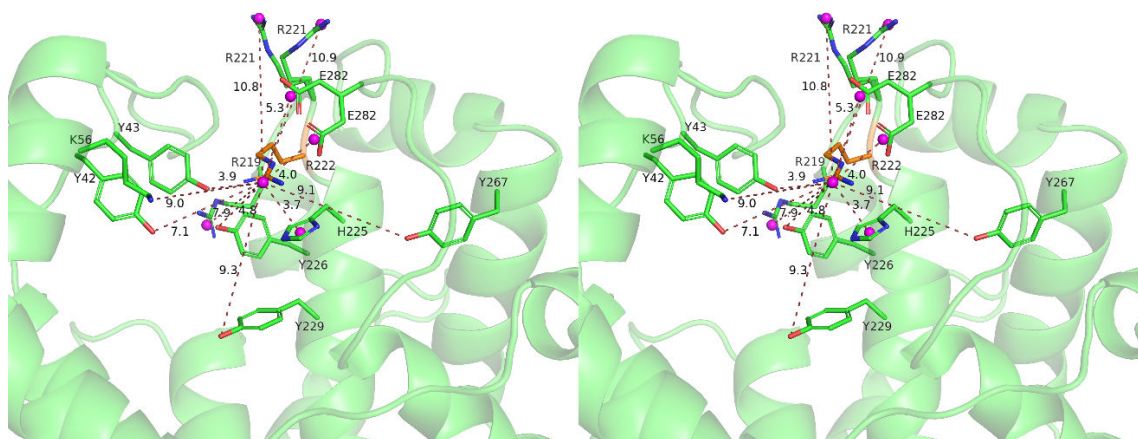


**Figure 4.10. The dynamics of PanPL L222R mutation.** A. The r. m. s. d. of the PanPL L222R structure in apo form (orange), PanPL L222R docked structure (blue), and the wildtype PanPL apo structure (green) and tetra ManA docked structure (yellow) throughout

the MD simulation. B. Comparison of initial structure with final structure during 100ns simulations: The starting structure at 0ns (cyan) and the structure at the 100ns (ruby) showing no open state during the trajectories. It means the closed state maintains itself throughout the simulation. The L222R mutation alters the dynamics property of the wildtype PanPL.

#### 4.9. Electrostatic network

The MD simulation trajectories of PanPL L222R couldn't access the open state. This suggests that the mutation has altered the intrinsic dynamics of the wildtype PanPL. The L222R mutation has arrested the catalytically competent closed state [Figure 4.10]. To understand the effect of mutation on the N-loop-lid dynamics, we studied the newly established inter-residue columbic interactions between the R222 residue and other flanking residues [Figure 4.12]. We modelled the electrostatic interactions using the list prepared by our PROPKA analysis [Table 4.4].



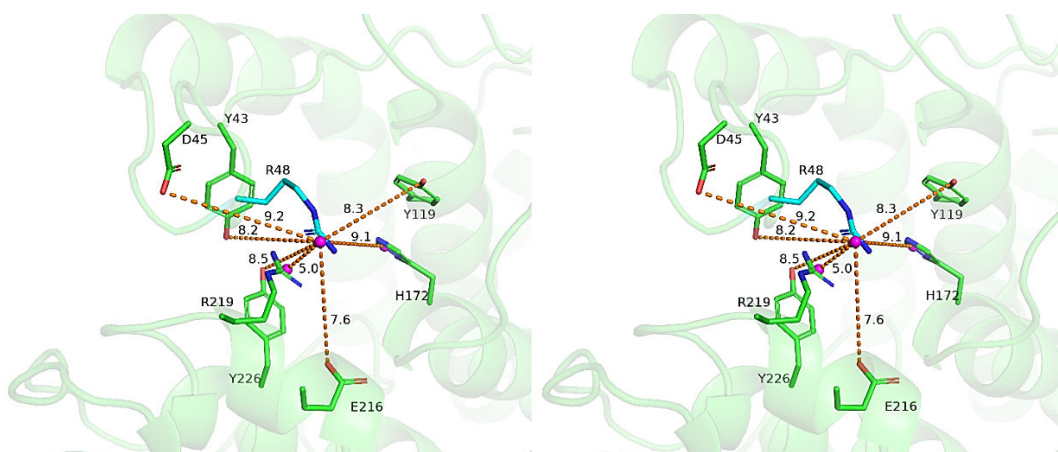
**Figure 4.11. The coulombic interaction network for R222 residue.** The figure shows the electrostatic network established between R222 with neighbouring residues. The dashed line labelled with distance in Å shows electrostatic interactions.

Table 4.4. list of residues interacting with R222 by coulombic interaction

Residue	Interacting residues	Distance(Å)
ARG 222	TYR 42	7.1
	TYR 226	4.8
	TYR 229	9.3
	TYR 267	9.1
	TYR 43	3.9
	GLU 282	4, 5.3
	HIS 225	3.7
	ARG 219	7.9
	ARG 221	10.8, 10.9
	LYS 56	9.0

#### 4.10. Electrostatic network of R48 residue

We modelled the electrostatic interaction network for R48 residue in wildtype PanPL. The mutation R48G disrupts the preorganised electrostatic network [Figure 4.12, Table 4.5]. By removing the charged residue, the overall positive charge of the catalytic tunnel is depleted.



**Figure 4.12. Electrostatic interaction network of R48.** Figure shows the network of R48 with the flanking residues in the PanPL wildtype apo structure. The dashed line labelled with the distance in Å represent the electrostatic interactions.

Table 4.5. list of residues interacting with R48 by coulombic interaction

Residue	Interacting residues	Distance(Å)
<b>ARG 48</b>	<b>TYR 43</b>	<b>9.2</b>
	<b>ASP 45</b>	<b>8.2</b>
	<b>TYR 119</b>	<b>8.3</b>
	<b>GLU 216</b>	<b>7.6</b>
	<b>TYR 226</b>	<b>8.5</b>
	<b>ARG 219</b>	<b>5.0</b>
	<b>HIS 172</b>	<b>9.1</b>

These interactions are lost because of R48G mutation

#### 4.11. Dynamics of PanPL R48G mutant

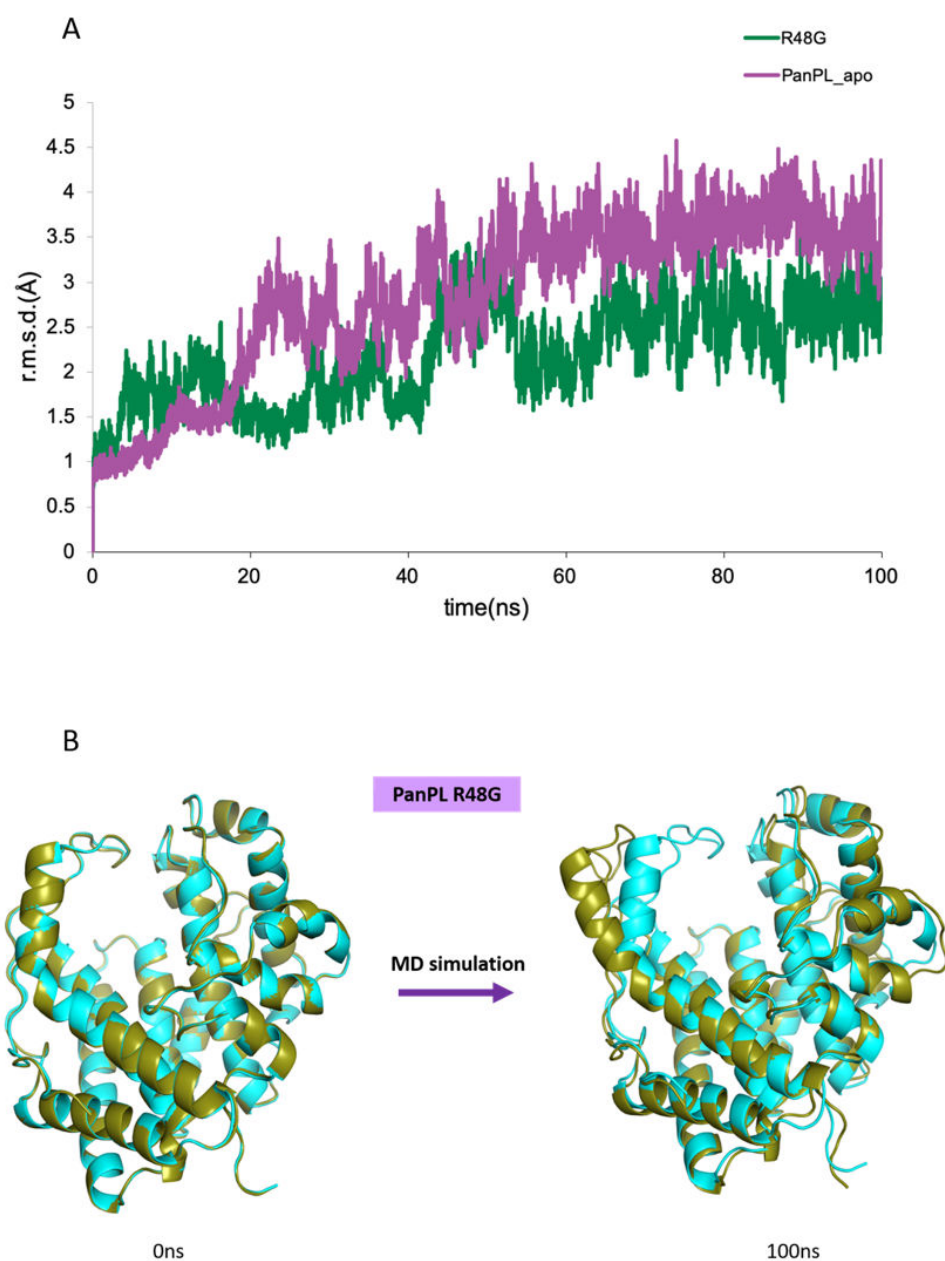
The MD simulation of PanPL R48G mutant structure for 100ns has been performed. The dynamic properties of PanPL R48G mutant shows the similar trend as the wildtype PanPL. The N-loop-lid opens during the simulation and the open state has been accessed. [Figure 4.13].

#### 4.12. PanPL R48G shifts the pKa of Y226

The biochemical characterization of R48G mutant shows a shift of optimum pH of activity for poly-ManA from pH 7 to pH 8.5, whereas for alginate the optimum pH remains pH 7. We wanted to see the effect of R48G mutation on the active site residues. Our PROPKA analysis predicts an increase in the pKa of active site residue Y226 (pKa of Y226 in PanPL pH 6.5. is 10.59 and pKa of Y226 in PanPL R48G is 15.23). Further, we modelled the electrostatic network of Y226 residue in the wildtype and the R48G mutant [Figure 4.14]. The interaction patterns are significantly different from each other [Table 4.6]. As the electrostatic interactions are important to determine the pKa of the residues, our study

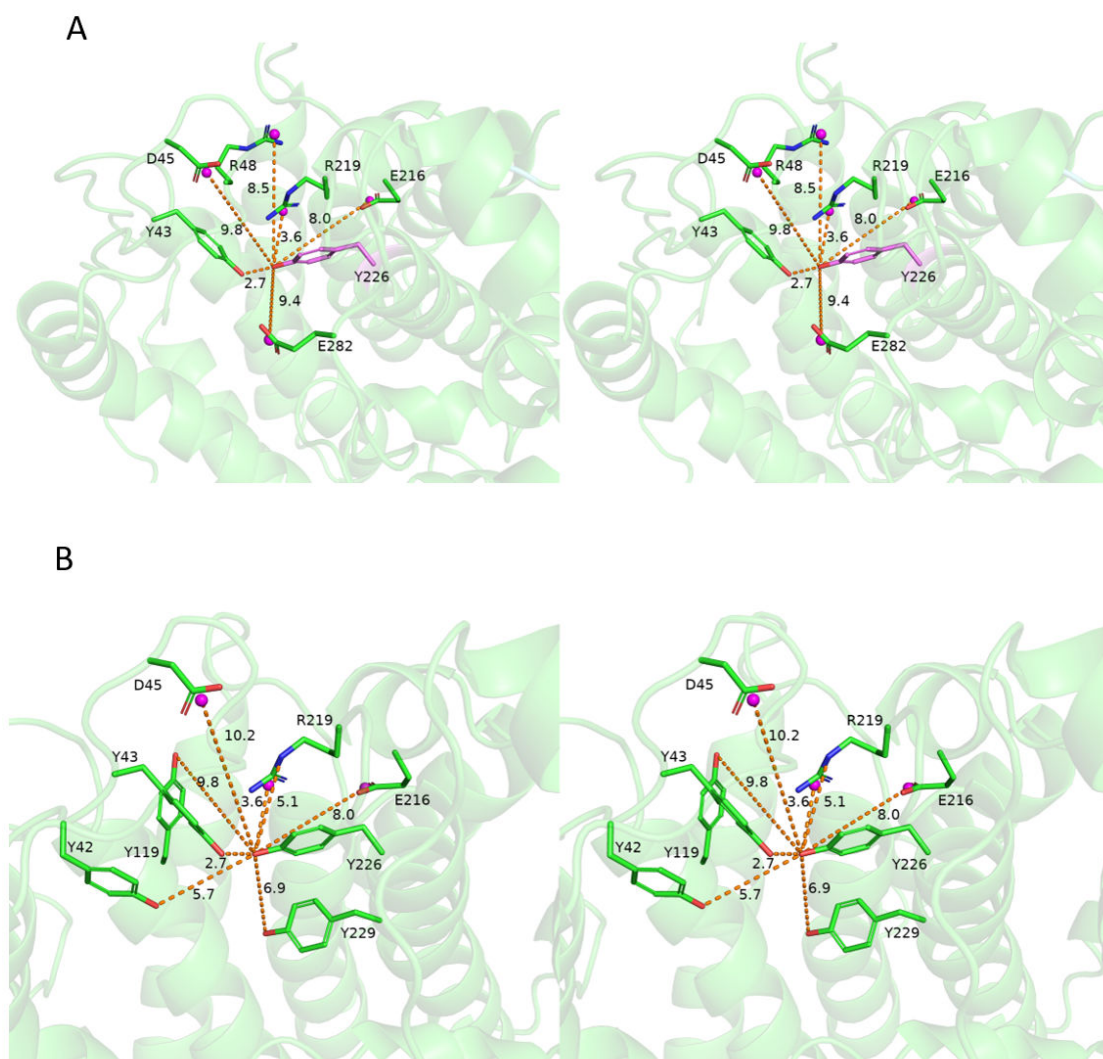


suggests the rewiring of the electrostatic network of the active site due to the mutation can play an important role in shift in optimum pH. Hence, we denote R48G as a pH switch.



**Figure 4.13. The r. m. s. d. of the PanPL R48G mutant** A. Figure shows comparison in r. m. s. d. of MD trajectories of PanPL R48G (green), and the wildtype PanPL apo structure (violet). B. The structure at 0ns cyan and 100 ns (smudge) during the MD simulation of PanPL R48G mutation are compared to show the N-loop-lid opening i.e., the open state is accessible.





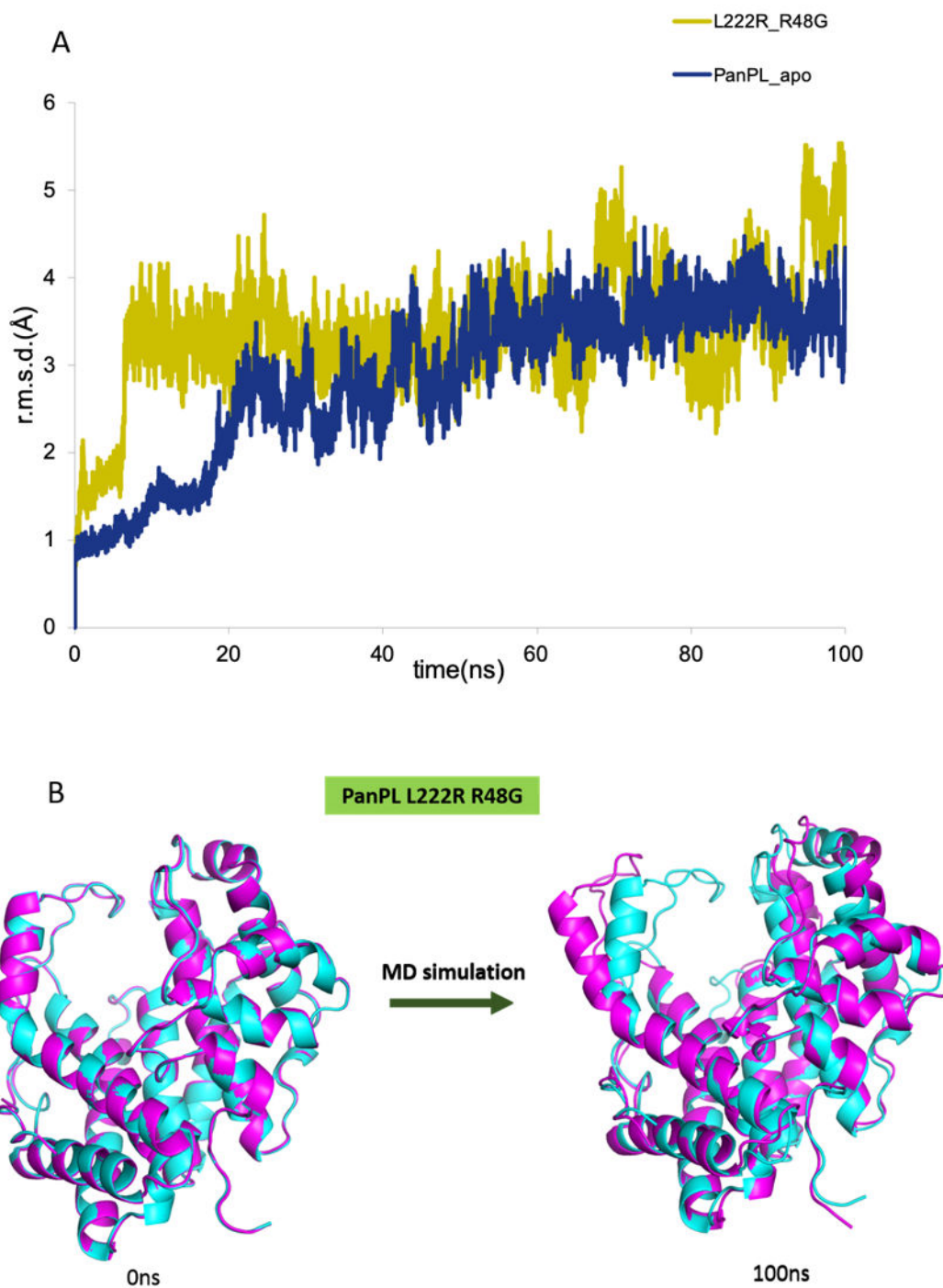
**Figure 4.14. The electrostatic interaction network of Y226** A. The coulombic interactions of Y226 with neighbouring residues in wildtype PanPL are shown in dash labelled with distance in Å, B. The coulombic interactions are shown in dash labelled with distance in Å for PanPL R48G. The electrostatic interaction patterns are different. The mutation R48G rewires the electrostatic network of Y226.

Table 4.6. list of residues interacting with Y226 by coulombic interaction

Enzyme	Residues	Distance from TYR 226 (Å)
<b>Wildtype PanPL</b>	TYR 43	2.7
	ASP 45	9.8
	ARG 48	8.5
	GLU 216	8.0
	GLU 282	9.4
	ARG 219	3.6
<b>PanPL R48G</b>	TYR 43	2.7
	ASP 45	9.8
	TYR 42	5.7
	TYR 119	9.8
	GLU 216	8.0
	TYR 229	6.9

### 4.13. Dynamics of L222R R48G mutant

The MD simulation for 100ns was performed for PanPL L222R R48G crystal structure. The double mutant PanPL L222R R48G shows similar dynamics as wildtype PanPL [Figure 4.15]. It accesses the open state during the simulation. Whereas the overall electrostatic interaction pattern maintains the interactions of the R222 residues and lost the interactions for the R48G mutation [Table 4.7].



**Figure 4.15. The r. m. s. d. of the PanPL L222R R48G** A. The figure shows the comparison between r. m. s. d. of the MD trajectories of PanPL L222R R48G structure (pale yellow), PanPL WT apo form (blue). B. The structure at 0ns cyan and 100 ns (magenta) during the MD simulation of PanPL L222R R48G mutation are compared to show the N-loop-lid opening i.e., the open state is accessible.

**Table 4.7. The list of newly formed and lost interactions due to mutation in PanPL L222R R48G structure**

Residue	Interacting residues	Distance(Å)	
ARG 222	TYR 42	7.1	These interactions are newly formed because of L222R mutation.
	TYR 226	4.8	
	TYR 229	9.3	
	TYR 267	9.1	
	TYR 43	3.9	
	GLU 282	4, 5.3	
	HIS 225	3.7	
	ARG 219	7.9	
	ARG 221	10.8, 10.9	
	LYS 56	9.0	
ARG 48	TYR 43	9.2	These interactions are lost because of R48G mutation.
	ASP 45	8.2	
	TYR 119	8.3	
	GLU 216	7.6	
	TYR 226	8.5	
	ARG 219	5.0	
	HIS 172	9.1	

#### 4.14. Discussion

Site Directed mutagenesis is considered as a powerful technique to edit the enzyme functionality directly at the sequence level. The comparative study of PL-5 crystal structures provided the information about the H-bond interactions between the N-loop-lid and an Arginine residue at the entry site of the catalytic tunnel [Figure 4.3]. The sequence analysis shows the Arginine is evolutionary conserved among PL-5 enzymes, whereas PanPL has lost the Arg during evolution [Figure 4.1]. There is another Arginine at the exit site of PanPL i.e., which attains alternative conformation as a function of pH in PanPL apo

crystals and highly dynamic in PanPL substrate bound crystal structure [Figure 3.9, Chapter 3]. The homologous enzyme Smlt1473 has a Glycine residue at the exit site instead of Arginine [Figure 4.4]. Based on the sequence and structural analysis, we created total three mutations PanPL L222R, PanPL R48G and PanPL L222R R48G. The mutant crystal structures have been solved. In the PanPL L222R structure, we observed the R222 accessing a new alternative conformation compared to other PL-5 structures and established a new H-bond interaction with Y43 [Figure 4.3.C].

#### **4.14.1. Mutation modulates the biochemistry**

The biochemistry of PanPL L222R shows an enhanced catalytic activity for both of its substrate alginate and poly ManA. The lower value of  $K_M$  compared to wildtype PanPL, for the poly-ManA substrate signifies that L222R mutant is a better binder [Table 4.1]. So, we were interested to evaluate the contribution of R222 in substrate binding. We performed docking of PanPL L222R crystal structure with tetra ManA substrate. Our docking study shows only one polar interaction between the substrate and the R222 [Figure 4.8, Table 4.3]. However, the binding energy decomposition analysis of the MMPBSA calculation showed a pronounced contribution of R222 to the substrate [Figure 4.9]. Our analysis suggests that replacing the hydrophobic L222 with the positively charged R222 at the entrance of the tunnel makes the mutant a better binder for the substrates by modulating electrostatic environment. The catalytic efficiency of PanPL L222R has been increased 14-fold for the poly-ManA substrate compared to PanPL wildtype. We suggest that the modulation of positive charge in the tunnel might play an important role in enhancing catalytic efficiency of the PanPL L222R for anionic poly ManA substrate. The next mutation R48G has been created at the exit site of the tunnel. The R48G mutation makes a wider exit site as G48 lacks side chain. Interestingly the biochemistry of PanPL R48G mutant shows a shift in the optimum pH for the poly-ManA substrate from pH 7 to pH 8.5

in contrast to wildtype. Our pKa prediction by PROPKA shows alteration in the pKa of the Y226 catalytic residue due to R48G mutation. We modelled the electrostatic network for Y226 for both PanPL wildtype and PanPL R48G mutant [Figure 4.14]. The networks show significant changes in the electrostatic interaction patterns. We observed an increasing trend in the pKa of Y226 in PROPKA calculation. This might have influenced the optimum pH of the catalysis to shift to pH 8.5 for poly- ManA. The R48G mutation acts like a pH switch for the enzyme. The double mutant PanPL L222R R48G has maintained the optimum pH for both of its substrates to pH 7. However, the allosteric behaviour of the double mutant has decreased significantly compared to the wildtype.

#### **4.14.2. Mutation alters the dynamics**

The wildtype PanPL has an intrinsic dynamic to access the open state during the MD simulations [Figure 3.28, Chapter 3]. The closed state is catalytically competent and the open state might transiently accessible due to inherent flexibility of N-loop-lid [Figure 3.37, Chapter 3]. Our MD simulation studies on the three mutant structures i.e., PanPL L222R, PanPL R48G and PanPL L222R R48G have been shown interesting dynamic behaviour. The PanPL L222R structure maintains the catalytically competent closed state throughout the simulation without accessing the open state [Figure 4.10]. However, the PanPL R48G and PanPL L222R R48G mutants show similar dynamic properties of PanPL wildtype [Figure 4.13; 4.15]. To understand the altered dynamics of PanPL L222R, we analysed the mutant crystal structures carefully. As discussed previously, there is a newly established H-bond formation with the Y43 residue of N-loop-lid and R222 residue in both PanPL L222R and PanPL L222R R48G crystal structures [Figure 4.7]. Despite of the H bond formation the PanPL L222R R48G mutant accesses the open state, whereas the PanPL L222R maintains the closed state. We suggest that only one H-bond formation is not enough to hold the loop in its closed state. In order to investigate the other stabilizing

factors for the closed state, we analysed the electrostatic network of the R222 residue. We observed the mutation L222R added a few extra columbic interactions in the N-loop-lid and the catalytic tunnel region in both PanPL L222R and PanPL L222R R48G mutant structures [Figure 4.11]. In addition to that, we listed the electrostatic interactions of the R48 in wildtype-PanPL and PanPL L222R structures [Figure 4.12], which has been disrupted by the R48G mutation in the PanPL R48G and PanPL L222R R48G structures. The careful observation shows that when both R222 and R48 are present (PanPL L222R structure) in the catalytic tunnel, the tunnel maintains the closed state, whereas if any of the Arginines is missing (PanPL, PanPL R48G, PanPL L222R R48G structures) the open state becomes accessible [Table 4.8]. Our study suggests that the newly wired electrostatic network of R222 and preorganised network of R48 in PanPL L222R structure help to hold the N-loop-lid like a clutch, and arrest the closed state. The stabilization of the catalytically competent closed state then contributes to the enhanced catalytic efficiency of the mutant enzyme.

Table 4.8. Summary of the PanPL mutants

Protein	R48	R222	Optimum pH for poly-ManA	Dynamics
PanPL	✓	✗	7	Open & Closed
PanPL_L222R	✓	✓	7	Closed Loop Clutcher
PanPL_R48G	✗	✗	8.5 pH Switch	Open & Closed
PanPL_L222R_R48G	✗	✓	7	Open & Closed

## 4.15. Conclusion

The desired mutation not only changes the electrostatic environment but also tweak the dynamic property of the enzyme. The dynamicity might further affect or enhance the catalytic turn over. How exactly the dynamics of an enzyme is related to catalysis is not yet understood clearly. In this chapter, we reported the 14-fold enhancement in catalytic activity of PanPL L222R mutant with an alteration of dynamic behaviour. The PanPL L222R acts as a loop clutcher and freezes the N-loop-lid in its closed state throughout the simulation without accessing the open state. This result provides the insight that not only the N-loop-lid length helps in selecting the mechanism, even the single mutation distal to the active site can alter the intrinsic flexibility to influence the mechanism via electrostatic network rearrangement. Introducing L222R mutation changed the electrostatic environment in the catalytic tunnel by increasing the positive charge which attracts the negatively charged substrate into the tunnel more efficiently. Our binding energy decomposition analysis on PanPL L222R- tetra ManA complex shows the additional contribution of R222 in substrate binding in comparison to wildtype PanPL and tetra ManA complex. We suggest, firstly the better substrate binding proficiency and secondly, the capability to maintain the catalytic competent closed state make PanPL L222R catalytically more efficient enzyme compared to wildtype. The R48G mutation at the exit site acts as a pH switch for poly-ManA substrate, where the optimum pH shifted from pH 7 to pH 8.5. Our analysis suggest that the R48G mutation disrupts pre-existing electrostatic network and rewires the network of the surrounding catalytic residue Y226 and shifts the pKa of Y226 residue, which alters the optimum pH for poly-ManA substrate. However, the PanPL L222R R48G mutant doesn't show any remarkable changes in its biochemistry or dynamic behaviour compared to wildtype.



Our study provides an insight to connect the dynamics and the catalysis aspects together by using electrostatic analysis of the enzymes.



## **CHAPTER 5: Future Directions**

## **5.1. Introduction**

This chapter briefs the scope of future directions and development of this thesis work. Polysaccharide lyases are well known biotechnological tools for their multifaceted potential to contribute to medical, chemical, food, and green energy industries(80,81). In this scenario, PanPL adds itself to the existing list of PLs to serve the industrial applications. However, dedicated research should be carried out for engineering multifunctional, stable, catalytically efficient, versatile enzymes to meet the demand of industrial scale.

## **5.2. PanPL as a model enzyme for research enthusiasts.**

Our study on the structure-function-dynamics of novel PanPL discovers an unique biochemical properties. PanPL shed the light into the hidden enzyme dynamics which reflects in the biochemistry as allostery. The distinct modes of dynamics of PanPL, the evolutionary aspect of selecting the mechanism and the allosteric behaviour of the enzyme make it as a unique enzyme. This inspires the enzymologist to incorporate the dynamics in the research as a new normal. The mutational study of PanPL directly establishes the significance of dynamics in enzyme efficiency. The story of PanPL updates the existing knowledge of enzymology. The knowledge generated by this thesis work will contribute to build a generalized model for the mechanism of PL-5 enzymes correlating sequence-structure-dynamics-mechanism in a holistic manner.

## **5.3. PanPL as a model enzyme for enzyme engineering**

- PanPL is a robust enzyme based on its structural stability across the pH spectrum. It shares a ~60% sequence identity with Smlt1473, the PL-5 enzyme with a unique pH regulated multi-substrate specificity. The catalytic tunnel of the Smlt1473 and PanPL, shares a similar amino acid distribution except a few residues, R48, L222,

H167 and a longer loop with three residue insertion at the entry site. Smlt1473 has a wider entry site compared to PanPL. This comparative analysis inspires to modify the catalytic tunnel of the PanPL based on Smlt1473 and to investigate the biochemical properties of engineered PanPL. The study of L222R, R48G, L222R R48G mutations was based on this rationale, which delineate several interesting facts described in chapter 4 of the thesis. The further study can be extended to incorporate all the Smlt1473-like mutation in PanPL and test the pH regulation on the substrate specificity for the modified PanPL. PanPL is an alginate specific PL. Engineering the catalytic tunnel by editing the amino acid composition might enable the modified enzyme to take up other anionic polysaccharides like poly hyaluronic acid, poly glucuronic acid and chondroitin sulphate etc as substrates. Bestowing the multi-substrate specificity to the engineered PanPL in a regulatory manner might be beneficial for industrial and medical applications.

- PanPL is endolytic and produces dimeric sugar as a major product. However, the mode of action of the enzyme on its substrate can be altered by suitable mutations in the catalytic tunnel. The PLs with an exolytic mode of action are considered potential biotechnological tool in bioethanol production as the monomeric end product can directly be converted to bioethanol. In this regard, engineering for an exolytic PanPL would enable the modified enzyme to perform saccharification of polysaccharides. The need for alternative energy sources might be assisted by creating such engineered enzymes.
- The oligosaccharides with variable chain lengths are used in pharmaceutical industries as raw material for several purposes like drug delivery and gelling agent. However, synthesizing these oligomers with specific length is hectic and costly. The controlled depolymerization of polysaccharide chain by an engineered PanPL,

will replace the cumbersome chemical synthesis of value-added oligosaccharides in industries in a greener and economical way.

- Enzymes are optimized for specific functions. However, they might possess a very low-level activity, i.e., promiscuity for non-native functions. The evolutionary selection process plays an important role in optimizing enzymes for the specific function. By utilizing the method of directed evolution, we can train the enzyme to adopt new functions after several generation of random mutation and screening looking for promiscuity for alternate functions. Inspired by such ideas, we can investigate the hidden alternative functionality embedded in PanPL by sequence and structure analysis. The remodelled enzyme can then be trained to adopt newly found non-native non-optimized functionalities using the directed evolution approach.

#### **5.4. Medical application of PanPL**

- Some hospital borne nosocomial bacteria form biofilm to confer multidrug resistance. A few of them have PLs to control the chain length of the oligomers used in biofilm production (32,82). It has been reported that inhibiting the PL function can disrupt the biofilm production machinery and make the bacteria more susceptible to drugs (82). The source organism for PanPL, i.e., *Pandoraea apista*, has been reported to be an emerging gram-negative, nosocomial pathogen isolated in the sputum of cystic fibrosis patients (83). However, the role of PanPL in bacterial pathogenesis has not been established yet. The role of PanPL in its niche should be established.
- The bacterial biofilm is mostly composed of alginate. PanPL, an alginate specific lyase might be used as an anti-biofilm drug to target specific superbugs. However,

clinical experiments should be carried out extensively to establish PanPL's anti-biofilm activity.



## **APPENDIX**

**Table A1: Loops and helices in PanPL structures**

<b>Loops</b>	<b>Residue range</b>	<b>Helices</b>	<b>Residue span</b>
L <sub>1</sub>	26-52	H <sub>i1</sub>	53-81
L <sub>2</sub>	82-85	H <sub>o1</sub>	86-101
L <sub>3</sub>	102-115	H <sub>i2</sub>	116-135
L <sub>4</sub>	136-140	H <sub>o2</sub>	141-161
L <sub>5</sub>	160-164	H <sub>i3</sub>	165-187
L <sub>6</sub>	188-189	H <sub>o3</sub>	190-206
L <sub>7</sub>	207-213	H <sub>i4</sub>	223-242
L <sub>8</sub>	219-222	H <sub>o4</sub>	270-277
L <sub>9</sub>	243-254	H <sub>i5</sub>	292-296
L <sub>10</sub>	255-270	H <sub>o5</sub>	307-310
L <sub>11</sub>	278-287	H <sub>o6</sub>	324-330
L <sub>12</sub>	288-291		
L <sub>13</sub>	297-306		
L <sub>14</sub>	311-323		
L <sub>15</sub>	330-336		



## REFERENCES

---

1. Moss G. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes by the Reactions they Catalyse [Internet]. International Union of Biochemistry and Molecular Biology. 2021 [cited 2023 Mar 7]. Available from: <https://iubmb.qmul.ac.uk/enzyme/>
2. Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft*. 1894;27(3):2985–93.
3. Garron ML, Cygler M. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology*. 2010;20(12):1547–73.
4. Warshel A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *Journal of Biological Chemistry* [Internet]. 1998;273(42):27035–8. Available from: <http://dx.doi.org/10.1074/jbc.273.42.27035>
5. Dash P, Acharya R. Distinct Modes of Hidden Structural Dynamics in the Functioning of an Allosteric Polysaccharide Lyase. *ACS Cent Sci* [Internet]. 2022 Jul 27;8(7):933–47. Available from: <https://doi.org/10.1021/acscentsci.2c00277>
6. Qi G, Hayward S. Database of ligand-induced domain movements in enzymes. *BMC Struct Biol*. 2009;9.
7. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Vol. 450, *Nature*. Nature Publishing Group; 2007. p. 964–72.
8. Shih WM, Gryczynski Z, Lakowicz JR, Spudich JA. A FRET-Based Sensor Reveals Large ATP Hydrolysis-Induced Conformational Changes and Three Distinct States of the Molecular Motor Myosin. Vol. 102, *Cell*. 2000.
9. Koder N, Noshiro D, Dora SK, Mori T, Habchi J, Blocquel D, et al. Structural and dynamics analysis of intrinsically disordered proteins by high-speed atomic force microscopy. *Nat Nanotechnol*. 2021 Feb 1;16(2):181–9.

10. Trajkovic S, Zhang X, Daunert S, Cai Y. Atomic force microscopy study of the conformational change in immobilized calmodulin. *Langmuir*. 2011 Sep 6;27(17):10793–9.
11. Ishima R, Torchia DA. Protein dynamics from NMR [Internet]. 2000. Available from: <http://structbio.nature.com>
12. Palmer AG. Enzyme dynamics from NMR spectroscopy. *Acc Chem Res*. 2015;48(2):457–65.
13. Sun Z, Liu Q, Qu G, Feng Y, Reetz MT. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem Rev*. 2019;
14. Tom Burnley B, Afonine P v., Adams PD, Gros P. Modelling dynamics in protein crystal structures by ensemble refinement. *Elife*. 2012;2012(1):1–29.
15. Skjaerven L, Yao XQ, Scarabelli G, Grant BJ. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics*. 2014;15(1):1–11.
16. Mackerell AD, Bashford D, Bellott L M, Dunbrack L R L, Evanseck JD, Field MJ, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins † [Internet]. 1998. Available from: <https://pubs.acs.org/sharingguidelines>
17. Karplus M, Kuriyan J. Molecular dynamics and protein function [Internet]. 2005. Available from: <https://www.pnas.org>
18. Callender R, Dyer RB. The dynamical nature of enzymatic catalysis. *Acc Chem Res*. 2015 Feb 17;48(2):407–13.
19. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Vol. 450, *Nature*. Nature Publishing Group; 2007. p. 964–72.
20. Bhabha G, Biel JT, Fraser JS. Keep on moving: Discovering and perturbing the conformational dynamics of enzymes. *Acc Chem Res*. 2015;48(2):423–30.
21. Petrovic D, Risso VA, Kamerlin SCL, Sanchez-Ruiz JM. Conformational dynamics and enzyme evolution. *J R Soc Interface*. 2018;15(144).

- 
22. Jackson CJ, Foo JL, Tokuriki N, Afriat L, Carr PD, Kim HK, et al. Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. *Proc Natl Acad Sci U S A*. 2009;106(51):21631–6.
23. Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, et al. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat Struct Mol Biol*. 2013;20(11):1243–9.
24. Vieille C, Zeikus GJ. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiology and Molecular Biology Reviews*. 2001 Mar;65(1):1–43.
25. Feller G, Gerday C. Psychrophilic enzymes: Hot topics in cold adaptation. *Nat Rev Microbiol*. 2003;1(3):200–8.
26. Cornish-Bowden A. Understanding allosteric and cooperative interactions in enzymes. Vol. 281, *FEBS Journal*. 2014. p. 621–32.
27. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. *Nature*. 2014;508(7496):331–9.
28. Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide [Internet]. Vol. 90, *Biochemistry Communicated by Peter B. Dervan*. 1993. Available from: <https://www.pnas.org>
29. Stemmer W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature*. 1994;370(6488):389–91.
30. Reetz MT. Directed Evolution of Artificial Metalloenzymes: A Universal Means to Tune the Selectivity of Transition Metal Catalysts? *Acc Chem Res*. 2019;
31. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal*. 2010 Dec 15;432(3):437–44.
32. Boyd A, Ghosh M, May TB, Keogh R, Chakrabarty AM. Sequence of the algL gene of *Pseudomonas aeruginosa* and purification of its alginate lyase product. *Gene*. 1993;131:1–8.

- 
33. Ertesvåg H, Erlien F, Skjåk-Braek G, Rehm BHA, Valla S. Biochemical Properties and Substrate Specificities of a Recombinantly Produced *Azotobacter vinelandii* Alginate Lyase [Internet]. Vol. 180, JOURNAL OF BACTERIOLOGY. 1998. Available from: <https://journals.asm.org/journal/jb>
34. MacDonald LC, Berger BW. A polysaccharide lyase from *Stenotrophomonas maltophilia* with a unique, pH-regulated substrate specificity. *Journal of Biological Chemistry*. 2014 Jan 3;289(1):312–25.
35. Dash P, Acharya R. Distinct Modes of Hidden Structural Dynamics in the Functioning of an Allosteric Polysaccharide Lyase. *ACS Cent Sci*. 2022 Jul 27;8(7):933–47.
36. Rahman MM, Wang L, Inoue A, Ojima T. CDNA cloning and bacterial expression of a PL-14 alginate lyase from a herbivorous marine snail *Littorina brevicula*. *Carbohydr Res*. 2012 Oct 1;360:69–77.
37. Sindelar M, Jilkova J, Kubala L, Velebny V, Turkova K. Hyaluronidases and hyaluronate lyases: From humans to bacteriophages. Vol. 208, *Colloids and Surfaces B: Biointerfaces*. Elsevier B.V.; 2021.
38. Zheng L, Xu Y, Li Q, Zhu B. Pectinolytic lyases: a comprehensive review of sources, category, property, structure, and catalytic mechanism of pectate lyases and pectin lyases. Vol. 8, *Bioresources and Bioprocessing*. Springer Science and Business Media Deutschland GmbH; 2021.
39. Cao J. The Pectin Lyases in *Arabidopsis thaliana*: Evolution, Selection and Expression Profiles. *PLoS One*. 2012 Oct 9;7(10).
40. Lunin V v., Li Y, Linhardt RJ, Miyazono H, Kyogashima M, Kaneko T, et al. High-resolution crystal structure of *Arthrobacter aurescens* chondroitin AC lyase: An enzyme-substrate complex defines the catalytic mechanism. *J Mol Biol*. 2004 Mar 19;337(2):367–86.
41. Guthrie EP, Salyers AA. Evidence that the *Bacteroides thetaiotaomicron* Chondroitin Lyase II Gene Is Adjacent to the Chondro-4-Sulfatase Gene and May Be Part of the Same Operon [Internet]. Vol. 169, JOURNAL OF BACTERIOLOGY. 1987. Available from: <https://journals.asm.org/journal/jb>

- 
42. Pojasek K, Shriver Z, Kiley P, Venkataraman G, Sasisekharan R. Recombinant expression, purification, and kinetic characterization of chondroitinase AC and chondroitinase B from *flavobacterium heparinum*. *Biochem Biophys Res Commun*. 2001;286(2):343–51.
43. Córdula CR, Lima MA, Shinjo SK, Gesteira TF, Pol-Fachin L, Coulson-Thomas VJ, et al. On the catalytic mechanism of polysaccharide lyases: Evidence of His and Tyr involvement in heparin lysis by heparinase i and the role of Ca<sup>2+</sup>. *Mol Biosyst*. 2014 Jan;10(1):54–64.
44. MacDonald LC, Berger BW. A polysaccharide lyase from *Stenotrophomonas maltophilia* with a unique, pH-regulated substrate specificity. *Journal of Biological Chemistry*. 2014;289(1):312–25.
45. Ogura K, Yamasaki M, Yamada T, Mikami B, Hashimoto W, Murata K. Crystal structure of family 14 polysaccharide lyase with pH-dependent modes of action. *Journal of Biological Chemistry*. 2009;284(51):35572–9.
46. Mikami B, Ban M, Suzuki S, Yoon HJ, Miyake O, Yamasaki M, et al. Induced-fit motion of a lid loop involved in catalysis in alginate lyase A1-III. *Acta Crystallographica Section D [Internet]*. 2012 Sep;68(9):1207–16. Available from: <https://doi.org/10.1107/S090744491202495X>
47. Ulaganathan T, Shi R, Yao D, Gu RX, Garron ML, Cherney M, et al. Conformational flexibility of PL12 family heparinases: Structure and substrate specificity of heparinase III from *Bacteroides thetaiotaomicron* (BT4657). *Glycobiology*. 2017;27(1):176–87.
48. Dong S, Wei T di, Chen XL, Li CY, Wang P, Xie B bin, et al. Molecular insight into the role of the N-terminal extension in the maturation, substrate recognition, and catalysis of a bacterial alginate lyase from polysaccharide lyase family 18. *Journal of Biological Chemistry*. 2014;289(43):29558–69.
49. Pandey S, Mahanta P, Berger BW, Acharya R. Structural insights into the mechanism of pH-selective substrate specificity of the polysaccharide lyase Smlt1473. *Journal of Biological Chemistry*. 2021 Oct 1;296(4).

- 
50. Pandey S, Mahanta P, Berger BW, Acharya R. Structural insights into the mechanism of pH-selective substrate specificity of the polysaccharide lyase Smlt1473. *Journal of Biological Chemistry* [Internet]. 2021;297(4):101014. Available from: <https://www.sciencedirect.com/science/article/pii/S0021925821008164>
51. Weissbach A, Hurwitz J. The Formation of 2-Keto-3-deoxyheptonic Acid in Extracts of *Escherichia coli* B I. IDENTIFICATION.
52. Dreifke<sup>1</sup> M, Fröba M, Armbrecht M. Eppendorf BioSpectrometer ® kinetic: activity measurements of enzymes immobilized on a mesoporous silica matrix. 2016.
53. Taylor G. Biological Crystallography The phase problem. *Acta Cryst* [Internet]. 2003;59:1881–90. Available from: <http://www.ytbl.york.ac.uk/~cowtan/fourier/>
54. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. {it Phaser} crystallographic software. *J Appl Crystallogr* [Internet]. 2007 Aug;40(4):658–74. Available from: <https://doi.org/10.1107/S0021889807021206>
55. Liebschner D, Afonine P V., Baker ML, Bunkoczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr D Struct Biol*. 2019;75:861–77.
56. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* [Internet]. 2004 Dec;60(12 Part 1):2126–32. Available from: <https://doi.org/10.1107/S0907444904019158>
57. Terwilliger TC, Grosse-Kunstleve RW, Afonine P V., Moriarty NW, Zwart PH, Hung LW, et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. In: *Acta Crystallographica Section D: Biological Crystallography*. 2007. p. 61–9.
58. Afonine P V., Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*. 2012 Apr;68(4):352–67.
59. Lebedev AA, Isupov MN. Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. *Acta Crystallogr D Biol Crystallogr*. 2014 Aug 26;70(9):2430–43.

- 
60. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by C $\alpha$  geometry:  $\phi$ ,  $\psi$  and C $\beta$  deviation. *Proteins: Structure, Function and Genetics*. 2003 Feb 15;50(3):437–50.
61. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004;32(WEB SERVER ISS.):665–7.
62. Olsson MHM, SØndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions. *J Chem Theory Comput*. 2011 Feb 8;7(2):525–37.
63. Baker NA, Sept D, Joseph S, Holst MJ, Andrew McCammon J. Electrostatics of nanosystems: Application to microtubules and the ribosome [Internet]. Vol. 98. 2001. Available from: [www.pnas.org/doi/10.1073/pnas.181342398](http://www.pnas.org/doi/10.1073/pnas.181342398)
64. Schrödinger L, & DW. Pymol. [Internet]. Internet; 2020 [cited 2023 Mar 19]. Available from: <http://www.pymol.org/pymol>
65. Combs SA, Deluca SL, Deluca SH, Lemmon GH, Nannemann DP, Nguyen ED, et al. Small-molecule ligand docking into comparative models with Rosetta. *Nat Protoc*. 2013;8(7):1277–98.
66. Burnley BT, Gros P. phenix.ensemble\_refinement: a test study of apo and holo BACE1. *Computational crystallography newsletter* [Internet]. 2013;4:51–8. Available from: <http://dspace.library.uu.nl/handle/1874/290786%5Cnpapers3://publication/uuid/B0417039-97CA-4DC2-9C50-015F922B547C>
67. Song Y, Dimaio F, Wang RYR, Kim D, Miles C, Brunette T, et al. High-resolution comparative modeling with RosettaCM. *Structure* [Internet]. 2013;21(10):1735–42. Available from: <http://dx.doi.org/10.1016/j.str.2013.08.005>
68. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* [Internet]. 2021;596(7873):583–9. Available from: <http://dx.doi.org/10.1038/s41586-021-03819-2>

- 
69. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun.* 1995;91(1-3):43-56.
70. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM.pdf. *J Comput Chem.* 2010;31(4):671-90.
71. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry A.* 2001;105(43):9954-60.
72. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem.* 2010 Mar;31(4):671-90.
73. Wolf A, Kirschner KN. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11-23S subdomain. *J Mol Model.* 2013;19(2):539-49.
74. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. Gmx\_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J Chem Theory Comput.* 2021 Oct 12;17(10):6281-91.
75. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values. *J Chem Theory Comput.* 2011 Jul 12;7(7):2284-95.
76. Macdonald LC, Berger BW. Insight into the role of substrate-binding residues in conferring substrate specificity for the multifunctional polysaccharide lyase Smlt1473. *Journal of Biological Chemistry [Internet].* 2014;289(26):18022-32. Available from: <http://dx.doi.org/10.1074/jbc.M114.571299>
77. Knott BC, Crowley MF, Himmel ME, Ståhlberg J, Beckham GT. Carbohydrate-protein interactions that drive processive polysaccharide translocation in enzymes revealed from a computational study of cellobiohydrolase processivity. *J Am Chem Soc.* 2014;136(24):8810-9.
78. Hubbard RE, Kamran Haider M. Hydrogen Bonds in Proteins: Role and Strength. *eLS.* 2010;(February).



- 
79. Worth CL, Blundell TL. On the evolutionary conservation of hydrogen bonds made by buried polar amino acids: The hidden joists, braces and trusses of protein architecture. *BMC Evol Biol.* 2010;10(1).
80. Cao S, Li L, Li Q, Jiang L, Zhu B, Yao Z. A novel alginate lyase and its domain functions for the preparation of unsaturated monosaccharides. *Appl Microbiol Biotechnol* [Internet]. 2023 Feb 16; Available from: <https://link.springer.com/10.1007/s00253-023-12424-4>
81. Eckersley E, Berger BW. An Engineered Polysaccharide Lyase to Combat Harmful Algal Blooms [Internet]. 2018. Available from: <http://www.elsevier.com/open-access/userlicense/1.0/>
82. Thiang Yian Wong, Preston LA, Schiller NL. Alginate lyase: Review of major sources and enzyme characteristics, structure-function analysis, biological roles, and applications. *Annu Rev Microbiol.* 2000;54:289–340.
83. Atkinson RM, LiPuma JJ, Rosenbluth DB, Dunne WM. Chronic colonization with *Pandora* apista in cystic fibrosis patients determined by repetitive-element-sequence PCR. *J Clin Microbiol.* 2006 Mar;44(3):833–6.