COMPUTATIONAL STUDIES ON THERMOSTABILITY OF ENDOGLUCANASE FROM *FUSARIUM OXYSPORUM*

BY

SHUKREE WAESOHO

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 activitywere 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

السليلوز هو العنصر الرئيسي في الخلية النباتية وبالتالي البوليمر الحيوي الأكثر وفرة على سطح الأرض. هو سلولاز مجموعة من الانزيمات التي تتحلل المواد السليلوزية وتنتمي إلى hydrolases O– عنصرا رئيسيا في سلولاز التي استخدمت في (EC 3.2.1.4) .Endoglucanase (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 والقيد الرئيسي لاستخدام سلولاز هو عدم وجود نشاط انزيم والاستقرار عند 60°Cحرارة أعلى فوق لديهم درجة الحرارة المثلى في 45-55 (GH7) في الأسرة 7endoglucanases ارتفاع درجات الحرارة. عموما يفقد تماما النشاط بعد تسخين (EGFO) من أوكسيسبورم فيوزاريومendoglucanase درجة مئوية، و جديدة بالحرارة من endoglucanase Cلمدة ثلاث ساعات. من أجل تصميم°60الانزيمات في تم استخدام تقنية المحاكاة لمعرفة العوامل المسؤولة عن (MD) أوكسيسبورم فيوزاريوم، الجزيئية ديناميات أليف endoglucanases واستخدمت.(EG) المعروفendoglucanases ديناميات الاستقرار الحراري لل insolens Humicola حرارة من endoglucanase و(EGFO)الحرارة المعتدلة من أوكسيسبورم فيوزاريوم مع هياكل الكريستال المعروفة ونشاط انزيم لمقارنة تصرفاتهم الديناميكية عند°40C درجة مئوية (EGHI) في الوسط المائي. وقد وجد أن جذر متوسط مربع الانحرافMD و°60C درجة مئوية باستخدام المحاكاة يميل إلى زيادة بسرعة أكبر عند ارتفاع درجات الحرارة، في حين أن EGFOالعمود الفقري ل (RMSD) منRMSD إما مشابهة أو النقصان لا يزال في ارتفاع درجة الحرارة. واللوالبEGHI لRMSD القيم العمود الفقري. والتشكل هيكل الثانوي في موقف مخلفات RMSD أيضا سلوك مماثلة لتلك التي EGFO ارتفاع درجة الحرارة، في حين يتم الاحتفاظ التشكل منEGFO 225حتي 231من تغييرات كبيرة في جسر التفاعلات والتفاعلات EGHI في هذه المواقف كما يتم زيادة درجة الحرارة. ويظهر الملح EGFO ، endoglucanaseمسعور في هذه المناطق. وبالتالي هذه العوامل هما حاسمة بالنسبة لاستقرار الحراري لل بهدف EGFOوتستخدم هذه المعلومات التي تم الحصول عليها لتنفيذ العديد من الطفرات في سيليكون على أكثر بالحرارة ووجدت أن السلوك الديناميكي من المسوخ المصممة حديثا تتفق endoglucanase تصميم

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

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).

.....
FaridahYusof
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

COMPUTATIONAL STUDIES ON THERMOSTABILITY OF ENDOGLUCANASE FROM *FUSARIUM OXYSPORUM*

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 Noorbatcha 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 Noorbatcha 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 Chemae) 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

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D	three-dimensional
BcX	*Bacillus circulans*xylanase
CHARMM	Chemistry at HARvard Molecular Mechanics
CMC	Carboxymethyl cellulose
Cel12A	Cellulase (clan A) from family 12
EG	Endoglucanase
EGFO	Endoglucanase from *Fusariumoxysporum*
EGHI	Endoglucanase from *Humicolarinsolens*
EGuia	Previous mutation of Endoglucanase in IIUM (UIA) laboratory
DNA	Deoxyribonucleic acid
GH	Glycoside Hydrolase
GUS	β-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

σ_i	solvation parameter
$\epsilon(r)$	poison-dependent dielectric
$\Psi(r)$	electrostatic potential
$\rho^f(r)$	charge density of solute
$\lambda(r)$	factor for the position-dependent accessibility of position *r*
ε_0	permittivity of free space
ε	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
ζ	collision frequency
γ	friction coefficient
F_i	force on particle *i*
m_i	mass of particle *i*
U	total potential energy
U_{bonded}	total bonded potential energy
$U_{un\text{bonded}}$	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 (Noorbatcha 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 filed 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 themostability 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.