

**COMPUTATIONAL STUDIES ON  
THERMOSTABILITY OF ENDOGLUCANASE  
FROM *FUSARIUM OXYSPORUM***

**BY**

**SHUKREE WAESHO**

**A dissertation submitted in partial fulfilment of  
the requirements for the degree of Master of  
Science (Biotechnology Engineering)**

**Kulliyyah of Engineering  
International Islamic University  
Malaysia**

**MAY 2013**

## ABSTRACT

Cellulose is the main component in plant cell and thus the most abundant biopolymer on earth. Cellulase is a group of enzymes that degrade cellulosic materials and belong to the O-glycoside hydrolases (EC 3.2.1.x). Endoglucanase (EC 3.2.1.4) is a key component in cellulase which has been used in various industries such as textiles, detergents, foods and animal feed, pulps and papers and recently in bio-fuel industries. Since most of the processes in many industries are carried out at higher temperature (above 60°C), the main limitation of cellulase utilization is the lack of enzyme activity and stability at higher temperatures. Generally, endoglucanases in glycoside hydrolase family 7 (GH 7) have the optimum temperature at 45-55°C and endoglucanase from *Fusarium oxysporum* (EGFO) completely loses activity after heating the enzymes at 60°C for three hours. In order to design a new thermostable endoglucanase from *Fusarium oxysporum*, molecular dynamics (MD) simulation technique was used to find out the dynamics factors responsible for the thermal stability of known endoglucanases (EG). Mesophilic endoglucanases from *Fusarium oxysporum* (EGFO) and thermophilic endoglucanase from *Humicola insolens* (EGHI) with known crystal structures and enzyme activity were used to compare their dynamical behaviors at 40°C and 60°C using MD simulation in aqueous media. It has been found that the Root Mean Square Deviation (RMSD) backbone of EGFO tends to increase more rapidly at higher temperatures, whereas the RMSD values for EGHI either remains similar or decreases at higher temperature. The RMSD helices of EGFO also have the behavior similar to that RMSD backbone. The secondary structure conformation at the residues position 225 to 231 of EGFO changes significantly at higher temperature, whereas conformation of EGFO at these positions is maintained as the temperature is increased. The EGHI shows salt-bridge interactions and hydrophobic interactions in these regions. Hence these two factors are crucial for the thermal stability of endoglucanase, this information obtained was used to carry out several *in silico* mutations on EGFO with the objective of designing more thermostable endoglucanase and found that the dynamic behavior of newly designed mutants are consistent with the conclusions. Therefore, the new quintuple mutant obtained by mutating at the positions T224E/G229A/S230F/S231E/N321R is predicted to be more thermostable than EGFO.

## خلاصة البحث

السليولوز هو العنصر الرئيسي في الخلية النباتية وبالتالي البوليمر الحيوي الأكثر وفرة على سطح الأرض. هو O-hydrolases سلولاز مجموعة من الانزيمات التي تتحلل المواد السليولوزية وتنتمي إلى عنصرا رئيسيا في سلولاز التي استخدمت في Endoglucanase (EC 3.2.1.4) (EC 3.2.1.x) غليكوزيدات مختلف الصناعات مثل المنسوجات والمنظفات والمواد الغذائية والأعلاف الحيوانية، والأوراق واللب مؤخرا الصناعات. منذ يتم تنفيذ معظم عمليات في العديد من الصناعات بها في درجة في إنتاج الوقود الحيوي والقيود الرئيسي لاستخدام سلولاز هو عدم وجود نشاط انزيم والاستقرار عند 60°C حرارة أعلى فوق لديهم درجة الحرارة المثلى في 45-55 (GH7) في الأسرة endoglucanases7 ارتفاع درجات الحرارة. عموما يفقد تماما النشاط بعد تسخين (EGFO) من أو كسيسبورم فيوزاريوم endoglucanase درجة مئوية، و جديدة بالحرارة من endoglucanase لمدة ثلاث ساعات. من أجل تصميم 60°C الانزيمات في تم استخدام تقنية المحاكاة لمعرفة العوامل المسؤولة عن (MD) أو كسيسبورم فيوزاريوم، الجزئية ديناميات أليف endoglucanases واستخدمت. (EG) المعروف endoglucanases ديناميات الاستقرار الحراري لل insolens Humicola حرارة من endoglucanase و (EGFO) الحرارة المعتدلة من أو كسيسبورم فيوزاريوم مع هياكل الكريستال المعروفة ونشاط انزيم لمقارنة تصرفاتهم الديناميكية عند 40°C درجة مئوية (EGHI) في الوسط المائي. وقد وجد أن جذر متوسط مربع الانحراف MD و 60°C درجة مئوية باستخدام المحاكاة يميل إلى زيادة بسرعة أكبر عند ارتفاع درجات الحرارة، في حين أن EGFO العمود الفقري ل (RMSD) من RMSD إما مشابهة أو النقصان لا يزال في ارتفاع درجة الحرارة. واللواجب EGHI ل RMSD القيم العمود الفقري. والتشكل هيكل الثانوي في موقف مخلفات RMSD أيضا سلوك مماثلة لتلك التي EGFO ارتفاع درجة الحرارة، في حين يتم الاحتفاظ التشكل من EGFO 225 حتى 231 من تغييرات كبيرة في جسر التفاعلات والتفاعلات EGHI في هذه المواقف كما يتم زيادة درجة الحرارة. ويظهر الملح EGFO ، endoglucanase مسعور في هذه المناطق. وبالتالي هذه العوامل هما حاسمة بالنسبة لاستقرار الحراري لل بهدف EGFO وتستخدم هذه المعلومات التي تم الحصول عليها لتنفيذ العديد من الطفرات في سيليكون على أكثر بالحرارة ووجدت أن السلوك الديناميكي من المسوخ المصممة حديثا تنفق endoglucanase تصميم

مع استنتاجات . لذلك، من المتوقع أن متحولة جديدة حصلت عليها خمس اضعاف تحور في EGFO thermostable. مواقع إلى أن تكون أكثر من T224E/G229A/S230F/S231E/N321R.

## APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion; it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Science (Biotechnology Engineering).

.....  
Ibrahim Ali Noorbatcha  
Supervisor

.....  
Hamzah Mohd Salleh  
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Science (Biotechnology Engineering).

.....  
Mohammed Saedi Jami  
Internal Examiner

.....  
Amir Feisal Merican  
External Examiner

This dissertation was submitted to the Department of Biotechnology Engineering and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology Engineering).

.....  
Faridah Yusof  
Head, Department of  
Biotechnology Engineering

This dissertation was submitted to the Kulliyyah of Engineering and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology Engineering).

.....  
Md Noor Bin Salleh  
Dean, Kulliyyah of  
Engineering

## DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Shukree Waesoho

Signature.....

Date.....

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

**DECLARATION OF COPYRIGHT AND AFFIRMATION  
OF FAIR USE OF UNPUBLISHED RESEARCH**

Copyright © 2013 by International Islamic University Malaysia. All rights reserved.

**COMPUTATIONAL STUDIES ON THERMOSTABILITY OF  
ENDOGLUCANASE FROM *FUSARIUM OXYSPORUM***

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the author and copyright holder except as provided below.

- i. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.
- ii. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.
- iii. The author of this dissertation will have the right to reproduce, rewrite and transmit copies (print or electronic) of full or any part of this dissertation for any purpose.
- iv. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

Affirmed by: Shukree Waesoho

.....  
Signature

.....  
Date

## ACKNOWLEDGEMENTS

My utmost gratitude goes to the Almighty, Allah (s.w.t.), for he only made this study possible, and gave me the knowledge and strength to carry it out to the best of my knowledge and ability. I would like to express my sincere gratitude to my supervisor Assoc. Prof. Dr. Ibrahim Ali Noorbacha for sharing his extensive knowledge on the subject matter, guiding me to successfully accomplish this research. I am thankful for having the chance to work under his supervision, as he was an inspiration for thirst for knowledge, dedication, and professionalism.

I would also like to thank my co-supervisor Assoc. Prof. Dr. Hamzah Mohd. Salleh for the time and effort that he spent to make this study success. Special thanks to anonymous internal and external examiners for sharing the knowledge and guiding me to successfully accomplish this dissertation.

I am extremely grateful to my parents (Mr. Rungsit (Abdul Rashid) Waesoho and Kim Mi Young), my sister (Miss.Sabirah Waesoho), and my family who have been strong pillars of support during this research, their boundless encouragement; wisdom and counseling have been the keys of all my achievements.

I also wish to express my special gratitude to Miss. Abir Hayeesa-i for her unconditional love and moral support. I am thankful to Marinee, Hazlin and other friends in Prince of Songkla University (PSU), University Kabangsaan Malaysia (UKM) and University Putra Malaysia (UPM) for motivating me during the course of my research and for sharing useful reference sources to contribution this thesis.

Very special thanks to my colleagues Mr. Anas Mufid Sultan, Mr. Muaz Abdul Hadi, Mr. Tamreen and my other friends in Biotechnology Engineering department at this university (IIUM) for sharing their thoughts and experience on the subject. Their assistance was great help and was highly appreciated. Special thank to Miss. Ayesha Masrur Khan who pioneer her research work under Assoc. Prof. Dr. Ibrahim Ali Noorbacha supervisor on molecular dynamics simulation.

I would like to thank the team at Research Computational Unit (RCU) in Research Management Office (RMO), for providing the tools required for running my computational experiments and Research Management Centre at IIUM as well as Ministry of Higher Education Malaysia for awarding the grant to support my research work.

Finally, special thank to my friends in Mahallah Zubair and Billal (especially, Mr.Rosuedi Chemaee) who have been encourage me to finish my thesis and last but not least, I would like to thank everyone who in one way or another contributed to my study.



# TABLE OF CONTENTS

Abstract .....	ii
Abstract (Arabic) .....	iii
Approval Page .....	iv
Declaration Page .....	v
Copyright Page .....	vi
Acknowledgements .....	vii
List of Tables .....	xi
List of Figures .....	xii
List of Abbreviations .....	xvi
List of Symbols .....	xvii
<b>CHAPTER ONE: INTRODUCTION .....</b>	<b>1</b>
1.1 Overview of Study .....	1
1.2 Problem Statement and Its Significance .....	3
1.3 Research Objectives .....	4
1.4 Research Methodology .....	4
1.5 Scope of Research .....	5
1.6 Dissertation Organization .....	5
<b>CHAPTER TWO: LITERATURE REVIEW .....</b>	<b>7</b>
2.1 Introduction .....	7
2.2 Utilizations of Enzymes for industrial applications .....	7
2.3 Endoglucanase and its Applications .....	9
2.4 Thermostable Enzymes .....	14
2.5 Overview of Experimental Study on Thermostability .....	16
2.5.1 Glycoside Hydrolase .....	18
2.5.2 Endoglucanase .....	19
2.6 Overview of Computational Study on Thermostability .....	20
2.6.1 Glycoside Hydrolase .....	24
2.6.2 Endoglucanase .....	26
2.7 Molecular Dynamics Simulation Theory .....	28
2.7.1 Equation of Motion .....	28
2.7.2 Force Field .....	29
2.7.2.1 Bonded Potential Energy Terms .....	30
2.7.2.2 Non-Bonded Potential Energy Terms .....	31
2.7.3 Numerical Methods .....	36
2.7.3.1 Finite Difference Method .....	36
2.7.3.2 Predictor-Corrector Integration Method .....	38
2.8 Summary .....	41
<b>CHAPTER THREE: MATERIALS AND METHODS .....</b>	<b>42</b>
3.1 Introduction .....	42
3.2 Software / Tools .....	44
3.2.1 Bioinformatics Tools .....	44

3.2.2 Homology modeling Tools .....	45
3.2.3 Validation Tools .....	45
3.2.4 Visualization Tools .....	45
3.2.5 Mutation Tools .....	46
3.2.6 MD Simulation Tools .....	46
3.2.7 Analysis Tools .....	46
3.2.8 Computers .....	46
3.3 Molecular dynamics simulation of selected endoglucanase .....	49
3.3.1 Structure Preparation .....	48
3.3.1.1 Sample Collection, Comparison and Validation .....	48
3.3.1.2 Generating Protein Structure File (PSF) .....	52
3.3.1.3 Solvation endoglucanase structures .....	53
3.3.1.4 Neutralization .....	56
3.3.2 Energy Minimization .....	58
3.3.3 Heating the simulation system .....	60
3.3.4 Equilibration the simulation system .....	62
3.3.4.1 Temperature Control .....	63
3.3.4.2 Pressure Control .....	64
3.3.5 MD Production Run.....	66
3.3.6 Dynamics Analysis.....	66
3.4 Summary .....	68
<b>CHAPTER FOUR: RESULTS AND DISCUSSION .....</b>	<b>69</b>
4.1 Introduction .....	69
4.2 Template Structures .....	70
4.2.1 Overview Comparison .....	70
4.2.1.1 Sequence Comparison .....	70
4.2.1.2 Secondary Structure Comparison .....	72
4.2.1.3 Comparison of Physicochemical Properties .....	77
4.2.2 MD simulation .....	81
4.2.2.1 Structural Features of the proteins .....	84
4.2.2.2 Root Mean Square Deviation (RMSD) .....	103
4.2.2.3 Radius of Gyration .....	111
4.2.2.4 Root Mean Square Fluctuation (RMSF) .....	114
4.2.2.5 Number of Hydrogen Bonds .....	119
4.2.2.6 Number of Salt Bridges .....	121
4.2.2.7 Hydrophobic Character .....	122
4.2.2.8 Summary of Dynamic Behavior Analysis .....	130
4.3 <i>In Silico</i> Mutation .....	132
4.3.1 List of the Mutation Points .....	132
4.3.2 Novel ThermostableEndoglucanase.....	133
4.3.3 Summary of <i>in Silico</i> Mutation .....	143
4.4 Summary.....	144
<b>CHAPTER FIVE: CONCLUSION AND RECOMMENDATION .....</b>	<b>145</b>
5.1 Conclusion .....	145
5.2 Recommendation .....	147
<b>BIBLIOGRAPHY.....</b>	<b>148</b>

<b>PUBLICATIONS .....</b>	<b>158</b>
APPENDIX I: THREE-LETTER AND ONE-LETTER AMINO ACID SYMBOLS .....	159
APPENDIX II: DESCRIPTION OF MD SIMULATION PARAMETERS ....	160
APPENDIX III: EXAMPLE OF NAMD CONFIGURATION FILES FOR PERFORMING ENERGY MINIMIZATION.....	161
APPENDIX IV: EXAMPLE OF NAMD CONFIGURATION FILES FOR PERFORMING HEATING .....	163
APPENDIX V: EXAMPLE OF NAMD CONFIGURATION FILES FOR PERFORMING EQUILIBRATION .....	165
APPENDIX VI: EXAMPLE OF NAMD CONFIGURATION FILES FOR PERFORMING PRODUCTION RUN .....	167
APPENDIX VII: EXAMPLE OF OUTPUT GRAPH OF MD SIMULATION: MINIMIZATION .....	169
APPENDIX VIII: EXAMPLE OF OUTPUT GRAPH OF MD SIMULATION: HEATING .....	170
APPENDIX IX: AN EXAMPLE OF OUTPUT GRAPH OF MD SIMULATION: EQUILIBRATION .....	171
APPENDIX X: EXAMPLE OF OUTPUT GRAPH OF MD SIMULATION: PRODUCTION .....	172

## LIST OF TABLES

<u>Table No.</u>		<u>Page No.</u>
2.1	Examples of industrially important enzymes utilizations	8
2.2	Potential energy terms used in molecular dynamics simulations	34
2.3	Some available molecular dynamics simulation packages	41
3.1	Endoglucanases used in this research	50
3.2	Validation scores (RMS Z-scores) of EGFO, EGHI and EGuia	51
3.3	Parameter adjustments for energy minimization	61
3.4	Parameter adjustments for heating the simulation systems	62
3.5	Parameter adjustments forequilibrating the simulation systems	65
4.1	Secondary structure elements of EGFO, EGuia and EGHI	73
4.2	Predicted physicochemical properties of EGFO, EGuia and EGHI	78
4.3	Comparisons of time consuming when running EGHI in 5000 step of production run by using a computer workstation	82
4.4	Comparison of time consuming when running EGHI in 5000 steps of production run by using a cluster	82
4.5	Temperature dependence of the RMSF values of residues 159 and 230 in EGFO, EGHI and EGuia	115
4.6	Number of salt bridge interactions accounting from the initial structure after equilibrium, comparison between EGFO, EGHI and EGuia	121
4.7	List of mutations selected for <i>in silico</i> experiment.	132

## LIST OF FIGURES

<u>Figure No.</u>		<u>Page No.</u>
2.1	Cellulose chain	10
2.2	Sheets structure of cellulose	11
2.3	Reaction scheme for glucoside catalysis by EGFO	15
2.4	Overview of approach to enhance the functionality and property of enzymes	18
2.5	Graph of van der Waals potential with and without the application of the switching function active, the potential is smoothly reduced to 0 at the cutoff distance	37
2.6	Graph showing an electrostatic potential with and without the application of the shifting function	37
2.7	Depiction of the difference between the cutoff distance and the pair list distance	38
3.1	Flow chart of overall research methodology	43
3.2	Flow chart of the strategy of performing generic MD simulation	44
3.3	Periodic boundary conditions	54
3.4	Structure of EGFO and TIP3P using the VMD program	56
4.1	multiple sequence alignment results of EGuia, EGFO and EGHI structures and secondary structure cartoon of EGFO(PDB) and EGHI(PDB)	71
4.2	Secondary structure of EGFO, EGHI and EGuia	75
4.3	Two positions (41 and 216) of EGFO, EGHI and EGuia	76
4.4	Percentage of the number of amino acids is calculated with respect to protein length of EGFO and EGHI	78
4.5	Percentage composition of amino acid residues is calculated with respect to protein length of (a) EGFO, (b) EGHI and (c) EGuia	80
4.6	Total energy during equilibration at constant temperature (313K and 333K) of EGFO, EGHI and EGuia	83

4.7	A side-view showing the open canyon formed by $\beta$ -sheets and the active site region on EGFO structure	85
4.8	A side-view showing the open canyon formed by $\beta$ -sheets and the active site region on EGHI structure	85
4.9	Bird's eye view of the active site canyon of EGFO and EGHI structure	87
4.10	Distances between Glu197 OE 2 (oxygen atom) as nucleophile and Glu202 OE 1 or OE2 as acid/base sits on EGFO and EGHI crystallized structures	88
4.11	Distances between Glu197 OE 2 (oxygen atom) as nucleophile and Glu202 OE 1 or OE2 as acid/base sits on EGuia crystallized structures	89
4.12	Variation in the distances between oxygen atoms OE1 in Glu202 and OE2 Glu197 of EGFO, EGHI and EGuia	90
4.13	Variation in the distances between oxygen atom OE2 in Glu202 and OE2 Glu197 of EGFO, EGHI and EGuia	91
4.14	Trajectory frames for EGFO at 0ns, 5ns and 10ns	94
4.15	Trajectory frames for EGHI at 0ns, 5ns and 10ns	95
4.16	Trajectory frames for EGuia at 0ns, 5ns and 10ns	96
4.17	Trajectory frames for EGFO at 0ns, 5ns and 10ns, highlighting residue 225 to 233	97
4.18	Trajectory frames for EGHI at 0ns, 5ns and 10ns, highlighting residue 225 to 233	98
4.19	Trajectory frames for EGuia at 0ns, 5ns and 10ns, highlighting residue 225 to 233	99
4.20	Enlarge view of specific regions (Residue 225 to 233) conformation change of EGFO and interact among residues and water molecule during simulation	100
4.21	Enlarge view of specific regions (Residue 225 to 233) conformation change of EGHI and interact among residues and water molecule during simulation	101

4.22	Enlarge view of specific regions (Residue 225 to 233) conformation change of EGuia and interact among residues and water molecule during simulation	102
4.23	Backbone RMSD of EGFO, EGHI and EGuia at 40°C and 60°C	105
4.24	RMSD for all $\beta$ -sheets of EGFO, EGHI and EGuia at 40°C and 60°C	106
4.25	RMSD for all helices of EGFO, EGHI and EGuia at 40°C and 60°C	107
4.26	RMSD for all turns of EGFO, EGHI and EGuia at 40°C and 60°C	108
4.27	RMSD for all coils of EGFO, EGHI and EGuia at 40°C and 60°C	109
4.28	Comparison of different RMSD values of the backbone and overall geometry of helices plotted as a function of MD simulation time	110
4.29	The radius of gyration as a function of MD simulation time of EGFO, EGHI and EGuia at 40°C and 60°C	112
4.30	Comparison of the radius of gyration values as a function of MD simulation time of EGFO, EGHI and EGuia at 40°C and 60°C	113
4.31	Comparison of RMSF values of EGFO, EGHI and EGuia at 40°C and 60°C	116
4.32	RMSF of EGFO, EGHI and EGuia at 40°C and 60°C	117
4.33	Difference between RMSF values at 313K and 333K of EGFO and EGHI	118
4.34	Difference between RMSF values at 313K and 333K of EGFO and EGuia	118
4.35	Difference between RMSF values at 313K and 333K of EGHI and EGuia	118
4.36	Average number of hydrogen bond during simulation for RMSD of EGFO, EGHI and EGuia at 40°C and 60°C	120
4.37	Average number of salt bridge interactions during simulation for RMSD of EGFO, EGHI and EGuia at 40°C and 60°C	122
4.38	Trajectory frames showing solvent accessible surface of EGFO at 0ns, 5ns and 10ns, highlighting residue 225 to 233	123
4.39	Trajectory frames showing solvent accessible surface of EGHI at 0ns, 5ns and 10ns, highlighting residue 225 to 233	124

4.40	Trajectory frames showing solvent accessible surface of EGuia at 0ns, 5ns and 10ns, highlighting residue 225 to 233	125
4.41	Enlarge view of specific regions (residues 225 to 233) based on solvent accessible surface for EGFO	126
4.42	Enlarge view of specific regions (residues 225 to 233) based on solvent accessible surface for EGHI	127
4.43	Enlarge view of specific regions (residues 225 to 233) based on solvent accessible surface for EGuia	128
4.44	Propose residues that might affect the thermostability of endoglucanase	131
4.45	Backbone RMSD of mutant1, mutant2 and mutant3 at 40°C and 60°C	135
4.46	Backbone RMSD of mutant4, mutant5 and mutant6 at 40°C and 60°C	136
4.47	RMSD for all helices of mutant1, mutant2 and mutant6 at 40°C and 60°C	137
4.48	Comparison the different RMSD values of backbone atoms plotted as a function of MD simulation time of EGFO, EGHI and mutant6	138
4.49	Comparison the different RMSD values of overall geometry of helices plotted as a function of MD simulation time of EGFO, EGHI and mutant6	138
4.50	RMSF of mutant6 at 40°C and 60°C	139
4.51	Difference between RMSF values at 313K and 333K of EGHI and mutant6	139
4.52	Difference between RMSF values at 313K and 333K of EGFO and mutant6	139
4.53	Trajectory frames for mutant6 at 0ns, 5ns and 10ns, highlighting residue 225 to 233	140
4.54	Trajectory frames showing solvent accessible surface of mutant6 at 0ns, 5ns and 10ns, highlighting residue 225 to 233	141
4.55	Enlarge view of specific regions (residues 225 to 233) based on solvent accessible surface for mutant6	142



## LIST OF ABBREVIATIONS

3D	three-dimensional
BcX	<i>Bacillus circulans</i> xylanase
CHARMM	Chemistry at HARvard Molecular Mechanics
CMC	Carboxymethyl cellulose
Cel12A	Cellulase (clan A) from family 12
EG	Endoglucanase
EGFO	Endoglucanase from <i>Fusariumoxysporum</i>
EGHI	Endoglucanase from <i>Humicola</i> <i>insolens</i>
EGuia	Previous mutation of Endoglucanase in IIUM (UIA) laboratory
DNA	Deoxyribonucleic acid
GH	Glycoside Hydrolase
GUS	$\beta$ -glucuronidase
IIUM	International Islamic University Malaysia
MD	Molecular dynamics
MC	Monte Carlo
NAMD	NAnoscalable Molecular Dynamics
PSF	Protein Structure File
PDB	Protein Data Bank
PME	Particle Mesh Ewald
pNPC	p-nitro-phenyl cellobiose
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
UIA	Universiti Islam Anterabangsa
VMD	Visual Molecular Dynamics

## LIST OF SYMBOLS

$\sigma_i$	solvation parameter
$\epsilon(r)$	position-dependent dielectric
$\Psi(r)$	electrostatic potential
$\rho^f(r)$	charge density of solute
$\lambda(r)$	factor for the position-dependent accessibility of position $r$
$\epsilon_0$	permittivity of free space
$\epsilon$	dielectric constant
$a_i$	acceleration of particle $i$
$k_c$	electrostatic constant
$q_i$	electrostatic charge of particle $i$
$r_{ij}$	distance between particles $i$ and $j$
$S_{ij}$	score of amino acids $i$ and $j$
$p_{ij}$	probability of amino acids $i$ and $j$ replacing each other in a homologous sequence
$\zeta$	collision frequency
$\gamma$	friction coefficient
$F_i$	force on particle $i$
$m_i$	mass of particle $i$
$U$	total potential energy
$U_{\text{bonded}}$	total bonded potential energy
$U_{\text{unbonded}}$	total unbonded potential energy

# CHAPTER ONE

## INTRODUCTION

### 1.1 OVERVIEW OF STUDY

Enzyme technology has become a preferred choice of technology in various chemical industries. The large-scale production of the variety of products using chemical technology generates a lot of hazardous waste, which is a serious problem that has severe impact on environment. In order to address this problem, various industries attempt to avoid using chemical technology. Cellulose is one of the most important industrial cellulosic materials that can be degraded by cellulolytic enzymes known as cellulase.

Endoglucanase is one of the key components of this complex multi-enzyme system (cellulase), which breaks internal bonds in the crystalline structure of cellulose and exposes individual cellulosic polysaccharide chains. Cellulase has been used in various industries such as textiles (Buchert and Heikinheimo, 1998 and Reily et al., 2004), detergents (Walsh, 2002), foods and animal feed (Galante et al., 1998 and Urlaub, 2002), pulps and papers (Suurnakki et al., 2004), and recently in bio-fuel industries (Kumar et al., 2009). One of the main limitations of most cellulase utilization is the lack of enzyme activity and stability at high temperatures. Since most of the processes in many industries are carried out at high temperatures (above 60°C) (Shuyan et al., 2006), it is clear that thermostable enzymes is very important requirement in industrial processes using enzyme.

Most of the thermophilic cellulolytic enzymes have been widely isolated from thermophilic and hyperthermophilic fungi, which displays optimal temperature between 50 and 80°C. Some of these enzymes such as endoglucanase from *Thermotoga neapolitana* have very high thermostability, (half lives of 130 min at 106°C) (Bok et al., 1998). However, the endoglucanase from a thermophilic pathogenic plant fungus (*Fusarium oxysporum*) is found to have an optimum activity at not more than 60°C and also lack stability at this temperature. Moreover, it has been observed from the literature review that endoglucanase from *Fusarium oxysporum* is not stable at 60°C (Vlasenko et al., 2010). On the other hand, Shuyan et al., (2006) have found a novel endoglucanase from *Fusarium oxysporum*, which has the optimal temperature at 75°C but the gene of the enzyme is different from the known endo-1,4-glucanase from *Fusarium oxysporum* (Gene bank ID no. AAA65586.1).

Two different protein engineering approaches to enhance the enzyme performance are rational design and directed evolution (Kazlauskas and Bornscheuer, 2009). Protein engineering via the directed evolution approach is expensive, while protein engineering via the computational mutation (*in silico*) can be carried out, in order to save the cost. Using the computational approach, the scope of mutation can be reduced, and it can be narrowed down to a specific point of mutation or region in the protein (Noorbach et al., 2009). For instance, to carry out a mutation on a structure with 400 residues, every single position of residues can be substituted with twenty amino acids. Thus, it is tedious to do the random mutations in the laboratory. Hence, the computer aided protein design can help in reducing cost of trial-and-error method adapted in random mutation. It also provides more insights on the role of various residues on the enzyme activity.

This study describes the research carried out in designing more thermostable endoglucanase from *Fusarium oxysporum*. Computational studies via the Molecular Dynamics (MD) simulation is used to examine the dynamic behavior of the available three-dimensional (3D) structures of endoglucanases for which experimental data on thermal stability is available (Vlasenko et al., 2010). The comparison of endoglucanase structure from two different sources at the different temperatures had been done on this research, and the thermostability rules derived from this work are applied to design a newly thermostable endoglucanase from *Fusarium oxysporum*.

## **1.2 PROBLEM STATEMENT AND ITS SIGNIFICANCE**

In terms of better understanding of cellulase usage in various applications, the characteristics of the enzyme (example: thermostable enzymes, enzymes active in extreme pH, enzymes using different types of substrates, etc.) are essential. *Fusarium oxysporum*, which is used for the production of endoglucanase, is available in the International Islamic University Malaysia (IIUM) laboratory. We have plans to produce endoglucanase using this fungus. Endoglucanase from *Fusarium oxysporum* lacks thermal stability at 60°C (Vlasenko et al., 2010). However, temperatures above 60°C are required in many industrial processes (Shuyan et al., 2006) and hence there is a need to improve the thermal stability of this enzyme.

As discussed before, a directed evolution process to obtain a thermostable enzyme is very time consuming, and hence computer aided design approach is selected in this research. Currently, the molecular dynamics simulation is the only method, which can be used to simulate behavior of the enzymes at higher temperatures, and hence provide a strategy to understand the factors that contribute to

the stability of the enzymes at higher temperatures. Hence we have employed the molecular dynamics simulation approach to compare dynamics behavior of endoglucanase from the glucoside hydrolase family 7 (GH 7) and predict the regions and mutation points crucial for thermal stability of endoglucanase from *Fusarium oxysporum*. Thus, the computational studies for increasing the thermostability of the enzyme will lead to novel theoretical design models for engineering endoglucanase to improve the thermostability of this enzyme.

### **1.3 RESEARCH OBJECTIVES**

The overall objective of this research is to improve the thermal stability of endoglucanase via computational methods. The specific objectives of this study are as follows:

- a) To simulate the structure of mesophilic and thermophilic endoglucanase.
- b) To identify structural and dynamic factors responsible for thermostability of endoglucanase.
- c) To computationally predict thermostable-mutation in endoglucanase from *Fusarium oxysporum*.

### **1.4 RESEARCH METHODOLOGY**

This research implements computational methods to achieve the objectives. Several methods and tools have been used to study the structural and dynamic properties of endoglucanase from GH 7. The mathematical algorithms, including force field calculations and numerical methods are applied in MD simulation. They are used to

study the changes and behavior of endoglucanase at different temperatures. The stability of enzymes would be related to time dependent displacement of the atomic position in the enzyme, which is represented by the Root Mean Square Deviation (RMSD) of the atoms and residues in the protein (backbone, beta-sheets, alpha-helix, turns, and coils). The radius of gyration of the molecule is also calculated to analyze the compactness of the structure at difference temperatures. Moreover, the structure conformation changes of endoglucanase during simulation were analyzed to understand the reason behind the observed changes.

## **1.5 SCOPE OF RESEARCH**

This research uses computational methods to study and design a novel thermostable endoglucanase from *Fusarium oxysporum*. MD simulation and computational mutation techniques are employed to the selected three-dimensional structures of endoglucanase from GH 7. Analysis of the obtained results would provide significant static and dynamics factors, and conditions responsible to the thermal stability of endoglucanase.

## **1.6 DISSERTATION ORGANIZATION**

Five chapters of this thesis describe the research work that has been done to study and design a novel themostable endoglucanase from *Fusarium oxysporum*. Chapter 1 provides a short overview followed by a description of the importance on thermal stability of endoglucanase and the ideas on improving thermal stability of endoglucanase via computational protein engineering. Furthermore, the research objectives, methodology and scope are explained in this chapter.

The previous studies related to this research are described in chapter 2 in the form of a literature review. This chapter also provides the information about the experimental and computational studies implemented on the thermostability of proteins. Moreover, a review on endoglucanase, its applications and several protein engineering approaches to design thermostable enzymes are presented in this chapter.

Chapter 3 of this thesis records the materials and methods used to perform the study. The list of tools and software is pointed out, and this chapter also explains in detail, the methodology used in this research. The molecular dynamics simulation steps are described in detail to show how the molecular dynamics simulation concept can be implemented for studying the thermostability of endoglucanase.

The results and findings obtained throughout this research are provided in chapter 4. This chapter gives a critical analysis and comprehensive discussion on our new findings related to the thermostability of endoglucanase. This includes the results of the molecular dynamics simulation of endoglucanase from GH 7, RMSD comparison, analysis of the number of hydrogen bonds and salt bridges, and analysis of radius of gyration of the enzyme.

The last chapter of this thesis (chapter 5) presents a summary of the overall results and findings of this research and its limitations. Moreover, the recommendations for future studies related to the designing of thermostable proteins in general and endoglucanase in particular are included in this chapter.