IN-SILICO STRUCTURAL ANALYSIS AFFECTING THERMOSTABILITY IN RECOMBINANT PSYCHROPHILIC CHITINASE (CHI II) FROM Glaciozyma antarctica PI12

AMALINA BINTI AMER SHAH

A thesis submitted in fulfilment of the requirement for the award of the degree of Master of Engineering (Bioprocess)

Faculty of Chemical and Energy Engineering
Universiti Teknologi Malaysia

JANUARY 2017
The day has finally come. When I can say it out loud
'I made it'

To mama & bapak,
this is for you
ACKNOWLEDGEMENT

Alhamdulillah, all praises to Allah, Whom I am grateful for His guidance. Without His grant, this thesis could be possible. His divine and complex creations never failed to amaze me. I am thankful to my supervisor, Prof Dr. Rosli Md Illias for his encouragement and guidance through all these years. Not to forget, Encik Yaakop for the technical support provided. I would like to express my gratitude to Ministry of Science, Technology and Innovation Malaysia for the scholarship and research funding.

My utmost appreciation to mama and bapak, my sibling, Liyana, Luqman and Amirah for their endless love, encouragement and prayer. To my best friends Anita, Hazwani and Suhailis for their endless motivation and support. My deep appreciation also goes to my labmates, for your help, knowledge and unwavering support through ups and downs. Thanks for this friendship. I wish you guys luck. Not to forget, Dr Aizi Nor Mazila. Thank you so much for your knowledge and guidance in my early journey. I also wish to show my appreciation to those who directly or indirectly involved in finishing this thesis.

To all of you, thank you so much. I am blessed
ABSTRACT

Cold-adapted enzymes are significant with structure flexibility and high catalytic activity at low temperature. High structural flexibility could be due to combination of several features such as weak intramolecular bonds, decreased compactness of hydrophobic core and reduced number of proline and arginine residues. However, to compensate the structural flexibility, cold-adapted enzymes are also thermolabile which causes them to be easily inactivated at elevated temperature. Therefore, it would be more interesting and beneficial if more stable cold-adapted enzymes are produced to fulfill the industrial needs. In this study, a novel cold-adapted chitinase (CHI II) from Glaciozyma antarctica PI12 was rationally designed to improve their thermostability thus make them more resistant to increased temperature. Four CHI II mutants were designed through rational design named as A157Q, I134P, mutant Loop and Y257R by manipulating the structural hydrophobicity, introduction of proline in the loop regions, introduction of arginine salt bridges and loop shortening. Mutant Loop was designed by removing 9 residues in loop regions thus makes loop involved became shorter. Stability of all mutants was first predicted through a computational approach where all structures were subjected to 10 ns molecular dynamics simulation at three temperatures; 273 K, 288 K and at 300 K. Based on the simulation, it was found that mutants I134P, mutant Loop and Y257R exhibited structural stability at 300 K. This conclusion was made based on low and stable root-mean square deviation (RMSD) value at 300 K in comparison to RMSD values at 288 K and 273 K. Low RMSD values indicated mutant structure experienced low structural deviation throughout the simulation. Besides, this observation is correlated with reduction of structure compactness (radius of gyration), reduced solvent accessible surface area and increased numbers of hydrogen and salt bridges. However, mutant A157Q experienced structure destabilization at 300 K. Substitution of helix-preferred residue, alanine with a thermolabile residue, glutamine had caused A157Q structure become loosely packed at 300 K indicating a thermal denaturation. To support the theoretical model, CHI II and all mutants were then cloned into Pichia pastoris expression vector pPICZαC and expressed in P. pastoris (GS115).
ABSTRAK


# TABLE OF CONTENT

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td></td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td></td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF SYMBOLS AND ABBREVIATIONS</td>
<td></td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td></td>
<td>xix</td>
</tr>
</tbody>
</table>

## 1 INTRODUCTION

1.1 Background of study | 1 |
1.2 Problem statements | 3 |
1.3 Objectives | 4 |
1.4 Scopes of study | 4 |

## 2 LITERATURE REVIEW

2.1 Psychrophiles, a Cold-Loving Microbes | 5 |
2.2 Cold Adapted Enzymes | 6 |
2.3 Thermostability Improvement through rational Design | 10 |
2.3.1 Proline Substitution to Improve Protein Thermostability 12
2.3.2 Improving Thermostability by Loop Shortening 13
2.3.3 Introduction of Salt Bridges to Improve Protein Thermostability 13
2.3.4 Substitution of Surface Alanine Residue to a More Hydrophilic Residues 14

2.4 Thermostability Studies Through Molecular Dynamic (MD) Simulation 14

2.5 Chitin 16

2.6 Chitinase, a Chitinolytic Enzyme 18
2.6.1 Chitinase Subfamiliy and Classifications 19
2.6.2 Functional Domain of Fungal Chitinases 23
2.6.3 Roles of Chitinases in Various Organisms 25

2.7 A novel Chitinase (CHI II) From Glaciozyma antarctica PI12 26

2.8 Chitinases and Their Applications 28
2.8.1 Bioconversion of Chitin in Production of Single Cell Proteins 28
2.8.2 Enzymatic Hydrolysis in Production of Chitooligosaccharides 29
2.8.3 Biocontrol Agents 30

2.9 Heterologous Expression in Pichia pastoris 30

3 MATERIALS AND METHODS 33

3.1 Introduction 33
3.2 Source of Sequences 35
3.3 Tools and Software 35
3.4 Mutation Design Strategy 38
3.5 Construction of A157Q, I134P and Y257R 3D Models 38
3.6 Homology Modeling of Mutant Loop 38
3.7 Molecular Dynamics (MD) Simulation 39
3.7.1 Preparation Stage 40
3.7.2 Production Stage 40
3.7.3 Trajectories Analysis for CHI II and Mutants 41
3.7.3.1 Root Mean Square Deviation (RMSD) 41
3.7.3.2 Root Mean Square Fluctuation (RMSF) 41
3.7.3.3 Radius of Gyration 41
3.7.3.4 Analysis of Secondary Structure for CHI II and it Mutants 42
3.7.3.5 Solvent Accessible Surface Area (SASA) 42
3.7.3.6 Analysis of Hydrogen Bonds and Salt Bridges 42
3.8 Cloning of CHI II and It Mutants into *Pichia pastoris* Expression System 42
3.9 Expression of Recombinant CHI II in *Pichia Pastoris* strains 45
3.10 Screening of Mutant Expression 45
3.10.1 SDS PAGE Analysis 45
3.10.2 Western Blotting 46
3.10.3 Chitinase Assay 47

4 RESULTS AND DISCUSSION 49
4.1 Mutants Design Strategy 49
4.1.1 Introduction of Salt Bridge in Mutant Y257R 49
4.1.2 Loop Shortening in Mutant Loop 52
4.1.3 Introduction of Proline in Loop Regions for Mutant I134P 55
4.1.4 Substitution of Alanine to Glutamine in Mutant A157Q 56
4.2 Construction of Mutants 3D Models 57
4.2.1 Construction of A157Q, I134P and Y257R 3D Models Using PyMol 58
4.2.2 Homology Modeling of Mutant Loop 59

4.3 Molecular Dynamics (MD) Simulation of CHI II and Mutants 64

4.4 Comparative Analysis of MD Simulation Trajectories 65

4.4.1 Root Mean Square Deviation (RMSD) 65

4.4.2 Root Mean Square Fluctuation (RMSF) 71

4.4.3 Radius of Gyration 75

4.4.4 Analysis of Secondary Structure for CHI II and Its Mutants 79

4.4.5 Solvent Accessible Surface Area (SASA) 84

4.4.6 Analysis of Hydrogen Bonds 85

4.4.7 Analysis of Salt Bridges 87

4.5 Cloning of CHI II and Its Mutants into *Pichia pastoris* 92

4.6 Expression of CHI II Mutants into *P. Pastoris* GS115 96

5 CONCLUSION 100

5.1 Conclusions 100

5.2 Recommendations 101

REFERENCES 102

APPENDIX A-F 119
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>List of tools and software used in this study</td>
<td>36</td>
</tr>
<tr>
<td>4.1</td>
<td>Model evaluation of CHI II and mutant A157Q, I134P and Y257R</td>
<td>59</td>
</tr>
<tr>
<td>4.2</td>
<td>The alignment search result against PDB database from various servers</td>
<td>60</td>
</tr>
<tr>
<td>4.3</td>
<td>The best model generated by Modeller 9.10 and evaluations</td>
<td>61</td>
</tr>
<tr>
<td>4.4</td>
<td>Evaluation of mutant Loop model (model 84) before and after energy minimization (EM)</td>
<td>62</td>
</tr>
<tr>
<td>4.5</td>
<td>Average radius of gyration for CHI II and all mutants at 273 K, 288 K and 300 K.</td>
<td>78</td>
</tr>
<tr>
<td>4.6</td>
<td>The average values of total solvent accessible surface area of CHI II and all mutants at three temperatures</td>
<td>85</td>
</tr>
<tr>
<td>4.7</td>
<td>The average number of hydrogen bonds in CHI II and mutants at 273 K, 288 K and 300 K</td>
<td>87</td>
</tr>
<tr>
<td>4.8</td>
<td>The average number of salt bridges in CHI II and mutants at three temperatures.</td>
<td>88</td>
</tr>
<tr>
<td>4.9</td>
<td>The list of salt bridges that are conserved in CHI II and mutants at 273 K, 288 K and 300 K.</td>
<td>90</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Factors contributed to the structural flexibility of cold-adapted enzymes</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>The structures of amino acids arginine, proline and glycine</td>
<td>9</td>
</tr>
<tr>
<td>2.3</td>
<td>Dry chitin flakes and the arrangement of chitin polymer</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>The arrangement of chitin microfibrils in three polymorph forms</td>
<td>17</td>
</tr>
<tr>
<td>2.5</td>
<td>Illustration of cleavage pattern of chitinolytic enzymes</td>
<td>19</td>
</tr>
<tr>
<td>2.6</td>
<td>The classification of chitinase GH18 and GH19 subfamilies and subgroup</td>
<td>21</td>
</tr>
<tr>
<td>2.7</td>
<td>The domain organization of fungal chitinases in GH18 family</td>
<td>25</td>
</tr>
<tr>
<td>2.8</td>
<td>The predicted structure of CHI II by using threading method</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Illustration of operational framework of this study</td>
<td>34</td>
</tr>
<tr>
<td>3.2</td>
<td>Three stages of MD simulation consists of preparation, production and analysis stages</td>
<td>39</td>
</tr>
<tr>
<td>3.3</td>
<td>The overview of OE-PCR strategy to obtain a mutant’s gene</td>
<td>43</td>
</tr>
<tr>
<td>4.1</td>
<td>The location of three salt bridges found in CHI I structures</td>
<td>51</td>
</tr>
<tr>
<td>4.2</td>
<td>Multiple sequence alignment showed the mutation site of Y257R</td>
<td>52</td>
</tr>
<tr>
<td>4.3</td>
<td>The structure alignment between CHI II and its template 1ITX</td>
<td>53</td>
</tr>
<tr>
<td>4.4</td>
<td>Multiple sequence alignment showed the mutation site of mutant Loop</td>
<td>55</td>
</tr>
<tr>
<td>4.5</td>
<td>Multiple sequence alignment showed the mutation site or I134P</td>
<td>56</td>
</tr>
</tbody>
</table>
4.6 Multiple sequence alignment showed the mutation site of A157Q

4.7 Ramachandran plot of mutant Loops model (a) before energy minimization and (b) after energy minimization

4.8 RMSD overlay of CHI II and four mutant at 273 K

4.9 RMSD overlay of CHI II and four mutants at 288 K

4.10 RMSD overlay of CHI II and four mutants at 300 K

4.11 An overview of CHI II and mutants structure snapshots in 5 ns interval at 300 K simulation

4.12 The overlay RMSF graph of CHI II and all mutants at 273 K

4.13 The overlay RMSF graph of CHI II and all mutants at 288 K

4.14 The overlay RMSF graph of CHI II and all mutants at 300 K

4.15 The overlay Rg graph of CHI II and all mutants at 273 K

4.16 The overlay Rg graph of CHI II and all mutants at (a) 288 K and (b) 300 K

4.17 Secondary structure assignment for CHI II and mutants at 273 K

4.18 Secondary structure assignment for CHI II and mutants at 288 K

4.19 Secondary structure assignment for CHI II and mutants at 300 K

4.20 The interaction maps between Arg57 and Asp205 which involved carboxylate group of Asp205 and guanidium group of Arg257.

4.21 The formation and breaking of salt bridge Asp205 and Arg257 which was introduced in Y257R at three temperatures.

4.22 PCR product of first PCR and second PCR reaction for all mutants.

4.23 Linearization of pPICZαC and A157Q constructed vector

4.24 Formation of A157Q-GS115 colonies on YPD and YPDS plate (a) YPDS plate with 100 µg/ml of zeocin, (b) YPD agar with 500 µg/ml of zeocin, (c) YPD agar with 1000 µg/ml of zeocin and (d) YPD agar with 2000 µg/ml of zeocin
4.25 PCR colony for all mutant gene integration into *P. pastoris* GS115.  
4.26 SDS-PAGE analysis of CHI II and mutant A157Q  
4.27 SDS-PAGE analysis of mutant Loop, I134P and Y257R.
# LIST OF SYMBOLS AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>percent</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>asterisk</td>
<td></td>
</tr>
<tr>
<td>.top</td>
<td>Topology file</td>
<td></td>
</tr>
<tr>
<td>~</td>
<td>About</td>
<td></td>
</tr>
<tr>
<td>&gt;</td>
<td>Greater</td>
<td></td>
</tr>
<tr>
<td>≥</td>
<td>Greater or equal to</td>
<td></td>
</tr>
<tr>
<td>°</td>
<td>Degree</td>
<td></td>
</tr>
<tr>
<td>°C</td>
<td>degree celcius</td>
<td></td>
</tr>
<tr>
<td>½</td>
<td>Half</td>
<td></td>
</tr>
<tr>
<td>3D</td>
<td>Three dimensional</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Alanine</td>
<td></td>
</tr>
<tr>
<td>Å</td>
<td>Angstrom</td>
<td></td>
</tr>
<tr>
<td>Arg or R</td>
<td>Arginine</td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Asparagine</td>
<td></td>
</tr>
<tr>
<td>Asp or D</td>
<td>Aspartic acid/aspartate</td>
<td></td>
</tr>
<tr>
<td>BD</td>
<td>Binding domain</td>
<td></td>
</tr>
<tr>
<td>C. congregates</td>
<td>Coprinellus congregates</td>
<td></td>
</tr>
<tr>
<td>C. immitis</td>
<td>Coccidiodes immitis</td>
<td></td>
</tr>
<tr>
<td>CAZy</td>
<td>Carbohydrate Active Enzyme database</td>
<td></td>
</tr>
<tr>
<td>CBD</td>
<td>Chitin binding domain</td>
<td></td>
</tr>
<tr>
<td>CID</td>
<td>Chitinase insertion domain</td>
<td></td>
</tr>
<tr>
<td>Cl⁻</td>
<td>Chloride ion</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
<td></td>
</tr>
<tr>
<td>Cα</td>
<td>Carbon alpha</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>DOPE</td>
<td>Discrete optimized protein energy</td>
<td></td>
</tr>
<tr>
<td>DSSP</td>
<td>Dictionary of secondary structure prediction</td>
<td></td>
</tr>
<tr>
<td>EC</td>
<td>Enzyme commission</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Phenylalanine</td>
<td></td>
</tr>
<tr>
<td>G. antarctica</td>
<td>Glaciozyma Antarctica</td>
<td></td>
</tr>
<tr>
<td>GH</td>
<td>Glycosyl hydrolase</td>
<td></td>
</tr>
<tr>
<td>GlcN</td>
<td>Glucosamine</td>
<td></td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetylglucosamine</td>
<td></td>
</tr>
<tr>
<td>Glu or E</td>
<td>Glutamic acid/glutamate</td>
<td></td>
</tr>
<tr>
<td>Glu or Q</td>
<td>Glutamine</td>
<td></td>
</tr>
<tr>
<td>Gly or G</td>
<td>Glycine</td>
<td></td>
</tr>
<tr>
<td>GPI</td>
<td>Glycosylphosphatidylinositol</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>Hydrogen</td>
<td></td>
</tr>
<tr>
<td>H. atriviridis</td>
<td>Hypocrea atroviridis</td>
<td></td>
</tr>
<tr>
<td>His or H</td>
<td>Histidine</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
<td></td>
</tr>
<tr>
<td>Ile or I</td>
<td>Isoleucine</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>Kelvin</td>
<td></td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodalton</td>
<td></td>
</tr>
<tr>
<td>kJ/mol</td>
<td>Kilojoule per mol</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>Leucine</td>
<td></td>
</tr>
<tr>
<td>Lys or K</td>
<td>Lysine</td>
<td></td>
</tr>
<tr>
<td>LysM</td>
<td>Lysin motif</td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
<td></td>
</tr>
<tr>
<td>MSA</td>
<td>Multiple sequence alignment</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Nitrogen</td>
<td></td>
</tr>
<tr>
<td>Na⁺</td>
<td>Sodium ion</td>
<td></td>
</tr>
<tr>
<td>Nm</td>
<td>Nanometer</td>
<td></td>
</tr>
<tr>
<td>Ns</td>
<td>Nanoseconds</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Oxygen</td>
<td></td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>pKa</td>
<td>Acid dissociation constant</td>
<td></td>
</tr>
<tr>
<td>PME</td>
<td>Particle mesh ewald</td>
<td></td>
</tr>
<tr>
<td>Pro or P</td>
<td>Proline</td>
<td></td>
</tr>
<tr>
<td>Rg</td>
<td>Radius of gyration</td>
<td></td>
</tr>
<tr>
<td>RMSD</td>
<td>Root mean square deviation</td>
<td></td>
</tr>
<tr>
<td>RMSF</td>
<td>Root mean square fluctuation</td>
<td></td>
</tr>
<tr>
<td>S. cerevisae</td>
<td>Saccharomyces cerevisae</td>
<td></td>
</tr>
<tr>
<td>SASA</td>
<td>Solvent accessible surface area</td>
<td></td>
</tr>
<tr>
<td>SCP</td>
<td>Single cell protein</td>
<td></td>
</tr>
<tr>
<td>Ser or S</td>
<td>Serine</td>
<td></td>
</tr>
<tr>
<td>sp.</td>
<td>Species</td>
<td></td>
</tr>
<tr>
<td>SPC</td>
<td>Simple point charge</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>Threonine</td>
<td></td>
</tr>
<tr>
<td>T. atroviride</td>
<td>Trichoderma atroviride</td>
<td></td>
</tr>
<tr>
<td>T. aurantiacus</td>
<td>Thermoascus aurantiacus</td>
<td></td>
</tr>
<tr>
<td>T. lanuginosus</td>
<td>Thermomyces lanuginosus</td>
<td></td>
</tr>
<tr>
<td>TIM</td>
<td>Triosephosphate isomerase</td>
<td></td>
</tr>
<tr>
<td>Tm</td>
<td>Midpoint temperature</td>
<td></td>
</tr>
<tr>
<td>Tyr or Y</td>
<td>Tyrosine</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>VMD</td>
<td>Visual molecular dynamic</td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>Alpha</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>γ</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>φ</td>
<td>Phi</td>
<td></td>
</tr>
<tr>
<td>ψ</td>
<td>Psi</td>
<td></td>
</tr>
<tr>
<td>µg/µl</td>
<td>Microgram per microliter</td>
<td></td>
</tr>
<tr>
<td>µl</td>
<td>Microliter</td>
<td></td>
</tr>
<tr>
<td>µmol</td>
<td>Micromole</td>
<td></td>
</tr>
<tr>
<td>BMGY</td>
<td>Buffered glycerol complex medium</td>
<td></td>
</tr>
<tr>
<td>BMMY</td>
<td>Buffered methanol complex medium</td>
<td></td>
</tr>
<tr>
<td>Bp</td>
<td>Base pair</td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>Molarity</td>
<td></td>
</tr>
<tr>
<td>ml</td>
<td>Milliliter</td>
<td></td>
</tr>
<tr>
<td>mM</td>
<td>Millimole</td>
<td></td>
</tr>
<tr>
<td>OE</td>
<td>Overlapping extension</td>
<td></td>
</tr>
<tr>
<td>P. pastoris</td>
<td><em>Pichia pastoris</em></td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>Unit activity</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>Voltage</td>
<td></td>
</tr>
<tr>
<td>v/v</td>
<td>Volume over volume</td>
<td></td>
</tr>
<tr>
<td>YPD</td>
<td>Yeast extract peptone dextrose</td>
<td></td>
</tr>
<tr>
<td>YPDS</td>
<td>Yeast extract peptone dextrose sorbitol</td>
<td></td>
</tr>
<tr>
<td>zeoR</td>
<td>Zeocin resistance</td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Tutorial for MD Simulation</td>
<td>119</td>
</tr>
<tr>
<td>B</td>
<td>Routine and Media</td>
<td>123</td>
</tr>
<tr>
<td>C</td>
<td>Primer Sequences</td>
<td>128</td>
</tr>
<tr>
<td>D</td>
<td>pPICZαC Vector Map</td>
<td>129</td>
</tr>
<tr>
<td>E</td>
<td>Nucleotide and Protein Sequences of CHI II and All mutants</td>
<td>130</td>
</tr>
<tr>
<td>F</td>
<td>Sequencing Result</td>
<td>132</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Background of Study

Earth’s surface is dominated by low temperature regions including polar region, mountains top and oceans which cover 80% of the biosphere where temperatures never exceed 5°C. Low temperatures are known to place severe physicochemical constraints on cellular function by negatively influencing cell integrity, water viscosity, solute diffusion rate, membrane fluidity, enzyme kinetics and macromolecular interactions (Marx et al., 2004). However, despite most other species cannot grow at this low temperature regions, an extremophiles known as psychrophiles are able to survive and inhabit this region. Then the question arises: how can psychrophiles survive, let alone thrive under this harsh conditions? The answer is, psychrophiles evolved and adapted to their environments by developing unique mechanisms to keep their cellular components stable and active.

The ability of psychrophiles to survive in cold regions is therefore dependent on numbers of adaptive strategies to successfully counteract those low temperature constraints (D’Amico et al., 2006). One of the strategies is by producing the cold-adapted enzymes that able to perform their catalysis efficiently under these extreme environmental conditions (D’Amico et al., 2006). For these reasons, cold-adapted enzymes have been considered as biotechnological potential due to their ability to perform catalysis at low temperatures thus offering advantages in the
environmental application and energy savings in industrial processes (Gianese et al., 2001).

While other enzymes are subject to cold denaturation and suffered the loss of activity at low temperatures, cold-adapted enzymes are resistant to cold denaturation with efficient catalytic activity. Their survival is correlated with their structural flexibility that was believed as a compensation for the freezing effect in cold habitats (Johns & Somero 2004). Structures flexibility of cold-adapted enzymes is the result of combination of several features such as increased numbers of hydrophobic side chains that are exposed to the solvent, a decrease in the compactness of hydrophobic core, a higher number of glycine and lysine residues, a reduced number of proline and arginine residues and weakening of intramolecular bonds (Rodrigues & Tiedje 2008). However, because of their structural flexibility, cold-adapted enzymes become less stable and also thermolabile which cause them to denature at elevated temperature (Siddiqui & Cavicchioli 2006).

Therefore, cold-adapted enzymes are often engineered either through rational design or directed evolution to improve its thermostability. Thermostability is defined as improved long-term survival under mild conditions and increased ability to remain active under harsh industrial condition but still retains its catalytic efficiency (Wijma et al., 2013). In this study, a cold-adapted chitinase named as CHI II was used as the subject understudied. Chitinase (EC 3.2.2.14) are categorized under glycosyl hydrolases (GH) family and can be found in wide range of organisms such as bacteria, fungi, yeasts, plants and mammals. Capabilities of chitinase to hydrolyse chitin to a low molecular weight chitooligomers cause them to have broad potential in industrial, agricultural and medicinal functions (Dahiya et al., 2006; Liu et al., 2013; Patil et al., 2000; Park & Kim 2010; Khan et al., 2015).

CHI II was previously isolated from Glaciozyma antarctica PI12 and it’s three-dimensional (3D) structures had been modeled by Ramli et al. (2011, 2012). Based on the structure analysis and primary sequence analysis, several characteristics related to cold adaptations were found in CHI II. CHI II was identified to have less number of salt bridges and arginine residues, increase in surface hydrophobicity and reduced
number of hydrogen bonds (Ramli et al. 2012). These characteristics were proved to be related to structural flexibility of CHI II which causes CHI II to be thermolabile and could not withstand elevated temperature and harsh environment. In addition to the wide potential of chitinases in industrial application and biotechnological application, it is best for CHI II to be engineered to improve their thermostability. Based on the information obtain from previous study on amino acid affecting thermostability of cold adapted chitinase and based on comparison studies between mesophilic, psychrophilic and thermophilic enzymes, rational design was used to design CHI II mutants (Mavromatis et al., 2003; Siddiqui & Cavicchioli 2006). Therefore, four mutants of CHI II will be designed through rational design and the effect of the mutation will be studied using an in-silico approach. In particular, the mutant’s structure stability will be studied through molecular dynamic (MD) simulation at three temperatures: 273 K, 288 K and 300 K. This is an indicator of CHI II mutants performance as it reflects the ability of the mutants to perform under conditions relevant to an industrial process where enzyme is continually affected by temperature elevation.

1.2 Problem Statement

About 80 000 metric tons biomass waste of marine invertebrate were produced every year and it was predicted that the oceans will be depleted of chitin if this insoluble biomass is not converted into simple and recyclable material (Patil et al., 2000). Capabilities of cold-adapted chitinases to have high catalytic efficiency and high flexibility (low stability) at low temperatures allows them to offer several novel opportunities in industrial application. Because of their inherent flexible structure, cold-adapted chitinase was correlated to be thermolabile as their reaction rates decrease when the temperature increases. Hence, this condition becomes a limitation for cold-adapted chitinase to be used in industrial application. Thus, production of cold-adapted enzymes chitinase with desired thermostability become an important aspect of industrial application which could also help to overcome chitin depletion. This can be achieved through mutagenesis of cold-adapted chitinase to improve its thermostability without compromising its structurally dependent cold-adapted properties (Cesarini et al., 2012).
1.3 Objectives

The main objective of this study is to analyse the effect of amino acids substitution, loop shortening and introduction of the salt bridge in the non-catalytic region on CHI II thermostability through *in-silico* approach.

1.4 Scopes of Study

The scope of this study are:

a) Design four CHI II mutants through rational design.

b) Construction of four mutants three-dimensional (3D) structures using mutagenesis plugin in PyMOL and homology modeling by using Modeller.

c) Performing the Molecular Dynamics (MD) simulation of CHI II and its four mutants at three different temperatures; 273 K, 288 K and 300 K.

d) Performing comparative trajectories analysis on CHI II and its mutants to study the effect of mutation on CHI II thermostability.
REFERENCES


APPENDIX A

APPENDIX A: TUTORIAL FOR MD SIMULATION

1) **Process the pdb file with pdb2gmx**

   `pdb2gmx -ignh -f inputabf.pdb -o abf.pdb -p abf.top -water spce`

   Pdb2gmx command converts the pdb file to a gromacs file and write the topology.

   Points to ponder:
   - What is the total mass of your protein?
   - What is the total charge of your protein?
   - Open the topology file (abf.top) using Gedit, see how the force define protein

2) **Set-up box for simulation**

   `editconf -bt cubic -f abf.pdb -o abf_bsolv.pdb -d 2.0`

   Editconf specify the simulation box. ‘-d’ sets the dimension of the box 2.0 nm (20Å).
   It should at no less than 0.9 nm for most system.

3) **Solvate the box**

   `genbox -cp abf_bsolv.pdb -cs spc216.gro -o abf_bion.pdb -p abf.top`

   Points to ponder:
   - Does the size of your box change after the solvation?
   - How many SOL molecules were added into your simulation box?
   - Any change to your topology file?

4) **Neutralize the protein system**

   `grompp -f ion.mdp -c abf_bion.pdb -p abf.top -o ion.tpr -maxwarn 5`

   `genion -s ion.tpr -o abf_b4em.pdb -nnname CL (-pname NA) -nn 2 (-np 2) -p abf.top`

   `–g ion.log`

   Type “13” for SOL

   Points to ponder:
• How many NA and CL had been added into the system?
• Is there any charge in the topology file?
• What do the flags used along the genion?

5) **Energy Minimization**

```bash
grompp -f em.mdp -c abf_b4em.pdb -p abf.top -o em.tpr -maxwarn 5
mdrun -v -deffnm em
```

Points to ponder:
• What do -v and -deffnm mean?
• How many steps does the system take to converge?
• How many output files are there?

The output will be em.gro. If the $F_{\text{max}}$ did not converge, repeat the step by changing the input and output file (em.mdp file did not change). Take output from first step (em.gro) as an input for second step (after `-c` command). The second step output will be em2.tpr (after `-o` command). Run the simulation by naming em2 at mdrun step.

6) **Position Restrained Molecular Dynamics (equilibration)**

```bash
grompp -f pr.mdp -c em.gro -p abf.top -o pr.tpr -maxwarn 5
```

Once the pr.tpr is generated successfully, run the position restrained MD

```bash
mdrun -v -deffnm pr
```

Points to ponder:
• Is there any note/warning when grompp is pre-processing the pr.mdp?

7) **Convert Gromacs File**

```bash
editconf -f file.gro -o file.pdb
```

8) **First Evaluation**

a) First evaluate the system and see the water molecules had been equilibrated or not

b) Compute the RMSD of the protein backbone and plot several graphs

```bash
g_rms -f pr.trr -s pr.tpr -o rmsd_pr
```
- Examine using GRACE. xmgrace rmsd_pr.xvg
- For least square fit and RMSD calculation, select group 4 (Backbone)
- The program will generate a plot for RMSD over time

c) Examine the temperature:
   
   \texttt{g\_energy -f pr.edr -o temperature_pr}

   - Select ‘14’ (Temperature)
   - Examine using GRACE. \texttt{xmgrace temperature_pr.xvg}

d) Use \texttt{g\_energy} to plot density\_pr.xvg and pressure\_pr.xvg, use \texttt{xmgrace} command to plot the graph.

e) System had been equilibrated and may proceed to the production stage when:
   
   a. The temperature plot stabilized/constant at 300K
   b. The average reading for density\_pr.xvg and pressure\_pr.xvg are 1000 kg/m\(^3\) and 1.05 bar respectively

9) Production Stage

\texttt{grompp -f md.mdp -c pr.gro -p abf.top -o md.tpr -maxwarn 5}

\texttt{mdrun -v -deffnm md}

10) Trajectories analysis

Time evolving coordinates of a system are called trajectories. Trajectory files (*.trr) are normally binary files that contain several sets of coordinates for the system.

a) Compress the trajectory

\texttt{trjconv -f md.trr -s md.tpr -o md.xtc -pbc nojump}

b) Analyse the energy output (same for potential energy, kinetic energy and total energy)

\texttt{g\_energy -f md.edr -o xxx and plot xmgrace -nxy xxx.xvg}

c) Measure radius of gyration and select ‘4’ (Backbone).

\texttt{g\_gyrate -f md.trr -s md.tpr -o abf\_gyrate.xvg}

d) Measure RMSD of the structure by and select ‘4’ (Backbone).

\texttt{g\_rms -s md.tpr -f md.trr -dt 10 -o md\_rmsd.xvg}

e) Compare RMSD to the NMR structure and select ‘4’ (Backbone)

\texttt{g\_rms -s em.tpr -f md.trr -o abf\_rmsd.xvg}

f) RMS fluctuation of atom positions and select ‘3’ (C-alpha).

\texttt{g\_rmsf -s md.tpr -f md.trr -b 200 -e 1000 -o abf\_rmsf}

g) RMSF to compute average structures and select ‘1’ (Protein).
h) Analyse the secondary structure of model by and select ‘1’ (Protein).

```
do_dssp -s md.tpr -f md.trr -o abf_ss.xpm -dt 10
```