การเพิ่มการแสดงออกของเอนไซม์สลายเซลลูโลสใน *Pichia pastoris* และการประยุกต์ในการหมักเอทานอลจากฟางข้าว

นายณษพัฒน์ บุญวิทยา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิทยาศาสตร์ชีวภาพ
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

OVEREXPRESSION OF CELLULOLYTIC ENZYMES IN *Pichia pastoris* AND ITS APPLICATIONS IN ETHANOL FERMENTATION OF RICE STRAW

Mr. Nassapat Boonvitthya

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Biological Science
Faculty of Science
Chulalongkorn University
Academic Year 2011

Thesis Title	Overexpression of Cellulolytic Enzymes in *Pichia pastoris* and Its Applications in Ethanol Fermentation of Rice Straw
By	Mr. Nassapat Boonvitthya
Field of Study	Biological Science
Thesis Advisor	Associate Professor Warawut Chulalaksananukul, Ph.D.
Thesis Co-advisor	Michael J. O'Donohue, Ph.D.
	Sophie Bozonnet, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

…….. Dean of the Faculty of Science
(Professor Supot Hannongbua, Dr., rer, nat)

THESIS COMMITTEE

…….. Chairman
(Assistant Professor Chumpol Khunwasi, Ph.D.)

……..……. Thesis Advisor
(Associate Professor Warawut Chulalaksananukul, Ph.D.)

…….. Thesis Co-advisor
(Michael J. O'Donohue, Ph.D.)

…….. Thesis Co-advisor
(Sophie Bozonnet, Ph.D.)

…….. Examiner
(Associate Professor Alisa Vangnai, Ph.D.)

…….. Examiner
(Chompunuch Virunanon, Ph.D.)

…….. External Examiner
(Vorakan Burapatana, Ph.D.)

ณษพัฒน์ บุญวิทยา : การเพิ่มการแสดงออกของเอนไซม์สลายเซลลูโลสใน *Pichia pastoris* และการประยุกต์ในการหมักเอทานอลจากฟางข้าว. (OVEREXPRESSION OF CELLULOLYTIC ENZYMES IN *Pichia pastoris* AND ITS APPLICATIONS IN ETHANOL FERMENTATION OF RICE STRAW) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : รศ. ดร. วรวุฒิ จุฬาลักษณานุกูล, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม : Michael J. O'Donohue, Ph.D., Sophie Bozonnet, Ph.D, 150 หน้า.

ในปัจจุบันนี้ความสำคัญของอุตสาหกรรมการกลั่นชีวภาพจากมวลชีวภาพที่มีศักยภาพทางเศรษฐกิจกำลังได้รับความสนใจเพิ่มขึ้นและกลายเป็นงานวิจัยที่มีความความสำคัญสูง อย่างไรก็ตามสิ่งมีชีวิตที่มีความสามารถในการย่อยสลายมวลชีวภาพผลิตเอนไซม์หลากหลายชนิดซึ่งยากต่อการทำให้บริสุทธิ์ และประยุกต์ใช้ในอุตสาหกรรมการกลั่นชีวภาพโดยเฉพาะอย่างยิ่งเมื่อคำนึงถึงค่าอุณหภูมิและพีเอชที่ทำให้ค่าแอคทิวิตีของเอนไซม์สูงสุด ดังนั้นยังคงมีความจำเป็นในการผลิตเอนไซม์เซลลูเลสจากเชื้อราในสิ่งมีชีวิตชนิดอื่นที่มีความสามารถในการผลิตเอนไซม์ได้มาก โดยที่สามารถถ่ายโอนยีนและทำให้เอนไซม์บริสุทธิ์ได้ง่ายเพื่อที่จะมอบความสามารถในการสลายเซลลูโลสให้กับจุลินทรีย์ที่ไม่มีความสามารถนี้และพัฒนาตัวเร่งปฏิกิริยาให้มีลักษณะบางอย่างดีกว่าตัวเร่งปฏิกิริยาแบบดั้งเดิมโดยผ่านการออกแบบอย่างมีเหตุผลหรือผ่านวิวัฒนาการแบบตรง ในวิทยานิพนธ์นี้ ยีนกลูแคนเบต้ากลูโคซิเดสจากเชื้อรา *Aspergillus oryzae* ประสบความสำเร็จในการแสดงออกภายใต้การควบคุมของโพรโมเตอร์ที่มีการแสดงออกแบบตลอดเวลาหรือแบบเหนี่ยวนำในยีสต์ *Pichia pastoris* เอนไซม์ที่ผลิตได้มีขนาดมวลโมเลกุลประมาณ 40 กิโลดาลตัน โดยมีความเหมือนของลำดับกรดอะมิโนกับฐานข้อมูล 96 เปอร์เซ็นต์ เอนไซม์ที่ผลิตได้มีค่าแอคทิวิตีสูงสุดที่ 2 ยูนิตต่อมิลลิลิตร ที่เวลา 42 ชั่วโมง ภายใต้การควบคุมแบบเหนี่ยวนำและมีความสามารถทนกลูโคสได้ที่ความเข้มข้น 365 มิลลิโมลาร์ ค่าคงที่ของมิคาลิส 0.56 มิลลิโมลาร์ และอัตราเร็วสูงสุด 10042 ไมโครโมลาร์ต่อนาทีต่อมิลลิกรัมโปรตีน นอกจากนี้ลำดับการเข้ารหัสของยีนเอนโดกลูคาเนส 2 และเซลโลไบโอไฮโดรเลส 2 จากเชื้อรา *Trichoderma reesei* สายพันธุ์ QM9414 ยังประสบความสำเร็จในการถูกโคลนเข้าสู่ยีสต์ *Yarrowia lipolytica* และทำการเปรียบเทียบระดับการแสดงออกกับยีสต์ *P. pastoris* เพื่อชี้ให้เห็นความเป็นไปได้ที่จะใช้ *Y. lipolytica* เป็นยีสต์ทางเลือกในการผลิตเอนไซม์ย่อยสลายเซลลูโลส ค่าแอคทิวิตีของเอนไซม์เอนโดกลูคาเนสและเซลโลไบโอไฮโดรเลสถูกทำให้สูงสุดในยีสต์ *Y. lipolytica* สายพันธุ์ Po1d โดยใช้โพรโมเตอร์ที่มีการแสดงออกแบบตลอดเวลาและใช้สัญญาณการหลั่งของยีน preproLip2 แอคทิวิตีของเอนไซม์เอนโดกลูคาเนสที่ได้จาก *Y. lipolytica* มีค่าน้อยกว่าที่ได้จาก *P. pastoris* ที่ถูกเหนี่ยวนำด้วยเอทานอลความเข้มข้น 3 เปอร์เซ็นต์ ประมาณ 7 เท่า ในขณะที่ระดับการแสดงออกของเอนไซม์เซลโลไบโอไฮโดรเลสที่ได้จาก *Y. lipolytica* และค่าแอคทิวิตีจำเพาะมีค่ามากกว่าที่ได้จาก *P. pastoris* ระดับของการเปลี่ยนแปลงของเอนไซม์ผ่านกระบวนการไกลโคซิเลชั่นมีผลเพียงเล็กน้อยต่อระดับการแสดงออกและคุณสมบัติของเอนไซม์ ภายหลังการทำวิวัฒนาการแบบตรงโดยใช้วิธีปฏิกิริยาลูกโซ่แบบสุ่ม และการทำให้เกิดการกลายแบบอิ่มตัวบริเวณจำเพาะ พบว่ากรดอะมิโนของยีนเอนโดกลูคาเนส 2 ที่ตำแหน่ง 257 ได้เปลี่ยนแปลงจากกรดอะมิโนทรีโอนีนเป็นกรดอะมิโนแอสปาเตตหรือแอสปาราจีน ซึ่งสามารถปรับปรุงคุณสมบัติการทนร้อนของเอนไซม์นี้ จากที่สามารถทนร้อนได้ที่อุณหภูมิ 70 องศาเซลเซียส เวลา 60 นาที เป็น 120 นาที ผลการทดลองนี้ชี้ให้เห็นว่า *Y. lipolytica* เป็นระบบที่มีศักยภาพและน่าสนใจสำหรับการแสดงออกของเอนไซม์และการคัดเลือกสายพันธุ์กลายอย่างมีประสิทธิภาพ

สาขาวิชา วิทยาศาสตร์ชีวภาพ ลายมือชื่อนิสิต.....

ปีการศึกษา 2554 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....

4973814523 : MAJOR BIOLOGICAL SCIENCE
KEYWORDS : *Trichoderma reesei* / *Aspergillus oryzae* / cellulases / heterologous expression / thermostability / *Yarrowia lipolytica* / *Pichia pastoris*

NASSAPAT BOONVITTHYA : OVEREXPRESSION OF CELLULOLYTIC ENZYMES IN *Pichia pastoris* AND ITS APPLICATIONS IN ETHANOL FERMENTATION OF RICE STRAW. ADVISOR : ASSOC. PROF. WARAWUT CHULALAKSANANUKUL, Ph.D., CO-ADVISOR : MICHAEL J. O'DONOHUE, Ph.D., SOPHIE BOZONNET, Ph.D, 150 pp.

In recent years, the importance of economically viable biorefineries from lignocellulosic biomass is increasing and become a high research priority. However, cellulolytic organisms produce multiple enzymes that are difficult to purify and adapt for biorefining process configurations, notably concerning temperature and pH optima. Therefore, there is still a need to produce fungal cellulases in heterologous systems allowing easy transformation and purification in order to confer cellulolytic capabilities to other non-cellulolytic microorganisms, and develop biocatalysts posing some characteristics over the traditional catalyst through rational or directed evolution strategies. In this context, a glucan 1,3-beta-glucosidase A gene (*exgA*) from *Aspergillus oryzae* was favorably expressed under the control of either constitutive or inducible promoter in *Pichia pastoris*. Recombinant ExgA had an apparent of molecular weight about 40 kDa having 96% amino acid sequence homology with *A. oryzae* ExgA. The resulted showed enzymatic activity was highest at 2 U/ml after 42 h for inducible expression, and tolerated to glucose inhibition with K_i, K_M and V_{max} were 365 mM, 0.56 mM, 10042 μmol min^{-1} mg of $protein^{-1}$, respectively. Moreover, the sequences encoding endoglucanase II (*egl*II) and cellobiohydrolase II (*cbh*II) from the fungus *Trichoderma reesei* QM9414 were also successfully cloned and expressed in *Yarrowia lipolytica* and *P. pastoris* expression system to point out the possibility to use *Y. lipolytica* as alternative cellulolytic yeast. Extracellular endoglucanase and cellobiohydrolase activity was maximized in *Y. lipolytica* Po1d strain using constitutive promoter and preproLip2 secretion signal. The endoglucanase activity was less than seven-times when compared to recombinant *P. pastoris* induced by 3.0% (v/v) methanol, whereas, the expression level of cellobiohydrolase from *Y. lipol*ytica was higher than in *P. pastoris*. The specific activity of both proteins was greater than their homologs produced by *P. pastoris*, and glycosylation level had little effect on their enzymatic activity and properties. After two rounds of directed evolution via error-prone PCR and site-saturation mutagenesis, variants T257N and T257D were the best thermostable EGII mutants. The thermostability of EGII mutants was improved which half of its activity was lost at 70 °C within 120 min. These results demonstrated that *Y. lipolytica* is potentially an excellent and attractive system for heterologous expression and high-throughput screening.

Field of Study : BIOLOGICAL SCIENCE Student's Signature .....

Academic Year : 2011 Advisor's Signature .....

Co-advisor's Signature .....

Co-advisor's Signature .....

ACKNOWLEDGEMENTS

I also would like to express my sincere gratitude to Assistant Professor Dr. Chumpol Khunwasi, Associate Professor Dr. Alisa Vangnai, Dr. Chompunuch Virunanon, and Dr. Vorakan Burapatana who are the chairman and members of thesis committee. They dedicated valuable time to make useful comments and suggestion.

I would like to express my gratitude and appreciation to my advisor Associate Professor Dr. Warawut Chulalaksananukul for accepting me as a student since Bachelor's degree and for his constant enthusiasm encouragement and kindly support.

My deepest gratitude goes to my co-advisor, Dr. Michael J. O'Donohue and Dr. Sophie Bozonnet, for their invaluable guidance, suggestion, reviewing and comment on manuscripts, encouragement when the things got rough, and support throughout the whole period in France.

A great deal of appreciation goes to my friends and colleagues in hemicellulase group: Claire Dumon, Régis Faure, Gélladine Bastien Uluis, Letian Song, Hélène Texier, Marie-Laure Desrousseaux, Susana Peremarti Barba, Grégory Arnal, Faten Arab, Bastien Bissaro, and Vinciane Borsenberger for their kindness help, guidance, and friendship during my stay in France. I also thanks to all members at EAD1, INSA-Toulouse, especially, Florence Bordes, Leticia Casas, Ivana Rivera, Romain Irague, and Cédric Montanier for their kindness help and friendship throughout my research time in France. I am thankful to my colleagues and lovely sisters in Biofuel by Biocatalyst Research Unit for closet friendship during my thesis.

I would like to acknowledge Chulalongkorn University Graduate Scholarship to Commemorate the 72nd Anniversary of His Majesty King Bhumibol Adulyadej, and Chulalongkorn University Dutsadi Phiphat Scholarship, which is collaboration between Chulalongkorn University and PTT public company limited, for financial support. My thankful also extend to Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP), INSA de Toulouse and Department of Botany, Faculty of Science, Chulalongkorn University for providing facilities and equipments.

Finally, the greatest gratitude is expressed to my lovely parents, my brother, and my dear, Siriporn, for their loves, support, confidence, and understanding during this long journey.

CONTENTS

PAGE

PAGE

PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BGL:	Beta-glucosidase
CBH:	Cellobiohydrolase
Cel:	Cellulase
CMC:	Carboxymethyl cellulose
DNA:	Deoxyribonucleic acid
DNS:	3,5-dinitrosalycilic acid
EG:	Endoglucanase
ExgA:	Glucan 1,3-beta-glucosidase A
dNTP:	Deoxyribonucleotide
EDTA:	Ethylene diamine tetraacetic acid
GH:	Glycoside hydrolase
LB:	Luria-Bertani culture media
LiAc:	Lithium acetate
OD:	Optical density
PAGE:	Polyacrylamide gel electrophoresis
PASC:	Phosphoric acid swollen cellulose
PCR:	Polymerase chain reaction
RNA:	Ribonucleic acid
SDS:	Sodium dodecyl sulfate
Tris:	Trishydroxymethylaminomethane

LIST OF PUBLICATIONS

This thesis is based on the following articles referred to in the text by Arabic letters.

A: Boonvitthya, N., Tanapong, P., Kanngan, P., Burapatana, V., and Chulalaksananukul, W. (2012). Cloning and expression of *Aspergillus oryzae* glucan 1,3-beta-glucosidase A in *Pichia pastoris*. (Submitted)

B: Boonvitthya, N., Bozonnet, S., Burapatana, V., O'Donohue, M. J., and Chulalaksananukul, W. (2012). Comparison of the heterologous expression of *Trichoderma reesei* endoglucanase II and cellobiohydrolase II in the yeasts *Pichia pastoris* and *Yarrowia lipolytica*. (In press)

C: Boonvitthya, N., Bozonnet, S., Burapatana, V., O'Donohue, M. J., and Chulalaksananukul, W. (2012). Directed evolution of *Trichoderma reesei* endoglucanase II to improve its thermostability in *Yarrowia lipolytica*. (Preparation)

CHAPTER I

INTRODUCTION

Biorefinery is the configuration of process that uses biological matter to produce various kinds of biological products. The use of biomass and especially plant waste materials for production of biofuels, chemicals, materials, heat, and power has increased significantly during the past decades. Cellulose, the main constituent of agricultural residues, is a linear homopolymer of repetitive anhydrous D-glucose units coupled to each other by β-1,4-glycosidic bonds. It has enormous potential as a renewable source for biorefinery because it is the most abundant polysaccharide on Earth approximately 4 x 10^9 tonnes per year. However, due to the complexity and recalcitrant structure of the polymer, it requires collective groups of enzyme known as cellulases to work synergistically for complete degradation, especially endoglucanases (EC 3.2.1.4; EGs), cellobiohydrolases (EC 3.2.1.91; CBHs) and β-glucosidases (EC 3.2.1.21; BGLs). EGs hydrolyze internal β-1,4-bonds that are located in amorphous regions, while CBHI and CBHII act on the free reducing and non-reducing extremities respectively, and produce cellodextrins as main products. The resultant soluble cellodextrins are then hydrolyzed by BGLs to produce glucose. The simple sugar can be further decomposed into alcohols, methane, and hydrogen which can be used for bioethanol and electricity. Moreover, cellulases can be applied to use in many industrial sectors such as food, brewery and wine, animal feed, textile and laundry, pulp and paper applications. Among cellulase producers, *Trichoderma reesei* (Ascomycota: *Hypocreales*) is an efficient colonizer of lignocellulosic biomass secreting extracellular protein up to 20 g/L. Seventeen cellulases from this fungus are discovered consisting of two cellobiohydrolases, eight endoglucanases, and seven β-glucosidases, but only four major cellulases (EGI, EGII, CBHI, and CBHII) are secreted in large quantities (90–95% of the total secreted cellulases). Nonetheless, *T. reesei* BGLs is partly mycelium-bound and presented in relatively small amount in the extracellular cellulase complex, and it is more easily inhibited from glucose compared to those of *Aspergillus* species. Moreover, commercial cellulases are still

too costly in industrial bioconversion, and process conditions are usually not suitable for enzymes to work effectively e.g. extreme pH, temperature, pressure, organic solvents, and inhibitors. Therefore, isolating and engineering of novel cellulases with relatively high activity and thermostability, as well as to make them resistant to extreme pH or tolerant to organic solvents are now considered as a big challenge for both academic purposes and industrial applications. In order to improve cellulase production, activity, and stability, development of these properties can be pursued by two strategies. The native cellulolytic strategy involves metabolic engineering, random mutagenesis to select desired properties, and co-culturing. The recombinant cellulolytic strategy means engineering of cellulase genes into non-cellulolytic microorganisms having high production background, excellent product properties, and eases to screen and characterize of interested proteins. Due to the low expression level and difficulty to study the properties of individual cellulase in the native hosts, heterologous expression is a powerful strategy to improve the yield of enzymatic production, as well as produce the enzyme in highly purified and well characterized form. *Yarrowia lipolytica* and *Pichia pastoris* are two of the non-conventional yeasts that most widely used for recombinant enzyme expression. They offer several advantages such as rapid growth in simple media to high cell density stage, ease of genetic manipulation, high expression level background, and ability to perform eukaryotic processing mechanisms. To improve enzymes for desired properties, directed evolution is a method used in protein engineering that mimic natural evolution. It is random and combinatorial approaches which do not require sufficient insight into the structure – function relationships to introduce mutations.

In this thesis the focus is on the heterologous expression of cellulases in two powerful yeast expression systems, and engineering the enzyme to be more stable at high temperature via directed evolution approach.

In second chapter, the structure of lignocellulosic substrates, the characteristics of microbial cellulolytic enzymes, the key features for heterologous expression of cellulases and their enzyme engineering are introduced, discussed and compared. Cloning and expression of *A. oryzae* glucan 1,3-beta-glucosidase A in *P. pastoris* was investigated under two expression promoters and presented in the third

chapter (Article A). *T. reesei* EGII and CBHII were heterologously expressed in *Y. lipolytica* and *P. pastoris* in order to compare the expression level between two yeast expression platforms, and study their basic biochemical properties (Article B). To improve the thermostability of *T. reesei* EGII, mutant *egl2* library was constructed and high-throughput screened under *Y. lipolytica* expression system by using error-prone PCR method and further changed the residue at position 257 via site-saturation mutagenesis (Article C). The recombinant enzymes produced by *P. pastoris* were applied to simultaneously hydrolyze and ferment of pretreated rice straw. Finally, the general conclusions of overall results during this thesis are written in the fifth chapter and open the way to prospects.

Rationales and thesis objectives

Due to the cost of cellulases that is still too high, the difficulty to study the biochemical properties of individual cellulase in its native host, and some enzymatic characters which are not suitable to operate in biorefinery applications, the objectives of this thesis are to overexpress the fungal cellulolytic enzymes through two non-conventional yeasts, *Y. lipolytica* and *P. pastoris*, and engineer the protein to improve the thermostability of *T. reesei* EGII by constructing *egl2* mutant library via error-prone PCR and site-saturation mutagenesis. In addition, the presented work also describes the characterization of recombinant and mutant cellulases. Finally, in order to test the efficiency of recombinant enzymes, they are used to hydrolyze the pretreated rice straw and simultaneously fermented to measure ethanol yield.

CHAPTER II

LITERATURE REVIEW

PART I Lignocellulosic materials

Lignocellulosic biomass is a renewable organic material which constitutes the major structural component of all plants. Plants form two types of cell wall that differ in function and in composition. Primary walls surround growing and dividing plant cells. These walls provide mechanical strength but must also expand to allow the cell to grow and divide. The much thicker and stronger secondary wall, which accounts for most of the carbohydrate in biomass, is deposited once the cell has ceased to grow. The secondary cell wall of the plant material is a dense network of polysaccharides consisting mainly of three polymers: cellulose, hemicellulose, and lignin. Coupling of adjacent cellulose chains by hydrogen bonds, hydrophobic interactions and Van der Waal's forces lead to a parallel alignment of crystalline structures known as microfibril. Linear cellulose microfibrils are bound tightly together by hydrogen bonds, to give the main building subunit for cell walls. Hemicellulose connects the cellulose fibres and lignin through a network of cross-linked fibres providing the protection to the cellulose-hemicellulose matrix. These two components are embedded in hydrophobic amorphous lignin playing a very important role in equipping strength to the plant structure, prevents the degradation from other cellular components, and in particular protection against water (Dashtban *et al*., 2009; Chandel *et al*., 2011). The structural model of lignocellulosic materials is shown in Figure II-1.

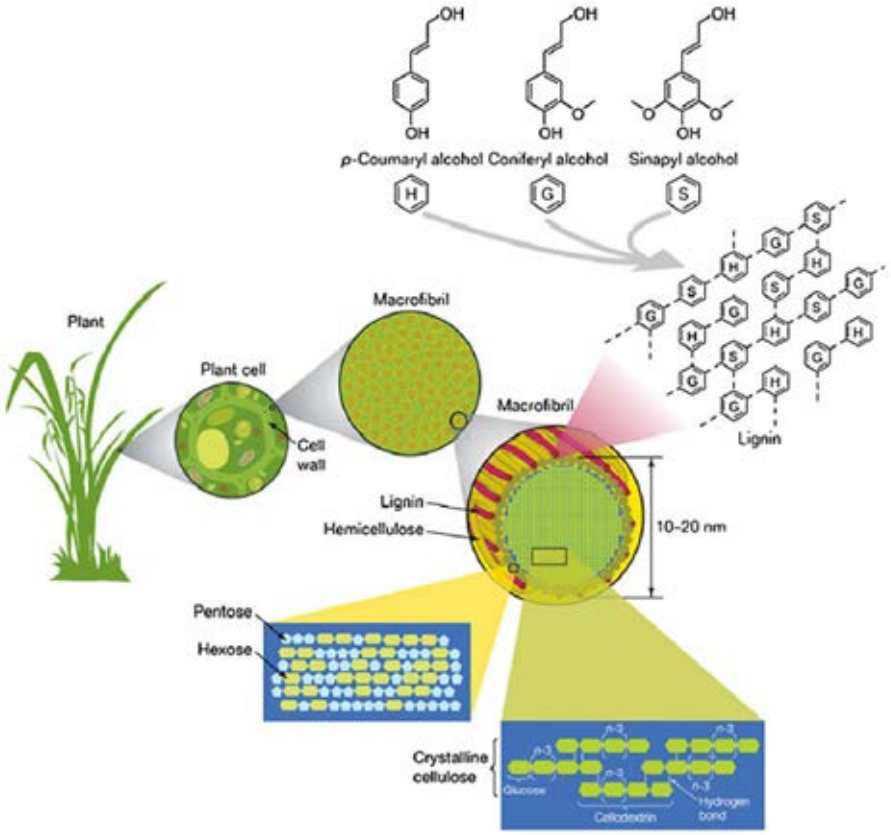


Figure II-1. A model of the main constituents of lignocellulosic material. Cellulose, a glucose subunit linked by β (1–4) hydrogen bond to form crystalline part. Hemicellulose composes of various 5- and 6-carbon sugars such as arabinose, galactose, glucose, mannose and xylose. Lignin is synthesized by polymerization of three major phenolic components, namely p-coumaryl alcohol (H), coniferyl alcohol (G) and sinapyl alcohol (S) (Rubin, 2008).

As a consequence, two major obstacles for the hydrolysis of lingocellulosic material are the recalcitrance of crystalline cellulose itself and the highly protective from lignin residue acting as a physical barrier for enzymatic degradation and water. Generally, plant structure contains 40 –50% cellulose, 20– 40% hemicellulose, 20-30% lignin by weight (Agbor *et al*., 2011). The proportions and composition of cellulose, hemicellulose, and lignin are different depending on the type, species and even source of the biomass (Table II-1).

Table II-1. Composition of some lignocellulosic materials (Sánchez, 2009)

Lignocellulosic residues	Lignin (%)	Hemicellulose (%)	Cellulose (%)	Ash (%)
Hardwood stems	18–25	24–40	40–55	NA
Softwood stems	25–35	25–35	45–50	NA
Nut shells	30–40	25–30	25–30	NA
Corn cobs	15	35	45	1.36
Paper	0–15	0	85–99	1.1–3.9
Rice straw	18	24	32.1	NA
Sorted refuse	20	20	60	NA
Leaves	0	80–85	15–20	NA
Cotton seeds hairs	0	5–20	80–95	NA
Newspaper	18–30	25–40	40–55	8.8–1.8
Waste paper from chemical pulps	5–10	10–20	60–70	NA
Primary wastewater solids	24–29	NA	8–15	NA
Swine waste	NA	28	6	NA
Solid cattle manure	2.7–5.7	1.4–3.3	1.6–4.7	NA
Coastal Bermuda grass	6.4	35.7	25	NA
Switch grass	12.0	31.4	45	NA
S32 rye grass (early leaf)	2.7	15.8	21.3	NA
S32 rye grass (seed setting)	7.3	25.7	26.7	NA
Orchard grass (medium maturity)	4.7	40	32	NA
Grasses (average values for grasses)	10–30	25–50	25–40	1.5
Sugar cane bagasse	19–24	27–32	32–44	4.5–9
Wheat straw	16–21	26–32	29–35	NA
Barley straw	14–15	24–29	31–34	5–7
Oat straw	16–19	27–38	31–37	6–8
Rye straw	16–19	27–30	33–35	2–5
Bamboo	21–31	15–26	26–43	1.7–5
Grass Esparto	17–19	27–32	33–38	6–8
Grass Sabai	22.0	23.9	NA	6.0
Grass Elephant	23.9	24	22	6
Bast fiber Seed flax	23	25	47	5
Bast fiber Kenaf	15–19	22–23	31–39	2–5
Bast fiber Jute	21–26	18–21	45–53	0.5–2
Leaf Fiber Abaca (Manila)	8.8	17.3	60.8	1.1
Leaf Fiber Sisal (agave)	7–9	21–24	43–56	0.6–1.1
Leaf Fiber Henequen	13.1	4–8	77.6	0.6–1
Coffee pulp	18.8	46.3	35	8.2
Banana waste	14	14.8	13.2	11.4
Yuca waste	NA	NA	NA	4.2

NA = Not available.

1. Cellulose

In 1838, Anselme Payen, a Frenchwood chemist first identified a fibrous substance commonly found in wood, cotton, and other plants. This constituent of plant material was therefore termed “cellulose” and was introduced into scientific

literature in 1839 (Klemm *et al*., 2005). Cellulose is the major component of lignocellulosic biomass, conferring structural integrity to plants, but can also be produced by some fungi, algae and bacteria. It is a homopolymer of β-D-glucopyranose residues linked via β-(1,4) glycosidic bonds (Chandrakant and Bisaria, 1998) (Figure II-2).

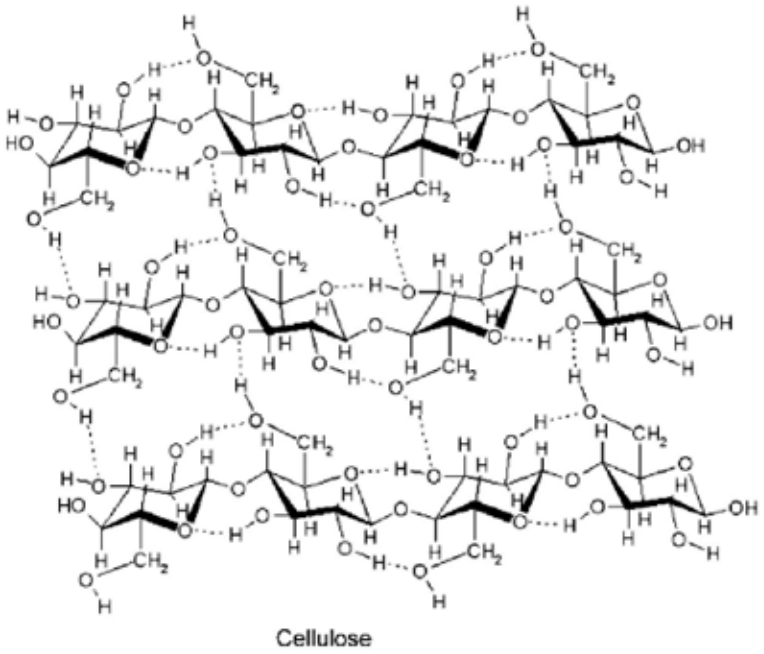


Figure II-2. Molecular structure of cellulose. Linear chains of β-1,4-linked glucoses assemble through hydrogen bonds (dashed lines) within a cellulose chain but also between two adjacent chains (Stöcker, 2008).

Cellulose smallest repeating unit is cellobiose, which consists of two glucose units. The degree of polymerization (DP) of cellulose chains in nature depends on the biomass and varies from 10,000 glucopyranose units in wood to 15,000 in cotton. The parallel linear cellulose chains (20-300) are joined together by covalent bonds, hydrogen bonding and Van der Waals forces to form elementary microfibril which is 40 Å wide, 30 Å tick and 100 Å long (Bidlack *et al*., 1992). The elementary microfibrils are aggregated together to form cellulose microfiber having infinite in length and 250 Å wide (Fan *et al*., 1982). Intra-molecular H-bonds determines "straightness" of the chain whereas inter-molecular H-bonds arrange the microfibril to

crystalline and amorphous region. These hydrogen bonds make the cellulose structure to be resistant from water, mild chemicals, and enzymatic decomposition (Laureano-perez *et al*., 2005). The different in molecular ordering between crystalline and amorphous regions is proposed by 2 scientific hypotheses. Larsson, *et al*. (1997), defined that most of the amorphous regions are located at the surface, whereas crystalline part settle in the core of the microfibril. However, Moiser *et al*. (1999) reported crystalline (organized) and amorphous (not-well organized) regions are repeated in horizontal dimension with the regions of high crystallinity defined to have approximately 200 glucose residues in length separated from amorphous regions. The crystallinity index can be measured by X-ray diffraction or solid-state NMR. Cellulose can be exist in seven crystalline form (polymorphs) called Iα, Iβ, II, III_I, III_{II}, IV_I, and IV_{II}. Native crystalline cellulose exists in two different crystalline forms: cellulose Iα and cellulose Iβ (O'Sullivan, 1997) (Figure II-3).

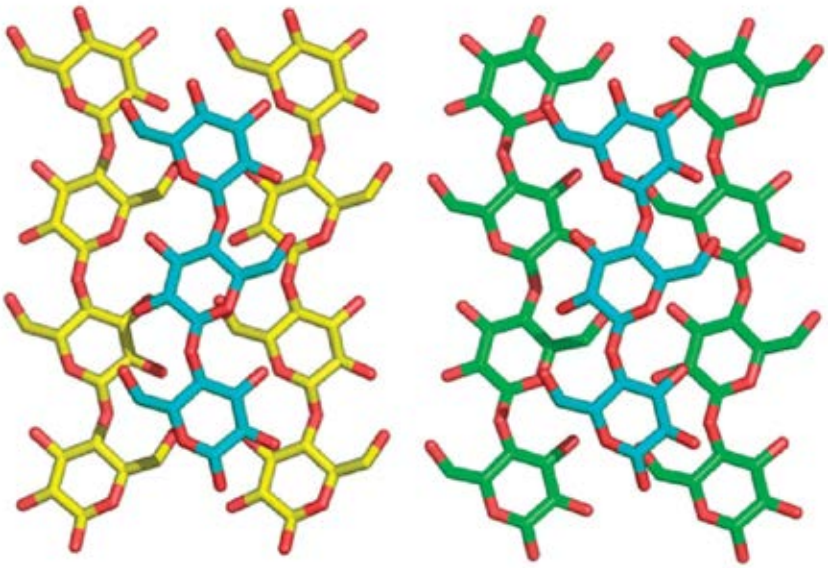


Figure II-3. Hydrophobic surface views of cellulose Iα (left) and Iβ (right). The first layer of cellulose chains is colored cyan. The second layer of the chains is colored yellow (cellulose Iα) and green (cellulose Iβ) (Igarashi *et al*., 2006).

Cellulose Iα polymorph is less stable, and so more reactive than Iβ. The proportions of these forms vary depending on the source of cellulose. Iα is more abundant in algal and bacterial celluloses origin whereas Iβ is more found in higher

plants. These two crystal structure are co-exist within a microfibril with different amount depending on the source of cellulose. Moreover, cellulose II is regenerated or macerated cellulose I. Cellulose III_I and III_{II} is a liquid ammonia treated cellulose whereas cellulose IV_I and IV_{II} is may be prepared by heating cellulose III_I and III_{II} in glycerol (O'Sullivan, 1997) (Figure II-4).

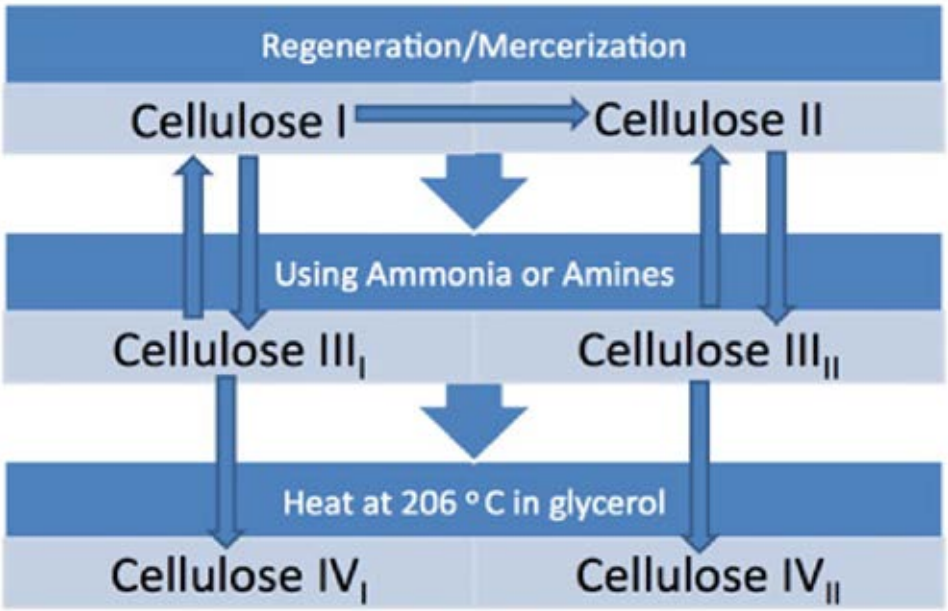


Figure II-4. Conversion of polymorphic cellulose. Cellulose I, is natural cellulose. Cellulose II is solvent pretreatment or mercerization in concentrated sodium hydroxide of cellulose I. Cellulose III_I and III_{II} are pretreated celluloses I and II with ammonia or amines, respectively. Heating to 206 °C in glycerol of polymorphs III_I and III_{II} results IV_I and IV_{II} forms (Agbor *et al*., 2011).

2. Hemicellulose

Hemicellulose, the second most abundant polymer of plant biomass, is branched and consists of heterosaccharide which may contain different carbohydrate monomers such as pentoses (β-D-xylose, α-L-arabinose), hexoses (β-D-mannose, β-D-glucose, α-D-galactose) and/or uronic acids (α-D-glucuronic, α-D-4-*O*-methylgalacturonic, and α-D-galacturonic acids) depending on their biological origin and different hemicelluloses structures can be found (Table II-2). Moreover, other sugar can be found in small amount such as α-L-rhamnose, α-L-fucose and the

hydroxyl groups of sugars partially substituted with acetyl groups (Hendriks and Zeeman, 2009; Gírio *et al*., 2010)

Table II-2. Composition of hemicelluloses in lignocellulosic materials (Gírio *et al*., 2010)

Raw material	Xyl	Ara	Man	Gal	Rha	UA	AcG
Softwoods							
Douglas fir	6.0	3.0	–	3.7	–	–	–
Pine	5.3–10.6	2.0–4.2	5.6–13.3	1.9–3.8	–	2.5–6.0	1.2–1.9
Spruce	5.3–10.2	1.0–1.2	9.4–15.0	1.9–4.3	0.3	1.8–5.8	1.2–2.4
Hardwoods							
Aspen	18–27.3	0.7–4.0	0.9–2.4	0.6–1.5	0.5	4.8–5.9	4.3
Birch	18.5–24.9	0.3–0.5	1.8–3.2	0.7–1.3	0.6	3.6–6.3	3.7–3.9
Black locust	16.7–18.4	0.4–0.5	1.1–2.2	0.8	–	4.7	2.7–3.8
Eucalypt	14–19.1	0.6–1	1–2.0	1–1.9	0.3–1	2	3–3.6
Maple	18.1–19.4	0.8–1	1.3–3.3	1.0	–	4.9	3.6–3.9
Oak	21.7	1.0	2.3	1.9	–	3	3.5
Poplar	17.7–21.2	0.9–1.4	3.3–3.5	1.1	–	2.3–3.7	0.5–3.9
Sweet gum	19.9	0.5	0.4	0.3	–	2.6	2.3
Sycamore	18.5	0.7	1.0	–	–	–	3.6
Willow	11.7–17.0	2.1	1.8–3.3	1.6–2.3	–	–	–
Agricultural and agro-industrial materials							
Almond shells	34.3	2.5	1.9	0.6	–	–	–
Barley straw	15	4.0	–	–	–	–	–
Brewery's spent grain	15	8	0	1	0	2	0.8
Cardoon	26.0	2.5	3.7	1.4	0.9	–	–
Corn cobs	28–35.3	3.2–5.0	–	1–1.2	1	3	1.9–3.8
Corn fibre	21.6	11.4	–	4.4	–	–	–
Corn stalks	25.7	4.1	<3.0	<2.5	–	–	–
Corn stover	14.8–25.2	2–3.6	0.3–0.4	0.8–2.2	–	–	1.7–1.9
Olive stones	2.0–3.7	1.1–1.2	0.2–0.3	0.5–0.7	0.3–0.5	1.2–2.2	–
Rice husks	17.7	1.9	–	–	–	–	1.62
Rice straw	14.8–23	2.7–4.5	1.8	0.4	–	–	–
Sugar cane bagasse	20.5–25.6	2.3–6.3	0.5–0.6	1.6	–	–	–
Wheat bran	16	9	0	1	0	2	0.4
Wheat straw	19.2–21.0	2.4–3.8	0–0.8	1.7–2.4	–	–	–

Unlike cellulose, it has lower molecular weight when compared to cellulose and it is chemically non-homogeneous with different linkages and short lateral chains substitutions that are easily hydrolyzed and more soluble in water. The individual hemicellulose chains are shorter than in cellulose; with a DP of approximately 80-200 (Saha, 2003; Peng *et al*., 2009). The major component of hemicelluloses is also different depending on its sources which can divide the plants into two groups according to the main sugar residues in the backbone (Table II-3). Normally, hemicelluloses are divided into two groups according to the main sugar residues i.e. xylans (β-D-xylose units) and mannans (β-D-mannose units) (Kuhad *et al*., 1997).

Table II-3. Main types of polysaccharides present in hemicelluloses (Gírio *et al*., 2010)

Polysaccharide type	Biological origen	Abbreviation	Amount[a]	Units		
				Backbone	Side chains	Linkage
Arabinogalactan	Softwoods	AG	1–3;35[b]	β-D-Galp	β-D-Galp α-L-Araf β-L-Arap	β-(1 → 6) α-(1 → 3) β-(1 → 3)
Xyloglucan	Hardwoods, grasses	XG	2–25	β-D-Glcp β-D-Xylp	β-D-Xylp β-D-Galp α-L-Araf α-L-Fucp Acetyl	β-(1 → 4) α-(1 → 3) β- (1 → 2) α-(1 → 2) α-(1 → 2)
Galactoglucomannan	Softwoods	GGM	10–25	β-D-Manp β-D-Glcp	β-D-Galp Acetyl	α-(1 → 6)
Glucomannan	Softwoods and hardwoods	GM	2–5	β-D-Manp β-D-Glcp		
Glucuronoxylan	Hardwoods	GX	15–30	β-D-Xylp	4-O-Me-α-D-GlcpA Acetyl	α-(1 → 2)
Arabinoglucuronoxylan	Grasses and cereals, softwoods	AGX	5–10	β-D-Xylp	4-O-Me-α-D-GlcpAβ-L-Araf	α-(1 → 2) α-(1 → 3)
Arabinoxylans	Cereals	AX	0.15–30	β-D-Xylp	α-L-Araf/Feruloy	α-(1 → 2) α-(1 → 3)
Glucuronoarabinoxylans	Grasses and cereals	GAX	15–30	β-D-Xylp	α-L-Araf 4-O-Me-α-D-GlcpA Acetyl	α-(1 → 2) α-(1 → 3)
Homoxylans	Algae	X		β-D-Xylp[c]		

Xylans in plant structure can be either glucuronoxylans (*O*-acetyl-4-*O*-methylglucuronoxylan; GX e.g. hardwood) or arabinoglucuronoxylans (arabino-4-*O*-methylglucuronoxylan; AGXs e.g. agricultural crop), in which homopolymeric backbone β-D-xylopyranose units are substituted at random interval with heteropolysaccharide of arabinose, rhamnose, glucuronic acid or its 4-*O*-methyl ether, acetic, ferulic, galacturonic, and *p*-coumaric acids. Mannans are typically found in softwoods. There are two main groups, namely glucomannan (both β-D-glucose and β-D-mannose units backbone) or galactoglucomannan (consist of β-D-glucose and β-D-mannose on the backbone and β-D-galactose on the side chain) as shown in Figure II-5 (Polizeli *et al*., 2005; Gírio *et al*., 2010).

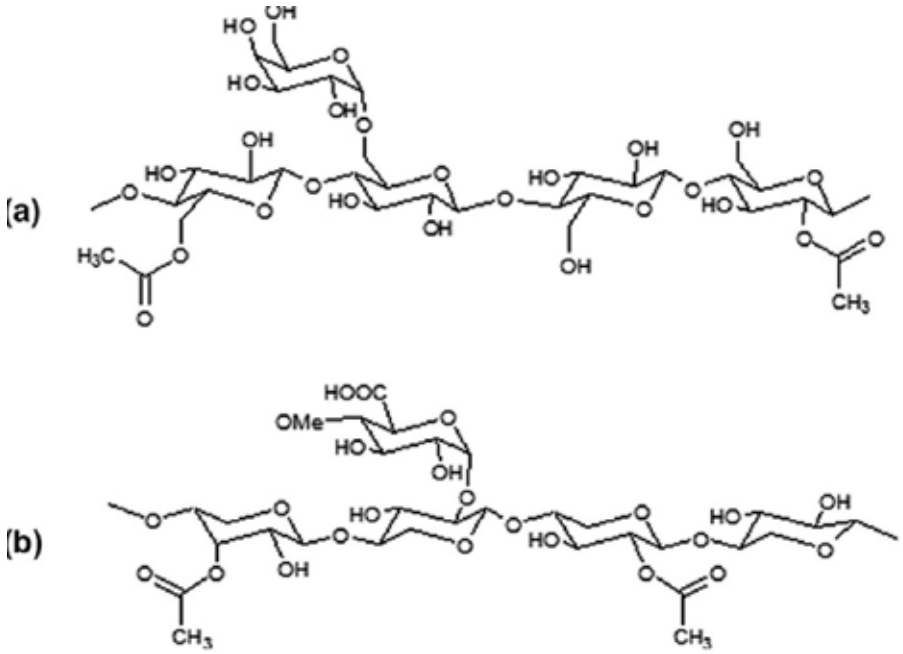


Figure II-5. Representative structure of (a) *O*-acetylated galactoglucomannan, and (b) *O*-acetyl-4-*O*-methylglucuronoxylan (Edlund *et al*., 2012).

Due to their branched structure, hemicelluloses are thermo-chemically sensitive and more easily soluble than cellulose, which xylan can be extracted by acids and base while glucomannan can hardly be extracted in an acid environment and needs a stronger alkaline environment (Lloyd and Wyman, 2005; Agbor *et al*., 2011). Hemicellulose coated cellulose fibril should be removed to reduce the structural constrains and increase cellulose digestibility. However, the hemicelluloses pretreatment could produce some toxic products to fermentation process such as furfurals and hydroxymethyl furfurals. Therefore, acid, water, and steam pretreatments are the most commonly applied technologies yielding a solubilisation of hemicelluloses for further bioprocessing by hemicellulase enzymes (Gírio *et al*., 2010; Agbor *et al*., 2011).

3. Lignin

Lignin is a complex amorphous hydrophobic polymer in secondary plant cell wall giving a rigid structure, water permeability to plants, and resistance to pathogen infections and oxidative stress. It is a network of phenyl propane units composed of three different types of phenolic monomers i.e. *p*-coumaryl alcohol, coniferyl alcohol

and sinapyl alcohol linked into irregular three dimensional structure (Figure II-6) (Chang, 2007; Stöcker, 2008).

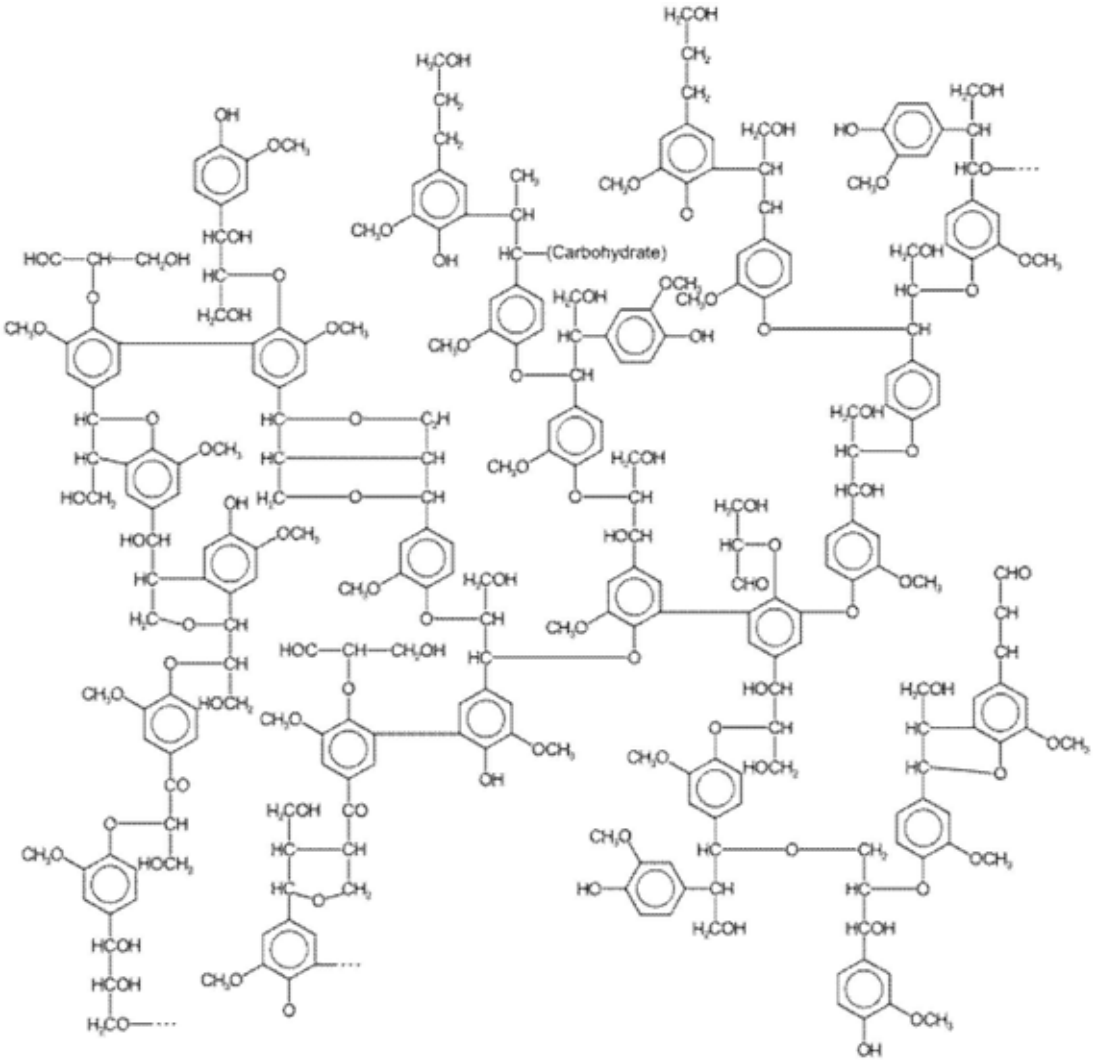


Figure II-6. A partial generalized structure of lignin (Stöcker, 2008).

Guaiacyl-coniferyl lignin is usually found in softwood whereas guaiacyl-syringyl lignin is predominantly in hardwood and grass lignin contains all of three structures. Lignin is synthesized by the polymerization reaction from enzyme peroxidase using free radical of phenylpropanoid monomers. The degree of polymerization is quite low about 50-500 DP. As it like cellulose and hemicelluloses, it presents vary in proportion according to species, cell type and stage of tissue development but normally softwoods contain more lignin than hardwoods. Lignin is concentrated between the outer layers of the fibers and covalently bound to side

groups on different hemicelluloses, forming a complex matrix and give structural rigidity. In the process of enzymatic hydrolysis of lignocelluloses, lignin is a physical barrier for enzymatic degradation. Moreover, it can be also bound irreversibly to cellulases and its derivatives could be toxic to microorganisms. As a consequence, it is necessary to remove by chemical pretreatments (called delignification). Three commonly chemical pretreatment methods are used: 1) oxidative delignification by using peroxidase enzyme 2) organic solvent process by using HCl or H_2SO_4 3) wet oxidation performed at high temperature and pressure. Delignification causes biomass swelling, disruption of lignin structure, increases in internal surface area, and increased accessibility of cellulolytic enzymes to cellulose fibers. Currently, there are no processing methods making it available to fermentation. However, it can be used to produce electricity due to its high energy content or produce other chemicals (Agbor *et al*., 2011; Kuhad *et al*., 1997; Goldschmidt, 2008).

4. Rice straw: a potential lignocellulosic feedstock

Rice (*Oryza sativa* L.) straw is the stem part of the rice plant that remains in the paddy. It is one of the promising renewable resources in many countries because of its abundant especially in Thailand. It can be easily collected from rice-processing sites and contain about 39% cellulose, 20% hemicellulose, and 14% lignin (Karimi *et al*., 2006). It contains high content of silicon dioxide (SiO_2) comparing with other crop residues whereas range of ash content is 13 to 20% according to the state of conservation of the straw after harvest. In general, the ash contains 75% SiO2, 10% K_2O, 3% Fe_2O_3, 3% P_2O_5, 1.3% CaO, and lesser amounts of sodium, sulfur, and magnesium. (Kadam *et al*., 2000).

However, it is considered to be agricultural waste since it cannot be converted into valuable products. The appropriate use of rice straw will enhance the value of this material, provide a solution for the removal of this abundant waste, and increase the economic yield of the process (Sun *et al*., 2000). By treating with sodium hydroxide or ammonium hydroxide and enhancing straw digestibility, rice straw is able to be used as animal feed. Furthermore, rice straw is one of the most appropriate productions for corrugated medium and newsprint. It can be converted into bioethanol, a clean-burning transportation-fuel oxygenate, through bioconversion

(Kadam *et al*., 2000). In 2010, there is about 20 million tons of rice straw in Thailand (Office of Agricultural Economics, 2012 : online). This amount of rice straw can potentially produce 7,173 million liters bioethanol per year. However, the strong crystalline structure of cellulose and the complex structure of lignin and hemicellulose with cellulose in rice straw limit its effective conversion (Chadha *et al*., 1995).

5. Pretreatment of lignocellulose

The first step in bioconversion of lignocellulosic biomass to bioethanol is size reduction and pretreatment. The goals of any pretreatment technology are: increase the accessible surface area of cellulose, breaking the lignin seal, and alteration or removal structural and compositional impediments to hydrolysis in order to improve the rate of enzymatic hydrolysis and increase yield of fermentable sugars from cellulose or hemicellulose (Balat *et al*., 2008; Weil *et al*., 1994). Pretreatment is able to be performed by four different categories: physical (grinding and milling), physio-chemical (wet oxidation, steam pretreatment/autohydrolysis, and hydrothermolysis), chemical (oxidizing agents, organic solvents, alkali, and dilute acid), biological pretreatment (Kumar *et al*., 2009).

Physical pretreatment can enhance the accessible surface areas and size of pores; in contrast, it also can lessen degrees of polymerization of cellulose and crystallinity. Different types of physical processes such as milling (e.g. ball milling, two-roll milling, hammer milling, colloid milling, and vibro energy milling) and irradiation (e.g. by gamma rays, electron beam or microwaves) can be used to improve the enzymatic hydrolysis or biodegradability of lignocellulosic waste materials (Taherzadeh and Karimi, 2008).

Combining both physical and chemical processes are pretreatment named physio-chemical processes. Among these processes, thermochemical treatments (steam explosion, steam disruption), ammonia fiber explosion (AFEX), liquid hot water (LHW), and CO_2 explosion are the most applicable physio-chemical pretreatments. In these treatments, chipped biomass is treated with high-pressure saturated steam, liquid ammonia or CO_2. Thenceforth, the pressure is quickly

decreased which makes the materials to converse by an explosive decompression (Mtui, 2009).

Chemical pretreatment is clearly originated by chemical reactions for disruption of the biomass structure. Chemicals ranging from alkali, oxidizing agents, acids, and salts have potential to degrade cellulose, hemicelluloses, and lignin. Ozone and hydrogen peroxide, powerful oxidizing agents, greatly have ability to remove lignin with no toxic residues product for the downstream processes. The reactions are conducted at room temperature and pressure (Mtui, 2009). Alkaline peroxide has been recognized as a powerful oxidizing agent, and it is quite selective toward the lignin structure (Gould, 1984). It could obviously decrease the crystallinity and lignin content. Moreover, it does not leave residues in the biomass, and the formation of secondary products is practically inexistent. In addition, it utilizes lower temperatures and pressures compared to other pretreatment technologies (Rabelo *et al*., 2008). Previously, alkaline peroxide pretreatment has been studied on various biomasses such as wheat straw (Curreli *et al*., 1997; Saha and Cotta, 2006; Chen *et al*., 2008), softwood (Yang *et al*., 2002), rice hull (Saha and Cotta, 2007), and sugarcane bagasse (Rabelo *et al*., 2008; Cheng *et al*., 2008). With these substrates, it significantly improved enzymatic hydrolysis (about 60-90% hydrolysis).

Most pretreatment technologies require expensive instruments or equipment that has high energy requirements, depending on the process. In particular, physical and thermochemical processes require abundant energy for biomass conversion. Biological treatment involves the use of whole organisms to treat the lignocelluloses and enhance enzymatic hydrolysis. The applied microorganisms usually degrade lignin and hemicellulose but very little part of cellulose, since cellulose is more resistance than the other parts of lignocelluloses to the biological attack. Several fungi, e.g. brown-, white- and soft-rot fungi, have been used for this purpose. White-rot fungi are among the most effective microorganisms for biological pretreatment of lignocelluloses. Low energy requirement, no chemical requirement, and mild environmental conditions are the main advantages of biological pretreatment. Nevertheless, the rate of biological hydrolysis is usually very low, so this pretreatment requires long residence times (Taherzadeh and Karimi, 2008; Kumar *et al*., 2009).

PART II Microbial cellulolytic enzymes

1. History and properties of cellulases

Cellulolytic enzymes or cellulases hydrolyse β-1,4-glycosidic linkages in cellulose. They were discovered due to the rapid deterioration of textiles in Southeast Asia during the Second World War (1939-1945). The U.S. army research program was initiated in order to identify the organisms responsible for the degradation and to determine the mechanisms by which they operate. Thirty years later, over 100 articles had been published, and their results were reviewed by Reese in 1976, emphasizing the extensive studies made on cellulase complexes, cellulase-mediated conversion of plant materials into digestible products and production of glucose, and the availability of mutant strains (Reese, 1976; Knowles *et al*., 1987). Microbial cellulolytic enzymes are crucial for balance of the global carbon cycle, as they are involved in energy and nutrient uptake by these microorganisms. They are also believed to participate in plant pathogenesis by fungi releasing hydrolytic enzymes to degrade plant cell walls (Have *et al*., 2002).

Cellulolytic enzymes can be either freely secreted extracellularly (and be designated as a non-complexed cellulase system) or attached to the cell wall of the microorganism (constituting then a complexed cellulase system or cellulosome). Non-complexed cellulase systems are mostly found in microorganisms that can penetrate the lignocellulosic materials by using their hyphal extension such as filamentous fungi (e.g. *Trichoderma reesei*, *Humicola grisea*) and actinomycete bacteria (e.g. *Streptomyces lividans*, *Thermobifida fusca*). These microorganisms hydrolyse the cellulosic substrates by releasing free enzymes which operate individually or in synergy. After that, hydrolysed products such as glucose and cellodextrins are taken up by the microorganisms and either used directly or further hydrolysed by intracellular hydrolases. This group of cellulases is most often used in the industries because secreted enzymes are usually easy to harvest (Lynd *et al*., 2002). In contrast, complexed cellulase systems are mainly found in anaerobic microorganisms such as *Clostridium thermocellum* that cannot penetrate lignocellulosic biomass. The different cellulolytic enzymes are organized into multi-enzyme complexes called cellulosomes. The cellulosome consists of individual enzyme molecules anchored

onto a common scaffolding protein by using cohesin-dockerin interactions (Figure II-7). A typical bacterial cellulosome is composed of ~50 protein molecules with a total molecular weight of about 2 to 6 million Daltons. The main function of the cellulosome is to maintain a tight interaction between the microorganism and the cellulosic substrate and therefore, to reduce enzyme diffusion and losses. Enzymes organized into cellulosomes result in an overall hydrolytic activity higher than the sum of the individual activities of the enzymes (Doi, 2008; Lynd *et al*., 2002).

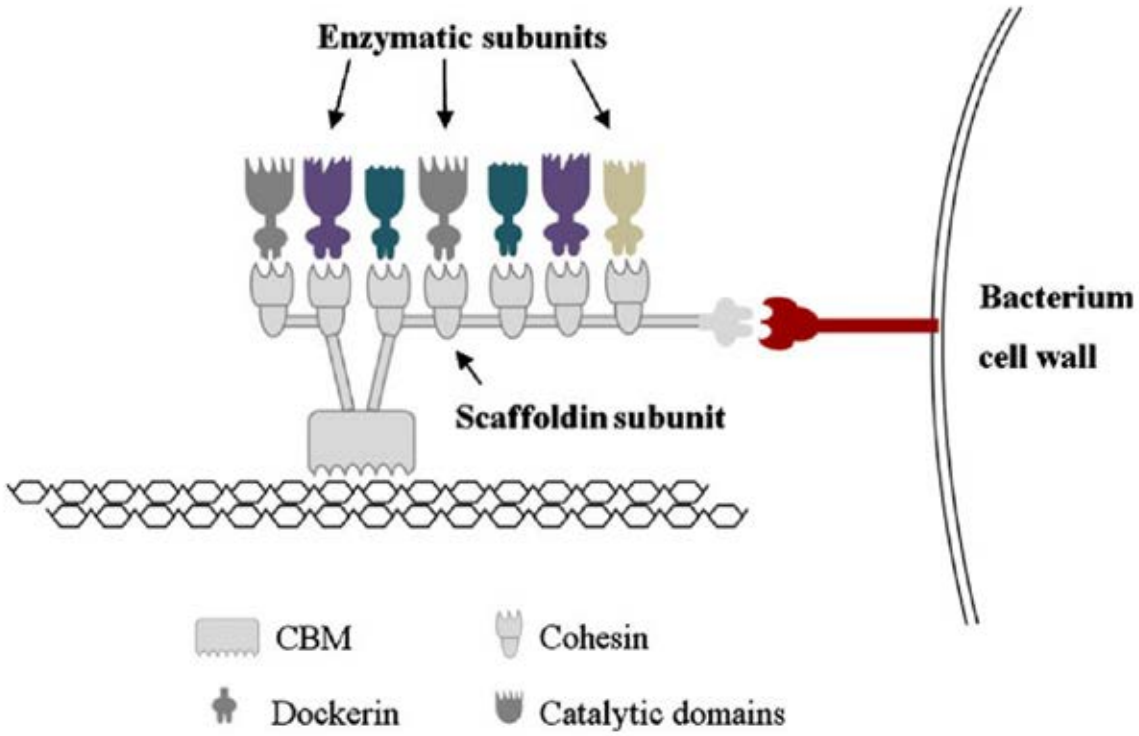


Figure II-7. Structure of *C. thermocellum* cellulosome (Chang and Yao, 2011).

2. Source of cellulases

There are a lot of microorganisms from bacterial and fungal origin that can produce cellulolytic enzymes. Bacteria generally degrade lignocellulosic materials slowly, so that, usually, they attack plant cell wall simultaneously with fungi because they lack some penetrating capability. Anaerobic cellulolytic bacteria are generally found in oil and decaying plant materials, rumen of various ruminants, sewage plants, and the gut of termites, where they live in symbiosis with their hosts and process the nutrients for them (Table II-4), and many belong to *Clostridium* species (Doi, 2008).

Table II-4. Some anaerobic bacteria with active cellulolytic enzymes (Doi, 2008)

Microorganism	Habitat
Acetivibrio cellulolyticus	Sewage sludge[1,2]
Bacteroides cellulosolvens	Sewage sludge[3]
Butyrivibrio fibrisolvens	Bovine rumen[4]
Clostridium acetobutylicum	Soil[5]
Clostridium aldrichii	Wood digester[6]
Clostridium cellobioparum	Soil[7]
Clostridium cellulofermentans	Dairy farm soil[8]
Clostridium cellulolyticum	Decayed grass[9]
Clostridium cellulovorans	Wood chips[10]
Clostridium herbivorans	Pig intestine[11]
Clostridium hungatei	Soil[12]
Clostridium josui	Compost[13]
Clostridium papyrosolvens	Paper mill[14]
Fibrobacter succinogenes	Rumen[15]
Ruminococcus albus	Rumen[16]
Ruminococcus flavefaciens	Rumen[17]

In contrast, aerobic cellulolytic bacteria are usually discovered in water, soil, on decaying plant materials, animal feces, and leaf litter (Table II-5). Actinomycetes and filamentous eubactria, such as *Cellulomonas* and *Streptomyces* are two common genera of cellulolytic aerobic bacteria (Kuhad *et al*., 1997).

Table II-5. Some aerobic cellulolytic bacteria (Doi, 2008)

Microorganism	Habitat
Bacillus megaterium	Soil[18]
Bacillus pumilus	Soil, dead plant[19]
Cellulomonas fimi	Soil[20]
Cellulomonas flavigena	Soil, leaf litter[21]
Cellulomonas gelida	Soil[22]
Cellulomonas iranensis	Forest humus soils[23]
Cellulomonas persica	Forest humus soils[23]
Cellulomonas uda	Sugar cane field[24]
Cellvibrio gilvus	Bovine feces[25]
Cellvibrio mixtus	Soil[26]
Pseudomonas fluorescens	Soil, water[27]
Streptomyces antibioticus	Soil[28]
Streptomyces cellulolyticus	Soil[29]
Streptomyces lividans	Soil[30]
Streptomyces reticuli	Soil[31]

Ascomycetes (e.g. *T. reesei*), white-rot fungal basidiomycetes (e.g. *Phanerochaete chrysosporium*), brown-rot fungal basidiomycetes (e.g. *Fomitopsis palustris*) are typical examples of aerobic fungi producing non-associated enzymes, which play an important role in the degradation of plant material in nature . On the other hand, anaerobic fungi produce multi-enzyme complexes of cellulases, hemicellulases, and ligninases similar to bacterial cellulosomes. Anaerobic fungi live in the gastrointestinal tract of large ruminant animals, are much less common as there exist only six recognized genera: *Anaeromyces*, *Caecomyces*, *Cyllamcyces*, *Neocallimastix*, *Orpinomyces* and *Piromyces* (TableI-6) (Doi, 2008; Dashtban *et al*., 2009).

Table II-6. Some aerobic and anaerobic fungi producing different lignocellulolytic enzymes and their substrates (Dashtban *et al*., 2009)

	Group	Fungal strain	Enzymes	Substrate
Aerobic fungi	Ascomycetes	*T. reesei*	Cellulases (CMCase, CBH, BGL), Hemicellulase (xylanase)	Wheat straw
		T. harzianum	Cellulases (CMCase, CBH), β-1,3-glucanases	Wheat bran, wheat straw
		A. niger	Cellulases, Xylanases	Sugar cane bagasse
(Extracellular lignocellulolytic enzymes)		*Pestalotiopsis sp.*	Cellulases (CMCase, CBH), Laccase	Forest litter of *Quercus variabilis*
	Basidiomycetes	*P. chrysosporium*	Cellulases (CMCase, CBH, BGL), CDH, LiP, MnP, Hemicellulase (xylanases)	Red oak, grape seeds, barley bran, woodchips
		F. palustris	Cellulases (CMCase, CBH, BGL)	Microcrystalline cellulose
Anaerobic rumen fungi (Chytridiomycetes)	*Anaeromyces*	*Anaeromyces mucronatus 543*	Cellulase (CMCase), Hemicellulase (xylanase)	Orchard grass hay
	Caecomyces	*Caecomyces communis*	Cellulases, Hemicellulases (xylanase, β-D-xylosidase)	Maize stem
	Cyllamyces	*Cyllamyces aberensis*	Cellulases, Xylanases	Grass silage
(Cell-wall associated lignocellulolytic enzymes, "cellulosome")	*Neocallimastix*	*Neocallimastix frontalis*	Cellulases, Hemicellulase (xylanase, β-galactosidase)	Cotton fiber, wheat straw
	Orpinomyces	*Orpinomyces sp.*	Cellulase (CMCase, CBH, β-glucosidase), Hemicellulases (xylanase, mannanases)	Wheat straw
	Piromyces	*Piromyces sp.*	Cellulases (CMCase, CBH, β-glucosidase) Hemicellulases (xylanase, mannanases)	Maize stem

3. Current industrial application of cellulases

Cellulolytic enzymes are being used in a variety of industrial applications and have been available for commercial use over the last few decades. In the early 1980s, they were first used in animal feed applications followed by food applications, and later in textile and laundry sectors but also for pulp and paper industries. At present, they are considered as the third largest group of enzymes used in industry accounting for approximately 20% of the world's enzyme market, mostly produced from *Trichoderma* sp. and *Aspergillus* sp. Due to the interest in commercial applications, cellulases have been studied extensively to identify new candidates in nature or to genetically modify known enzymes to have suitable properties and characteristics to fit industrial applications and processes. Details of the cellulase applications are widely discussed by Bhat (2000), Sukumaran and colleagues (2005), and Kamakar and Ray (2011).

3.1 Food application

The process of fruit and vegetable juice production requires an extraction, a clarification and a stabilization step. In the early 1930s, the principal difficulties in this process were to produce and filter the juice to an acceptable clarity. Thus, macerating enzymes (a combination of pectinases, cellulases, and hemicellulases) helped to overcome this problem. The macerating enzymes also increased juice yield, reduced the processing time, and improved the extraction of valuable nutrients. A list of macerating enzymes used nowadays is shown in Table II-7.

Table II-7. Macerating enzymes in food biotechnology (Bhat, 2000)

Enzyme	Function	Application
Cellulases and hemicellulases	Partial or complete hydrolysis of cell wall polysaccharides and substituted celluloses	Improvement in soaking efficiency; homogeneous water absorption by cereals; the nutritive quality of fermented foods; the rehydrability of dried vegetables and soups; the production of oligosaccharides as functional food ingredients and low-calorie food substitutents and biomass conversion
β-Glucanases and mannanases	Solubilization of fungal and bacterial cell wall	Food safety and preservation
Xylanases and endoglucanases	Hydrolysis of arabinoxylan and starch	Separation and isolation of starch and gluten from wheat flour
Pectin esterase with no polygalacturonase and pectin lyase activities	Fruit processing	Production of high quality tomato ketchup and fruit pulps
Rhamnogalacturonase	Cloud stability	Production of cloud stable apple juice
Rhamnogalacturonan acetyl esterase and galactanase		
Cellulase and pectinase	Release of antioxidants from fruit and vegetable pomace	Controlling coronary heart disease and atherosclerosis; reducing food spoilage
Endo-mannanase	Modification of guar gum	Production of water-soluble dietary fibres to enrich the fibre content of foods

3.2 Brewery and wine industries

Beer brewing process involves in its first steps the malting of barley grain, that is to say the enzymatic hydrolysis of the seed reserves into degradation products, such as sugars; it is then followed by the fermentation of the resulting wort by yeast; quite similarly, the wine process requires the hydrolysis of cellulosic polymers in grapes to make pentose/hexose sugars available for fermentation by yeast. In both cases, the addition of endoglucanase II and cellobiohydrolase II, two well known cellulases, contributes to increase plant cell wall degradation and accessibility to storage sugars, and thus, has been proven to improve filterability, wort viscosity, wine qualities, and

overall production efficiency. The applications of cellulases in brewery and wine industries are summarized in Table II-8.

Table II-8. Cellulase applications in brewery and wine biotechnology (Bhat, 2000)

Enzyme/micro-organism	Function	Application
β-Glucanase/glucanolytic yeast	Hydrolysis of β-1,3, and β-1,4 glucan; reducing the viscosity and releasing reducing sugars during primary fermentation	Improvement in primary fermentation, filtration and quality of beer
Pectin esterase	De-esterification and gelling of pectins	Improvement in the clarification of cider
Macerating enzymes (cellulases, hemicellulases and pectinases)	Hydrolysis of plant cell wall polysaccharides	Improvement in skin maceration and colour extraction of grapes; quality, stability, filtration and clarification of wines
β-Glucosidase	Modification of aromatic residues	Improvement in the aroma of wines

3.3 Animal feed biotechnology

Nowadays, cellulases are used in poultry, pig, ruminant and fish farming to improve feed utilization, resulting for example in increased milk yield and body-weight gain. They are used to eliminate Anti-Nutritional Factors (ANF), natural or synthetic substances found in the human diet or animal feed that have the potential to adversely affect health and growth, or to emulsify feed materials in order to improve the nutritional value of feed, or to partially hydrolyze plant material to improve fodder preservation. Some applications of cellulase use in animal feed are summarized in Table II-9.

Table II-9. Cellulases in animal feed biotechnology (Bhat, 2000)

Enzyme	Function	Application
Cellulases and hemicellulases	Partial hydrolysis of lignocellulosic materials; dehulling of cereal grains; hydrolysis of β-glucans; decrease in intestinal viscosity; better emulsification and flexibility of feed materials	Improvement in the nutritional quality of animal feed and thus the performance of ruminants and monogastrics
β-Glucanase and xylanase	Hydrolysis of cereal β-glucans and arabinoxylans, decrease in intestinal viscosity and release of nutrients from grains	Improvement in the feed digestion and absorption, weight gain by broiler chickens and hens
Hemicellulase with high xylanase actvity	Increase the nutritive quality of pig feeds	Reduction in the cost of pig feeds and the use of less expensive feeds for pigs
Cellulases, hemicellulases and pectinases	Partial hydrolysis of plant cell wall during silage and fodder preservation; expression of preferred genes in ruminant and monogastric animals for high feed conversion efficiency	Production and preservation of high quality fodder for ruminants; improving the quality of grass silage; production of transgenic animals

3.4 Textile and laundry industries

The largest current industrial use of cellulolytic enzymes is in the textile industry because of their ability to modify cellulosic fibres in a controlled and desired manner. Cellulases play the roles of “bio-stoning” and “bio-polishing” agents as they increase smoothness, softness, luster, and color brightness of treated fabrics. Moreover, they are also used in washing powder to enhance detergent performance (Table II-10).

Table II-10. Cellulase functions and applications in textile and laundry technology (Bhat, 2000)

Enzyme	Function	Application
Cellulase, preferably neutral and endoglucanase rich	Removal of excess dye from denim fabrics; soften the cotton fabrics without damaging the fibre	Bio-stoning of denim fabrics; production of high quality and environmentally friendly washing powders
Cellulase, preferably acid and endoglucanase rich	Removal of excess microfibrils from the surface of cotton and non-denim fabrics	Bio-polishing of cotton and non-denim fabrics
Cellulase, preferably endoglucanase rich	Restoration of softness and colour brightness of cotton fabrics	Production of high quality fabrics

3.5 Pulp and paper industries

In the pulp and paper industry, cellulolytic enzymes are used for example to increase hand sheet strength, to participate in deinking of recycled fibers, but also play an important role in improving drainage and runnability of paper mills (Table II-11).

Table II-11. Cellulase and hemicellulase applications in pulp and paper biotechnology (Bhat, 2000)

Enzyme	Function	Application
Cellulases and hemicellulases	Modification of coarse mechanical pulp and hand-sheet strength properties; partial hydrolysis of carbohydrate molecules and the release of ink from fibre surfaces; hydrolysis of colloidal materials in paper mill drainage	Bio-mechanical pulping; modification of fibre properties; de-inking of recycled fibres; improving draining and runnability of paper mills
Xylanases, mananases, β-xylosidase and α-L-arabinofuranosidase	Hydrolysis of re-precipitated xylan or removal of xylan from lignin-carbohydrate complexes; removal of glucomannan	Bio-bleaching of kraft pulps; reduction in chlorine requirement in subsequent bleaching and environmental pollution
Purified cellulase and hemicellulase components	Partial or complete hydrolysis of pulp fibres	Bio-characterization of pulp fibres

3.6 Biofuel and its process configuration

Cellulosic materials contribute the major part in forest, agricultural fields, and agro-industry. Therefore, there are a lot of unutilized and that cause the environmental problems. Biofuel from the fermentation of hydrolyzed cellulosic materials can be accomplished by using cellulases. For efficient bioethanol production, extensive research have to be done in order to identify efficient cellulase systems and process conditions along with the studies of biochemical and genetic improvement of microorganism utilized the sugars (Bhat, 2000). Generally, the process configuration to produce bioethanol can be divided into four categories. The process which enzymatic hydrolysis can be performed separately from fermentation is called separate hydrolysis and fermentation (SHF). The primary advantage of SHF is that hydrolysis and fermentation occur at optimum conditions, but the disadvantage is that cellulolytic enzymes are end-product inhibited so that the rate of hydrolysis progressively reduced when glucose and cellobiose are accumulated (Viikari *et al*., 2007). Second, simultaneous saccharification and fermentation (SSF) offers benefits such as increase ethanol yields by reducing the product inhibition exerted by saccharification products and also elimination the need of separate reactors for saccharification and fermentation, which results in cost reductions (Krishna and Chowdary, 2000). In order to improve the economics of any biomass conversion process, both hexoses and pentoses found in hydrolysate must be converted into desired end product (Chu and Lee, 2007). *S. cerevisiae*, the most well-known hexose-fermenting yeast, has several advantages owing to its high ethanol production from hexose and high tolerance to bioethanol and other inhibitory compounds (Balat *et al*., 2008). Unfortunately, it is unable to efficiently utilize pentose. Thus, the initial efforts to produce a commercially viable ethanologen have focused on co-fermentation of hexose and pentose, called simultaneous saccharification and co-fermentation (SSCF), using *S. cerevisiae* and pentose-fermenting yeast such as *Pichia stipitis* (Gray *et al*., 2006). Previously, Laplace and colleagues showed the successful co-fermentation of glucose and xylose by respiratory deficient mutant of *S. cerevisiae* co-cultivated with *P. stipitis* which produces ethanol about 0.42 g/g of sugar (100% glucose and 69% xylose conversion, respectively) (Laplace *et al*., 1993). Finally, consolidated bioprocessing (CBP) is a process which required enzymes and ethanol fermentation

are produced by a single microorganisms community, in a single reactor. CBP seems the logical endpoint in the evolution of biomass conversion technology. Application of CBP implies no capital or operating costs for dedicated enzyme production, and compatible enzyme and fermentation systems (Hamelinck *et al*., 2005).

3.7 Agricultural and biotechnology research

In agriculture, cellulases are used to control plant pathogen, produce protoplast, prevent plant disease, and enhance plant growth. In the field of biotechnology, the strong promoter of cellobiohydrolase I can be used for heterologous expression of many proteins (Table II-12).

Table II-12. Application and function of cellulases in agricultural and biotechnology research (Bhat, 2000)

Enzyme/microorganism	Function	Application
Mixture of cellulases, hemicellulases and pectinases	Solubilization of plant or fungal cell walls	Production of plant or fungal protoplasts, hybrid and mutant strains
Cellulases and related enzymes, preferably β-1,3 and 1,6 glucanases; *Trichoderma* sp. and *Geocladium*	Inhibition of spore germination, germ tube elongation and fungal growth	Bio-control of plant pathogenes and diseases
Trichoderma sp., *Geocladium* sp., *Chaetomium* sp., *Penicillium* sp., *Rhizopus nigricans, Fusarium roseum*	Enhancing seed germination, plant growth and flowering; improving root system; increasing the crop yields.	Agriculture
CBD of cellulases and cellulosomes; dockerins, cohesins and linkers of cellulosome	Affinity tag, affinity systems, conjugation and gene fusion	Affinity purification, immobilization and fusion of proteins, enzymes and antibodies; production of hybrid molecules for various applications
Cellobiohydrolase I promoter from *T. reesei* and glucoamylase promoter from *A. niger*	Expression of heterologous proteins and enzymes	Production of high levels of proteins, enzymes and antibodies
Native enzymes, subunits of cellulosome or recombinant enzymes	Improving the efficiency of a specific application	Production of designer cellulosomes

4. **Cellulase classification**

Cellulolytic enzymes belong to the *O*-Glycosyl hydrolase family (EC 3.2.1.-) of the International Union of Biochemistry and Molecular Biology (IUBMB), which classifies enzymes based on the type of chemical reaction they catalyse. *O*-glycosyl hydrolase hydrolyze the glycosidic bond between two or more carbohydrates, or between a carbohydrate and non-carbohydrate molecule. The enzymes are further classified into sub-families based on their substrate specificities. Cellulases have been classified into three distinct classes, which complete the hydrolysis of cellulosic materials. Endoglucanases (1,4- β-D-glucan glucanohydrolase, EGs, E.C. 3.2.1.4) are endo-acting enzymes hydrolysing internal β-1,4-glycosidic bonds of amorphous cellulose chain. Exoglucanases (1,4-β-D-glucan cellobiohydrolase, CBHs, E.C. 3.2.1.91) are exo-acting enzymes that cleaved β-1,4-glycosidic bonds from free chain ends, in crystalline regions of cellulose, producing mainly cellobiose but also cello-oligosaccharides as end products. Cellobiohydrolases are very important in cellulose degradation because they solubilise crystalline cellulose efficiently. The third class of enzymes is constituted by β-glucosidases (BGs, E.C. 3.2.1.21) working in synergy with CBH and EG to hydrolyse glucose oligomers and cellobiose to form monomeric glucose (Okada *et al*., 1998b; Ito *et al*., 2004) (Figure II-8).

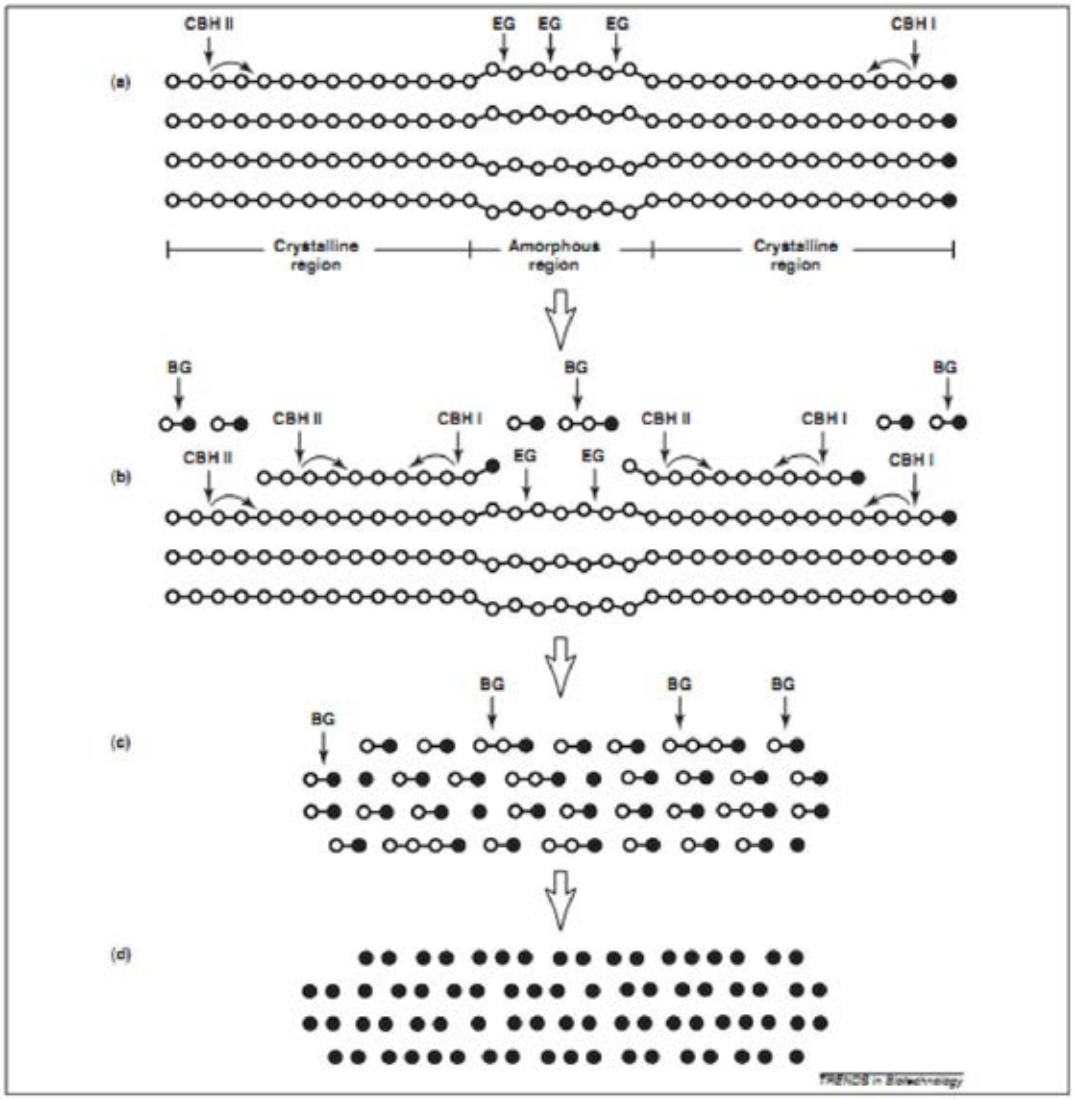


Figure II-8. Mode of action of cellulases on cellulose chains. (a) Initial cellulose consisting of crystalline and amorphous regions. (b) Partially hydrolyzed cellulose. (c) Outer solution containing cellobiose (disaccharide) as a major intermediate product, together with minor amounts of higher oligosaccharides and glucose. (d) Final glucose syrup. The open circles represent anhydroglucose residues in cellulose and oligosaccharides; the solid circles represent reducing ends of cellulose and oligosaccharides or glucose (Gusakov, 2011).

In the early 1990s, Henrissat and coworkers proposed a new classification for "Carbohydrate Active Enzymes": glycosyl-hydrolases were then grouped into families based on amino acid sequence similarities and hydrophobic cluster analysis

(HCA), to better describe the conservation of their structural fold and catalytic mechanism (Table II-13). This new classification system facilitates the prediction of three-dimensional fold and stereo-specificity of hydrolysis (retaining or inverting, see catalytic domain section) which are conserved between each member of a family (Figure II-9), e.g. all enzymes in family 5 are retaining, while all enzymes in family 6 are inverting enzymes. In addition, some families can be grouped in "clan" when they share a common fold. Until now, there are approximately 130 families with 14 clans (clan A-N) classified by this system (Davies and Henrissat, 1995).

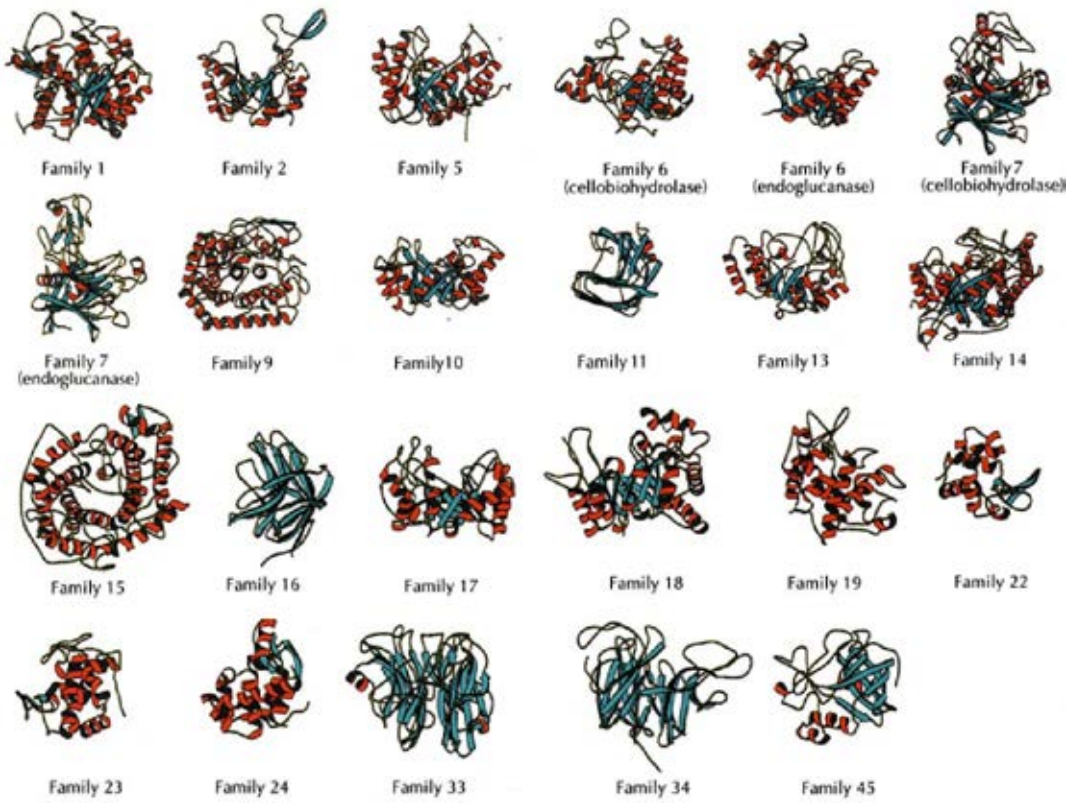


Figure II-9. Ribbon diagrams showing the overall fold of catalytic domains in 23 CAZy families (Davies and Henrissat, 1995).

Table II-13. Classification of cellulases into families and clan based on structural similarity (Bayer *et al*., 1998)

Fold	Clan*	Family	Former cellulase-based classification	Enzyme type (substrate specificity)	Taxonomic distribution	Number of enzymes†	Mechanism
Structures determined, clans established:							
$(\beta/\alpha)_8$	GH-A	1	–	Mainly β-glucosidases and related glycosyl hydrolases	Very broad: bacteria, fungi, plants and animals	71	Retaining
$(\beta/\alpha)_8$	GH-A	5	Family A (five subfamilies)	Mainly endoglucanases	Bacteria and fungi	119	Retaining
$(\beta/\alpha)_8$	GH-A	10	Family F	Mainly xylanases	Bacteria and fungi	59	Retaining
$(\beta/\alpha)_8$	GH-A	17	–	β-Glucosidases, endo-1,3-β glucosidases and lichenases	Fungi and plants	59	Retaining
$(\beta/\alpha)_8$‡	GH-A	26	Family I	Mainly endo-1,4-β mannosidases	Bacteria	14	Retaining
$(\beta/\alpha)_8$‡	GH-A	39	–	β-Xylosidases	Bacteria	7	Retaining
β-jelly roll	GH-B	7	Family C	Endoglucanases and cellobiohydrolases	Fungi	17	Retaining
β-jelly roll	GH-B	16	–	Mainly β-glucanases (lichenases and laminarinases)	Bacteria	71	Retaining
β-jelly roll	GH-C	11	Family G	Mainly xylanases	Bacteria and fungi	65	Retaining
β-jelly roll	GH-C	12	Family H	Endoglucanases	Bacteria and fungi	14	Retaining
Structures determined, clans not established:							
Distorted (β/α) barrel	–	6	Family B	Endoglucanases and cellobiohydrolases	Bacteria and fungi	18	Inverting
$(\alpha/\alpha)_6$	–	8	Family D	Mainly endoglucanases	Bacteria	10	Inverting
$(\alpha/\alpha)_6$	–	9	Family E (two subfamilies)	Mainly endoglucanases	Bacteria and fungi	45	Inverting
β barrel	–	45	Family K	Endoglucanases	Bacteria and fungi	5	Inverting
$(\alpha/\alpha)_6$	–	48	Family L	Processive endoglucanases and/or cellobiohydrolases	Bacteria	6	Inverting
Structures not determined:							
–	–	3	–	Mainly β-glucosidases	Bacteria and fungi	58	Retaining
–	–	44	Family J	Endoglucanases	Bacteria	5	Inverting
–	–	51	–	Endoglucanases and arabinofuranosidases	Bacteria and fungi	9	Retaining
–	–	52	–	β-Xylosidases	Bacteria	2	ND
–	–	55	–	Exo and endo-1,3-glucanases	Fungi	2	ND
–	–	61	–	Endoglucanases	Fungi	1	ND

Many carbohydrate-active enzymes also possess a carbohydrate binding domain (CBD), which role is to deliver the catalytic domain to its substrate. So far, about 64 distinct families have been described on the basis of sequence homology (Davies and Henrissat, 1995; Bayer *et al*., 1998). Updated information is available on Carbohydrate Active Enzyme Database server (Centre national de la recherche scientifique, 2007 : online).

5. Modular structure of cellulases and their mechanisms

Most fungal cellulases possess two independent domains: a large catalytic domain (CD) and a smaller carbohydrate binding module (CBM), connected by a short flexible glycosylated linker (30-44 amino acids) (Gilkes *et al*., 1991; Davies and Henrissat, 1995). X-ray scattering and transmission electron microscopy studies show they adopt a “tadpole” shape, with the catalytic core forming a “large ellipsoidal

head" while the cellulose binding domain forms an "elongated cylindrical tail" (Figure II-10) (Lee and Brown, 1997; Srisodsuk *et al*., 1993).

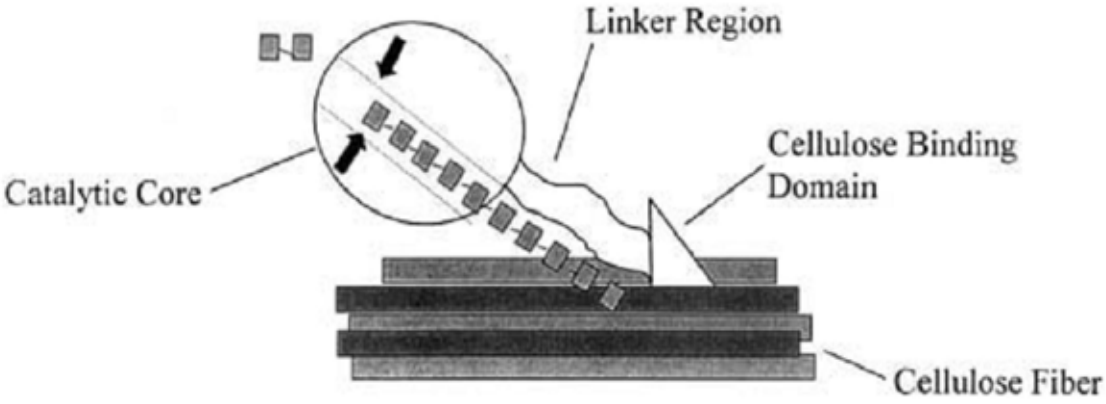


Figure II-10. Modular structure of cellulases (Mosier *et al*., 1999). Arrows representing the active site within the catalytic domain show the ability of cellobiohydrolase to remain bound to the substrate while the hydrolyzed product, cellobiose, is released.

5.1 Catalytic domain (CD)

5.1.1 3D structure

The catalytic domain is the largest module and responsible for the hydrolysis of cellulose chains. Active sites can adopt three different topologies: i) a pocket-like structure normally found in glycosidase that hydrolyse disaccharide (β-glucosidases); ii) a tunnel-like structure, usually found in exo-enzymes, which leads to the formation of cellobiose as only product, because steric hindrance in the tunnel prevents cellulose chain to turn so that cleavage can only happen every second glucose unit; iii) and a groove or cleft-like structure commonly found in endo-acting enzymes. This open structure results in a less strict binding when compared to tunnel structure, so that many hydrolysis products can be released from cellulose chains (Mosier *et al*., 1999) (Figure II-11).

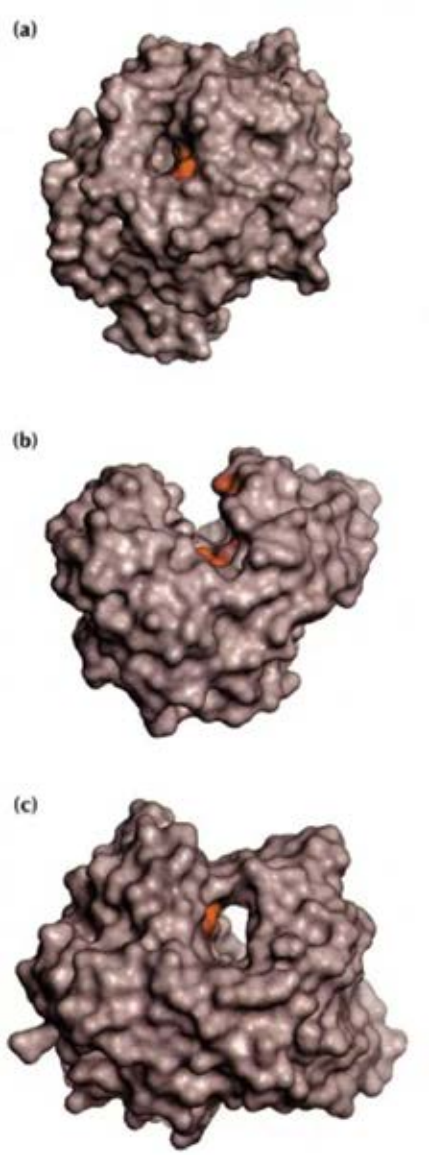


Figure II-11. Three types of active sites found in glycosyl hydrolase (a) pocket-like structure (b) cleft-like structure (c) tunnel-like structure. The proposed catalytic residues are shaded in red (Davies and Henrissat, 1995).

5.1.2 Catalytic mechanism

In glycosyl-hydrolases, the enzymatic hydrolysis proceeds via two different mechanisms: either inverting or retaining. In both cases, the hydrolysis occurs via a general acid catalysis which requires a proton donor/acid and a nucleophile/base. Both mechanisms rely on a pair of essential carboxylic acid residues (aspartic or glutamic residues) located on opposite sides of the catalytic site.

The inverting catalysis occurs via a single displacement mechanism, and leads to an inversion of the configuration of the anomeric carbon of the substrate after cleavage, i.e. the inversion of a β-glycosidic linkage to an α-configuration during the hydrolysis of cellulose chain for example. A first acidic residue donates a proton to the leaving glycosidic oxygen while the general base residue, located on the opposite side of the sugar unit, activates a water molecule that will participate to a nucleophile attack on the anomeric carbon (Figure II-12). These results general base to become proton donor and general acid to become nucleophile for the next reaction (Davies and Henrissat, 1995; Mosier *et al*., 1999).

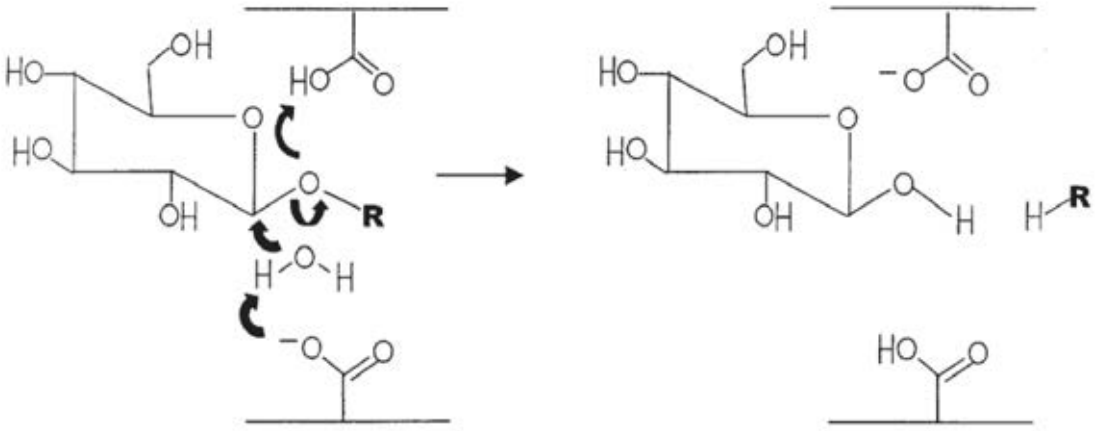


Figure II-12. Catalytic mechanism of inverting enzymes (Mosier *et al*., 1999).

In contrast, retaining catalysis, a double-displacement mechanism, leads to a net retention of the configuration of the anomeric carbon after cleavage. In the first chemical step, aglycosyl-enzyme intermediate displaying an inverted anomeric configuration is formed by the simultaneous deprotonation of the general acid and the nucleophile attacked by the opposite general base residue on the anomeric carbon. In the second step, a water molecule donates a proton to the general base (previously the general acid) and simultaneously hydroxylates the anomeric carbon involved in the covalent glycosyl-enzyme intermediate, inverting the anomeric configuration again (Figure II-13). This is because the distance between the general acid and the general base in the active site of retaining enzymes is shorter than in inverting enzymes, that retaining enzymes can form glycosyl-enzyme intermediate whereas a water molecule

performs the attack directly in inverting enzymes (Davies and Henrissat, 1995; Mosier *et al*., 1999).

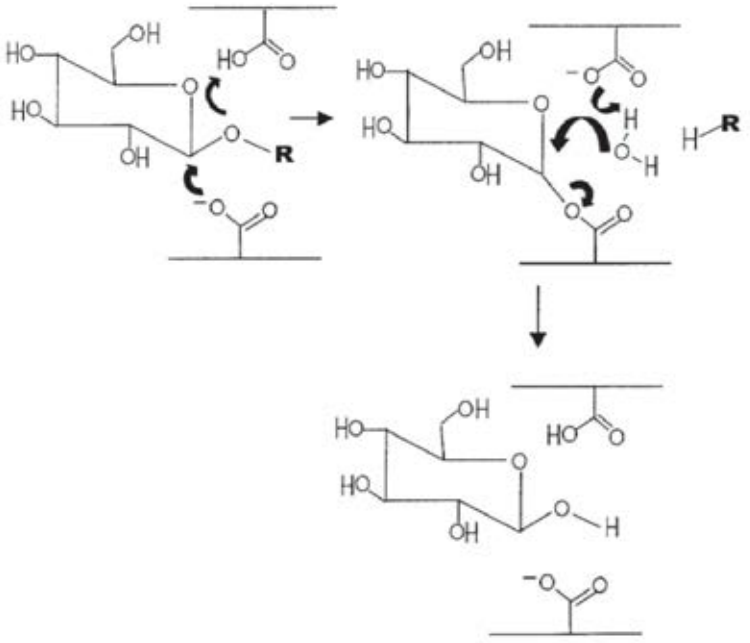


Figure II-13. Catalytic mechanism of retaining enzymes (Mosier *et al*., 1999).

5.2 Carbohydrate binding module (CBM)

Most fungal cellulases possess a carbohydrate-binding module which is located at either the N- or C-terminus of the protein, and is connected to the catalytic domain via a glycosylated linker region. It consists of approximately 35-200 amino acid residues believed to: 1) improve the binding between CD and crystalline cellulose, 2) allow a strong and continuous adsorption on the substrate during the hydrolysis, 3) increase the local concentration of enzyme on the substrate surface, 4) increase the time that the enzyme spends near the substrate, 5) enhance the hydrolysis activity on insoluble cellulose substrate (Linder and Teeri, 1996; Gray *et al*., 2006; Igarashi *et al*., 2009). When a CBM is absent, CD can still adsorb onto cellulose but has low affinity. An enzyme presenting only a CD lost almost all of its hydrolytic activity on insoluble substrates whereas activity on soluble substrates was not affected (Karlsson *et al*., 2002; Shoseyov *et al*., 2006). *T. reesei* cellulases (CBHI, CBHII, EGI, EGII) all possess CBMs which belong to family I (Centre national de la recherche scientifique, 2007 : online). CBMI members are characterised by three

conserved amino acids (two tyrosine and one tryptophan residue) which are responsible for tight binding to cellulose chains (Mosier *et al*., 1999).

5.3 Linker region

The linker region is a flexible peptide sequence which role is to interconnect the CD and its CBM. This region is usually heavily *O*-glycosylated and varies in size between 6 to over 59 amino acid residues depending on the source of the enzyme. Its amino acid sequence is typically rich in prolines, serines and threonines. Its flexibility facilitates the independent function of the CD and the CBM while its hyper-glycosylation protects the fungal cellulases from proteolytic enzymes (Gilkes *et al*., 1991; Harrison *et al*., 1998). The importance of the linker region was studied by Srisodsuk and coworker in 1993, who found that when the linker region was deleted by mutation, the rate of crystalline cellulose degradation was dramatically reduced. However, flexibility and heterogeneous glycosylation of the linker region make it difficult to obtain crystal of entire fungal cellulases (Mosier *et al*., 1999).

6. *Trichoderma reesei* cellulases

Since 1976, more than 14,000 fungal species, comprising cellulosic degrading ones, have been isolated, but only a few of them have been chosen for in depth studies. Cellulolytic fungi, especially soft-rot fungi such as *Trichoderma*, *Humicola*, and *Penicillium* species and white-rot fungi such as *Phanerochaete*, and *Pycnoporus* species are among the most studied species (Covert *et al*., 1992; Kumar *et al*., 2008; Wood *et al*., 1989; Wood and Garcia-Canpayo, 1990).

T. reesei (teleomorph: *Hypocrea jecorina*) is one of the best characterized fungi and the most efficient producer of cellulases and hemicellulases. The first description of genus *Trichoderma* was in 1794 by Persoon, and in 1825, the sexual state *Hypocrea* was proposed. It was described as a cellulolytic fungus in the 1950s at Solomon Islands by E.T. Reese because it had been related to the rapid destruction of cotton fatigues and tents of the US Army. Through microscopic observation, they discovered *T. viride* strain QM 6a. Extensive studies were started during the Second World War, and resulted in the selection of 2 hypercellulolytic strains named *T. viride*

QM9414 and MCG77. In 1977, the fungus was renamed *T. reesei* to honor the discovery made by Reese (Gusakov, 2011; Schuster and Schmoll, 2010).

Nowadays, the hyper-secreting mutant Rut C-30 strain is one of the most powerful and best characterized *T. reesei* strains, with an extracellular protein production of up to 19 g/L (Table II-14). It was the result of three steps of random mutagenesis (first UV, then N-nitroguanidine, and finally UV) performed by Rutgers University. Another mutant strain, CL847, was shown to yield a higher production of protein, about 40 g/L, while Rut-C30 may be more cellulase-production oriented (Peterson and Nevalainen, 2012; Herpoël-Gimbert *et al*., 2008).

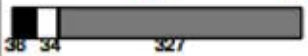
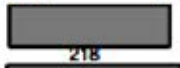
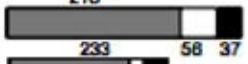
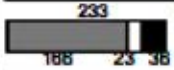
Table II-14. Extracellular protein production and cellulase activity from *T. reesei* QM6a and its derived strains (Peterson and Nevalainen, 2012)

Strain	Soluble protein (mg ml^{-1})	FPU (U ml^{-1})	Productivity (FPU l^{-1} h^{-1})	CMC (U ml^{-1})	βGl (U ml^{-1})
QM6a	7	5	15	88	0.3
QM9414	14	10	30	109	0.6
MCG77*	16	11	33	104	0.9
NG-14	21	15	45	133	0.6
RUT-C30	19	14	42	150	0.3

T. reesei secretes large amounts of all cellulases needed for degradation of crystalline cellulose. To date, 17 cellulases encoding genes have been discovered and cloned: they consist in two cellobiohydrolases, eight endoglucanases, and seven β-glucosidases (Table II-15). However, only four cellulases are secreted in large quantities (90–95% of the total secreted cellulases) by this fungus: CBH I (Cel7A), CBH II (Cel6A), EG I (Cel7B) and EG II (Cel5A), accounting for 40-60%, 12-20%, 5-10%, 1-10% of total protein production, respectively (Gusakov, 2011; Sukumaran *et al*., 2005; Rosgaard *et al*., 2007). Although up to 8 β-glucosidases genes have been found in *T. reesei* genome, they only constitute ~0.5% of the secreted cellulases, as most of them are intracellular: their action is thus insufficient to complete the degradation of the formed cellobiose to glucose. As a consequence, cellobiose accumulates and inhibits the endo- and exo-glucanases by feedback. Therefore, β-glucosidases (usually from *Aspergillus* sp.) are usually supplemented into *T. reesei* cellulase preparations to provide a more efficient saccharification of cellulosic

substrates (Merino and Cherry, 2007; Gusakov, 2011; Kumar *et al*., 2008). *Aspergilli* are indeed good producers of β-glucosidases and other accessory enzymes. Nonetheless, β-glucosidases from *T. reesei* have been recently improved by Novozymes (C-Tec cocktail) and Genencor International Inc. (Accellerase cocktail) (Berlin *et al*., 2005; de Vries and Visser, 2001).

Table II-15. Cellulolytic system and molecular properties of cellulase genes in *T. reesei* (Miettinen-Oinonen, 2004)

Enzyme	Gene	GenBank™ accession no.	Length	Molecular mass, kDa[11]	Structural organisation[3]	Ref.
CEL7A (CBHI)	*cbh1/cel7a*	P00725[12]	513	59-68		Shoemaker *et al.*, 1983, Teeri *et al.*, 1983
CEL6A (CBHII)	*cbh2/cel6a*	M16190	471	50-58		Teeri *et al.*, 1987
CEL7B (EGI)	*egl1/cel7b*	M15665	459	50-55		Penttilä *et al.*, 1986
CEL5A (EGII)	*egl2/cel5a*	M19373	418	48		Saloheimo *et al.*, 1988
CEL12A (EGIII)	*egl3/cel12a*	AB003694	234	25		Ward *et al.*, 1993, Okada *et al.*, 1998
CEL61A (EGIV)	*egl4/cel61a*	Y11113	344	34[2]		Saloheimo *et al.*, 1997
CEL45A (EGV)	*egl5/cel45a*	Z33381	242	23[2]		Saloheimo *et al.*, 1994
EGVI[4]				95-105		Bower *et al.*, 1998a
CEL74A (EG)[5,8]	*cel74a*	AY281371	838	87[2]	Contains CBM[10]	Foreman *et al.*, 2003
CEL61B (EG)[5,8]	*cel61b*	AY281372	249	27[2]	No CBM[10]	Foreman *et al.*, 2003
CEL5B (EG)[5,9]	*cel5b*	AY281373	438	47		Foreman *et al.*, 2003
CEL3A (BGLI)	*bgl1/cel3a*	U09580	744	75		Barnett *et al.*, 1991, Mach, 1993
CEL1A (BGLII)	*bgl2/cel1a*	AB003110	466	52		Takashima *et al.*, 1999, Saloheimo *et al.*, 2002a
CEL3B (BGL)[5,6]	*cel3b*	AY281374	874	94[2]		Foreman *et al.*, 2003
CEL3C (BGL)[5,7]	*cel3c*	AY281375	833	91[2]		Foreman *et al.*, 2003
CEL1B (BGL)[5,7]	*cel1b*	AY281377	484	55[2]		Foreman *et al.*, 2003
CEL3D (BGL)[5,7]	*cel3d*	AY281378	700	77[2]		Foreman *et al.*, 2003
CEL3E (BGL)[5,6]	*cel3e*	AY281379	765	83[2]		Foreman *et al.*, 2003

[1]Length of the protein including the signal peptide, amino acid residues, [2] Molecular mass calculated from the amino acid sequence, [3] – CD, β – linker, ■ – CBM [4] Described on the protein level,

6.1 T. reesei endoglucanases

Endoglucanases (EGs) usually cleave randomly at internal glycosidic linkages in amorphous zones of the cellulose chain. They are also known as carboxymethylcellulases (CMCases) because they have a significant activity on carboxymethyl-cellulose (CMC) which is often used as a substrate analog to measure enzymatic activity. Hydrolysis leads to the production of cellulose chains with a lower degree of polymerization, and provides new free chain ends that are accessible to the action of cellobiohydrolases. *T. reesei* endoglucanases are only sparsely glcosylated, and their active site adopts an open binding cleft topology. Their pH optima are normally between 4.0 and 5.0 and temperature optima from 50 to 70 °C. With the exception of EGIII, they all possess a CBM, which is not vital for the hydrolysis activity (Dashtban *et al*., 2009; Boisset *et al*., 2000).

6.1.1 Endoglucanase I (EGI) or Cel7B

EGI represents about 5-10% of total secreted cellulases. It belongs to family 7 of the glycosyl hydrolase classification, and contains 437 amino acid residues resulting in a molecular weight of 48 kDa. The isoelectric point of EGI is between 4 and 6. Moreover, the enzyme proceeds with a retaining mechanism (Claeyssens *et al*., 1990; Biely *et al*., 1991). The CBM of EGI is located at the C-terminal end of the protein. The three-dimensional (3-D) structure of EGI catalytic domain revealed a β-jelly-roll fold (Figure II-14) where four loops that usually form a tunnel are completely deleted, resulting in an open groove-shaped active site. Thus, the open cleft allows cellulose chains to enter the active site for subsequent cleavage of internal bonds (Kleywegt *et al*., 1997). EGI exhibits a broad substrate specificity, as it offers a high activity on substituted cellulose [carboxymethylcellulose (CMC) and hydroxylethylcellulose (HEC)] but also on xylan, although EGI and xylanases have little sequence similarity. The main products from the hydrolysis of cellulosic substrates are cellobiose and glucose with a little cellotriose when phosphoric acid swollen cellulose is used as substrate (Karlsson *et al*., 2002).

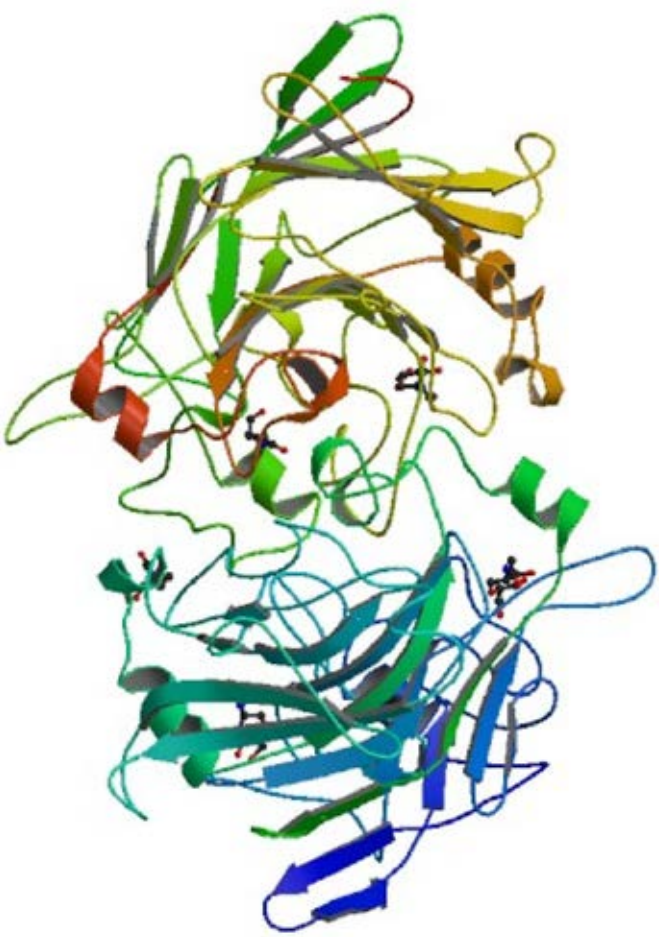


Figure II-14. β-jellyroll structural folds of *T. reesei* endoglucanase I (Cel7B) (Worldwide Protein Data Bank, 2012a : online).

6.1.2 Endoglucanase II (EGII) or Cel5A

EGII represents about 1-10% of total cellulasessecreted by *T. reesei*. It belongs to glycosyl-hydrolase family 5 and contains 397 amino acid residues, resulting in a molecular weight of about 48 kDa. The hydrolysis of glycosidic bond proceeds via a double-replacement mechanism, and thus leads to the retention of the anomeric configuration (Viikari *et al*., 2007; Rabinovich *et al*., 2002). In contrast to EGI, the CBM of EGII is located at its N-terminus. Its isoelectric point is of 5.5 (Saloheimo *et al*., 1988). The 3-D structure of EGII was elucidated by Lee and coworkers in 2011, and corresponds to a $(\beta/\alpha)_8$-barrel fold with an active site in the form of an open cleft (Lee *et al*., 2011) (Figure II-15).

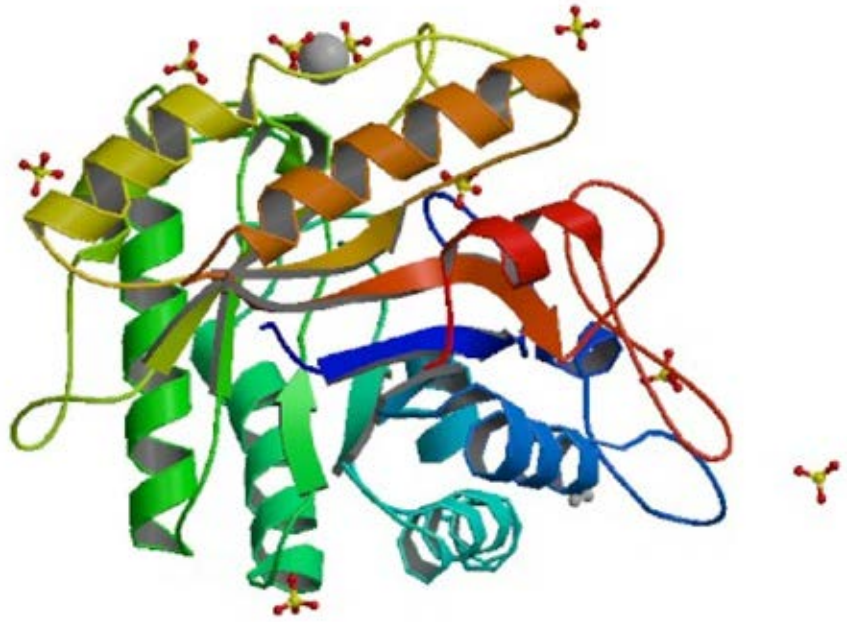


Figure II-15. $(\beta/\alpha)_8$–TIM barrel fold of *T. reesei* endoglucanase II (Cel5A) (Worldwide Protein Data Bank, 2012b : online).

EGII hydrolyses linear and substituted cellulose, Avicel, and PASC, releasing glucose and cellobiose as the main products together with traces of cellotriose (Karlsson *et al*., 2002). EGII role is very important for the overall endoglucanase activity produced by *T. reesei*, because when EGII is absent from the secreted complex, the global endoglucanase activity decreases by approximately 55%, whereas the absence of EGI only causes a 25% decrease (Suominen *et al*., 1993).

6.1.3 Other T. reesei endoglucanases

EGIII (Cel12A), previously known as EGIV, is a small non-glycosylated protein and represents less than 1% of total secreted cellulases. It consists of 218 amino acid residues, resulting in a protein with a molecular weight of about 25 kDa, and has an isoelectric point of 7.5. In contrast to the other endoglucanases, it does not possess any CBM. EGIII belongs to CAZy family 12, its structure is made of 15 antiparallel β-sheets and a single α-helix, assembling in a β-jelly roll (Figure II-16), and its active site is in groove shape (Sandgren *et al*., 2001). It is active on substituted cellulose but not on avicel and filter paper. Hydrolysis

products range from mono- to tetra-saccharides with cellobiose as the main product (Karlsson *et al*., 2002).

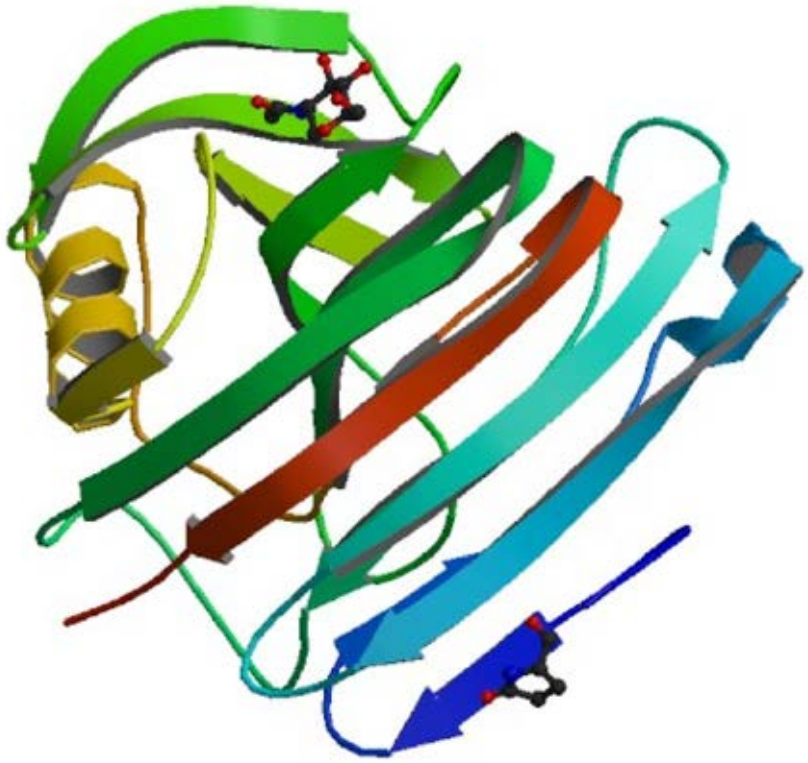


Figure II-16. The β-jelly roll fold of *T. reesei* endoglucanase III (Cel12A) (Worldwide Protein Data Bank, 2012c : online).

EGIV (Cel61A) contains 344 amino acid residues, and has a molecular weight of about 34 kDa, and an isoelectric point of 6.0. It prefers to cut shorter oligosaccharides such as cellotetraose and cellopentaose, yielding cellobiose as the main product. Moreover, EGIV has a unique characteristics because of its long *O*-glycosylated linker region connecting the catalytic domain to its C-terminal CBM (Saloheimo *et al*., 1997), but its crystal structure has not yet been determined.

EGV (Cel45A) contains 242 amino acid residues, resulting in a molecular weight of about 23 kDa, and has an isoelectric point of 2.9. Its crystal structure has not been solved yet either. It hydrolyzes amorphous cellulose using an inverting mechanism, and produces cellotetraose as the main products with low levels of cellotriose and cellopentaose (Saloheimo *et al*., 1994; Karlsson *et al*., 2002).

6.2 T. reesei cellobiohydrolases

Cellobiohydrolases (CBHs) are processive enzymes which hydrolyse β-1,4-glycosidic bonds of microcrystalline cellulose from chain ends, liberating cellobiose as the major product, with trace amount of glucose and cellotriose, during initial stages of hydrolysis (Dashtban *et al*., 2009; Rouvinen *et al*., 1990). There are two main types of cellobiohydrolases. CBHI type enzymes (such as Cel7A in *T. reesei*) cleave from the non-reducing end of cellulose chains and prefer to bind crystalline regions, whereas CBHII enzymes (such as Cel6A in *T. reesei*) cut from the reducing end and does not show any preference between crystalline or amorphous regions (Dashtban *et al*., 2009; Kuhad *et al*., 1997). Both types of cellobiohydrolases are very slow at decreasing the degree of polymerization of cellulose. Moreover, they have low substrate specificity on substituted cellulose, and their end product, cellobiose, can act as a competitive inhibitor, limiting the efficiency of the enzymes (Lynd *et al*., 2002). CBHs have also low levels of glycosylation. Their pH optima range between 4.0 and 5.0, and their temperature optima between 37 to 60 °C. In addition, their active site is located inside a tunnel (Kuhad *et al*., 1997; Dashtban *et al*., 2009).

6.2.1 Cellobiohydrolase I (CBHI) or Cel7A

CBHI is the major secreted cellulases, synthesized by *T. reesei* and represents about 60% of the total amount of proteins. It belongs to family 7 of the glycosyl hydrolase classification and contains 497 amino acid residues resulting in a molecular weight of about 54 kDa. It is a glycoprotein containing 6% carbohydrate and has an isoelectric point of about 3.9. CBHI processively hydrolyzes crystalline cellulose with a retaining mechanism (Medve *et al*., 1998; Barr *et al*., 1996; Claeyssens *et al*., 1990). CBHI possess a CBM at its C-terminal end, and its 3D structure has been shown to be a β-jelly roll (Figure II-17) containing four surface loops forming a tunnel with a length of 50 A˚ and covering ten glucose binding sites (Lynd *et al*., 2002).

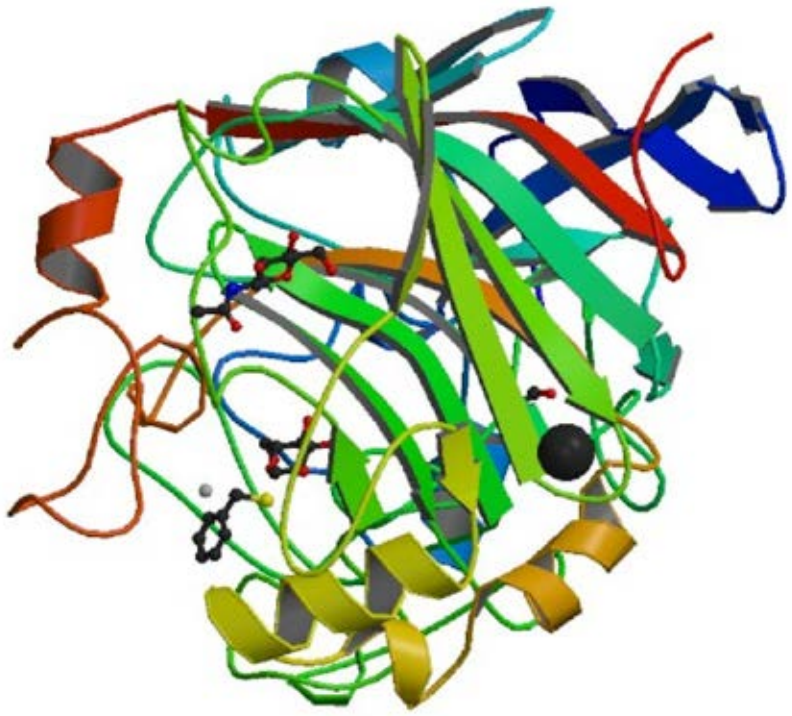


Figure II-17. Ribbon diagram representing the β-jelly roll fold of *T. reesei* CBHI (Cel7A) (Worldwide Protein Data Bank, 2012d : online).

6.2.2 Cellobiohydrolase II (CBHII) or Cel6A

CBHII are the second principal component of *T. reesei* cellulase system, representing 20% of the total secreted cellulases. It is grouped into GH family 6 and contains 447 amino acid residues, resulting in a molecular weight about 50 kDa. Its linker region is highly *O*-glycosylated and connects the CD with a N-terminal CBM. Its isoelectric point is between 5.1 and 6.3 (Mosier *et al*., 1999; Lynd *et al*., 2002; Teeri *et al*., 1987). In contrast to CBHI, CBHII processively hydrolyzes microcrystalline cellulose (but has no activity on substituted cellulose) with an inverting mechanism (Claeyssens *et al*., 1990; Lynd *et al*., 2002). From all the cellulases produced by *T. reesei*, it has the highest specific activity against crystalline cellulose (Messner *et al*., 1991). The 3D structure of CBHII shows a distorted (α/β) barrel shape (Figure II-18), in which two surface loops protect a six glucose long active site in a 20 Å long tunnel (Sandgren *et al*., 2001; Rouvinen *et al*., 1990; Teeri, 1997).

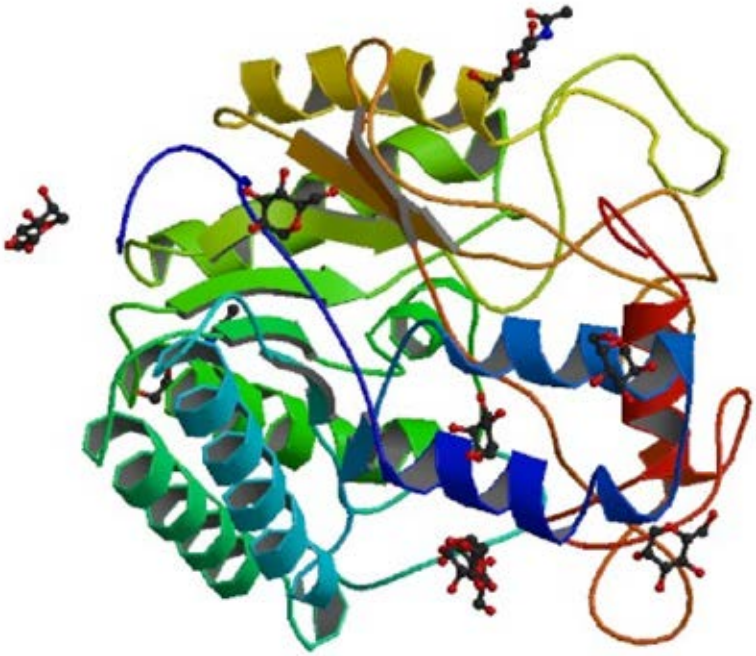


Figure II-18. Ribbon diagram representing the distorted (α/β) barrel fold of *T. reesei* CBHII (Cel6A) (Worldwide Protein Data Bank, 2012e : online).

6.3 T. reesei β-glucosidases

Beta-glucosidases (BGLs) hydrolyze mainly cellobiose and cellodextrin to yield glucose as product, though it acts as a competitive inhibitor of many cellulolytic enzymes (Dashtban *et al*., 2009). Those from *T. reesei* belong to glycosyl hydrolase families 1 (BGLII) and 3 (BGLI). Family 1 enzymes are usually found in bacteria, plants and mammals, whereas family 3 members are often found in fungi, bacteria, and plants, and lack CBM (Henrissat, 1991). BGLs from both families hydrolyze β-1,4-glycosidic bonds using the retaining mechanism. Moreover, they have broad substrate specificity since they can degrade cellooligosaccharides, aryl- and/or alkyl-β-D-glucosides (Dan *et al*., 2000; Rooyen *et al*., 2005). The pH optima of BGLs depend on the origin of the enzymes from 4 to 5.5 whereas the temperature optima range from 45 to 75 °C. They can be produced as intracellular, cell-wall associated, or extracellular proteins (Dashtban *et al*., 2009). Although *T. reesei* produces at least 7 β-glucosidases, these represent a small portion of the total extracellular proteins (about 1%), thus limiting the complete hydrolysis of cellulosic substrate, allowing enough glucose production for fungal growth but not for saccharification processes.

Moreover, *T. reesei* BGLs are more sensitive to glucose inhibition than *Aspergillus* species ones. Therefore, β-glucosidases from *Aspergillus* species are most often used on an industrial scale to complement *T. reesei* cellulolytic cocktail (Lynd *et al*., 2002).

7. β-glucosidases from *Aspergillus oryzae*

Aspegillus oryzae is a generally recognized as safe (GRAS) microorganism and is very commonly used in the preparation of traditional Japanese fermented products such as soy sauce, soybean paste and rice wine. It has the ability to produce various useful enzymes and the sequencing of its genome was completed by a Japanese research group in 2005 (Adachi *et al*., 2008; Kaya *et al*., 2008). *A. oryzae* secretes two different β-glucosidases: the major form has a high molecular mass, about 130 kDa, and is highly inhibited by glucose whereas the minor form (representing less than 18% of total β-glucosidase activity) has a lower molecular weight (40 kDa) and is regarded as a highly glucose-tolerant β-glucosidase (HGT-BGL). The production of HGT-BGL is highly induced by quercetin, a phenolic flavonoid found in plant cell walls (Riou *et al*., 1998). It belongs to GH family 5 and contains 405 amino acid residues resulting in a molecular weight of about 44 kDa. Moreover, it is a monomeric protein with a pI of 4.2. It is optimally active at 50 °C and pH 5.0. Very interestingly, HGT-BGL has a broad substrate specificity because it can hydrolyse β(1,3), β(1,4), β(1,6), α(1,3), α(1,4), an α(1,6) glucosidic linkages, as well as aryl- and alkyl-β-glycosides. *A. oryzae* HGT-BGL Ki (glucose inhibition constants) value is about 1.36 M which is 13-2700 fold higher than most microbial BGLs (Ki values ranging from 0.5 mM to 100 mM) and 2.5-fold higher than *A. niger* BGLs (Ki value is about 0.543 M). It also has a specific activity of about 1,066 U per mg protein on pNP-glucopyranoside (one unit of β-glucosidase activity corresponded to release of 1 mmol of p-nitrophenol min^{-1}), which is the most efficient fungal β-glucosidase described so far (Langston *et al*., 2006; Mega and Matsushima, 1979; Riou *et al*., 1998). However, these β-glucosidases are mostly found associated to cell-surface, so that in recent years, several studies have attempted to express β-glucosidase in yeast expression systems to provide a high, stable and extracellular enzyme production (Tokuhiro *et al*., 2008).

8. Synergism between cellulases

Concerning cellulose degradation, cellulases cannot work efficiently independently but need to work cooperatively. Synergism refers to the simultaneous activity of two or more enzymes, resulting in a higher collective activity than the sum of the activities of the individual enzymes (Mansfield *et al*., 1999). Various types of cellulases have different mode of actions: they attack the substrate at different locations and create new chain ends, which can be attacked in turn by other enzymes. The degree of synergism is dependent on i) the individual enzymes that constitute the cocktail, ii) enzyme loading, iii) substrate saturation and iv) properties of substrate (crystalline vs. soluble cellulose) (Medve *et al*., 1998; Nidetzky *et al*., 1994). Three types of synergism have been described (Mosier *et al*., 1999). First, in the endo-exo synergism, endoglucanases decrease the degree of polymerization of cellulose and generate new chain ends for subsequent attack by cellobiohydrolases. Enzymatic collaboration of EGI and CBHI has the highest degree of synergism which has been found (Henrissat *et al*., 1985). Moreover, other studies report the synergistic hydrolysis of crystalline cellulose by EGIII and CBHI (Okada *et al*., 1998a), or EGI and CBHII (Medve *et al*., 1998). Second, the exo-exo synergism concerns the collaborative work of type I cellobiohydrolase (CBHI), which hydrolyse cellulose chain from its reducing end and type-2 cellobiohydrolases (CBHII), which cleave the substrate from its non-reducing end (Medve *et al*., 1994; Fujita *et al*., 2002). Some authors, however, have suggested that CBHII could have some endoglucanase activity (Nutt *et al*., 1998; Boisset *et al*., 2000). Finally, intramolecular synergism refers to the influence of the carbohydrate binding module on its connected catalytic domain, which activity is higher than when independently produced (Mosier *et al*., 1999; Zhang and Lynd, 2004).

9. Cellulase substrates and assay methods

9.1 Substrates

Substrates for cellulase activity assays can be divided into two groups based on their solubility in water (Table II-16). Soluble substrates are usually employed to measure individual cellulase activities: they include cellodextrins (i.e. cello-oligosaccharides with a degree of polymerization lower than 6), chromogenic and fluorogenic substrates derived

from cellodextrins, such as *p*-nitrophenyl(pNP)-β-D-glycosides and methylumbelliferyl (MUF)-β-D-glycosides long and substituted cellulose, such as carboxymethyl cellulose (CMC) and hydroxyethyl cellulose (HEC), Remazol-Brilliant-Blue R dyed soluble CMC (also known as AZO-CM Cellulose) which is often used for agar plate screening and zymogram analysis. On the other hand, insoluble substrates consist of pure celluloses (e.g. cotton, Whatman No.1 filter paper, Avicel microcrystalline cellulose, bacterial cellulose, phosphoric acid-or alkali-treated amorphous cellulose, and regenerated amorphous cellulose) and "impure" celluloses (e.g. pretreated celluloses, dyed cellulose or α-cellulose, a solid residue of lignocellulose obtained after extraction with strong alkali and mainly composed of cellulose with a small amount of hemicellulose). An extensive review of substrates and of their preparations is proposed by Zhang and coworkers (2006).

Table II-16. Substrates commonly employed for the measurement of cellulase activity and their associated detection methods (Zhang *et al*., 2006)

Substrate	Detection[a]	Enzymes
Soluble		
Short chain (low DP)		
Cellodextrins	RS, HPLC; TLC	Endo, Exo, BG
Radio-labeled cellodextrins	TLC plus liquid scintillation	Endo, Exo, BG
Cellodextrin derivatives		
β-methylumbelliferyl-oligosaccharides	Fluorophore liberation, TLC	Endo, Exo, BG
p-nitrophenol-oligosaccharides	Chromophore liberation, TLC	Endo, Exo, BG
Long chain cellulose derivatives		
Carboxymethyl cellulose (CMC), hydroxyethyl cellulose (HEC)	RS; viscosity	Endo
Dyed CMC	Dye liberation	Endo
Insoluble		
Crystalline cellulose-		
Cotton, microcrystalline cellulose (Avicel),	RS, TSS, HPLC	Total, Endo, Exo
Valonia cellulose, bacterial cellulose	RS, TSS, HPLC	
Amorphous cellulose - PASC, alkali-swollen cellulose RAC	RS, TSS, HPLC, TLC	Total, Endo, Exo
Dyed cellulose	Dye liberation	Total, Endo
Fluorescent cellulose	Fluorophore liberation	Total
Chromogenic and fluorephoric derivatives		
Trinitrophenyl-carboxymethylcellulose (TNP-CMC)	Chromophore liberation	Endo
Fluram-cellulose	Fluorophore liberation	Endo, Total
Practical cellulose-containing substrates		
α-cellulose, pretreated lignocellulosic biomass	HPLC, RS	Total

* RS, reducing sugars; TSS, total soluble sugars.

9.2 Assay method

Assays perfomed to determine cellulase activity are classified into three main groups: 1) assays quantifying the formation of hydrolysis products, 2) assays measuring the decrease of substrate quantity, 3) and assays following the changes in physical properties of the substrate. Most common methods rely on the measurement

of the concentration of reducing sugars at the end-point of hydrolysis. Reducing sugars designate sugar molecules possessing a free aldehyde group which can be oxidized to form a carboxylic group in the presence of oxidizing agents. This free aldehyde group can be found on every sugar molecules possessing a non-linked anomeric carbon. Upon hydrolysis of a poly- or oligosaccharide, a new free reducing end is liberated. (Dashtban *et al*., 2009). The details of reducing sugars assays are described in Table II-17.

Table II-17. Common colorimetric sugar assays and their sensitivity (Zhang *et al*., 2006)

Method		Sample (mL)	Reagent (mL)	G amount (μg/sample)	G concn. (mg/L)	Ref.
Reducing Sugar Assay						
DNS	Micro	1–3	3	20–600	6.7–600	Miller, 1959
DNS	Macro	0.5	3	100–2500	200–5000	Ghose, 1987
Nelson-Somogyi	Micro	1–5	2+2	1–10	0.2–10	Somogyi, 1952
Nelson-Somogyi	Macro	2	2+2	10–600	5–300	Somogyi, 1952
Nelson	Semi-Micro	2	2	5–100	2.5–50	Nelson, 1944
Ferricyanide-1		1–3	1+5	1–9	0.3–9	Park and Johnson, 1949
Ferricyanide-2		1	0.25	0.18–1.8	0.18–1.8	Kidby and Davidson, 1973
PAHBAH	Micro	0.5	1.5	0.5–5	1–10	Lever, 1972
PAHBAH	Macro	0.01	3	5–50	500–5000	Lever, 1972
BCA		0.5	0.5	0.2–4.5	0.4–9	Waffenschmidt and Janeicke, 1987
Modified BCA		1	1	0.4–9	0.4–9	Zhang and Lynd, 2005b
Total Sugar Assay						
Phenol-H_2SO_4		1	1+5	5–100	10–100	Dubois et al., 1956; Zhang and Lynd, 2005b
Anthrone-H_2SO_4		1	1+5	5–100	10–100	Roe, 1955; Viles and Silverman, 1949
Enzymatic Glucose Assay						
Glucose-HK/PGHD kit		0.01	1	2–50	200–5000	Sigma Kit
Glucose-HK/PGHD kit		0.5	0.5	2–50	4–100	Zhang and Lynd, 2004a

The DNS (3,5-dinitrosalicylic acid) and Nelson-Somogyi methods are the 2 most common colorimetric sugar assays performed to measure cellulase activity because they have a broad range detection of sugar concentration and don't suffer from interference by other components of the reaction. It is, however, known that DNS reagent may partly hydrolyze cellulosic substrates (Ghose, 1987; Nelson, 1944; Somogyi; 1952). The PAHBAH and BCA methods benefit from higher sensitivity but proteins in the reaction may interfere in the analysis (Lever, 1972; Waffenschmidt and Jaenicke, 1987). The main problem of these quantification methods is that they rely on glucose samples used for standard concentrations while measured reducing ends come from either cellulose or cellodextrin molecules, thus causing a difference in sensitivity. Glucose oxidase kit and HPLC are good methods to overcome the described problems as they strictly measure the formation of glucose, but they require

more steps to convert cello-oligosaccharide to glucose (Zhang *et al*., 2006; Zhang *et al*., 2009a). Phenol-H_2SO_4 and anthrone-H_2SO_4 methods are used to measure total carbohydrate concentrations: both methods are very sensitive, do not suffer from protein interference and are strictly stoichiometric as all poly- and oligo-saccharides are completely acid-hydrolysed to form sugar monomers (Dubois *et al*., 1956; Roe, 1955).

9.3 Cellulase assays

When working with cellulase cocktails, assays can be performed to either 1) determine the total activity of the enzyme consortium 2) or measure the activities of each individual cellulase.

Total cellulase activity can be measured by using various cellulosic substrates such as Whatman No. 1 filter paper, Avicel microcrystalline cellulose, dyed cellulose, ^{14}C-labelled cellulose, cotton linters, bacterial cellulose, algal cellulose. Filter paper assay (FPA) is the most widely used method since 1984 and has been standardised by Ghose (1987) who proposed to uses 50 mg or 1×6 cm strip of Whatman No. 1 filter paper as substrate. Released reducing sugars are estimated as glucose-equivalent by the DNS method. One FPU (filter paper unit) is defined as the amount of enzyme that releases 1 micromole of glucose equivalent per minute under the assay conditions. However, FPAs is time-consuming, labor-intensive, requires reagents in large quantities but most of all, extra-addition of β-glucosidase to enable reliable measurements (Ghose, 1987; Kuhad *et al*., 1997; Dashtban *et al*., 2009).

Individual endoglucanase activity is usually measured by using substituted celluloses such as carboxymethyl cellulose (CMC) and hydroxyethyl cellulose (HEC) because they cannot be hydrolyzed by exoglucanases. The hydrolysis of the substrate can be monitored via viscometric methods, CMC plate-clearing assays, reducing sugar measurements, chromatography methods. Chromogenic or fluorogenic CMC can also be used. Measurement of released reducing sugars from CMC (CMCase activity) is often done at 2.0% (w/v) concentration, pH 4.8 and 50 °C. When working with substituted cellulose, an important factor concerns the degree of substitution of the substrate. This value should be lower than 0.7 (7 carboxymethyl groups per 10

anhydroglucose units) because cellulases can only access non-substituted glucose units. Moreover, adding dye to CMC is known as “zymograms” used for detection of endoglucanases in polyacrylamide gels where the endoglucanases are presented (Sharrock, 1988; Dashtban *et al*., 2009).

Measurement of exoglucanase activity is more difficult. Avicel is the substrate that is usually used for measurement of CBHs activity because it has a low degree of polymerization and a high ratio of reducing ends over total β-glucosidic bonds, resulting in a low accessibility to EGs. Unfortunately, Avicel still contains some cellulose chains with amorphous regions and soluble cellodextrins that can be either hydrolyzed by endo- or exo- acting enzymes (Dashtban *et al*., 2009). Therefore, there are no specific substrates for exoglucanases activity determination within a cellulase mixture. Chromogenic substrate like *p*-nitrophenyl-β-D-cellobioside and fluorogenic substrate like 4-methylumbelliferyl-β-D-lactoside are effective substrates for assaying CBHI activity in the presence of β-glucosidase inhibitor (D-glucono-1,5-δ-lactone). These substrates however, cannot be hydrolysed by CBHII which prefers longer glucose chains (Dashtban *et al*., 2009; Zhang *et al*., 2006).

β-Glucosidases assays are very sensitive and simple methods which employ chromogenic or fluorogenic substrates such as *p*-nitrophenol-β-D-glucophyranoside (pNP-Glc) and 4-methylumbelliferyl-β-D-glucopyranoside (MUF-Glc) which are the two most commonly used substrates for β-glucosidase assay. Hydrolysis reactions are usually done at a substrate concentration between 1 and 5 mM, and are stopped by addition of sodium tetraborate. Absorbance is then read at 405 nm. Soluble cellodextrins with DP up to 6, can also be used to measure β-glucosidases, providing that the liberated glucose can be evaluated by the glucose oxidase/peroxidase method (Kuhad *et al*., 1997; Dashtban *et al*., 2009).

10. Thermostable cellulases

Enzymes showing a tolerance to high temperatures are potential interesting tools to be used in harsh industrial processes. For an enzyme, thermostability refers to the residual activity that remains after incubation of the enzyme at a chosen temperature along the studied time. In biorefineries, cellulosic biomass needs to be

pretreated by heating to make it more suitable for enzymatic saccharification. Thus, the use of thermostable cellulases would lead to less processing time and save energy cost by eliminating the cooling step after thermal pretreatment (Rastogi *et al*., 2010; Liu *et al*., 2011a). According to Arrhenius law, a temperature increase by 10 °C in the bioreactor will double the reaction rate and decrease the amount of enzyme required for the catalytic conversion (higher specific activity). Moreover, working at elevated temperatures permits to reach a higher solubility for concentrated substrates, to reduce the risk of microbial contamination, to improve fermentation yields and qualities, and to increase the flexibility of the process configuration (Shyamala *et al*., 2011; Yang *et al*., 2010). Several thermophilic bacteria such as *Clostridia*, *Thermotoga*, *Rhodothermus* sp. and thermophilic fungi such as *Talaromyces emersonii*, *Thermoascus aurantiacus*, *Chaetomium thermophilum* and *Myceliophthora thermophila*, have been isolated from various environments to produce thermostable cellulases (Table II-18) (Maheshwari *et al*., 2000; Viikari *et al*., 2007). Protein engineering was also used to increase the thermostability of various cellulases (see PART IV of the Introduction). A good general conformational structure-involving for example a compact packing or a reduced unfolding entropy associated to intermolecular interactions facilitating protein folding and maintaining a catalytically active conformation are important factors affecting thermostability (Li *et al*., 2005).

Table II-18. Sources of microoraganisms producing thermostable cellulases (Viikari *et al*., 2007)

Organisms	Enzymes	Characteristics of enzymes			
		MW (SDS PAGE) (kDa)	pH optimum	T optimum (°C)	Stability
Acidothermus cellulolyticus	Endoglucanase I	57.420–74.580	5.0	83	Inactivated at 110 °C
Anaerocellum thermophilum	Endoglucanase	230	5–6	95–100	Half-life 40 min at 100 °C
Bacillus sp. KSM-S237	Endoglucanase	86	8.6–9.0	45	30% of activity remained after 10 min at 100 °C
Caldocellum saccharolyticum	Endoglucanase	na			na
Caldocellulosiruptor saccharolyticus	Endoglucanases Exoglucanases	na	7.0	68–70	na
Chaetomium thermophilum	Endoglucanase	68	4.0	60	Stable at 60 °C > 60 min, half-life 7 min at 90 °C
Cladosporium sp.	Endoglucanase Exoglucanase	na	4–6	60	Stable at 60 °C for 24 h
Clostridium stercorarium	Endoglucanase	100	6.0–6.5	90	Stable for several days
Clostridium stercorarium	Exoglucanase	87	5–6	75	Stable at 70 °C for 3 days
Clostridium thermocellum	Endoglucanase	83	6.6	70	33% of activity remained after 50 h at 60 °C
Clostridium thermocellum	Endoglucanase	76	7.0	70	50% of activity remained after 48 h at 60 °C
Melanocarpus albomyces	Endoglucanase	20	6–7	70	70% of activity remained after 60 min at 80 °C
Rhodothermus marinus	Endoglucanase	49	7.0	95	50% of activity remained after 3.5 h at 100 °C, 80% after 16 h at 90 °C

Table II-18. Sources of microoraganisms producing thermostable cellulases (Continued)

Organisms	Enzymes	Characteristics of enzymes			
		MW (SDS PAGE) (kDa)	pH optimum	T optimum (°C)	Stability
Rhizomucor pusillus	Endoglucanase	na	5.5	60	Stable at 60 °C for 84 h
Sporotrichum sp.	Endoglucanase	33	4.5–5	70	Stable at 70 °C for 30 min
Streptomyces sp.	Exoglucanase	44	4	60	30% of activity remained after 10 min at 100 °C
Streptomyces sp.	Endoglucanase	26	4	60	20% of activity remained after 10 min at 100 °C
Talaromyces emersonii	Exoglucanase (CBH IA)	66	3.6	78	Half-life 34 min at 80 °C
Thermoascus aurantiacus	Endoglucanase		4.5	75	Half-life 98 h at 70 °C and 4.1 h at 75 °C
Thermomonospora curvata	Endoglucanase	100	6.0–6.5	70–73	Stable for 48 h at 60 °C, half-life 30 min at 80 °C
Thermotoga sp.	Exoglucanase	36	6.8–7.8	100–105	Half-life 70 min at 108 °C, 2 min at 117 °C
Thermotoga maritima	Endoglucanase	27	6.0–7.5	95	Half-life 2 h at 95 °C
Thermotoga neapolitana	Endoglucanase (CelA)	29	6.0	95	Half-life > 240 min at 100 °C
	Endoglucanase (CelB)	30	6.0–6.6	106	

PART III Heterologous expression of cellulases

Nowadays, commercial cellulases are still too costly to be largely employed in industrial bioconversion process. Therefore, isolating and engineering novel cellulases with high activity and thermostability for cellulosic ethanol production represent a big challenge (Hong *et al*., 2007; Badieyan *et al*., 2012). To achieve this, one strategy could be to improve natural cellulolytic strain, using metabolic engineering, random mutagenesis on selected gene sequences to optimize desired properties, and co-culturing different cellulose-degrading microorganisms. Alternatively, cellulase genes can be recombinantly produced in non-cellulolytic microorganisms having high protein production capabilities (Lynd *et al*., 2002; Lynd *et al*., 2005). Heterologous expression is a powerful strategy to improve the yield of enzyme production, as well as to produce enzymes in highly purified and well-characterized form. That's why many different cellulase genes have been already cloned and expressed in bacteria, fungi and yeast expression platforms (Dashtban *et al*., 2009). Ultimately, such engineered strains designed to produce a heterologous cellulose system, they should be able to grow on and ferment hydrolysed sugars resulting from lignocellulose degradation (Lynd *et al*., 2005).

1. Bacterial expression systems

Heterologous expression of proteins in bacterial strains has been used since the late 1970s for overexpression of human insulin and somatostatin. Bacteria are usually attractive hosts because of their rapid growth, cheap cultivation medium, and well-established genetic tools, including large selections of plasmids for "gene transportation" and mutant host strains (Makino *et al*., 2011). Concerning cellulases, their heterologous production has been mainly done in high ethanol producing bacteria such as mutant strains of *E. coli*, *Klebsiella oxytoca*, or *Zymomonas mobilis* (Lynd *et al*., 2005).

The gram-negative bacterium *E. coli* is the most widely used microorganism for large production of proteins on industrial scale and a number of *E. coli* strains have been developed to facilitate heterologous expression. In microbiology, most strains derive from either *E. coli* B strains (for example, *E. coli*

BL21 (DE3) or Rosetta strains, designed to enhance rare codon usage) or from *E. coli* K-12 (for example, Origami strains which are used to facilitate disulfide bond formation in the cytoplasm) (Terpe, 2006). Bacterial cellulases have been successfully expressed in *E. coli* as stated in many reports. For instance, an alkaline endoglucanase from *Bacillus akibai* was cloned into vector pET-28a (+) and expressed efficiently in *E. coli* BL21 (DE3); optimization of culture conditions could lead to a considerable increase of extracellular endoglucanase production, from 5.3 to 18.5 U/ml of culture medium, by addition of 5g/L EDTA (Liu *et al*., 2011b). Moreover, *T. reesei* EGI, EGII, EGIII genes were also expressed in *E. coli* strains, Rosetta-gami B (DE3) pLacI or Origami B (DE3) pLacI, resulting in best yields of 6.9, 72, and 50 mg/L, respectively (Nakazawa *et al*., 2008). However, when *T. reesei* CBHII gene was expressed in *E. coli* strains BL21 (DE3), Origami, or Origami B, the enzyme was tightly associated to the outer membrane (Abdeljabar *et al*., 2012).

Z. mobilis is also of interest for recombinant protein production because it is a host of choice for consolidated bioprocessing. Indeed, when *Z. mobilis,* transformed with a plasmid containing a gene coding for *Enterobacter cloacae* endoglucanase, was used to hydrolyse carboxymethyl cellulose or 4% NaOH-pretreated bagasse, up to 5.5% and 4% ethanol (v/v), could directly be produced, respectively (Vasan *et al*., 2011). Furthermore, an ethanologenic derivative of *Klebsiella oxytoca* engineered to produce endoglucanases CelY and CelZ from *Erwinia chrysanthemi* could reach an ethanol production of up to 27 g/L, providing the addition of 10 ml/L of Genencor Spezyme (Zhou *et al*., 2001).

Nonetheless, bacterial recombinant expression has also some disadvantages. First, prokaryotic bacteria do not possess post-translational modification machineries for glycosylation or phosphorylation for example required for proper folding and function of proteins (Yin *et al*., 2007). Moreover, *E. coli* often produce and accumulate endotoxins, which are pyrogenic in mammals and humans. Another problem concerns difference in codon usage which can be critical to express eukaryotic genes: a bias in the frequency of codon usage can lead to a lack of some tRNAs during the translation process, and thus cause premature termination, frameshift errors, or amino-acid misincorporation. Furthermore, cytoplasmic proteins

produced by bacteria usually are produced as insoluble and most often non exploitable aggregates called inclusion bodies (Terpe, 2006).

2. Yeast expression systems

Since eukaryotic genes are difficultly expressed in bacterial systems, heterologous production systems have been developed in eukaryotic hosts such as yeasts, filamentous fungi, insects, and mammalian cells. They offer post-translation modification possibilities, which are very important process for many recombinant proteins to function efficiently.

Yeasts are hosts of choice, offering rapid growth in simple media to reach high cell-density (130 g/L), ease in genetic manipulations, ability to perform eukaryotic processing mechanisms such as disulfide bond formation and glycosylation (Romanos *et al*., 1992).

Fungi have been used industrially for many years but they are time-consuming to grow, offer low secretion rates, their genetic characterization is incomplete and cloning tools are still insufficient. Insect and mammalian cell lines are suitable for very complex, high-molecular weight proteins, e.g. blood clotting factors, multimeric antibodies. However, they produce proteins with low yield, require high operational costs, and are difficult to scale up (Sodoyer, 2004; Schmidt, 2004).

2.1 Conventional Saccharomyces cerevisiae yeast

In the early 1980s, after enough information had been accumulated about its physiology and its genetics baker's yeast *S. Cerevisiae* started to be used to produce recombinant proteins, and first, human interferon (Cregg *et al*., 2009; Dominíguez *et al*., 1998). *S. cerevisiae* had been used for centuries in food and beverage industry because it is considered as GRAS (Generally Regarded As Safe). Its survival strategy is based upon the production of ethanol levels often exceeding 10% v/v, using C_6 sugars as carbon sources. This level of ethanol can eliminate most of other microbes, which usually cannot tolerate more than 4% v/v. As a consequence, this combination of ethanol production capability and tolerance, has led *S. cerevisiae* to be considered as an interesting CBP microorganism. Many cellulase genes have been successfully expressed in *S. cerevisiae*, with EGs being usually produced in

larger quantities than CBHs (Table II-19). All major *T. reesei* EGs (EGI-V) and CBHs (CBHI, CBHII) are efficiently secreted by recombinant *S. cerevisiae* (van Zyl *et al*., 2007). In another study, a display strategy has been applied to expose at the same time on the cell-surface of *S. cerevisiae* three groups of cellulases (*T. reesei* EGII, CBHII, and *Aspergillus aculeatus* BGL1), resulting in a mutant yeast that can directly utilize amorphous cellulose and simultaneously ferment the hydrolysed sugars to produce ethanol (Fujita *et al*., 2002; Fujita *et al*., 2004). Nevertheless, *S. cerevisiae* suffers from several disadvantages. First, its secretion efficiency and production yield are usually quite low, with a maximum of 1–5% of total protein recovered in the extracellular medium, since proteins are often retained in the periplasm or remain bound to the cell wall. Then, hyperglycosylation is frequent and its level can affect protein stability. On a genetic point of view, mutants can be instable due to loss of plasmid during cultivations (Dominíguez *et al*., 1998).

Table II-19. Heterologous expression of cellulases in *S. cerevisiae* (van Zyl *et al*., 2007)

Organism & gene/enzyme	Titer (mg/L)	% cell pro-tein	Substrate(s) activity was detected against (values indicate activity measured per L culture broth)	Specific activity (U/mg)
CBHI				
Trichoderma reesei CBHI	2	1.5	MUC, AC	NR
	5	*0.123*	MUL, BMCC	0.26 (on BMCC)
	0.22	0.006	0.06 U/L (PASC), 0.06 U/L (BMCC)	0.22 (on PASC)
Aspergillus niger CBHB	NR	NR	0.035 U/L (AC), 0.03 U/L (BMCC)	NR
Phanerochaete chrysosporium CBH1–4	NR	NR	12 U/L, ~ 3.3 U/g DCW (BBG), 10 U/g DCW	NR
	NR	NR	(PNPC) 22 U/g DCW (AC)	NR
	NR	NR	18 U/g DCW (PNPC)	NR
	NR	NR	0.035 U/L (AC),	NR
EG				
Trichoderma reesei EGI	NR	0.5	CMC	15 (on CMC)
	10	*0.09*	MUC	NR
	0.66	0.25	BBG, lichenan, CMC, HEC, MUL, MUC	NR
	5	*0.12*	72 U/g DCW (HEC)	60 (on HEC)
Trichoderma reesei EGII	NR	NR	3.64 U/g DCW (AC)	NR
Trichoderma reesei EGIII	NR	NR	BBG, lichenan, CMC, HEC	NR
Trichoderma reesei EGV	NR	NR	BBG, HEC	NR
Trichoderma reesei EGIV	NR	NR	BBG, AC, CMC	NR
Aspergillus niger eng1	*2.8*	*0.07*	574 U/L (CMC)	204 (on CMC)
Aspergillus aculeatus CMCase	NR	NR	0.5 U/L, *~ 0.06 U/g DCW* (CMC)	NR
Aspergillus aculeatus F1-CMCase	NR	NR	60 U/L (CMC)	NR
	NR	NR	CMC, IOSC	11 (on IOSC)
Cellulomonas fimi Eng (*cenA*)	*13*	NR	293 U/L (low viscosity CMC)	NR
BGL				
Kluyveromyces fragilis BGL	NR	15	PNPG, C2	64.4 (on PNPG)
Aspergillus aculeatus BGLI	NR	NR	BGL1 = 21.3 U/g DCW (PNPG)	NR
	1	*0.02*	IOSC	25 (on IOSC)
Saccharomycopsis fibuligera BGLI	*10*	*0.25*	PNPG, C2, C3, C4	43.3, 20.1, 26.2, 27.1 (as for activity)
Saccharomycopsis fibuligera BGLII	18.9	*0.47*	PNPG, C2, C3, C4	168, 0.8, 1.7, 1.5 (as for activity)
	NR	NR	115 000 U/L, ~ 12 800 U/g DCW (PNPG)	NR
	NR	NR	112 U/g DCW (PNPG)	NR
	NR	NR	19 U/g DCW (PNPG)	NR

2.2 Non-conventional yeasts

Besides *S. cerevisiae*, a number of non-conventional yeasts have been developed for foreign gene expression such as *Pichia pastoris*, *P. methanolica*, *Hansenula polymorpha*, *Kluyveromyces lactis*, *Yarrowia lipolytica*, *Arxula adeninivorans* or *Schizosaccharomyces pombe* (Dominíguez *et al*., 1998; Muller *et al*., 1998). However, it is important to keep in mind that the level of expression of a gene is affected by many factors, including the efficiency of its transcription or the stability of the resulting protein for example. Therefore, a number of recombinant hosts should be tested to compare the level of expression and thus to improve chance of success (Romanos *et al*., 1992).

2.2.1 Methylotrophic and lactose-utilizing yeasts

P. pastoris was discovered in 1969 by Koichi Ogata because of its ability to utilize methanol as the sole carbon source for single-cell protein production: it thus belongs to the family of methylotrophic microorganisms. Initially exploited by Phillips Petroleum Company in collaboration with the Salk Institute Biotechnology/Industrial Associates, Inc. (SIBIA), this yeast has been the most widely used non-conventional yeast for protein production since the 1980s. A series of vectors, strains and methods were indeed developed based on the strong regulated *AOX*1 (alcohol oxydase I) gene promoter. Then, in 1993, *P. pastoris* was developed to be an expression kit (Invitrogen) for academic purpose. This yeast has long been considered as the most powerful yeast because its secretion efficiency is very high and its cultivation does not require complicated process developments (Higgins and Cregg, 1998). In 2004, there had been more than 400 proteins recombinantly produced by this yeast (Cereghino and Cregg, 2000). The highest yields of recombinant protein production by *P. pastoris* were reported for murine collagen (15 g/L) and tetanus toxin fragment C (12 g/L), to be compared to a 1 g/L production in *S. cerevisiae*. Interestingly, only a few copies of the gene can drive an expression level equivalent to 50 gene-copies in *S. cerevisiae*. Nowadays, *P. pastoris* is used in industrial processes for the production of recombinant trypsin by Roche for example (Schmidt, 2004).

H. polymorpha is also methylotrophic yeast, which can grow at temperatures up to 50 °C and is able to secrete large amount of proteins with fermentation times shorter than *P. pastoris*. Heterologous expression systems largely employ the powerful *MOX* (methanol oxydase) promoter (Dominíguez *et al*., 1998; Müller *et al*., 1998). Moreover, when autonomously replicating sequence (ARS) based vectors are used, the gene expression cassette is integrated with a very high frequency, reaching more than 100 copies (Dominíguez *et al*., 1998; Cregg *et al*., 2009). Some limitations concern homologous integration in the *MOX* site which is more difficult than for the *Pichia* system, and the high glycosylation level, corresponding to up to a 20 kDa increase for some proteins (Dominíguez *et al*., 1998; Müller *et al*., 1998). The highest yield of recombinant protein production by *H. polymorpha* approximately 13.5 g/L was reported for phytase (Mayer *et al*., 1999). Currently, *H. polymorpha* is used for the industrial production of hepatitis B vaccine by Rhein-Biotech (Schmidt, 2004).

Unlike other methylotrophic yeasts, *K. lactis* is lactose-utilizing yeast and considered as a GRAS microorganism. It was the first yeast developed for heterologous protein expression after *S. cerevisiae* and was mainly used to produce β-galactosidase, the main enzyme used for the preparation of low-lactose milk, and rennet enzymes for cheese processing. For recombinant expression in *K. lactis*, strong promoters from lactose-assimilation pathways such as lactose permease and galactosidase promoters are used to drive the secretion of recombinant proteins. The highest yield of recombinant protein production by *K. lactis* was reported for human serum albumin, and is approximately of 3 g/L. At present, it is used in industrial applications for bovine prochymosin production by Gist –Brocades and it can be purchased as a commercial kit for protein expression from New England Biolabs (Schmidt, 2004; Cregg *et al*., 2009).

2.2.2 Dimorphic yeasts

Dimorphic yeasts are characterized by their two stages of growth: a vegetative and a mycelial growth phase. It has been observed that the level of protein secretion is higher during the mycelia than during the vegetative growth phase. *A.*

adeninivorans is dimorphic yeast which can grow at temperature as high 48 °C and shows an osmotolerance of up to 3.4 M NaCl (Cregg *et al*., 2009).

Y. lipolytica also belongs to the group of dimorphic yeasts, and is considered as one of the most attractive non-conventional yeasts because of its good transformation efficiency, gene integration stability, capacity for high level of protein production, with no excessive post-translational glycosylation problems, and a good adaptability to high-throughput screening methodologies (Sodoyer, 2004; Cregg *et al*., 2009; Müller *et al*., 1998; Bordes *et al*., 2007). Recombinant genes, such as those for anaphylatoxin C5a and bovine prochymosin, were initially expressed under the control of the alkaline extracellular protease (*XPR2*) promoter. However, this promoter requires large amount of peptone for full level of induction. Therefore, numbers of promoters have been developed to achieve high expression level in more ordinary medium (Müller *et al*., 1998; Dominíguez *et al*., 1998). In 1998, Müller and coworkers compared the expression level of six enzymes (galactanase, lipase, polygalacturonase, xylanase and two cellulases) between *Y. lipolytica* and four other yeasts (*S. cerevisiae*, *H. polymorpha*, *K. lactis*, *S. pombe*). Moreover, they also developed novel promoter *TEF* and *RPS7* instead of *XPR2* promoter. They concluded that with the using novel *TEF* promoter, *Y. lipolytica* is an attractive host for heterologous gene expression with its secretion ability is 4.5-100 times higher than *S. cerevisiae* (Müller *et al*., 1998).

3. *Pichia pastoris* expression platform

3.1 Characteristics

P. pastoris has been widely used for heterologous protein production thanks to (Li *et al*., 2007; Romanos, 1995; Macauley-Patrick *et al*., 2005):

1) its rapid growth rate to high cell density stage (>130 g/L dry cell weight) in inexpensive and non-complex medium;
2) its very high expression level (~gram per litre of culture or accounting for 5-40% of total protein), even when the product is toxic to the cell;
3) its ease of genetic manipulation, like *S. cerevisiae;*
4) the availability of an efficient host and vector system;

5) the existence of tightly controlled (repressed by glucose or other carbon sources) and highly inducible promoters;
6) the possibility of post translation modification processes such as polypeptide folding, glycosylation, methylation, acylation, sulphation, phosphorylation, lipidation, proteolytic adjustment, and targeting to subcellular compartments
7) the absence of known proteases concomitant to a very low level of endogenous proteins, allowing easy purification.

In addition, it has been used as a model organism because it is well-studied and has similar growth conditions to *S. cerevisiae* and tolerances without using of special equipment. The steps to a successful expression in *P. pastoris* are: 1) insertion of the gene of interest into an expression vector, 2) integration of the resulting expression cassette into host genome, 3) screening of potential clones for expression of studied enzymes. The developments of this system have enabled a wide choice of vectors, promoters, secretion signals, selectable markers, strains, and now available in commercial kits (Li *et al*., 2007).

3.2 Expression vectors

All *P. pastoris* expression vectors are integrating vectors because episomal plasmids are not stable. They are shuttle vectors that can work in both *E. coli* and *P. pastoris*, because they contain at the same time an origin of replication for plasmid maintenance in *E. coli*, multi-cloning sites, a promoter to drive the expression of the gene of interest in the yeast, and selectable markers functional in one or both organisms (antibiotic resistance genes, auxotrophy markers). In addition, some vectors also contain a signal sequence that needs to be fused to the sequence of the protein of interest to facilitate its secretion. Besides the gene, a 3'*AOX* or *HIS* flanking region can also be found which role is to direct the integration of the expression cassette at the corresponding loci in the yeast chromosome by single crossover insertion or gene replacement, a very stable mechanism for gene maintenance without any needs of selective pressure (Figure II-19) (Romanos, 1995; Cregg *et al*., 2000). First generation of *P. pastoris* vectors contain the histidine dehydrogenase gene (*HIS4*) or the GS418 (kanamycin) resistance gene for selection of

desired transformants in *P. pastoris*. However, these vectors were very large (9.0 - 9.3 kb thus compromising cloning process and causing the integrants to be less genetically stable. Therefore, another set of vectors (pPICZ and pGAPZ series from Invitrogen) was created, which uses a zeocin resistance gene (only 375 bp) for selection of transformants (Daly and Hearn, 2005). The list of expression vectors for *P. pastoris* heterologous expression is presented in Table II-20.

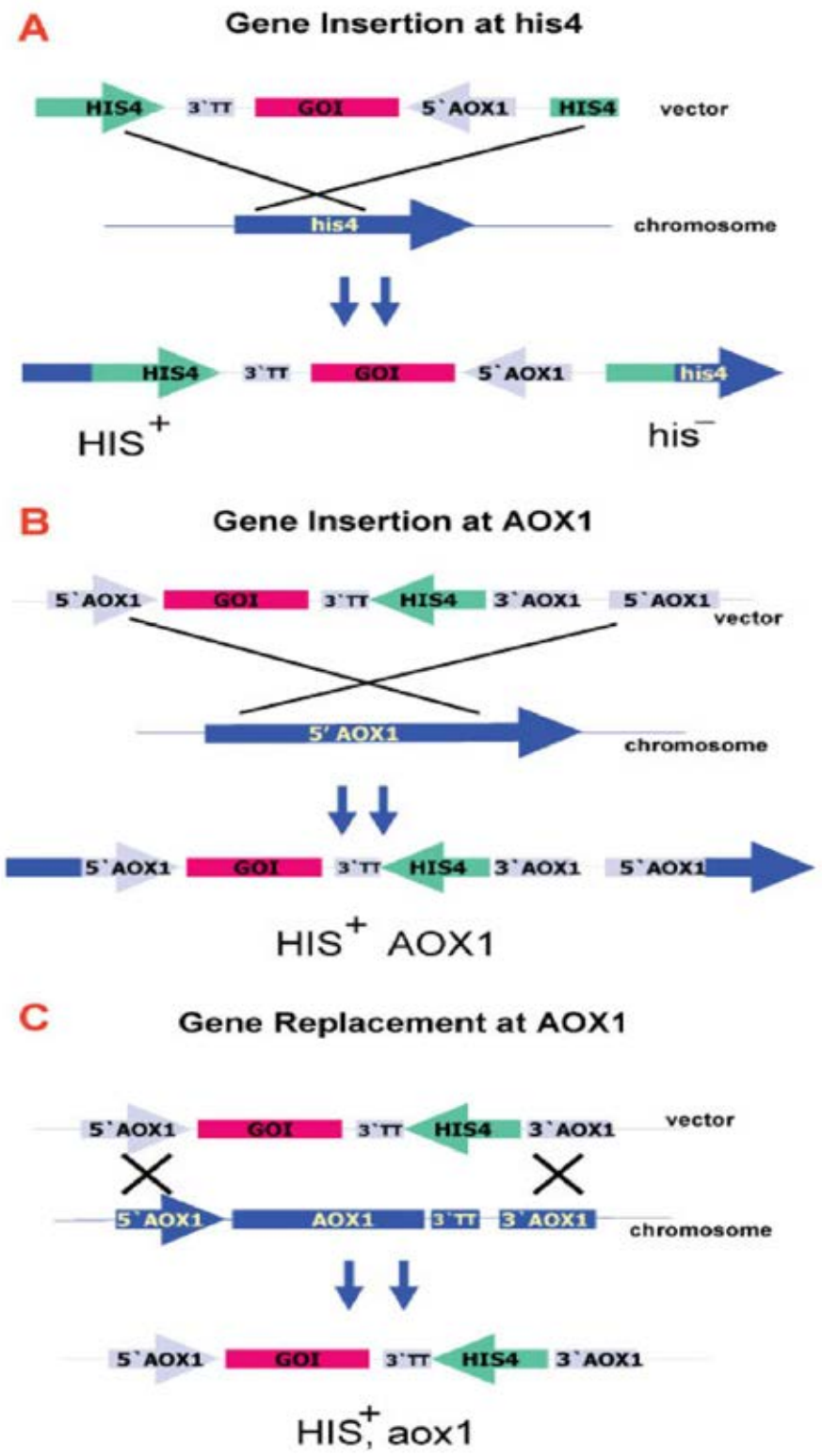


Figure II-19. Mechanisms of integration of expression cassettes in *P. pastoris* genome. (A) insertion of gene at *his4* locus. (B) insertion of gene at *AOX1* locus. (C) gene replacement at *AOX1* locus. GOI: gene of insertion (Daly and Hearn, 2005).

Table II-20. *P. pastoris* expression vectors (Higgins and Cregg, 1998)

Vector name	Selectable markers	Features
Intracellular		
pHIL-D2	*HIS4*	*Not*I sites for *AOX1* gene replacement
pAO815	*HIS4*	Expression cassette bounded by *Bam*HI and *Bgl*II sites for generation of multicopy expression vector
pPIC3K	*HIS4* and *kan*^{r}	Multiple cloning sites for insertion of foreign genes; G418 selection for multicopy strains
pPICZ	*ble*^{r}	Multiple cloning sites for insertion of foreign genes; Zeocin selection for multicopy strains; potential for fusion of foreign protein to His_6 and *myc* epitope tags
pHWO10	*HIS4*	Expression controlled by constitutive *GAPp*
pGAPZ	*ble*^{r}	Expression controlled by constitutive *GAPp*; multiple cloning site or insertion of foreign genes; Zeocin selection for multicopy strains; potential for fusion of foreign protein to His_6 and *myc* epitope tags
Secretion		
pHIL-S1	*HIS4*	*AOX1p* fused to *PHO1* secretion signal; *Xho*I, *Eco*RI, and *Bam*HI sites available for insertion of foreign genes
pPIC9K	*HIS4* and *kan*^{r}	*AOX1p* fused to α-MF prepro signal sequence; *Xho*I (not unique), *Eco*RI, *Not*I, *Sna*BI and *Avr*II sites available for insertion of foreign genes; G418 selection for multicopy strains
pPICZα	*ble*^{r}	*AOX1p* fused to α-MF prepro signal sequence; multiple cloning site for insertion of foreign genes; Zeocin selection for multicopy strains; potential for fusion of foreign protein to His_6 and *myc* epitope tags
pGAPZα	*ble*^{r}	Expression controlled by constitutive *GAPp*; *GAPp* fused to α-MF prepro signal sequence; multiple cloning site for insertion of foreign genes; Zeocin selection for multicopy strains; potential for fusion of foreign protein to His_6 and *myc* epitope tags

3.3 Promoters

The *AOX* promoter is the most widely used promoter in *P. pastoris*, being involved in the regulation of expression of the central alcohol oxidasegene. *AOX* is homo-octamer enzyme which function is to oxidize methanol to generate formaldehyde (a central intermediate of the methanol catabolism pathway) and H_2O_2.

It is synthesized in large amounts (up to 30% of total protein). *P. pastoris* has two alcohol oxidase genes, *AOX1* and *AOX2*, the former being expressed at higher level than the latter. *AOX1* is repressed by carbon sources such as glycerol and glucose and is tightly regulated and strongly induced by methanol. This characteristic is useful for expression of proteins which are toxic to the host and minimizes the possibility to select non-expressing mutants/contaminants. In contrast, *AOX2* is expressed at lower level than *AOX1* (10 – 20 times), so its application is more limited. Despite many advantages, it is important to remark that methanol is highly flammable and toxic, therefore unsuitable for use in food applications. Moreover, cultivations and protein productions require high oxygen consumption and are performed as two-stage fermentation (Bollok *et al*., 2009; Cereghino and Cregg, 2000; Macauley-Patrick *et al*., 2005; Zhang *et al*., 2009b).

Research to discover methanol-free promoters have been going on for a few years. When glucose is used as carbon source, the strong constitutive glyceraldehyde 3-phosphate dehydrogenase (GAP) promoter has been shown to give production levels of recombinant protein similar to those of the *AOX1* promoter, without any need of induction, so that the expression process is much facilitated. However, this systemis not suitable for the production of proteins that is toxic to the yeast (Bollok *et al*., 2009; Zhang *et al*., 2009b). The glutathione-dependent formaldehyde dehydrogenase (*FLD1*) is also a key enzyme not only in the methanol metabolism, where it is involved in formaldehyde detoxification to form carbon dioxide and energy through NADH synthesis, but also in the assimilation of amines such as methylamine as nitrogen source. As *P. pastoris* cannot grow on methylamine alone as a nitrogen and carbon source, media containing either glucose and methylamine or methanol and ammonium sulfatewere shown to result in good induction of the *FLD1*-promoter (comprise up to 20% of total cellular protein),

whereas it is repressed by glucose and ammonium sulfate as combined carbon and nitrogen sources, respectively. Therefore, this promoter provides the flexibility to use either methanol or methylamine, an inexpensive non-toxic nitrogen source (Cregg *et al*., 2000; Cereghino *et al*., 2001; Gellissen *et al*., 2005).

In some cases, expression of genes under *AOX1*, *GAP*, or *FLD1* promoters is too strong and leads to misfolded, unprocessed or mislocalized recombinant proteins,. As a consequence, the use of moderate strength promoters like *PEX8* (*PER3*) and *YPT1* constitutes an interesting alternative for special applications. *PEX8* is induced by methanol and increases expression level up to 5-fold. The *YPT1* promoter is constitutively expressed in glucose, methanol or mannitol containing media, and usually gives expression levels ten times lower than the *GAP* promoter (Macauley-Patrick *et al*., 2005; Li *et al*., 2005).

Other promoters used for recombinant protein production in *P. pastoris* are those regulating the expression of: *TEF1*, a translation elongation factor gene which is associated to cell growth-and expressed in glucose containing media; *DHAS,* a dihydroxyacetone synthase gene involved in methanol metabolic pathway; and *ICL1*, an isocitrate lyase gene (Bollok *et al*., 2009; Daly and Hearn, 2005; Macauley-Patrick *et al*., 2005).

3.4 Secretion signals

Proteins that are normally secreted in their native host usually possess a signal sequence which directs their extracellular transport. However, this sequence may or may not function properly in the heterologous host. Therefore, it is often required to design and test two constructions, one with the native signal peptide and another that fuses an efficient yeast signal sequence with the sequence of the mature native protein (Cregg, 2007). In general, yeast signal sequences are more likely to be successful and yield 2-3 fold higher extracellular production than native signal peptide. The most popular signal sequence in *P. pastoris* is the α-factor prepropeptide (αMF) from *S. cerevisiae*, which is used to secrete a wide range of enzymes. It consists of a pre-sequence of 19-amino acid followed by a pro-region of 66 residues, mainly hydrophilic and containing three N-linked glycosylation sites and one Kex2

endopeptidase site (Romanos, 1995; Cereghino and Cregg, 2000). Alternative signal sequences exist such as those from *H. Polymorpha* acid phosphatase PHO1, *S. cerevisiae* invertase SUC2, and *Phaseolus vulgaris* phytohaemagglutinin PHA-E (Macauley-Patrick *et al*., 2005).

3.5 Selection markers

There are two groups of selectable markers used in *P. pastoris* and other yeast expression systems. First, nutritional markers like *HIS4* (histidinol dehydrogenase gene) are used to complement a deleted biosynthetic pathway in auxotrophic *P. pastoris* mutants. After that, a number of auxotrophic markers have been developed such as *ARG4* (argininosuccinate lyase gene), *URA3* (orotidine 5'-phosphate decarboxylase) and *ADE1* (PR-amidoimidazolesuccinocarboximide synthase). Alternatively, drug resistance markers such as *Micromonospora rhodorangea* geneticin or *Streptoalloteichus hindustanus* zeocin resistance genes, *Aspergillus tererus* blasticidin S deaminase gene or *Sorangium cellulosum*acetyl-CoA carboxylase gene are also used to select recombinant clones. Among these markers, the zeocin resistant gene is the most widely used because of its small size, and its easy use in both *E. coli* and *P. pastoris* (Li *et al*., 2007; Macauley-Patrick *et al*., 2005; Cereghino *et al*., 2001).

3.6 Production strains

Mut^+ (methanol utilization plus phenotype) strains have the ability to grow normally on methanol and are the most widely used strains: they include wild-type strains NRRL-Y 11430 and X-33, and mutant strain GS115 (*his4*). Some deriving strains having deletion of one or both *AOX* gene may have expression level better than wild-type. Moreover, they require less methanol which is toxic and flammability reagent (Cereghino and Cregg, 2000). Mut^s (methanol utilization slow phenotype) strains, such as KM71, grow slowly on methanol because the *AOX1* gene is disrupted and replaced with the *S. cerevisiae ARG4* gene, so that the strain can only use the alternative weaker *AOX2* gene to utilize methanol. Finally, Mut^- (methanol utilization minus phenotype) strains such as MC100-3 cannot grow on methanol because both *AOX* genes are deleted. Since some sensitive recombinant proteins are degraded by protease, protease deficient host strains, such as the SMD1163, SMD1165 and

SMD1168 strains, have been developed to provide suitable environment for secreted proteins. However, they show a decreased viability and a slower growth rate, and are more difficult to transform (Li *et al*., 2007; Higgins and Cregg, 1998; Cereghino and Cregg, 2000).

4. *Yarrowia lipolytica* expression platform

4.1 Characteristics

Y. lipolytica, a non-pathogenic, dimorphic, and oleaginous yeast, is considered as GRAS (Generally Regarded As Safe) by FDA (Food and Drug Administration). It was early classified as *Candida lipolytica* in the deuteromycetous group until its sexual stage was discovered. After that, it was reclassified to *Endomycopsis lipolytica*, *Saccharomycopsis lipolytica*, and finally as *Y. lipolytica*. The name *Y. lipolytica* refers to its ability to degrade *n*-paraffins and oils. Therefore, it was usually found in protein or oil containing natural habitat such as cheese, yogurt, sewage, and poultry products (Bankar *et al*., 2009; Kerscher *et al*., 2002). *Y. lipolytica* is an obligatory aerobe growing at temperatures which do not exceed 32-34 °C, and uses glucose, polyalcohols (ethanol up to 3%), acetate (up to 1%), hydrophobic substrates (1-alkenes, polymethylated, chlorinated alkanes, fatty acids, oils) as carbon sources (Barth and Gaillardin, 1997; Gellissen *et al*., 2005). It can accumulate lipids to more than 50% of its dry cell weight, notably through the action of liposan, a complex of carbohydrates and proteins located on its cell surface and possessing emulsifying properties. Then, internalised lipids are broken down by lipid metabolism to meet needs for growth, or accumulate in an unchanged or modified form (Beopoulos *et al*., 2009). The interesting characteristic of *Y. lipolytica* is its ability to secrete large amounts of metabolites, organic acids (2-ketoglutaric acid and citric acid), single-cell protein, single-cell oil, aroma, and extracellular proteins (up to 1- 2 g/L of lipases, proteases, esterases, and phosphatases). Moreover, it is also used as model organism for studying dimorphism, salt tolerance, recombinant protein production, and lipid accumulation (Bankar *et al*., 2009). Recently, *Y. lipolytica* genome was sequenced in the framework of the "Génolevures" project at GENOSCOPE, Centre National de Sequençage, Evry, France. The size of its genome was shown to be 20.5 Mb (twice as large as the genome of *S. cerevisiae*), organized

into 6 chromosomes. Moreover, results also revealed that the organism is only distantly related to the conventional yeast *S. cerevisiae* and other hemiascomycetous yeast, but shares common properties with filamentous fungi because of its high GC content (50% compared to 38% for other ascomycota), high frequency of introns (13% compared to 4.8% for *S. cerevisiae*), high number of protein-coding sequences (1.15-fold higher than in *S. cerevisiae*), low level of gene similarity with other yeasts (50-60% at amino acid level), unusual structure of its rDNA genes, and unusual types of transposable elements (Gellissen *et al*., 2005; Kerscher *et al*., 2002; Casaregola *et al*., 2000). At present, *Y. lipolytica* has become a reliable and popular system for heterologous protein expression in both academic purposes and commercial applications because of high transformation efficiency, low glycosylation level, plasmid stability, high secretion efficiency, good product yield, performance reproducibility, and facilitated high-throughput screening (Müller *et al*., 1998; Bordes *et al*., 2007).

4.2 Expression vectors

Most vectors used in *Y. lipolytica* expression system are integrative shuttle-vectors like those used for *P. pastoris*. Integrative vectors used to transport foreign genes function by simple cross-over at homologous site such as *LEU2*, *URA3* (see paragraph 4.5) and are introduced into host cell using the lithium acetate transformation method (Le Dall *et al*., 1994). Linearization of the plasmid within the homology region enhances considerably integration efficiency, which can reach up to 10^6 transformants/µg of DNA and results in a stable integration as the gene is retained in the genome after more than 100 generations under non-selective conditions (Madzak *et al*., 2004; Xuan *et al*., 1988). Moreover, this linearization strategy can permit to avoid integrating bacterial DNA coming from the shuttle vector in the yeast and is therefore designated as auto-cloning strategy (Madzak *et al*., 2004). These vectors contain zeta regions (long terminal repeats of *Y. lipolytica* retrotransposon Ylt1) which promote non-homologous integration. They also offer the possibility of multiple integration events, hence enhancing expression level. Although the transformation efficiency of zeta based vectors is lower than classical integrative vectors, especially those using the defective *ura3d4* marker (10^2 transformants/µg

DNA), they are still interesting to use because of the stability of the integrants (Pignède *et al*., 2000; Madzak *et al*., 1999; Le Dall *et al*., 1994).

4.3 Promoters

The promoter from alkaline extracellular protease *XPR2* gene was the first used for recombinant gene expression in *Y. lipolytica*. It is a strong inducible promoter, and used to produce many proteins. Nonetheless, it only works at pH over 6 and requires large amount of peptone for full induction. This complicated regulation lead to the development of novel promoters, such as the hydrid promoter *hp4d*, constructed by fusing four repeats of upstream activating sequences (UAS1) of the *LEU2* promoter. It is a semi-constitutive and growth-phase-dependent promoter which is poorly affected by environmental conditions (pH, carbon and nitrogen sources, peptones) (Blanchin-Roland *et al*., 1994; Madzak, *et al*., 2004). Translation elongation factor-1α (*TEF*) and ribosomal protein S7 (*RPS7*) were reported as alternative promoters to drive expression in *Y. lipolytica*. They are strong constitutive promoters, leading to increased numbers of positive clones after transformation and producing sufficient amounts of protein for characterization (Müller *et al*., 1998). In addition, inducible promoters for *Y. lipolytica* expression system were also developed with the view to produce potentially toxic proteins. These promoters usually come from genes involved in the lipid accumulation pathway, such as those for isocitrate lyase (*ICL*), acyl-CoA oxidase (*POX*), 3-oxo-acyl-CoA thiolase (*POT*) which are induced by fatty acids and alkanes (Madzak, *et al*., 2004).

4.4 Secretion signals

The *XPR* prepro sequence is the most commonly used signal peptide in *Y. lipolytica*. In *Y. lipolytica*, this sequence facilitates the translocation of an extracellular alkaline protease and is removed in the late golgi apparatus. Alternative signal sequences also exist such as the *XPR2* pre region alone, *LIP2* prepro-peptide, and a hybrid sequence between *XPR2* and *LIP2* prepro regions (Müller *et al*., 1998; Madzak *et al*., 2004).

4.5 Selection markers

Y. lipolytica is resistant to most commonly used drugs with the exception of a few antibiotics such as bleomycin/phleomycin (Gaillardin and Ribet, 1987) and hygromycin B (Otero and Gaillardin, 1996). Therefore, nutritional markers such as *URA3* (uracile) and *LEU2* (leucine) became popular auxotrophic markers, with some studies using *LYS5* (α-aminoadipate reductase), *ADE1* (phosphoribosyl-aminoimidazole-succinocarboxamide synthetase), and *GUT2* (glycerol 3-phosphate dehydrogenase) as alternative nutritional markers. *SUC2* (invertase gene) could also be used as a selection marker as it enables the yeast to use sucrose as the sole carbon source. This marker permitted to develop a new strain of *Y. lipolytica*, Po1d, capable of growing on sucrose and molasses (Barth and Gaillardin, 1997; Madzak *et al*., 2000).

4.6 Production strains

The genetic background of a host strain is very important factor for heterologous expression. Nowadays, Po1d is the most widely used *Y. lipolytica* strain because of its high level of secretion, its ability to use sucrose and molasses as a sole carbon source, and because the gene of its main alkaline extracellular protease has been deleted (Nicaud *et al*., 1989). Derivative strains of Po1d such as Po1f, Po1g and Po1h were constructed to eliminate all sources of remaining proteases or to enable non-homologous integration from auto-cloning expression vectors (Madzak *et al*., 2004). Recently, the *Y. lipolytica* JMY1212 strain was developed for directed evolution and high-throughput screening projects. The strain contains a zeta docking platform located after the *LEU2* gene, which enables a targeted integration of the gene of interest in the *Yarrowia* genome, and therefore greatly increase result reproducibility for screening process (Bordes *et al*., 2007).

PART IV Enzyme engineering

Enzymes and proteins in general have naturally evolved over billion of years to work efficiently in specific conditions (pH and temperature) and to adapt to various substrates or substrate concentrations. Nowadays, industrial processes usually require conditions in which most native enzymes cannot work efficiently, e.g. extreme pH, temperature or pressure, presence of organic solvents or inhibitors. Although the industrial process can be adapted to provide suitable conditions for the action of enzymes, finding catalysts able to work in these extreme conditions is more relevant. For example, working with highly thermostable enzymes provides a greater flexibility in the process configuration and speed, leads to energy savings, often increases product solubility but decreases the reaction viscosity, and prevents microbial contamination (Dalby, 2007; Eijsink *et al*., 2005). However, at elevated temperatures, the 3D-structure of most enzymes usually unfolds causing inactivation, so they cannot work for the desired tasks. Thus, enzymes need to evolve to avoid these obstacles. As a consequence, genetic techniques have been developed to mimic and accelerate natural evolution, namely rational design and directed evolution (Eijsink *et al*., 2005; Maki *et al*., 2009). The details of both strategies are easily summarized in Figure II-20.

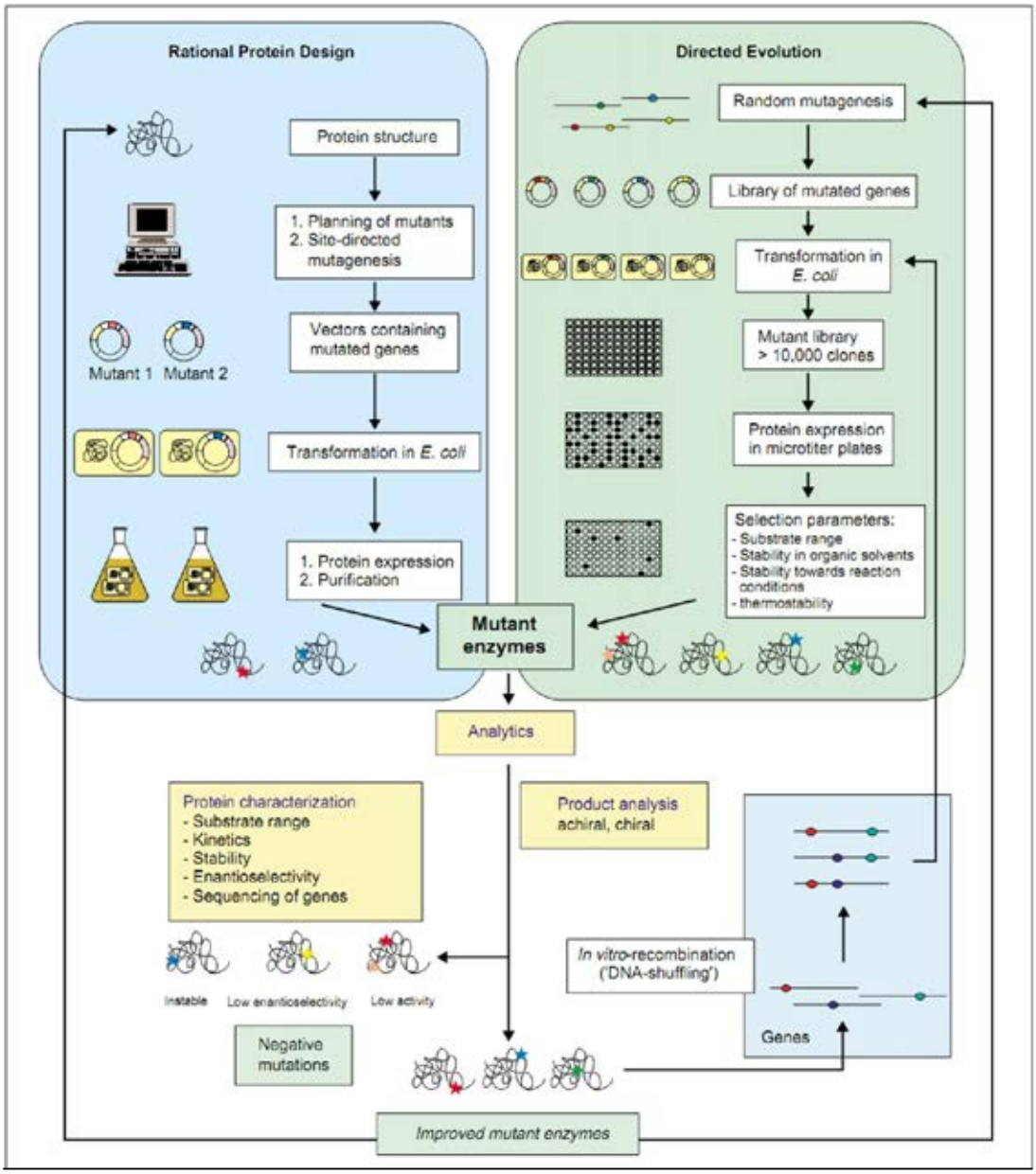


Figure II-20. Comparison of rational protein design and directed evolution strategies (Bornscheuer and Pohl, 2001).

1. Rational design

Rational design was the first approach used to engineer proteins after the development of recombinant DNA technologies. This approach relies on a detailed knowledge of the relationships between the structure and the function of enzymes. It involves the mutagenesis of defined residues or regions, supposedly participating to the studied properties. Site-directed mutagenesis is a widely used method to analyse

the role of individual amino acids on certain protein features (Zhang *et al*., 2006; Maki *et al*., 2009). This technique has been applied to improve specific properties of cellulases. For example, site-directed mutagenesis was used to identify the catalytic residues of *T. reesei* Cel12A: residue replacement of glutamines at position 116 and 200 reduced the specific activity of more than 98% (Okada *et al*., 2000). In addition, a thermostable and pH stable xylanase II mutant from *T. reesei* was also constructed via the engineering a disulfide bridge between residues 2 and 28, both of which have been mutated from a threonine to cysteine residue (Fenel *et al*., 2004). Nevertheless, the small number of available 3D protein structures and limited knowledge of structure-function relationships of many poteins restrain the success of using rational design.

2. Directed evolution

Direct evolution is a powerful method to generate protein diversity, and is based on the introduction of random mutations into a gene sequence to create libraries of mutant enzymes, screened under evolutionary pressure using high-throughput methods to select for the best protein variants. These steps are repeated until satisfying mutants are obtained. Alternatively, positive mutations obtained in several variants can be recombined by gene shuffling. The critical step remains screening the libraries, as it is laborious and costly, and needs to be specific for the property which is to be improved (Dalby, 2007; Eijsink *et al*., 2005). Directed evolution approaches are divided into three categories to generate genetic diversity of enzymes: random point mutation, randomization at target areas, and recombination of homologous genes or variants.

2.1 Error-prone polymerase chain reaction

Error-prone PCR is the most commonly used method to generate mutations randomly in one single gene. Mutations are introduced by using a low-fidelity DNA polymerase, whose malfunction can be increased by the presence of Mn^{2+} or Mg^{2+}, nucleotide analogues, or biased dNTP concentrations. These cause misincorporation of incorrect nucleotides and cause the random mutation of the parental gene. The mutated PCR product is then cloned into an expression vector and screen for desired properties after transformation (Sen *et al*., 2007). A critical point to generate useful libraries from error-prone PCR method concerns the mutation

frequency. If the mutation frequency is too low, the library will mainly be constituted by wild-type parents whereas if the mutation frequency is too high, it will produce too many inactive clones. Normally, it is recommended that the mutation frequency should be between 1.5 and 5 mutations per gene (Miyazaki *et al*., 2000). Error-prone PCR was adopted to improve the stability and the specific activity of *T. reesei* EGIII, cloned in *E. coli*. After screening of 9000 mutants, the best mutant protein showed a 130-fold production increase, significant broader pH stability (4.4–8.8), and 1.4-fold increased k_{cat} compared to wild-type EGIII (Nakazawa *et al*., 2009).

2.2 Site saturation mutagenesis

Saturation mutagenesis is a semi-rational approach substituting a codon encoding wild-type amino acid in a protein with all other 19 amino acids. It is a powerful method to characterize relationships between structure and function based on previous knowledge about the function of the target region. This procedure is simple, fast, efficient, and reduce the libraries size to only 93 clones (O'Donohue and Kneale, 1994; Zheng *et al*., 2004; Steffens and Williams, 2007). Previously, Qin and coworkers applied this approach with error-prone PCR and DNA shuffling to improve pH stability of *T. reesei* EGII. After screening of 60,000 transformant clones from the low mutation frequency epPCR library, and 80,000 transformant clones from the high mutation frequency, with two round of DNA shuffling, the best mutant can shift optimal pH to basic side from 4.8 to 5.8 and increase catalytic efficiency 4.5 fold compared with the wild-type enzyme (Qin *et al*., 2008a).

2.3 DNA shuffling

DNA shuffling is an *in vitro* recombination process which allows to recombine beneficial mutations from many positive variants into one or more genes, thus eliminating deleterious and even neutral mutation., where 2 or more of genes having high sequence similarity (usually more than 70% but it can also success to do in genes having low sequence similarity). After DNaseI digestion of the parental genes, small purified fragments are reassembled in a PCR reaction without primers. Full length PCR products are then re-amplified in a PCR reaction with primers annealing at the extremities of the parental genes. After cloning and transformation, a new library is created and screened for optimized mutants (Figure II-21) (Eijsink *et al*.,

2005; Sen *et al*., 2007). DNA shuffling was used by Kim and colleagues to increase the activity of *Bacillus subtilis* endoglucanase up to 5 times using a screening method based on CM-cellulose degradation on solid medium stained with Congo red (Kim *et al*., 2000).

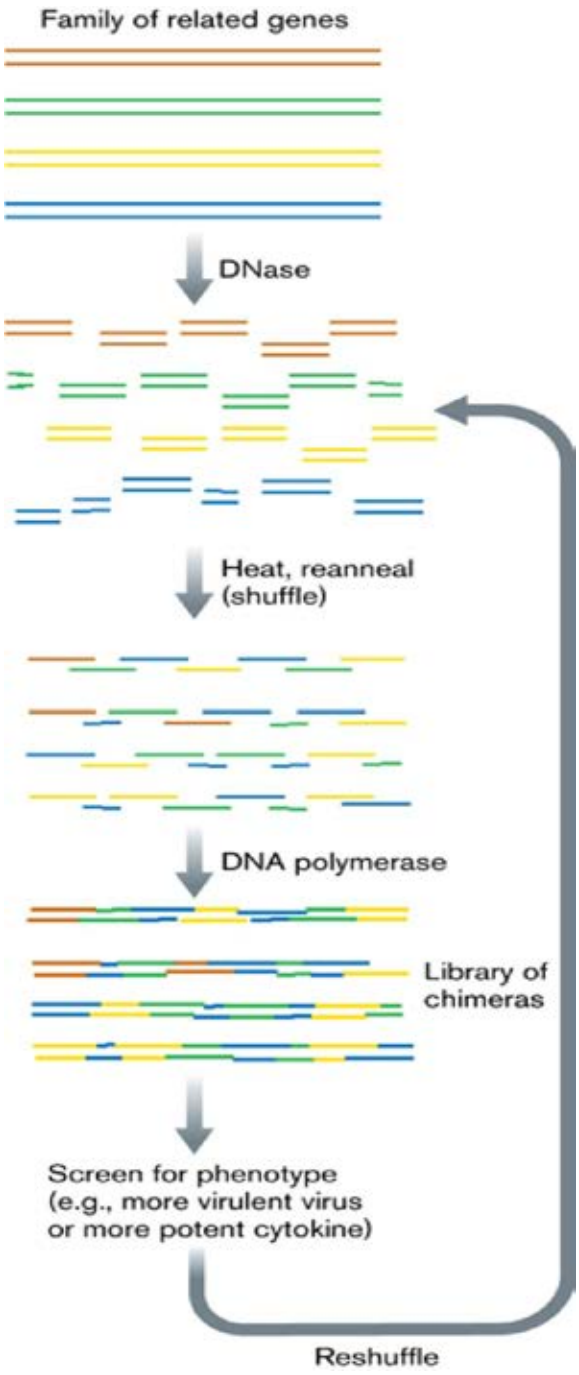


Figure II-21. DNA shuffling methodology (W.W. Norton and Company Inc., 2010 : online).

CHAPTER III

MATERIALS AND METHODS

PART I Cloning and expression of *Aspergillus oryzae* glucan 1,3-beta-glucosidase A in *Pichia pastoris*

1.1 Organism and culture conditions

The bacterial strain *Escherichia coli* DH5α (Gibco) was grown at 37 °C in Luria-Bertani broth (LB) supplemented with 25 μg zeocin/ml (Invitrogen). *A. oryzae* ATCC 11489, obtained from Microbiological Resource Center, Thailand Institute of Scientific and Technological Research (TISTR), was cultivated at 28 °C, 120 rpm for 14 days in a shaker flask using minimal medium containing 0.5% w/v quercetin (Sigma) as the inducer (Riou *et al*., 1998). *P. pastoris* X-33$^{(mut+)}$ cells (Invitrogen) was cultured at 30 °C, 200 rpm in yeast extract-peptone-dextrose (YPD) liquid medium and transformants were selected on YPDS agar (YPD with 1 M sorbitol) medium containing 100-2000 μg zeocin/ml.

1.2 RNA and DNA manipulations

The total RNA from 14-day cultured *A. oryzae* mycelia was isolated using the MasterPure™ Complete DNA and RNA Purification Kit (Epicenter), following the supplier's instructions. The RNA was purified using a RNeasy Mini Kit (Qiagen) and cDNA synthesis was performed using the One-Step RT-PCR Kit (Qiagen). The product of reverse transcription (total cDNA) was used as a template for amplifying the β-glucosidase gene. The primers were designed according to mature peptide sequences of the *A. oryzae exgA* (GenBank accession code AJ566365), but with additional 5' sequences (italic) containing the *Eco*R1 or *Not*1 consensus sites (underlined) to facilitate in-frame directional cloning, as follows: ExgA-F (5'-*G<u>GAATTC</u>*AGTCGCTTGGACCCTCGCG-3') and ExgA-R (5'-*TACCTTTT<u>GCGGCCGC</u>*GCACTGCTTAGGGAACTCCCG-3'). Reaction mixtures contained 1x Phusion HF buffer, 200 μM each dNTPs, 0.5 μM each of flanking primers, 10 ng of template and 0.02 U of Phusion DNA polymerase (Finnzyme) in a

total volume of 50 µl. Thermal cycling parameters were 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 59 °C for 30 s and 72 °C for 1 min, and then finally at 72 °C for 10 min. The amplicon was cut with *EcoR*I and *Not*I and the resulting *EcoR*I /*Not*I fragment was cloned into pPICZαA or pGAPZαA (also cut with the same enzymes) to yield pPICZαA_*exgA* and pGAPZαA_*exgA*, respectively, that were then separately transformed into *E. coli* DH5α. The selected positive clones were then commercially sequenced using the α-factor sequencing and 3'*AOX* primers (Invitrogen) by PACIFIC SCIENCE (Bangkok, Thailand).

1.3 Yeast transformation and selection

The transformation of *P. pastoris* X-33$^{(mut+)}$ with each plasmid construct was achieved using the lithium acetate and dithiothreitol method, as previously described (Wu and Letchworth, 2004). About 10 µg of *Pme*I linearized plasmid pPICZαA_*exgA* or *Avr*II linearized plasmid pGAPZαA_*exgA* combined with 25 µg of salmon sperm DNA (Invitrogen), were used to transform a 90 µl aliquot of competent *P. pastoris* X-33$^{(mut+)}$ cells using Gene Pulser Xcell (Bio-Rad), and immediately recovered in 1 ml of 1 M sorbitol for 2 h at 30 °C. Transformants were selected on YPDS agar containing 100-2000 µg zeocin/ml. Genomic DNA from the selected transformants was extracted by the E.Z.N.A.® Yeast DNA Kit (Omega), amplified by using α-factor sequencing and 3'*AOX* primers, and sequenced to verify the expression cassettes integrated into the yeast genome by homologous recombination.

1.4 Enzyme production and activity assay

The expression of *A. oryzae* rExgA in *P. pastoris* was achieved according to the supplier's instructions (Invitrogen). The constitutive expression (under *GAP*) was cultured in YPD medium whereas inducible expression (under *AOX*1) was employed by using buffered glycerol complex medium (BMGY) and buffered minimal methanol medium (BMMY), to the latter of which was added methanol daily to a final concentration between 0.5% and 3% (v/v) to study the effect of the methanol concentration upon the recombinant protein expression level. All enzyme expressions were performed in baffled shaker flasks at 30 °C, 200 rpm, for 5 days and aliquots were taken for analysis at periodic intervals.

The rExgA enzyme activity was measured using pNPβG as the substrate, as previously described (Araujo *et al*., 1983). The reaction mixture contained 0.5 ml of 2 mM pNPβG in 50 mM acetate buffer (pH 5.0) and an appropriate dilution of the rExgA enzyme solution. The reaction was conducted at 35 °C for 30 min and then the reaction was stopped by adding 1 ml of 1 M Na_2CO_3. The color that developed as a result of *p*-nitrophenol liberation was measured spectrophotometrically by monitoring the absorbance at 400 nm. One unit (U) of enzyme activity was defined as the amount of enzyme that released 1 μmole of *p*-nitrophenol ml^{-1} min^{-1}. All enzymatic activity measurements were performed in triplicate unless otherwise stated.

1.5 Enzyme purification

Yeast cells from liquid cultures (200 ml) were sedimented by centrifugation at 4500 x g, 4 °C for 10 min. The supernatant was concentrated by ultrafiltration through a Macrosep polyethersulfone (10 kDa cutoff) membrane (PALL). Concentrated proteins were applied to 3 ml of a TALON Metal Affinity Resin (Clontech). Washing of the resin and elution of the proteins was achieved according to the manufacturer's instructions, using imidazole buffer. Protein concentration of the crude and purified enzyme was spectrophotometrically measured at an absorbance of 595 nm using Bio-Rad Protein Assay (Bio-Rad) and bovine serum albumin as the standard.

1.6 SDS-PAGE analysis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli (1970) using 5% and 12% (v/v) polyacrylamide stacking and resolving gels, respectively, running in Tris-glycine buffer. Samples were mixed with equivalent 2X sample buffer (Invitrogen). Precision Plus Protein Standards (Bio-Rad) ranging from 10 to 250 kDa were used as the reference proteins. After running, the gel was stained using a PageBlue™ Protein Staining Solution (Fermentas). The molecular masses were estimated from the migration distance and logarithm of molecular weights.

1.7 Enzyme characterization

The evaluation of the optimum pH and temperature for the enriched rExgA enzyme preparation were performed as described above, using 100 mM

citrate/phosphate buffer pH 3.0-8.0 at 35 °C for evaluation of the optimum pH and at 20-60 °C using the buffer at pH 5.0 for determining the optimum temperature. The pH stability was examined by pre-incubating the enzyme in the absence of a substrate at 20 °C for 30 min in the above buffer (pH 3.0-8.0) before determining the activity. For thermostability determination, the enzymes were pre-incubated at pH 5.0 at several defined temperatures for 30 min. The enzyme kinetics of rExgA was determined using 0.06-4 mM pNPβG as the substrate. The inhibition of glucose on the enzyme activity was measured by adding 125-500 mM glucose into the reaction mixture with pNPβG as the substrate. The kinetic constants, K_M and V_{max} for the enzyme were carried out using the curve fitting procedure of SigmaPlot (SPSS, 2000). The inhibition constant (K_i) was obtained at the intersection of the line of a Dixon plot. Assays were always performed in triplicate and the mean values with their standard deviation (SD) are presented.

PART II Comparison of the heterologous expression of *Trichoderma reesei* endoglucanase II and cellobiohydrolase II in the yeasts *Pichia pastoris* and *Yarrowia lipolytica*

2.1 Chemicals

Unless otherwise stated, all chemicals were analytical grade and purchased from Sigma-Aldrich (St. Louis, MO, USA). Specific enzymes for DNA manipulations were purchased from New England Biolabs (Ipswich, MA, USA). Oligonucleotides were synthesized by Eurogentec (Angers, France) and components for yeast culture media were supplied by Difco (Difco, Paris, France), except kanamycin which was purchased from Euromedex (Souffelweyersheim, France), and zeocin and yeast nitrogen base (YNB, without amino acids and with ammonium sulphate) were purchased from Invitrogen (San Diego, CA, USA).

2.2 Strains, plasmids and culture conditions

The relevant genotypic information describing the various microbial strains used in this study is summarized in Table III-1. Likewise, succinct descriptions of all plasmids used or created in this study, are listed in Table III-2. The bacterial strain *E.*

coli DH5α (purchased from Gibco BRL, Rockville, MD, USA) was grown at 37 °C in Luria-Bertani medium, supplemented with kanamycin (50 µg/ml) or zeocin (25 µg/ml) depending on the plasmid construct being used. *T. reesei* QM 9414, obtained from Microbiological Resource Center, Thailand Institute of Scientific and Technological Research (TISTR), was cultivated at 28 °C for 7 days in shaker flask using Mandels medium with 15 mM cellobiose as the carbon source (Mandels and Weber, 1969). *P. pastoris* X-33 (purchased from Invitrogen, San Diego, CA, USA) was cultured at 30 °C, 200 rpm in baffled shaker flasks containing YPD (1% w/v yeast extract, 2% w/v peptone, 2% w/v glucose). For the selection of *P. pastoris* transformants in Petri dishes, solid YPDS agar medium (YPD plus 1.5% w/v agar and 1 M sorbitol) containing 100-2000 µg/ml zeocin was used and plates were incubated at 30 °C for 3 days. For heterologous protein expression in *P. pastoris*, a two stage strategy was employed. First, yeast biomass was generated in shaker flask cultures at 30 °C, 200 rpm growing in BMGY liquid medium (1% w/v yeast extract, 2% w/v peptone, 100 mM potassium phosphate buffer, pH 6.0, 1.34% w/v YNB, 4 x 10^{-5} % w/v biotin, 1% v/v glycerol) and suspended in sterile liquid BMMY medium (as per BMGY except with 0.5% or 3% v/v methanol instead of the glycerol). Shaking in baffled flasks at 30 °C, 200 rpm was pursued for 24 h before recovering the culture supernatant at 120 h for protein purification.

Table III-1. Bacterial and yeast strains used in this study

Strain[1]	Relevant features[2]	Source or reference
Bacterial strain		
E. coli DH5α	*supE*44 Δ*lac*U169 (*Φ*80*lacZ*ΔM15) *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1	(Sambrook *et al*., 1989)
P. pastoris yeast strains		
X-33	Wild type	Invitrogen
PnEGII	X-33 transformed with pPICZ_natEGII	This work
PfEGII	X-33 transformed with pPICZαA_fusEGII	This work
PnCBHII	X-33 transformed with pPICZ_natCBHII	This work
PfCBHII	X-33 transformed with pPICZαA_fusCBHII	This work
Y. lipolytica yeast strains		
Zeta	Ura-, Leu+, ΔAEP, Suc+	(Nicaud *et al*., 2002)
Po1d	Ura-, Leu-, ΔAEP, Suc+	(Le Dall *et al*., 1994)
YZ-nEGII (POX)	Zeta strain transformed with JMP62UraPOX_natEGII	This work
YZ-fEGII (POX)	Zeta strain transformed with JMP61UraPOX_fusEGII	This work
YZ-nEGII (TEF)	Zeta transformed with JMP62UraTEF_natEGII	This work
YZ-fEGII (TEF)	Zeta strain transformed with JMP62UraTEF_fusEGII	This work
YP-nEGII (POX)	Po1d strain transformed with JMP62UraPOX_natEGII	This work

[1]P = *P. pastoris*; Y = *Y. lipolytica*; n = native; f = fusion; Z = zeta strain; P = Po1d strain

[2]Details of the transforming plasmid are given in Table III-2

Table III-1. Bacterial and yeast strains used in this study (Continued)

Strain	Relevant features[1]	Source or reference
YP-fEGII (POX)	Po1d strain transformed with JMP61UraPOX_fusEGII	This work
YP-nEGII (TEF)	Po1d strain transformed with JMP62UraTEF_natEGII	This work
YP-fEGII (TEF)	Po1d strain transformed with JMP62UraTEF_fusEGII	This work
YP-fEGII (double POX)	Po1d strain transformed with JMP61UraPOX_fusEGII and JMP61LeuPOX_fusEGII	This work
YP-fEGII (double TEF)	Po1d strain transformed with JMP62UraTEF_fusEGII and JMP62LeuTEF_fusEGII	This work
YZ-nCBHII (POX)	Zeta strain transformed with JMP62UraPOX_natCBHII	This work
YZ-fCBHII (POX)	Zeta strain transformed with JMP61UraPOX_fusCBHII	This work
YZ-nCBHII (TEF)	Zeta strain transformed with JMP62UraTEF_natCBHII	This work
YZ-fCBHII (TEF)	Zeta strain transformed with JMP62UraTEF_fusCBHII	This work
YP-nCBHII (POX)	Po1d strain transformed with JMP62UraPOX_natCBHII	This work
YP-fCBHII (POX)	Po1d strain transformed with JMP61UraPOX_fusCBHII	This work
YP-nCBHII (TEF)	Po1d strain transformed with JMP62UraTEF_natCBHII	This work
YP-fCBHII (TEF)	Po1d strain transformed with JMP62UraTEF_fusCBHII	This work
YP-fCBHII (double POX)	Po1d strain transformed with JMP61UraPOX_fusCBHII and JMP61LeuPOX_fusCBHII	This work
YP-fCBHII (double TEF)	Po1d strain transformed with JMP62UraTEF_fusCBHII and JMP62LeuTEF_fusCBHII	This work

Table III-2. Descriptions of all plasmids used or created in this study

Construction plasmid	Description			Primer pair (forward/reverse)[1]	Yeast host[2]
	Gene	Promoter	Signal sequence		
pPICZA_natEGII	EGII	AOX1 / methanol inducible	native	natP_EGII-F / natP_EGII-R	*PP*
pPICZαA_fusEGII	EGII	AOX1 / methanol inducible	Sc-aSP[3]	fusP_EGII-F / fusP_EGII-R	*PP*
pPICZA_natCBHII	CBHII	AOX1 / methanol inducible	native	natP_CBHII-F / natP_CBHII-R	*PP*
pPICZαA_fusCBHII	CBHII	AOX1 / methanol inducible	Sc-aSP	fusP_CBHII-F / fusP_CBHII-R	*PP*
JMP62UraPOX_natEGII JMP62LeuPOX_natEGII	EGII	POX2 / oleic acid inducible	native	natY_EGII-F / comY_EGII-R	*YL*
JMP61UraPOX_fusEGII JMP61LeuPOX_fusEGII	EGII	POX2 / oleic acid inducible	preproLip2	fusY_EGII-F / comY_EGII-R	*YL*
JMP62UraPOX_natCBHII JMP62LeuPOX_natCBHII	CBHII	POX2 / oleic acid inducible	native	natY_CBHII-F / comY_CBHII-R	*YL*
JMP61UraPOX_fusCBHII JMP61LeuPOX_fusCBHII	CBHII	POX2 / oleic acid inducible	preproLip2	fusY_CBHII-F / comY_CBHII-R	*YL*
JMP62UraTEF_natEGII JMP62LeuTEF_natEGII	EGII	TEF / constitutive	native	natY_EGII-F / comY_EGII-R	*YL*
JMP62UraTEF_fusEGII JMP62LeuTEF_fusEGII	EGII	TEF / constitutive	preproLip2	BamY_Lip2 / comY_EGII-R	*YL*
JMP62UraTEF_natCBHII JMP62LeuTEF_natCBHII	CBHII	TEF / constitutive	native	natY_CBHII-F / comY_CBHII-R	*YL*
JMP62UraTEF_fusCBHII JMP62LeuTEF_fusCBHII	CBHII	TEF / constitutive	preproLip2	BamY_Lip2 / comY_CBHII-R	*YL*

[1] For primer sequences see Table III-3

[2] *PP* = *P. pastoris* and *YL* = *Y. lipolytica*

[3] Sc-aSP = *S. cerevisiae* alpha factor signal sequence

The growth media and culture conditions for all *Y. lipolytica* strains have been previously described (Bordes *et al*., 2007). Briefly, parental *Y. lipolytica* strains were cultured in YPD at 28 °C, 120 rpm in baffled shaker flasks or in Petri dishes containing YPD supplemented with 1.5 % (w/v) agar. To select for complementation (Ura^+ and Leu^+) of auxotrophic mutant strains, transformants were grown on solid YNB medium (0.17% w/v YNB, 1% glucose w/v, 0.5% w/v ammonium chloride, 0.2% w/v casamino acids and 100 mM sodium-potassium phosphate buffer, pH 6.8), supplemented with the appropriate auxotrophic requirements. For heterologous protein expression, recombinant *Y. lipolytica* strains were grown in YT medium (1% w/v yeast extract, 2% w/v tryptone and 100 mM phosphate buffer, pH 6.8), supplemented with either 2% w/v glucose (YTD) or with 2% w/v oleic acid (YTO), depending on the nature of the promoter being used.

2.3 RNA and DNA manipulations

The total RNA from 7-day cultured *T. reesei* mycelia was isolated using the MasterPure™ Complete DNA and RNA Purification Kit (Epicenter Biotechnologies, Madison, WI), following the supplier's instructions. The RNA was purified using a RNeasy Mini Kit (Qiagen, Valencia, California) and cDNA synthesis was performed using the One-Step RT-PCR Kit (Qiagen, Valencia, California) and gene-specific primers, designed using the reported cDNA sequences of the *T. reesei egl*2 and *cbh*2 (GenBank accession codes DQ178347 and M55080, respectively) (Table III-3).

The *Eco*RI /*Not*I fragments of the *egl*2 and *cbh*2 amplified products were cloned into pPICZ to yield pPICZ_natEGII and pPICZ_natCBHII, respectively. To clone the *egl*2 and *cbh*2 genes in other yeast expression vectors, the polymerase chain reaction (PCR) was used in combination with specifically designed primers (Table III-3). Each DNA fragment was amplified by PCR using the corresponding primer pairs with pPICZ_natEGII and pPICZ_natCBHII as the templates. Reaction mixtures contained 1x Phusion HF buffer, 200 µM each dNTPs, 0.5 µM each of flanking primers, 10 ng of template and 0.02 U of Phusion DNA polymerase in a total volume of 50 µl. Thermal cycling parameters were 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 59 °C for 30 s and 72 °C for 1 min, and then finally at 72 °C for 10 min. All DNA sequence reactions were performed using Sanger technology and a 9X ABI 3730XL (GATC BIOTECH, Germany).

Table III-3. PCR primers used in this study

Primer name	Sequence 5'-3' *	Restriction site
natP_EGII-F	TTT **GAA TTC** ATG AAC AAG TCC GTG GCT CC	*Eco*RI
natP_EGII-R	TTT TCC TTT A**GC GGC CGC** TGT ACT TTC TTG CGA GAC ACG AGC	*Not*I
fusP_EGII-F	CCC **GAA TTC** GCA CAG CAG ACT GTC TGG	*Eco*RI
fusP_EGII-R	TTT TCC TTT A**GC GGC CGC** CTT TCT TGC GAG ACA CGA GC	*Not*I
natP_CBHII-F	TTT **GAA TTC** ATG ATT GTC GGC ATT CTC ACC AC	*Eco*RI
natP_CBHII-R	TTA AGC TTT A**GC GGC CGC** TAT ACA GGA ACG ATG GGT TTG CGT	*Not*I
fusP_CBHII-F	CCG **GAA TTC** GCC CAG GCT TGC TCA AGC	*Eco*RI
fusP_CBHII-R	TTA AGC TTT A**GC GGC CGC** CAG GAA CGA TGG GTT TGC GT	*Not*I
natY_EGII-F	CGT **GGA TCC** ATG AAC AAG TCC GTG GCT CC	*Bam*HI
comY_EGII-R	GGG **CCT AGG** CTA CTT TCT TGC GAG ACA CGA GC	*Avr*II
fusY_EGII-F	GAG GCC GCA GTT CTC CAG AAG CGA GCA CAG CAG ACT GTC TGG	None
natY_CBHII-F	CCT **GGA TCC** ATG ATT GTC GGC ATT CTC ACC AC	*Bam*HI
comY_CBHII-R	CGA **CCT AGG** TTA CAG GAA CGA TGG GTT TGC GT	*Avr*II
fusY_CBHII-F	GAG GCC GCA GTT CTC CAG AAG CGA GCC CAG GCT TGC TCA AGC GTC	None
Lip2start	GCC ATG **AAG CTT** TCC ACC ATC CTT TTC ACA GCC TGC GCT AC	*Hind*III
Lip2prorev	TCG CTT CTG GAG AAC TGC GGC CTC AGA AG	None
BamY_Lip2	CG**G GAT CC**A TGA AGC TTT CCA CCA TCC TTT TC	*Bam*HI

*Relevant restriction sites are underlined.

2.4 Yeast transformation and selection

The transformation of *P. pastoris* X-33 with each plasmid construct was achieved using the lithium acetate and dithiothreitol method, as previously described (Wu and Letchworth, 2004). Briefly, a fresh colony of *P. pastoris* X-33 was used to inoculate and grow in 200 ml of sterile YPD at 30 °C with shaking at 200 rpm until an OD_{600} of 1.3 to 1.5 was attained. At this point, cells were harvested by centrifugation at 2000 x *g*, 4 °C for 10 min and suspended in 50 ml of YPD containing 160 mM HEPES (pH 8.0), 20 mM dithiothreitol and 100 mM lithium acetate and held at 30 °C for 15 min. Cells were washed twice with 150 ml of water, using a centrifugation step at 2000 x *g*, 4 °C for 10 min after each wash, and finally suspended in 0.4 ml of 1 M sorbitol. Plasmids linearized using *Pme*I (about 10 µg combined with 25 µg of salmon sperm DNA (Invitrogen, San Diego, CA, USA) were used to transform into 90 µl of competent *P. pastoris* X-33 cells using an electroporator (Model 2510, Eppendorf), at 1500 V and immediately recovered in 1 ml of 1 M sorbitol for 2 h at 30 °C. Transformants were selected on YPDS agar containing 100-2000 µg/ml zeocin.

To transform *Y. lipolytica* strains, the lithium acetate method previously described (Le Dall *et al*., 1994) was used. Briefly, the cells were grown at 28 °C in 20 ml of YPD containing 50 mM citrate buffer (pH 4) until a cell density of between 8 x 10^7 to 1.5 x 10^8 cells/ml was attained. Cells were collected by centrifugation at 1160 x *g*, 4 °C for 5 min and washed twice using 10 ml of TE (50 mM Tris.HCl, 1 mM EDTA, pH 8.0) and a centrifugation step after each wash (see previous conditions), and then treated with 20 ml of 100 mM lithium acetate at 28 °C for 1 h without shaking. The cells were then centrifuged at 510 x *g*, 4 °C for 2 min and suspended in 2 ml of 100 mM lithium acetate. Plasmid DNA linearized by *Not*I (approximately 500 ng combined with 25 µg of salmon sperm DNA) was incubated in 100 µl of *Y. lipolytica* competent cells at 28 °C for 15 min without shaking, and then 700 µl of 40% (w/v) PEG 4000 was added and incubated at 200 rpm, 28 °C for 1 h. After that, the cells were transformed by heat shock at 39 °C for 10 min and immediately recovered in 1.2 ml of 100 mM lithium acetate. In case of multi-copy expression, 10%

(v/v) DMSO was added before the heat shock treatment. Transformants were selected on YNB agar containing the appropriate nutritional complements.

2.5 Enzyme production and activity assay

The expression of the EGII and CBHII enzymes in *P. pastoris* was achieved according to the supplier's instructions (Invitrogen guideline manual version H). Briefly, expression was performed in baffled shaker flasks that were incubated at 30 °C, 200 rpm, for 5 days. To induce expression, methanol was added daily to a final concentration of either 0.5% or 3% (v/v), the two concentrations (low and high) being used to compare the effect of the methanol concentration upon the recombinant protein expression level. For the expression of recombinant enzymes in *Y. lipolytica*, a single colony of an appropriate transformant was used to inoculate 5 ml YTD medium in a baffled shaker flask. The culture was incubated for 16 h at 30 °C with shaking at 120 rpm before taking a 2 ml aliquot, which was used to inoculate 20 ml of fresh YTD (for expression using the *TEF* promoter) or YTO (for expression using the *POX* promoter).

The EG activity was measured as previously described (Ghose, 1987). The reaction mixture contained 0.5 ml of 2% (w/v) carboxymethyl cellulose (CMC) in 50 mM citrate buffer (pH 4.8) and 0.5 ml of diluted enzyme solution. The reaction was conducted at 50 °C for 30 min and then reducing sugars were quantified using the dinitrosalicylic acid (DNS) reagent. CBH activity was evaluated using phosphoric acid swollen cellulose (PASC) that was prepared from Avicel PH-101 as previously described (Wood, 1988). PASC at 1% (w/v) in 50 mM acetate buffer (pH 4.8) was incubated with the enzyme solution at 50 °C for 1 h and then centrifuged at 17 760 x *g* for 3 min to sediment the substrate. The level of soluble reducing sugars in the supernatant was quantified using the Nelson-Somogyi method (Nelson, 1944; Somogyi, 1952). One unit (U) of enzyme activity was defined as the amount of enzyme that released 1 µmole of reducing sugars per ml per min. All enzymatic activity measurements were performed in triplicate unless otherwise stated.

2.6 Enzyme purification

Yeast cells from liquid cultures (350 ml) were sedimented by centrifugation at 4500 x *g*, 4 °C for 10 min. The supernatant was concentrated by ultrafiltration using an Amicon polyethersulfone (10 kDa cutoff) membrane (Millipore Corporation, USA). For enzymes produced by *P. pastoris*, concentrated proteins were applied to 3 ml of a TALON Metal Affinity Resin (Clontech, Takara-Bio, Kyoto, Japan). Washing of the resin and elution of the proteins was achieved according to the manufacturer's instructions, using imidazole buffer. For enzymes produced by *Y. lipolytica*, the buffer containing the concentrated proteins was exchanged with 10 mM citrate buffer (pH 3.5) using the same type of ultrafiltration centrifugal column as above, before being loaded onto a 8 ml mono S column (Amersham, UK). After protein loading, the column was washed with 10 mM citrate buffer (pH 3.5) and the bound proteins were eluted with a linear salt gradient of 0 to 1 M NaCl in the same buffer at a flow rate of 5 ml/min. Fractions containing the target protein were pooled and monitored by SDS-PAGE and enzyme activity measurement. Protein concentration of the crude and purified enzyme was spectrophotometrically measured at 280 nm using molar extinction coefficients of 78 000 and 92 000 cm^{-1} M^{-1} for EGII and CBHII, respectively (Kipper *et al*., 2005).

2.7 SDS-PAGE, Western blot analysis, and zymogram activity staining

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using NuPAGE 10% Bis-Tris pre-cast gels and MOPS/SDS running buffer (Invitrogen, San Diego, CA, USA). Precision Plus Protein Standards (Bio-Rad, Hercules, CA) ranging from 10 to 250 kDa were used as the reference proteins. After running, the gel was stained using a Colloidal Blue Staining Kit (Invitrogen, San Diego, CA, USA). The molecular masses were estimated from the migration distance and logarithm of molecular weights. Western blots were prepared by electroblotting SDS-PAGE resolved samples from the gels onto nitrocellulose membranes (Whatman, Middlesex, UK) at 30 V for 1 h according to the manufacturer's instructions. The membrane was blocked with 1 x PBS, 5% w/v non-fat dry milk, 0.05% v/v Tween-20 for 1 h, and then incubated with anti-His (C-term) primary antibody (diluted 1:5000) (Invitrogen, San Diego, CA, USA) for 1 h. The membrane

was washed three times with 1 x PBS containing 0.05% v/v Tween-20, each time for 15 min and incubated with anti-mouse IgG conjugated with alkaline phosphatase secondary antibody (diluted 1:30000) (Sigma, St. Louis, MO, USA) for 1 h. The membrane was washed three times with 1 x TBS containing 0.05% v/v Tween-20, each time for 15 min. Detection was performed using BCIP/NBT (Sigma, St. Louis, MO, USA) as a substrate. For visualization of CMCase activity in the gel, the staining method previously described (Choi *et al*., 2009) was performed, using the following minor modifications. The protein was separated on a 10% (w/v) acrylamide separating gel with 0.4% (w/v) Azo-CMC (Megazyme International Ltd., Victoria, Australia) at 120 V for 75 min. The gel was washed three times with 1% v/v TritonX-100, each time for 20 min at room temperature to remove the SDS. The proteins were then renatured in 50 mM sodium phosphate buffer (pH 6.5) at 4 °C for 1 h and then the gel was transferred to the new buffer and incubated at 37 °C for 4-5 h.

2.8 Protein deglycosylation and N-terminal amino acid sequencing

The enzyme solution (about 20 µg) was denatured in denaturing buffer (NEB, Ipswich, MA, USA) at 100 °C for 10 min and then treated with 1 µl (1000 U) of endoglycosidase H (NEB, Ipswich, MA, USA) in a total volume of 30 µl at 37 °C for 2 h. The sample (20 µl) was then analyzed using SDS-PAGE/Western blotting. The N-terminal acid sequence of the purified recombinant fusion EGII and CBHII species were determined by Edman sequencing (PROTEODYNAMICS, France).

2.9 Enzyme characterization

The evaluation of the optimum pH and temperature for the various EGII and CBHII enzyme were performed as described above, using 50 mM citrate/phosphate buffer pH 4.0-8.0 at 50 °C for evaluation of the optimum pH and at 30-100 °C at pH 4.8 for determining the optimum temperature. The hydrolysis time of CMC or PASC was 30 or 60 min. The pH stability was examined by pre-incubating the enzyme in the absence of a substrate at 50 °C for 30 min in the above buffer (pH 3-8) before determining the activity. For thermostability determination, the enzymes were pre-incubated at pH 5.0 at several defined temperatures for 30 min. The enzyme kinetics of EGII was determined using 0.1-2.5% (w/v) CMC diluted in 50 mM citrate buffer

(pH 4.8) as the substrate. The kinetic constants, K_M and V_{max} for the enzyme were carried out using the curve fitting procedure of SigmaPlot (SPSS, 2000). Assays were always performed in triplicate and the mean values with their standard deviation (SD) are presented.

PART III Directed evolution of *Trichoderma reesei* endoglucanase II to improve its thermostability in *Yarrowia lipolytica*

3.1 General materials and chemicals

Unless otherwise stated, all chemicals were analytical grade and purchased from Sigma-Aldrich (St. Louis, MO, USA). Specific enzymes for DNA manipulations were purchased from New England Biolabs (Ipswich, MA, USA). Oligonucleotides were synthesized by Eurogentec (Angers, France) and components for yeast culture media were supplied by Difco (Difco, Paris, France), except kanamycin which was purchased from Euromedex (Souffelweyersheim, France). Sterile 96-well cell culture microtiter plates, 96- deep well plates, and sealing tapes were purchased from Corning Corp. (NY, USA), and other polypropylene microtiter plates were from Evergreen Scientific (Los Angeles, CA, USA).

3.2 Strains, plasmids and culture conditions

The constitutive expression vector JMP62UraExTEF_fusEGII (Boonvitthya *et al*., 2012a) was constructed by inserting the cDNA encoding mature *T. reesei egl2* in framed with prepro *LIP2* signal peptide at *Bam*HI and *Avr*II restriction sites. The bacterial strain *Escherichia coli* DH5α (purchased from Gibco BRL, Rockville, MD, USA) was grown at 37 °C in Luria-Bertani medium, supplemented with 50 μg kanamycin/ml when required. The growth media and culture conditions for all *Y. lipolytica* strain zeta have been previously described (Bordes *et al*., 2007). Briefly, *Y. lipolytica* strain zeta were cultured in YPD at 28 °C, 120 rpm in baffled shaker flasks or in Petri dishes containing YPD supplemented with 1.5 % (w/v) agar. Selection of Ura^+ transformants were performed on solid YNBcasaD plates (0.17% w/v YNB, 1% glucose w/v, 0.5% w/v ammonium chloride, 0.2% w/v casamino acids, 1.5 % w/v agar and 100 mM sodium-potassium phosphate buffer, pH 6.8). For growth in

screening process, mutant *Y. lipolytica* were grown in YT medium (1% w/v yeast extract, 2% w/v tryptone and 100 mM phosphate buffer, pH 6.8), supplemented with either 0.25% w/v glucose (YTD) or with 2% w/v glycerol (YTG).

3.3 Directed evolution of the enzyme

A random mutagenesis library of *egl2* mutants was performed by using 3-step PCR method (Figure III-1). First, the high-fidelity PCR was done by using two primers amplified from 5'-zeta region to prepro *LIP2* signal peptide: PCR1dL (5'-CCGCTGTCGGGAACCGCGTTCAGGTGGAACAGGACACC-3') and Lip2prorev (5'-TCGCTTCTGGAGAACTGCGGCCTCAGAAG-3'). Reaction mixtures contained 1x Phusion HF buffer, 200 µM each dNTPs, 0.5 µM each of flanking primers, 10 ng of JMP62UraExTEF_fusEGII, and 0.02 U of Phusion DNA polymerase (Finnzyme) in a total volume of 50 µl. Thermal cycling parameters were 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 68 °C for 30 s and 72 °C for 1.5 min, and then finally at 72 °C for 10 min. Second, the full-length of *egl2* with 3' zeta region was randomly mutated by using the GeneMorph II EZClone Domain Mutagenesis Kit (Stratagene, La Jolla, CA) following the manufacturer's instructions for 30 cycles. To achieve low mutation frequencies with 4.5 base substitutions per kilobase, 2 µg of JMP62UraExTEF_fusEGII was used as the template. Two primers, Lip2pro (5'-CTTCTGAGGCCGCAGTTCTCCAGAAGCGA-3') and PCR2rL (5'-CCGCAC TGAGGGCTTTGTGAGGAGGTAACGCCG-3'), were also designed. Thermal cycling parameters were 95 °C for 2 min followed by 30 cycles of 95 °C for 1 min, 60 °C for 1 min and 72 °C for 2 min, and then finally at 72 °C for 10 min. The third PCR was done to fuse between the first and second amplified PCR products by using PCR1dL and PCR2rL, and generated mutant *egl2* library containing zeta docking platform. Reaction mixtures contained the same compositions and concentration as described in the first PCR. Thermal cycling parameters were 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 68 °C for 30 s and 72 °C for 2.5 min, and then finally at 72 °C for 10 min.

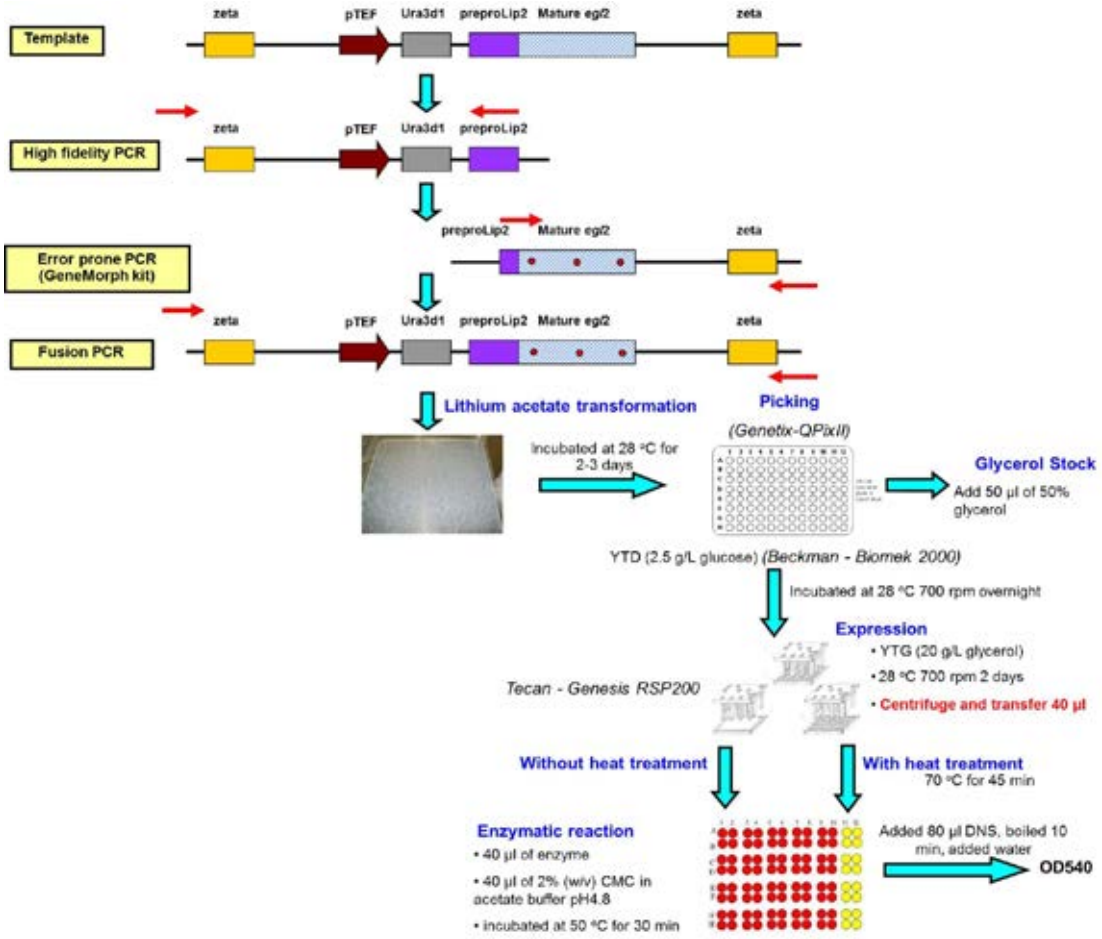


Figure III-1. Diagram of the random mutagenesis library of *egl2* mutants by using 3-step PCR method and its screening process.

Site-saturation mutagenesis on residues Tyr257 was performed using the QuikChange® Lightning Site-Directed Mutagenesis Kit, according to the recommendations provided in the instruction manual (Stratagene, La Jolla, CA). The mutagenic primers were designed for change the residue to other 19 amino acid residues. Thermal cycling parameters were 95 °C for 2 min followed by 30 cycles of 95 °C for 20 s, 60 °C for 10 s and 68 °C for 2.5 min, and then finally at 68 °C for 5 min. The changed amino acid residues were checked by DNA sequencing (PACIFIC SCIENCE, Bangkok, Thailand).

3.4 Yeast transformation

The transformation of *Y. lipolytica* strain zeta was achieved using the lithium acetate method, as previously described (Le Dall *et al*., 1994). Briefly, the cells were grown at 28 °C in 20 ml of YPD containing 50 mM citrate buffer (pH 4) until a cell density of between 8 x 10^7 to 1.5 x 10^8 cells/ml was attained. Cells were collected by centrifugation at 1160 x g, 4 °C for 5 min and washed twice using 10 ml of TE (50 mM Tris.HCl, 1 mM EDTA, pH 8.0) and a centrifugation step after each wash (see previous conditions), and then treated with 20 ml of 100 mM lithium acetate at 28 °C for 1 h without shaking. The cells were then centrifuged at 510 x g, 4 °C for 2 min and suspended in 2 ml of 100 mM lithium acetate. The amplified product from third PCR (approximately 500 ng combined with 25 µg of salmon sperm DNA) was incubated in 100 µl of *Y. lipolytica* competent cells at 28 °C for 15 min without shaking, and then 700 µl of 40% (w/v) PEG 4000 was added and incubated at 200 rpm, 28 °C for 1 h. After that, the cells were transformed by heat shock at 39 °C for 10 min and immediately recovered in 1.2 ml of 100 mM lithium acetate. Ura^+ transformants were selected onto YNBcasaD in Q-trays and incubated at 28 °C for 48 h.

3.5 Screening for thermostability Cel5A variants

The distribution of sterile YTD medium into microtiter plates and YTG medium into deep well plates were performed using a Biomek® 2000 Laboratory Automation Workstation (Beckman, USA). Colonies (including *Y. lipolytica* transformed with wild type JMP62UraExTEF_fusEGII as the control) were picked using an automated colony picker QpixII (Genetix, UK) as previously described (Bordes *et al*., 2007) to inoculate 96-well microplates containing 200 µl of YTD medium for preculture. The microplates were sealed with air permeable membranes and incubated at 28 °C, 200 rpm for 24 h. Then, 100 µl of the cultures were inoculated into 96-deep well plates containing 1 ml of YTG medium for enzyme expression and incubated at 28 °C, 700 rpm for 48 h. Deep-well plates were centrifuged at 5000 x g, 4 °C for 5 min and 40 µl of supernatant was distributed into two microplates using a liquid handling station (Genesis RSP200, Tecan, Hampshire, UK) for testing their thermostability (one with heat treatment and another without heat treatment). The heat treatment microplates were sealed with an aluminium film and incubated at 70 °C for

45 min. After that, both microplates were added by 40 µl of 2% w/v carboxymethyl cellulose (CMC) in 50 mM citrate buffer (pH4.8), and incubated at 50 °C for 30 min. The reducing sugars before and after heat treatment were performed by micro-DNS assays. Enzymatic reaction plates were added by 80 µl of DNS reagent, incubated at 95 °C for 10 min, and cooling in an ice bath. Colour development was transferred into polystyrene microplates and the absorbance was read at 540 nm. Finally, the colonies having ratio of OD_{540} after heat treatment higher than average plus two standard deviation of wild type were selected for further studied.

3.6 Purification of the recombinant Cel5A variants from *Y. lipolytica*

Yeast cells from liquid cultures (400 ml) were sedimented by centrifugation at 4500 x g, 4 °C for 10 min. The supernatant was concentrated by ultrafiltration using Macrosep polyethersulfone (10 kDa cutoff) membrane (PALL). Concentrated protein was exchanged with 10 mM citrate buffer (pH 3.5) using the same type of ultrafiltration centrifugal column as above, before being loaded onto Bio-Scale™ Mini UNOsphere™ S Cartridges (Bio-Rad, Hercules, CA). After protein loading, the column was washed with 10 mM citrate buffer (pH 3.5) and the bound proteins were eluted with a stepwise salt concentration from 0 to 1 M NaCl in the same buffer at a flow rate of 5 ml/min. Fractions containing the target protein were pooled and monitored by SDS-PAGE and enzyme activity measurement. Protein concentration of the crude and purified enzyme was spectrophotometrically measured at 595 nm using Bio-Rad Protein Assay (Bio-Rad, Hercules, CA) and bovine serum albumin as the standard.

3.7 Activity assays and kinetics

The EG activity was measured as previously described (Ghose, 1987). The reaction mixture contained 0.5 ml of 2% (w/v) CMC in 50 mM citrate buffer (pH 4.8) and 0.5 ml of diluted enzyme solution. The reaction was conducted at 50 °C for 30 min and then reducing sugars were quantified using the DNS reagent. Thermostability curve were studied by incubating the purified enzymes at 70 °C until 5 h and taking the aliquots for analysis the residual activity as described above at periodic intervals.

PART IV Application of the recombinant cellulases in ethanol fermentation of rice straw

4.1 Organism and culture conditions

Recombinant *P. pastoris* producing *T. reesei* EGII and CBHII (Boonvitthya *et al*., 2012a), and *A. oryzae* ExgA (Boonvitthya *et al*., 2012b) were maintained at 30 °C for 48 h on YPD agar (1% w/v yeast extract, 2% w/v peptone, 2% w/v glucose, and 1.5% w/v agar). For heterologous protein expression in *P. pastoris*, a two stage strategy was employed. First, yeast biomass was generated in shaker flask cultures at 30 °C, 200 rpm growing in BMGY liquid medium (1% w/v yeast extract, 2% w/v peptone, 100 mM potassium phosphate buffer, pH 6.0, 1.34% w/v YNB, 4 x 10^{-5} % w/v biotin, and 1% v/v glycerol) and suspended in sterile liquid BMMY medium (as per BMGY except with 0.5% [for recombinant ExgA] or 3% v/v methanol [for recombinant EGII and CBHII] instead of the glycerol). Shaking in baffled flasks at 30 °C, 200 rpm was performed for 120, 120, 42 h in order to produce recombinant EGII, CBHII, and ExgA, respectively. *S. cerevisiae* TISTR 5339 was obtained from Microbiological Resource Center, Thailand Institute of Scientific and Technological Research (TISTR), and *P. stipitis* CBS 5773 was obtained from The Centraal bureau voor Schimmelcultures (CBS), The Netherlands. Yeast strains were maintained at 30 °C for 48 h on YM agar (0.3% w/v yeast extract, 0.3% w/v malt extract, 0.5% w/v peptone, 1% w/v glucose, and 1.5% w/v agar). *S. cerevisiae* and *P. stipitis* were precultured as inocula for fermentation in Erlenmeyer flasks containing 50 ml of 0.3% (w/v) yeast extract, 0.3% (w/v) malt extract, 0.5% (w/v) $(NH_4)_2SO_4$, and 2% (w/v) glucose for *S. cerevisiae* or 2% (w/v) xylose for *P. stipitis*. Flasks were incubated at 30 °C and shaken at 200 rpm for 24 h. The fermentation medium contains 0.3% (w/v) yeast extract, 0.3% (w/v) malt extract, 0.5% (w/v) ammonium sulphate, and 50 mM sodium citrate buffer (pH 5.0). Shaking in baffled flasks at 35°C, 150 rpm was performed for 168 h (Laplace *et al*., 1993).

4.2 Enzyme production and activity assays

The expression of the EGII, CBHII, and ExgA enzymes in *P. pastoris* was achieved according to the supplier's instructions (Invitrogen guideline manual version H).

Briefly, expression was performed in baffled shaker flasks that were incubated at 30 °C, 200 rpm, for 5 days. To induce expression, methanol was added daily to a final concentration of either 0.5% or 3% (v/v).

The EG activity was measured as previously described (Ghose, 1987). The reaction mixture contained 0.5 ml of 2% (w/v) carboxymethyl cellulose (CMC) in 50 mM citrate buffer (pH 4.8) and 0.5 ml of diluted enzyme solution. The reaction was conducted at 50 °C for 30 min and then reducing sugars were quantified using the dinitrosalicylic acid (DNS) reagent. CBH activity was evaluated using phosphoric acid swollen cellulose (PASC) that was prepared from Avicel PH-101 as previously described (Wood, 1988). PASC at 1% (w/v) in 50 mM acetate buffer (pH 4.8) was incubated with the enzyme solution at 50 °C for 1 h and then centrifuged at 17 760 x *g* for 3 min to sediment the substrate. The level of soluble reducing sugars in the supernatant was quantified using the Nelson-Somogyi method (Nelson, 1944; Somogyi, 1952). One unit (U) of enzyme activity was defined as the amount of enzyme that released 1 µmole of reducing sugars per ml per min.

The ExgA activity was measured by using *p*-nitrophenyl-β-D-glucopyranoside (pNPβG) as previously described (Araujo *et al*., 1983). The reaction mixture contained 0.5 ml of 2 mM pNPβG in 50 mM acetate buffer (pH 5.0) and appropriate diluted enzyme solution. The reaction was conducted at 35 °C for 30 min and then the reaction was stopped by adding 1 ml of 1 M Na_2CO_3. The color that developed as a result of *p*-nitrophenol liberation was measured at 400 nm. One unit (U) of enzyme activity was defined as the amount of enzyme that released 1 µmole of *p*-nitrophenol ml^{-1} min^{-1}. All enzymatic activity measurements were performed in triplicate unless otherwise stated.

4.3 Pretreatment

The milled rice straw was pretreated by alkaline peroxide method. 1.2 g of rice straw was suspended in 7.5% (v/v) H_2O_2 and NaOH was then added to adjust the pH to 11.5. The pretreated samples were then incubated at 35°C, shaken at 250 rpm for 24 hours. Finally, concentrated HCl was added to adjust the pH to 4.8 before enzymatic hydrolysis and co-fermentation step (Saha and Cotta, 2007).

4.4 Enzymatic hydrolysis and co-fermentation

The fermentation medium was added to the pretreated whole substrate slurry and autoclaved. Recombinant enzymes producing in *P. pastoris*, cellulase (Celluclast 1.5L) and xylanase (Viscostar 150L) were filtered and sterilized through a 0.2 μm filters. In this study, the SSCF experiment was performed by using produced recombinant enzymes (12 U of EGII/g substrate, 12 U of CBHII/g substrate, 6 U of ExgA/g substrate) supplemented with 1200 U of xylanase/g substrate. Then, 3.3 ml (5% v/v) of each *S. cerevisiae* and *P. stipitis* was added. Final volume was adjusted to 65 ml by sterilized distilled H_2O. The samples were incubated with shaking at 35°C, 150 rpm for 7 days and ethanol production was determined by gas chromatography (Shimadzu, Japan).

CHAPTER IV

RESULTS AND DICUSSION

PART I Cloning and expression of *Aspergillus oryzae* glucan 1,3-beta-glucosidase A in *Pichia pastoris*

1.1 Construction of expression vectors and sequencing of *exgA* cDNA from *A. oryzae*

RT-PCR of the *exgA* produced a single amplified DNA fragment of 1173 bp encoding the mature peptide region. Sequencing analysis showed that the cDNA was homologous to the reported sequence (GenBank accession code AJ566365) with 97% identity and the predicted amino acid sequence showed 96% homology with *A. oryzae* ExgA (UniProtKB/Swiss-Prot code Q7Z9L3). The amplified PCR product was ligated in frame with the α-factor signal sequence of pPICZαA and pGAPZαA by introducing an artificial *EcoR*I restriction site at 5' end and *Not*I restriction site at 3' end of the cDNA, and the recombinant plasmids were subsequently transformed into *P. pastoris X-33*. The correctness of the open reading frame and integration of expression cassettes into the yeast genome were also confirmed by PCR analysis and DNA sequencing. The DNA sequence for the *exgA* gene from this isolate used in this study, and predicted amino acid sequence is deposited at GenBank (NCBI) with the previous accession code.

1.2 Heterologous expression and purification of ExgA in *P. pastoris*

Extracellular expression levels of the rExgA in *P. pastoris* was tested in 50 ml of YPD for constitutive expression and BMMY medium, induced daily with 0.5-3% (v/v) methanol, for inducible expression. Both constructs were successfully expressed in *P. pastoris* and secreted the enzyme into the culture medium. Constitutive expression of rExgA was correlated with the increased cell growth (data not shown) and reached a maximal extracellular expression level in the culture medium of 1.4 U/ml after 60 h (Figure IV-1). With respect to the inducible expression, the enzyme

was optimally induced with 0.5% (v/v) methanol (data not shown), reaching 2 U/ml after 42 h (Figure IV-1).

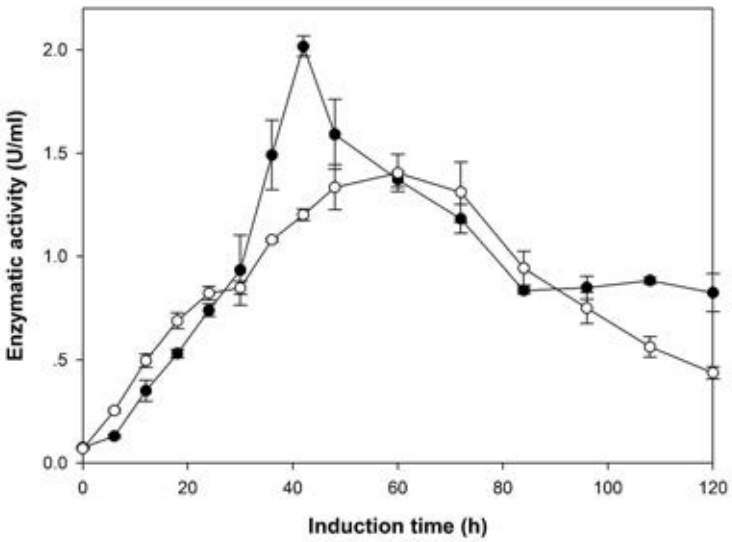


Figure IV-1. Time course of the production of rExgA from the *GAP* constitutive (*open circle*) and *AOX*1 0.5% (v/v) methanol-inducible (*solid circle*) promoter expression systems in *P. pastoris*. Data are shown as the mean ± 1 SD and are derived from three independent repeats.

Thus a 1.43-fold higher yield was obtained and sooner using the methanol-*AOX*1 inducible system and so this was used for further work. The rExgA was one-step purified from 200 ml of yeast culture medium (BMMY) using TALON Metal Affinity Resin (Clontech) column chromatography. The specific activity produced by *P. pastoris* was 289 U/mg with an 18% yield of the total activity. SDS-PAGE analysis indicated that the molecular mass of the enriched rExgA produced from *P. pastoris* was 40 kDa, which is similar to that reported for the mature peptide from *A. oryzae* ExgA (Figure IV-2), and that it was apparently enriched to homogeneity (single band only).

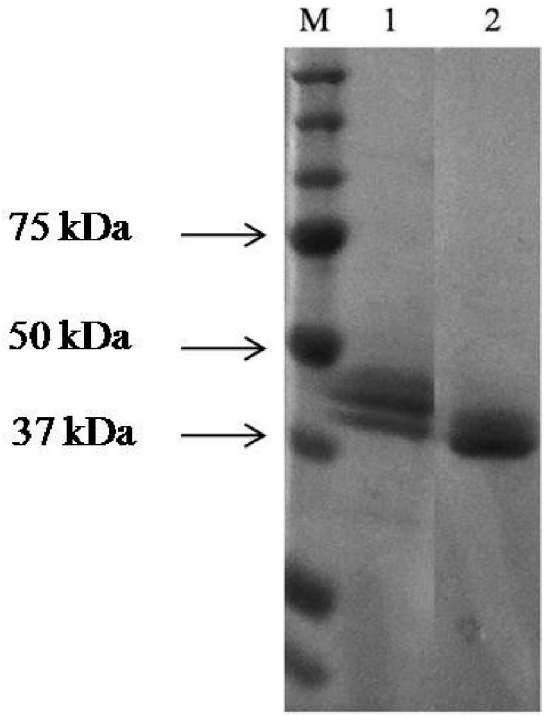


Figure IV-2. SDS-PAGE analysis of recombinant ExgA produced from *P. pastoris*: lane M, molecular weight marker; lane 1, crude enzyme; lane 2, purified enzyme.

1.3 Characterization of recombinant ExgA

The enriched rExgA enzyme exhibited a maximum activity at pH 5.0-6.5 but displayed at least 80% activity over the range pH 5-8 (Figure IV-3). The optimum temperature of *A. oryzae* ExgA assayed in the citric/phosphate (pH 5.0) buffer was 35 °C and retained more than 50% of activity after incubation at 45 °C for 30 min before declining sharply, with no activity being detectable at 50 °C (Figure IV-4). Therefore, it is a thermal sensitive enzyme which should be evaluated for potential improvement by genetic engineering.

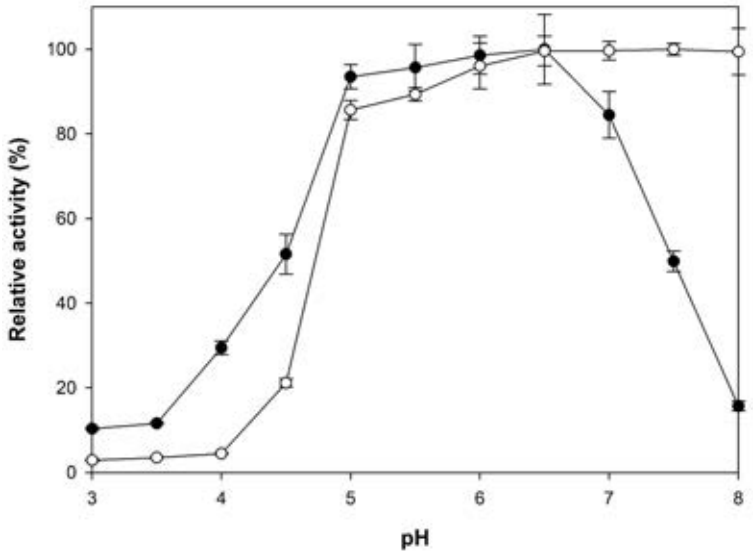


Figure IV-3. Enzymatic activity (*solid circle*) and stability (*open circle*) of recombinant ExgA produced from *P. pastoris* at various pH values. Data are shown as the mean ± 1 SD and are derived from three independent repeats.

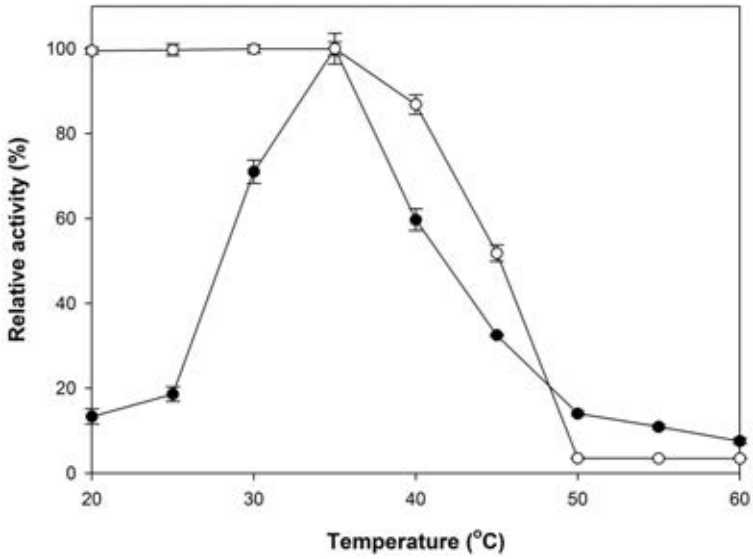


Figure IV-4. Enzymatic activity (*solid circle*) and stability (*open circle*) of recombinant ExgA produced from *P. pastoris* at various temperatures. Data are shown as the mean ± 1 SD and are derived from three independent repeats.

Using the pH and temperature optima for the enzyme, the kinetic parameters K_M, V_{max}, and k_{cat} of the enzyme for pNPβG were found to be 0.56 mM, 10042 μmol min^{-1} mg^{-1} of protein and 410 s^{-1}, respectively. Glucose acted as a competitive

inhibitor of pNPβG hydrolysis by rExgA with an inhibition constant (K_i) of 365 mM, as obtained from the intersection of different lines on a Dixon plot for different glucose and substrate (pNPβG) concentrations (Figure IV-5). These biochemical properties showed that rExgA produced from *P. pastoris* was active enzyme with higher V_{max} value although the glucose inhibition constant was lower than previous report (Riou *et al*., 1998).

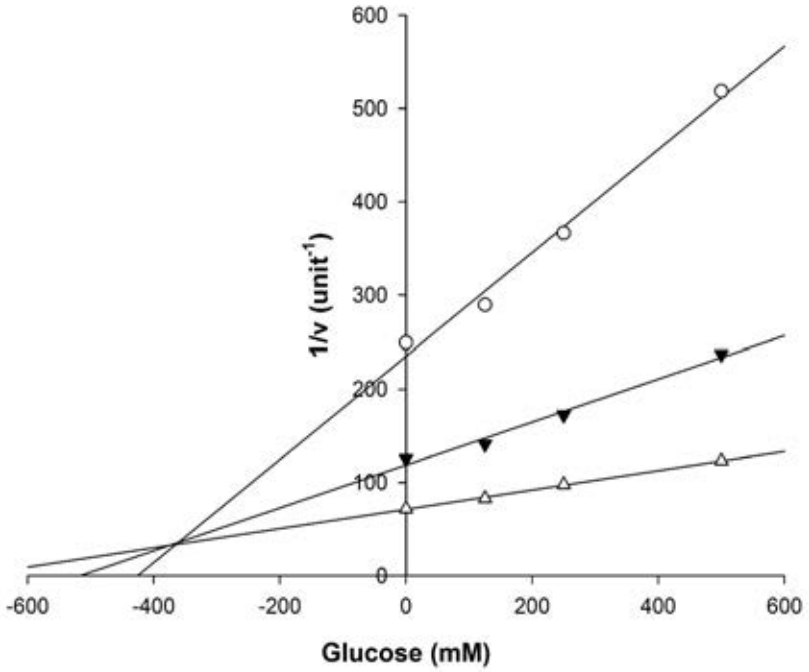


Figure IV-5. Dixon plot of the inhibitory effect of glucose on pNPβG hydrolysis by the enriched rExgA enzyme produced from *P. pastoris*. The reaction time was 30 min at pH 5.0 and 35 °C. The enzyme was used at 0.3 U/ml. The pNPβG concentrations used were 0.5 (*open circle*), 1.0 (*solid triangle*) and 2.0 (*open triangle*) mM.

PART II Comparison of the heterologous expression of *Trichoderma reesei* endoglucanase II and cellobiohydrolase II in the yeasts *Pichia pastoris* and *Yarrowia lipolytica*

2.1 Comparison of the activity of EGII and CBHII produced in *P. pastoris* and *Y. lipolytica*

To determine which yeast expression system is best adapted for the high yield production of EGII and CBHII, three clones of each construction were randomly selected and cultured in shake flasks. For recombinant expression in *P. pastoris*, EGII activity was detected in the culture broth shortly after induction with methanol (0.5 % v/v), after which it increased exponentially, reaching 8.8 and 9.2 U/ml (CMCase activity) after 96 and 120 h respectively (Figure IV-6).

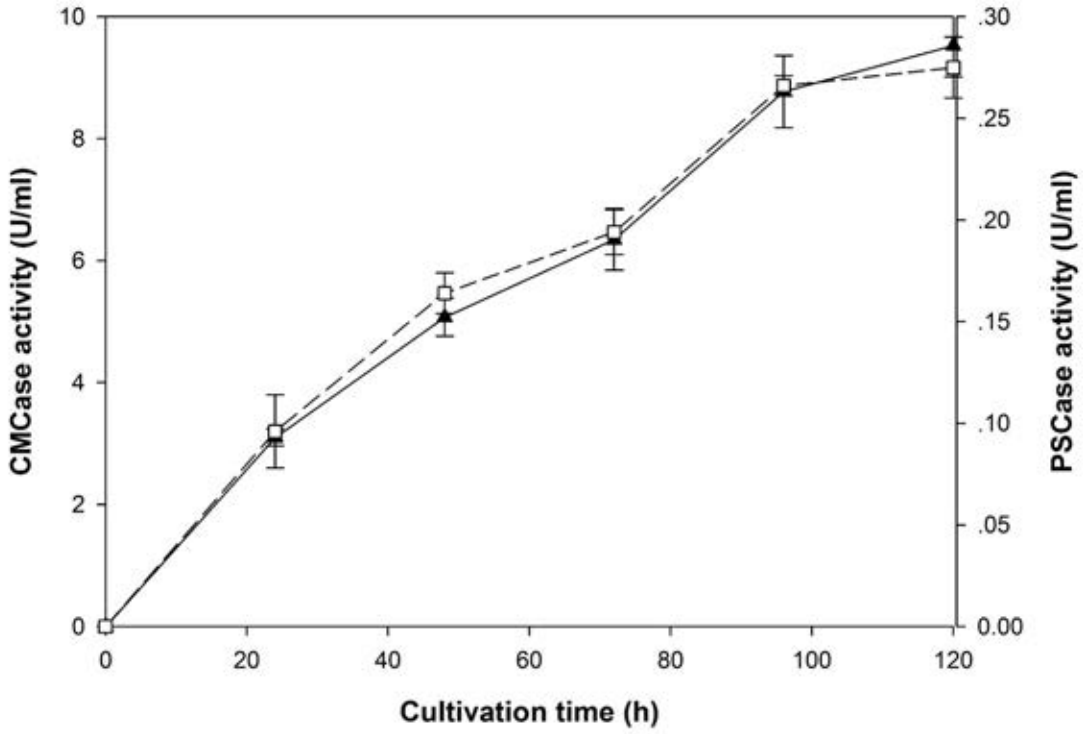


Figure IV-6. Time course production of EGII (*solid line, solid triangle*) and CBHII (*dashed line, open square*) in *P. pastoris* with 0.5% (v/v) methanol induction. Data are shown as the mean ± 1 SD and are derived from three independent repeats.

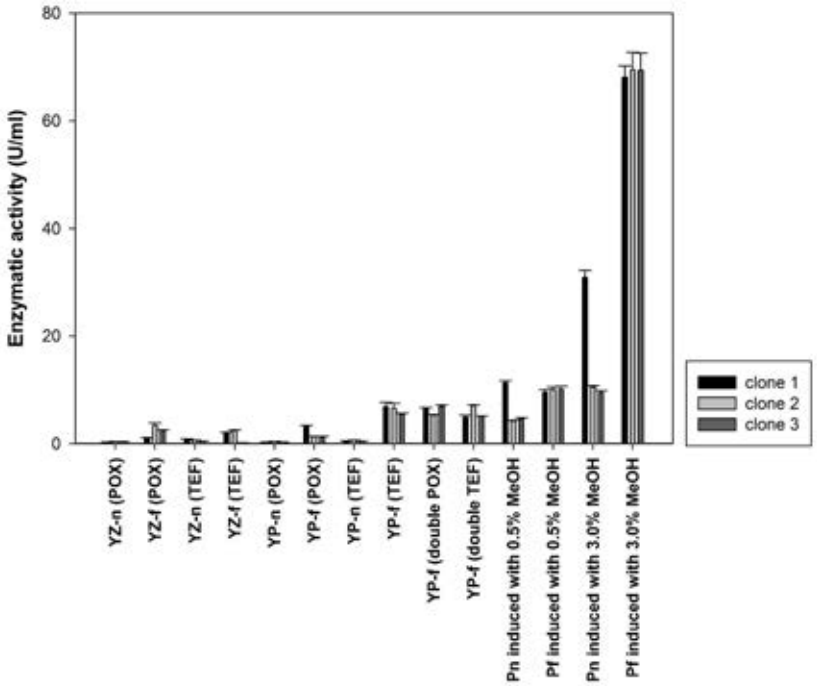
The production of CBHII followed a similar pattern, but was maximal at 0.27 U/ml (PSCase activity). Consequently, samples for enzyme analysis were removed 120 h after induction in all cases. Analysis of the production of EGII and CBHII revealed that the use of the heterologous preproLip2 (in *Y. lipolytica*) or the *S. cerevisiae* α-factor (in *P. pastoris*) signal peptides led to a higher level of recombinant enzyme production and secretion efficiency compared to the use of the native signal peptides of the enzymes (Figure IV-7). Moreover, the *Y. lipolytica* Po1d strain, bearing a single copy of *egl*2 or *cbh*2, produced higher amounts of EGII or CBHII than the equivalent zeta strain, as did the constitutive TEF promoter when compared to the inducible POX promoter. Comparing *P. pastoris* and *Y. lipolytica* revealed that the production of EGII in *P. pastoris* was on average 10 U/ml, while the best *Y. lipolytica* transformant only produced 6 U/ml. It is noteworthy, that the introduction of a second TEF-controlled *egl*2 gene copy into *Y. lipolytica* did not improve maximum production (Figure IV-7A). However, interestingly, in *P. pastoris* the use of a higher methanol concentration (3% v/v) for induction maximized production of EGII (up to 70 U/ml), but only when this enzyme was produced using the α-factor signal peptide, but not when using the EGII native signal peptide (Figure IV-7A).

Regarding CBHII production, the *Y. lipolytica* Po1d strain harboring a single copy of *cbh*2 linked to the preproLip2 signal peptide and controlled by the TEF promoter yielded an average extracellular PSCase activity of 0.16 U/ml, whereas *P. pastoris* induced with 3% (v/v) methanol produced up to 0.25 U/ml, a 2.8-fold increase compared to induction using 0.5% v/v methanol. However, when a second copy of the *cbh*2 gene was present in *Y. lipolytica* Po1d, the yield of recombinant protein (0.36 U/ml) was increased almost 2.2-fold, and was higher than the yield obtained using *P. pastoris* (Figure IV-7B).

2.2 Purification of recombinant proteins

The purification of EGII and CBHII from *P. pastoris* culture supernatants was performed using TALON Metal Affinity (Clontech) chromatography, whereas their purification from *Y. lipolytica* culture supernatants required the use of ion exchange chromatography (Mono S, GE Healthcare). The results of these purifications are summarized in Table IV-1. The total amount of purified enzyme varied from 15-27 mg/L protein, except for CBHII produced by *Y. lipolytica* whose yield was higher, being on average 50 mg/L protein.

A



B

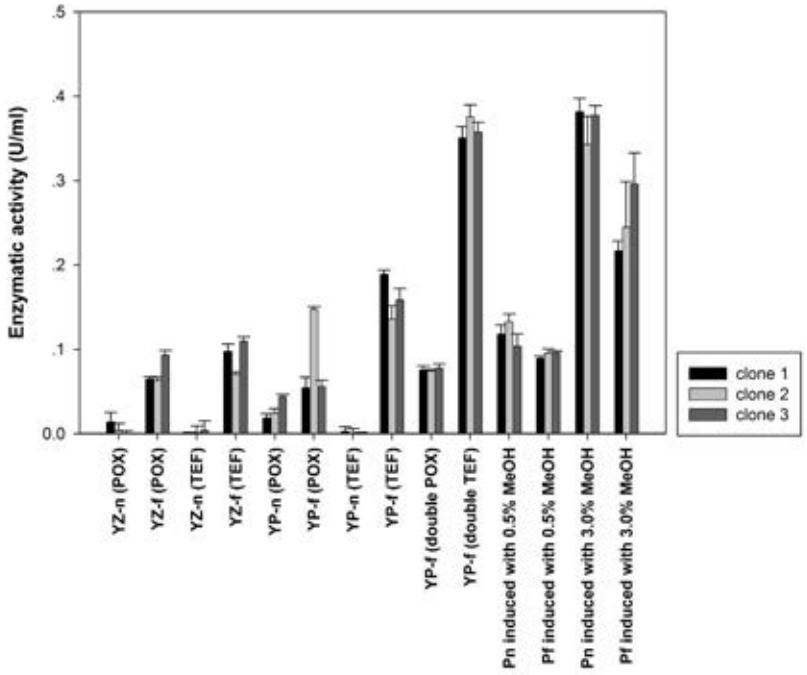


Figure IV-7. Extracellular enzymatic activity (U/ml) of all (A) EGII (as CMCase activity) and (B) CBHII (as PSCase activity) constructs produced from *P. pastoris* and *Y. lipolytica*. Data are shown as the mean $\pm$ 1 SD and are derived from three independent repeats.

[1]P = *P. pastoris*; Y = *Y. lipolytica*; n = native; f = fusion; Z = zeta strain; P = Po1d strain

SDS-PAGE analysis indicated that the molecular mass of the purified EGII produced from *P. pastoris* was 58 kDa, which is higher that of the purified EGII produced by *Y. lipolytica* (55 kDa). Similarly, the apparent molecular weight of the *P. pastoris* CBHII (approximately 65 kDa) also appeared to be slightly higher than that of the *Y. lipolytica* CBHII (*P. pastoris* recombinant enzymes, Figure IV-8A and *Y. lipolytica* recombinant enzymes, Figure IV-8B). Finally, it is important to note that the specific activities of EGII and CBHII produced by *Y. lipolytica* were 1.5- and 1.4-fold higher than those produced in *P. pastoris* (Table IV-1).

2.3 Enzymatic deglycosylation

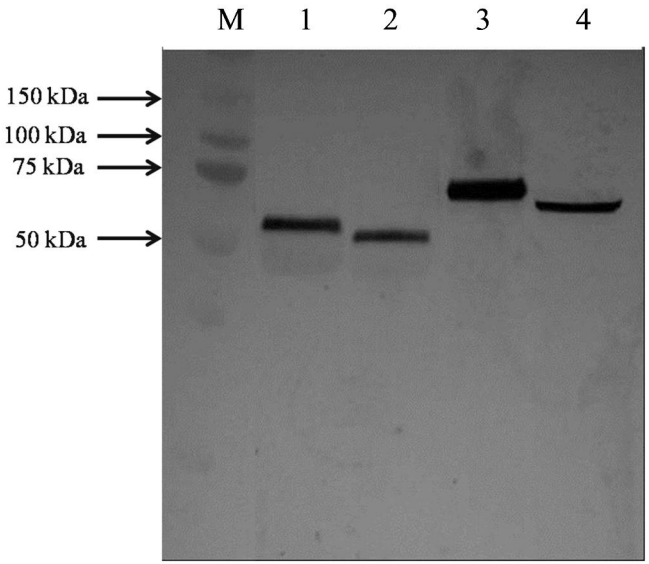
Electrophoretic analyses of EGII and CBHII from *P. pastoris* and *Y. lipolytica* revealed that molecular weights of these proteins were higher than those of their wild type counterparts (wt EGII, 48 kDa and wt CBHII, 50-58 kDa), produced by *T. reesei* (Saloheimo *et al*., 1988; Teeri *et al*., 1987), even though amino acid sequencing revealed the N-terminal sequence motifs (AQQTVV and AQACSS, for EGII and CBHII respectively) were identical to those found in the mature wild type homologs from *T. reesei*. Therefore, it was suspected that the molecular weight differences might arise from protein hyperglycosylation mediated by the yeast production hosts. To test this postulate, endoglycosidase H was used to remove asparagine-linked *N*-glycosyl motifs from yeast-produced EGII and CBHII. Upon deglycosylation, the molecular weight of the *P. pastoris* EGII decreased from 58 to 55 kDa, while that of the *P. pastoris* CBHII decreased from 65 kDa to 61 kDa (Figure IV-8A). Interestingly, during this analysis we observed the *P. pastoris* CBHII produced using the native peptide signal displayed two isoforms, the biggest of these presenting molecular weights, before and after deglycosylation, of 68 and 65 kDa respectively (data not shown). Similarly, deglycosylation led to a reduction in the molecular weights of the *Y. lipolytica*-produced EGII (55 to 52 kDa) and CBHII (63 to 60 kDa) (Figure IV-8B) and zymogram analysis further confirmed the deglycosylated EGII produced by *P. pastoris* has a slightly higher molecular weight than its counterpart produced by *Y. lipolytic*a (Figure IV-9). Nevertheless, despite these physical changes, measurement of CMCase and PSCase activities revealed that deglycosylation did not radically alter specific activity, although in the case of EGII a modest loss of activity (7.8%) was measured, which is in good agreement with previous observations (Qin *et al*., 2008b).

Table IV-1. Purification of extracellular EGII and CBHII proteins produced from the culture media of transformed *P. pastoris* and *Y. lipolytica*

Enzyme, production host and purification method	Total protein (mg)[1]	Total activity (U)	Yield (%)	Specific activity (U/mg)	Purification (fold)
P. pastoris EGII purified by TALON His-tag[2]	7.2	1160.5	32.7	161.2	5.6
Y. lipolytica EGII purified by Mono S	5.3	1276.5	55.7	243.1	5.7
P. pastoris CBHII purified by TALON His-tag[2]	9.6	16.5	26.2	1.72	4.4
Y. lipolytica CBHII purified by Mono S	17.5	42	40	2.4	2.5

[1]From 350 ml culture media

A



B

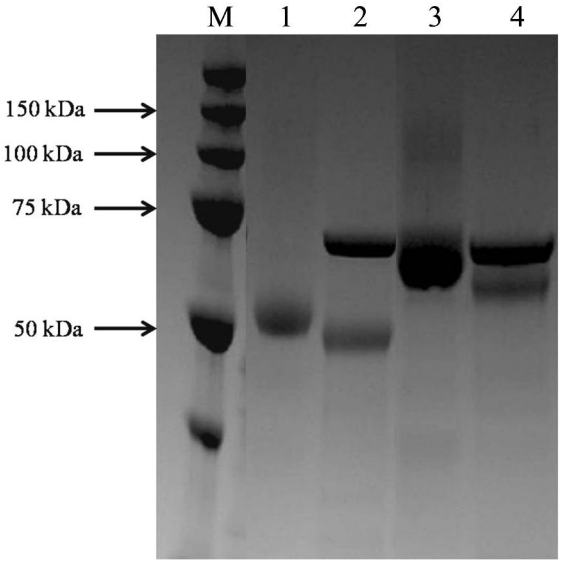


Figure IV-8. SDS-PAGE analysis of deglycosylated EGII and CBHII produced from (A) *P. pastoris* (Western blot detection) and (B) *Y. lipolytica*: lane M, molecular weight marker; lane 1, EGII; lane 2, endo-H treated EGII; lane 3, CBHII; lane 4, endo-H treated CBHII.

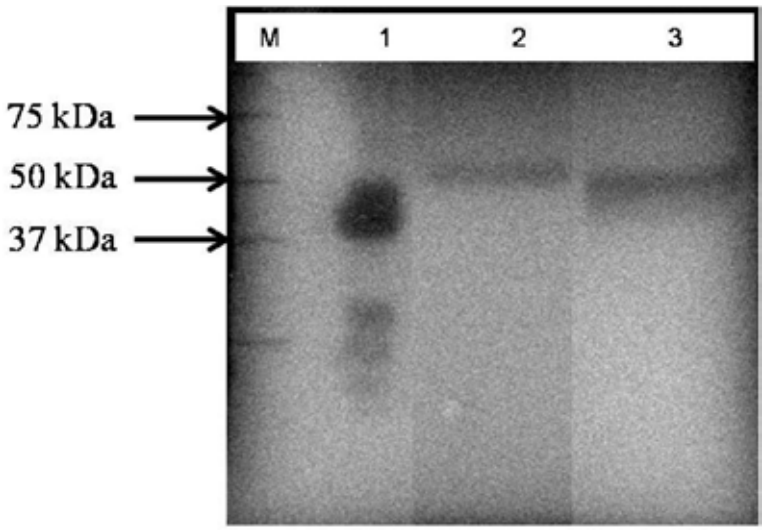


Figure IV-9. Zymogram analysis of EGII produced by *P. pastoris* and *Y. lipolytica*: lane M, molecular weight marker; lane 1, commercial Accelerase 1500 (control); lane 2, EGII from *P. pastoris*; lane 3, EGII from *Y. lipolytica*.

2.4 Characterization of EGII and CBHII proteins produced in *P. pastoris* and *Y. lipolytica*

Examination of the effects of pH and temperature on the different enzymes revealed that EGII produced in either *P. pastoris* or *Y. lipolytica* exhibited maximum activity at pH 5.0, and displayed at least 60% activity over the range pH 4-7 (Figure IV-10A). Similarly, the two CBHII variants displayed identical pH optima (5.5-6.0) for activity, although small differences at higher (*Y. lipolytica* CBHII less active) and lower pH values (*P. pastoris* CBHII less active) were observed (Figure IV-10B). Regarding pH stability, no differences were evidenced for the EGII variants, with 100% stability being observed at pH 4.0, followed by a rapid decline to < 20% residual activity after 30 minutes exposure at pH >6.5 (Figure IV-10A). Both CBHII enzymes displayed almost identical stability (100% stable in the range pH 5-6), undergoing rapid inactivation at alkaline pH and showing some instability below pH 5 (Figure IV-10B). The temperature dependency of the activity of the EGII variants was described by an approximate bell shaped curve, displaying a maximum at 60 $^{\circ}$C, and a steeper slope on the right-hand limb, indicating quite fast thermal denaturation. Thermostability of the two EGII species was also identical, being stable up to 60 $^{\circ}$C before declining sharply, with no activity being detectable after 30 minutes at 80 $^{\circ}$C (Figure IV-11A). For the CBHII enzymes, optimum activity was observed around 55 $^{\circ}$C,

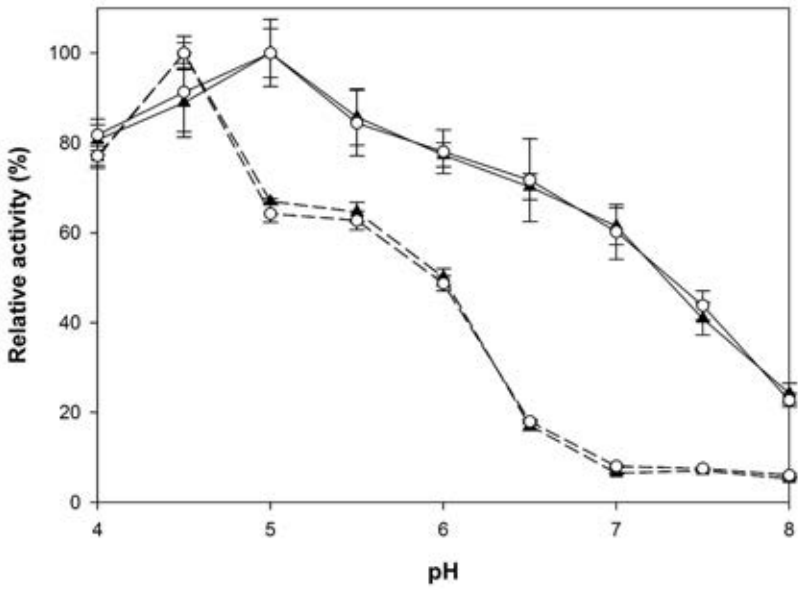
followed by a sharp decline at temperatures > 60 °C (Figure IV-11B). Likewise, both CBHII variants displayed good stability until 40°C, beyond which activity declined with 50% of initial activity being measured after 30 min at 60 °C (Figure IV-11B).

Using the pH and temperature optima for the different enzymes, it was possible to determine the kinetic parameters K_M and V_{max}. Regarding the EGII variants, V_{max} values were nearly identical (approximately 0.57 and 0.62 µmol/mL/min), whereas the *Y. lipolytica* EGII was characterized by a lower $K_{M(app)}$ value (0.72 mmol/L) compared to *P. pastoris* EGII (0.95 mmol/L).

Discussion

In this study, two cellulase genes, *egl*2 and *cbh*2, from the filamentous fungus *T. reesei* QM9414 were successfully expressed and secreted as heterologous extracellular proteins at a high level in the non-cellulolytic yeasts, *Y. lipolytica* and *P. pastoris*. The main purpose of this study was to evaluate the utility of *Y. lipolytica* as a host strain for EGII and CBHII production, thus a variety of constructions were tested. Unsurprisingly, the *Y. lipolytica* strain Po1d proved to be the best host for expression. This strain is the choice host for protein production, because an extracellular alkaline protease encoding gene has been deleted, the genetic background is adapted for high level protein expression and the production of invertases enables this yeast to use sucrose or molasses as a carbon source (Madzak *et al*., 2004). In contrast, the *Y. lipolytica* zeta strain, which is a useful tool for the construction of mutant gene libraries, was logically ill-adapted for optimal expression of EGII and CBHII (Nicaud *et al*., 1989). Regarding the promoter, *TEF* performed better than the inducible acyl-CoA oxidase2 promoter (*POX*2). Though this finding is interesting, it is already known that *TEF* is a strong constitutive promoter that can lead to the production of large amounts of recombinant protein (up to 3-10% of total soluble protein) (Cavallius *et al*., 1993). Moreover, it is convenient, because no induction is needed, although its use is not universal, because constitutive expression is clearly a disadvantage when heterologous proteins are toxic (van Zyl *et al*., 2007). Nevertheless, this was not the case for EGII and CBHII. Regarding the choice of signal peptide, it is interesting to note that the expression of CBHII and EGII was always better when their native signal peptides were replaced by ones better adapted to the yeast hosts, even though the fungal signal peptides were recognized and processed by both *Y. lipolytica* and *P. pastoris*.

A



B

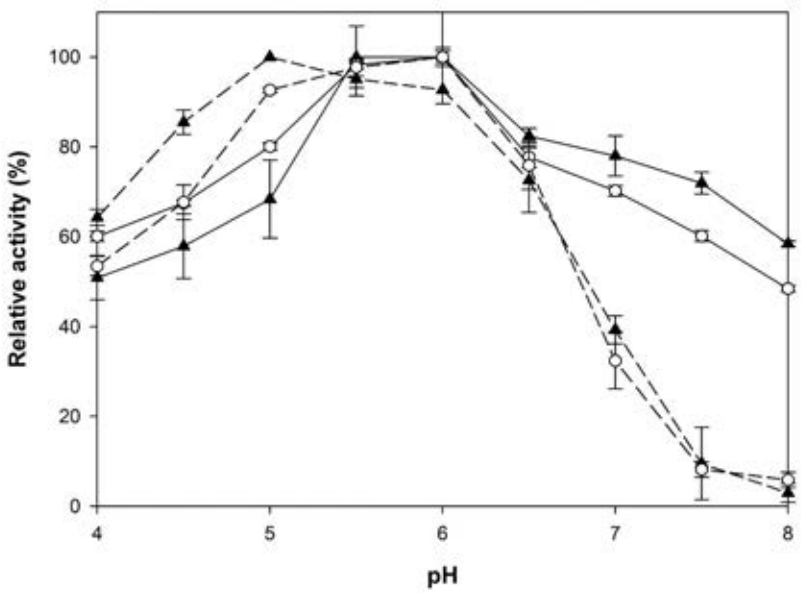


Figure IV-10. Enzymatic activity (*solid line*) and stability (*dashed line*) of the (A) EGII and (B) CBHII enzymes produced from *P. pastoris* (solid triangle) and *Y. lipolytica* (open circle) at various pH values. Data are shown as the mean $\pm$ 1 SD and are derived from three independent repeats.

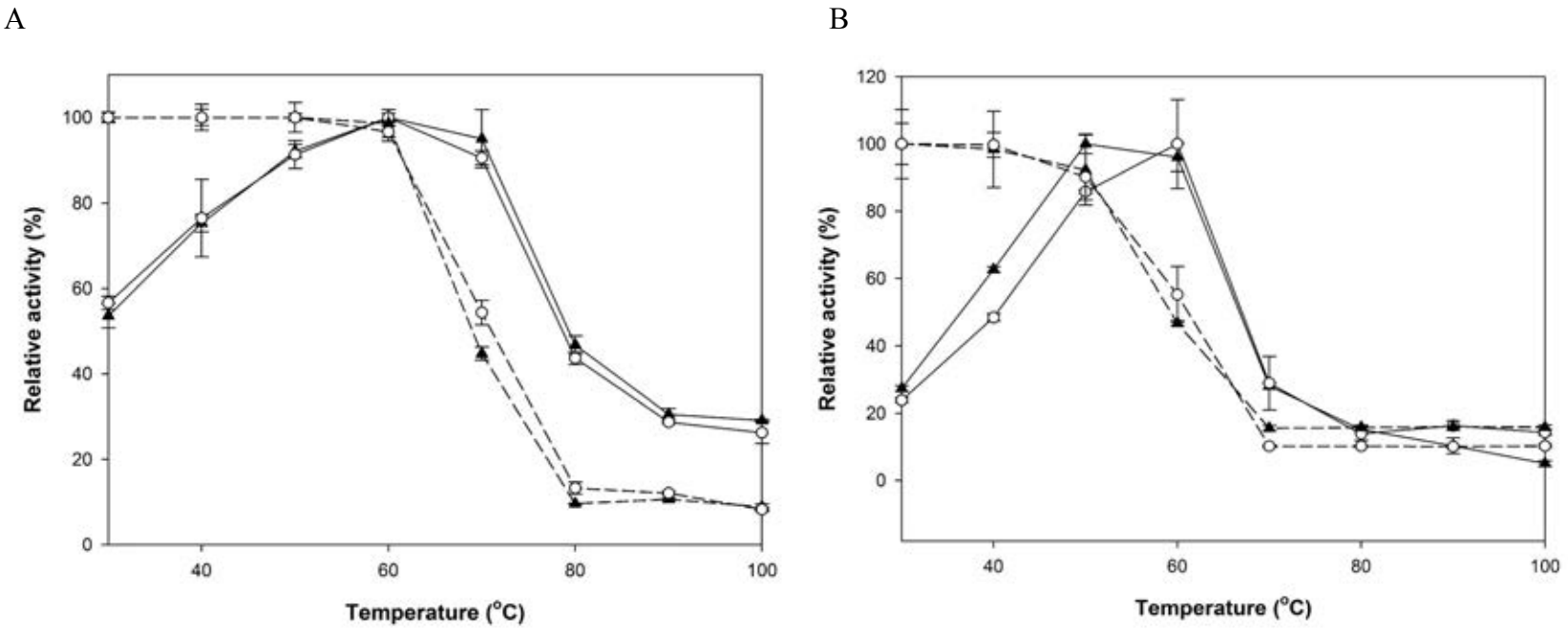


Figure IV-11. Enzymatic activity (*solid line*) and stability (*dashed line*) of the (A) EGII and (B) CBHII enzymes produced from *P. pastoris* (solid triangle) and *Y. lipolytica* (open circle) at various temperatures. Data are shown as the mean $\pm$ 1 SD and are derived from three independent repeats.

For expression in *P. pastoris*, we preferred to use the inducible alcohol oxidase1 (*AOX*1) promoter, because other reports, and our own data (not shown), suggest that this promoter is better than the constitutive glyceraldehyde-3-phosphate dehydrogenase (*GAP*) promoter (Boer *et al*., 2000). Accordingly, inducible expression in *P. pastoris* expression was preferable to constitutive expression in *Y. lipolytica*, at least when considering EGII. However, constitutive expression in *Y. lipolytica* was superior for CBHII, which indicates that factors other than simple promoter potential intervene to determine the best choice of host for a given protein.

The specific activity of CBHII and EGII were highest when produced in *Y. lipolytica*, being higher than that of EGII expressed by *S. cerevisiae* (213.1 U/mg protein) or CBHII produced by *S. pombe* (1.32-1.56 U/mg protein) (Okada *et al*., 1998b; Qin *et al*., 2008b). Moreover, the overall yield of protein in *Y. lipolytica* was quite high compared to other systems. Like many fungal cellulases, wild type EGII and CBHII are glycoproteins, which may be secreted as multiple molecular mass (Kubicek, 1992). In our study, recombinant EGII and CBHII produced by *Y. lipolytica* exhibited a lower level of glycosylation than homologs produced either by *P. pastoris* (this study) or *S. cerevisiae* (Qin *et al*., 2008b; Ilmén *et al*., 2011; Penttilä *et al*., 1988). The glycosylation pattern of *Y. lipolytica* usually contains 8-10 mannose residues, which is closer to that found in native proteins, compared to *S. cerevisiae* (50-150 mannose residues) or *P. pastoris* (8-14 mannose residues). The advantage of limited glycosylation is that the properties of recombinant proteins should not be adversely affected, when compared to wild type counterparts (Qin *et al*., 2008b; Madzak *et al*., 2004; Masárová *et al*., 2001). Typically, this was the case in our study, because the essential biochemical characteristics (pH and temperature optima etc.) were highly similar to those reported for wild type EGII and CBHII (Okada *et al*., 1998b; Qin *et al*., 2008b).

PART III Directed evolution of *Trichoderma reesei* endoglucanase II to improve its thermostability in *Yarrowia lipolytica*

3.1 Construction of mutant Cel5A libraries using error-prone PCR, screening, and characterization

In this study, thermostability of *T. reesei* Cel5A was improved by directed evolution in *Y. lipolytica*. Random mutagenesis was carried out by using 3-step PCR method. The error-prone PCR library of *egl2* containing zeta docking platform was successfully amplified (Figure IV-12).

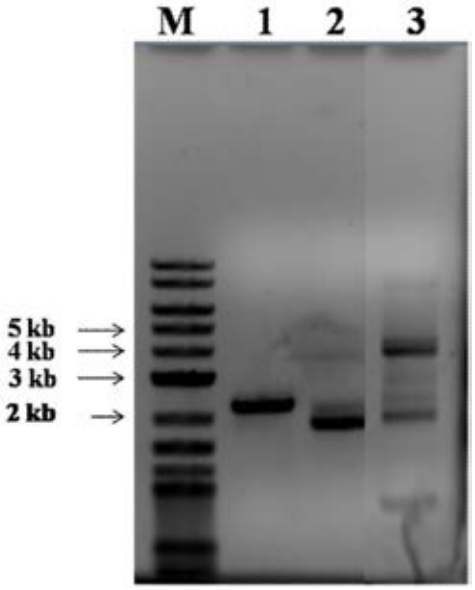


Figure IV-12. Agarose gel electrophoresis of amplified mutant *egl*2 containing zeta docking region: M, 1 kb molecular weight marker; lane 1. first high fidelity PCR; lane 2, second error-prone PCR using GeneMorph kit; lane 3, third fusion PCR.

After transformation of *egl2* library fragments into the yeast, mutant *Y. lipolytica* was spreaded onto YNBcasaD in Q-trays and showed homogeneous size of colonies in the plates (Figure IV-13) which were the results from the integration of a zeta docking platform in the *LEU2* locus of the genome. Thus, integration of the expression cassette could occur at a targeted locus, entailing the homologous and monocopy integration always at the same position in the genome (Bordes *et al*., 2007).

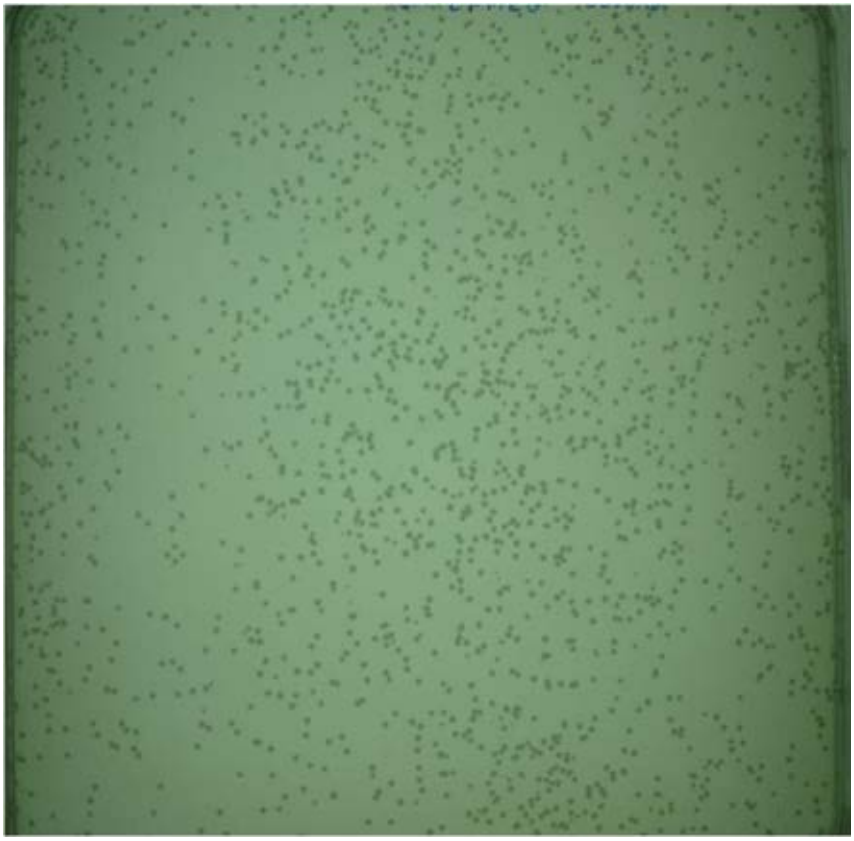


Figure IV-13. Homogeneous size of mutant *Y. lipolytica* colonies spreaded onto YNBcasaD in Q-trays plates.

During the screening process, two deep-well plates of *Y. lipolytica* culture transformed with wild type JMP62UraExTEF_fusEGII were included as an internal control. The $\mu \pm \sigma$ values of the wild-type controls in the three microplates were 0.53 ± 0.06. Among the 4,000 clones from the mutant library, seventeen mutants displaying an activity after heat treatment greater, between, or less than $\mu \pm 2\sigma$ of wild-type activity were named "G", "M", "O" mutants, respectively (Table IV-2), and selected for further DNA sequencing and performed in twenty-four wells culture to confirm the thermostability result. DNA sequence analysis of seventeen mutant clones revealed an average mutation rate of 2.2 base substitutions per kilobase, resulting in one to three amino acid substitutions and one silent mutations, and a transition/transversion ratio was about 1.4 indicating that the mutations were relatively unbiased in this study. Twenty-four wells culture result confirmed their thermostable property, and the most three thermostable clones (G1, G4, G7) showed the ratio of OD_{540} after heat treatment higher than 0.76.

Table IV-2. Screening results of mutant *egl2* library using error-prone PCR and their DNA sequences

Name of mutant	OD 540 before heat treatment	Ratio of OD 540 after heat treatment	Number of mutant base	Mutant1	Mutant2	Mutant3	Mutant4
G1 (28B4)	0,83±0,03	0,81±0,05	4	I(att)32F(ttt)	S(tca)44S(tct)	D(ttg)169N(tta)	T(cac)337S(cag)
G2 (29B3)	0,84±0,05	0,68±0,05	1	D(gat)298N(aat)			
G3 (34C2)	0,85±0,05	0,71±0,05	2	S(agc)172T(acc)	G(ggt)202D(gat)		
G4 (48B5)	0,85±0,05	0,83±0,03	1	T(acc)257N(aac)			
G5 (30B6)	0,71±0,07	0,66±0,11	3	V(gtc)25I(atc)	V(gtc)142V(gtt)	I(atc)204N(aac)	
G6 (39D11)	0,67±0,03	0,74±0,08	1	N(aat)196D(gat)			
G7 (45F3)	0,89±0,03	0,76±0,04	2	I(atc)63N(aac)	V(gtc)142I(atc)	Q(caa)368Q(cag)	
G8 (38G6)	0.88±0.04	0.67±0.05	2	F(ttc)125L(ctc)	L(ctt)168V(gtt)		
G9 (38G7)	0.81±0.03	0.70±0.02	4	N(aac)129I(atc)	Q(cag)229E(gag)	V(gtc)294V(gtt)	Y(tat)369F(ttt)
G10 (42F10)	0.80±0.03	0.71±0.03	2	G(ggt)202D(gat)	Q(cag)221Q(caa)		
G11 (48E5)	0.88±0.03	0.78±0.05	1	T(acc)257N(aac)			
M1 (35E7)	0,91±0,06	0,43±0,11	1	P(ccc)132H(cac)			
M2 (39B8)	0.87±0.05	0.62±0.05	3	G(ggt)31D(gat)	H(cac)320H(cat)	A(gcc)321R(acc)	
M3 (44C4)	0.85±0.03	0.57±0.03	0				
M4 (48B6)	0.86±0.04	0.64±0.03	3	G(ggc)43G(ggt)	S(agc)83N(aac)	T(act)324T(acc)	
O1 (45E5)	0,21±0,01		3	P(ccc)88L(ctc)	R(cga)199P(cca)	N(aac)342N(aat)	
O2 (45E9)	0,25±0,03		2	N(aat)39Y(tat)	Q(cag)345E(gag)		
UraExTEF_fusEGII	0,63±0,1	0,53±0,06					

The most three thermostable clones and one mutant having the highest OD_{540} before heat treatment were subjected to study their thermostability profiles. Three thermostable mutants showed more than 50% of residual activity after incubated the enzyme at 70 °C for 2 h whereas the wild-type and M1 mutant were lost almost of their enzymatic activity (Figure IV-14). DNA sequencing revealed G1 mutant had three amino acid substitutions and one silent mutation, G4 had only single amino acid substitution, and G7 mutant had two amino acid substitutions and one silent mutation. As the best thermostable mutant and only single amino acid was change (Threonine to Asparagine), the mutant G4 was subjected to further evolve via site-saturation mutagenesis.

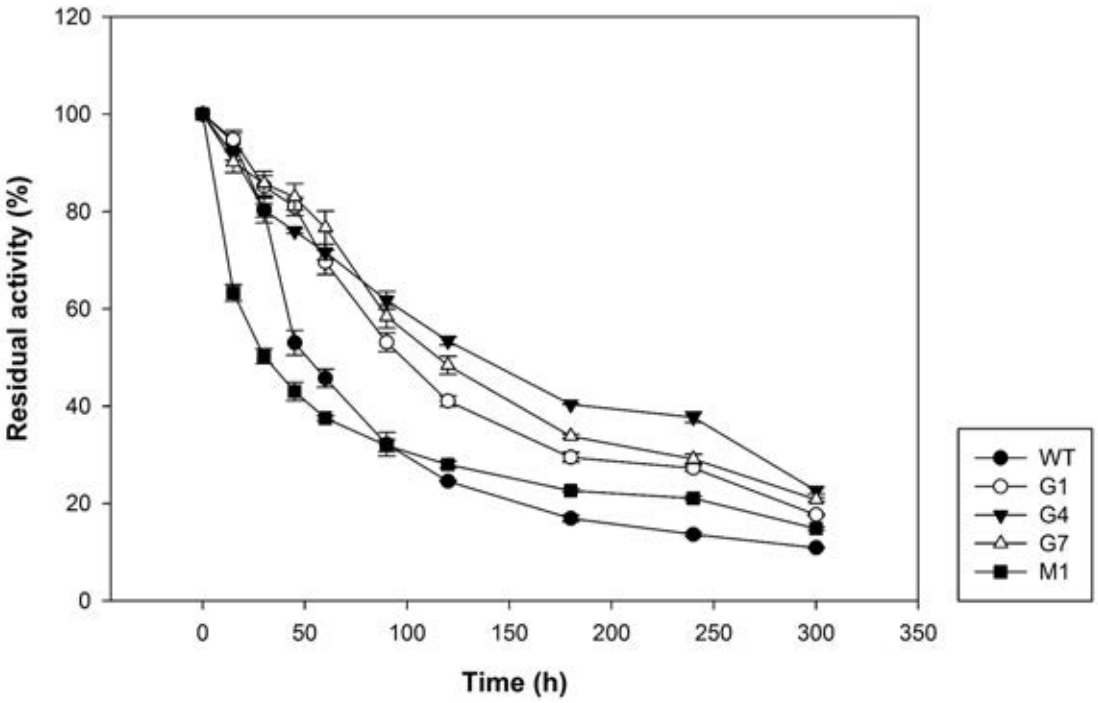


Figure IV-14. Thermostability profile at 70 °C of purified wild-type EGII and their mutants evolved via error-prone PCR .

3.2 Site-saturation mutagenesis and characterization of the variants

Among 4,000 clones screened in the first-round of evolution, G4 mutant was the best clone having good thermostable profile. The amino acid at position 257 was pinpointed as potentially interesting locations. Eighteen primers were designed to change the residue using QuikChange® Lightning Site-Directed Mutagenesis Kit.

DNA sequencing was performed to confirm the change of this residue to other remaining 18 amino acids before transformed the mutants into *Y. lipolytica*. After second round of evolution, thermostable profile divided the changes of amino acid into four groups. The first group was lost 50% of its activity at 15 min containing the change of the residue to phenylalanine, isoleucine, and tyrosine. Cysteine, histidine, glycine, proline, and glutamine were second group which lost 50% of its activity within 30 min. Loss of the same percentage within 60 min was the third group containing wild-type threonine and alanine mutant. Finally, aspartic acid and asparagines were the best residues at this position which half of its activity was lost within 120 min (Figure IV-15). However, there are remaining eight amino acid residues that have not activity after transformation.

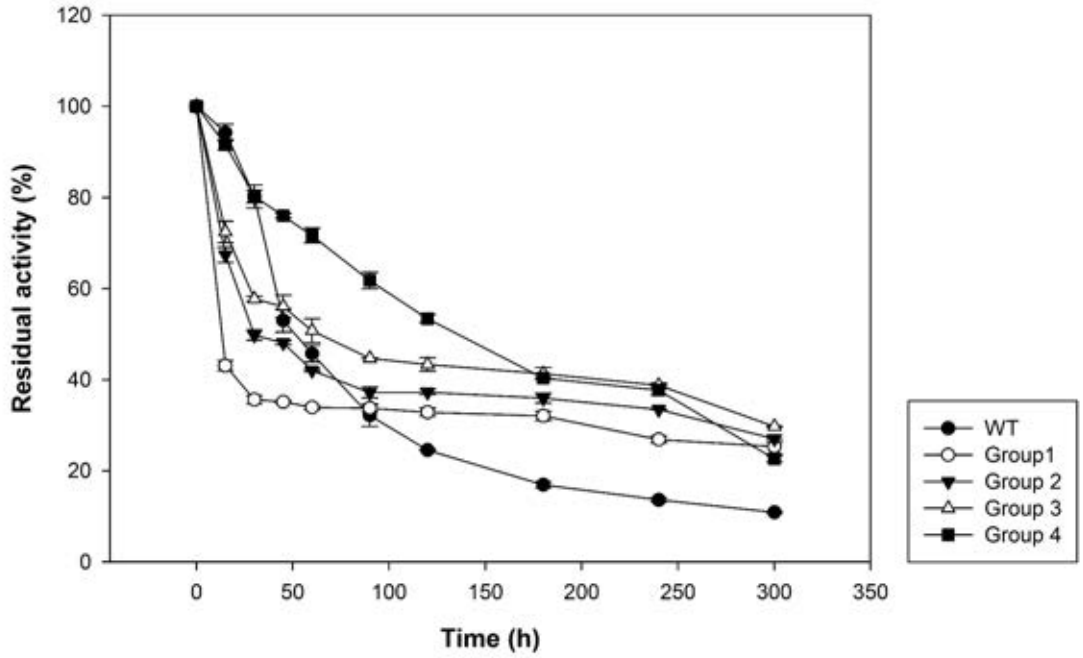


Figure IV-15. Thermostability profile at 70 °C of purified wild-type EGII and their mutants evolved by site-saturation mutagenesis.

PART IV Application of the recombinant cellulases in ethanol fermentation of rice straw

4.1 Enzyme production

The enzymatic activity of endoglucanase, cellobiohydrolase, glucan 1,3-beta-glucosidase A produced from the recombinant *P. pastoris* were assayed using CMC, PASC, pNPβG as the substrates, and showed the activities of 70±3.22, 25±2.21 (after concentration), 2.9±0.19, respectively. The commercial cellulase and xylanase were also measured using filter paper and xylan as the substrates, respectively. The results revealed that cellulase and xylanase showed the activities of 82±4.15 and 296±20.82 U/ml, respectively.

4.2 Enzymatic hydrolysis and co-fermentation

Hydrolysis of the main composition, cellulose and hemicellulose, into glucose and xylose as the substrates of fermentation is necessarily required for the production of ethanol. In this research, the mixture of 2 types of yeasts were used, *S. cerevisiae* and *P. stipitis* since the former organism can use only glucose while the latter can catalyze both glucose and xylose as the substrates. Nonetheless, *P. stipitis* showed slower rate of fermentation than *S. cerevisiae* (Taniguchi *et al*., 1997). There are various methods to achieve the fermentation but SSCF has been shown to be more efficient than the others from the fact that the raw materials can be hydrolyzed to hexose and pentose for further ethanol fermentation in single step. The gradual release of glucose from the hydrolysis can be continuously used for the fermentation resulting in the low concentration of glucose which yields better fermentation for xylose (Ohgren *et al*., 2006).

When the alkaline peroxide pretreated rice straw was fermented by SSCF using 2 types of yeasts; *S. cerevisiae* and *P. stipitis* together with cellulase enzymes obtained from recombinant enzyme production, and commercial xylanase at 35°C for 5 days, the highest yield of ethanol was obtained about 0.86 g/L or 0.71 g/g substrate.

This indicated that the rice straw may contain certain chemicals such as tannin or silica in the quantities that could possibly affect the activities of the enzyme and the growth of the yeasts resulting in the reduction of the ethanol production obtained from

fermentation. Moreover, the presence of other substances toxic to the microorganisms in the system, such as the derivatives of lignin which are the phenolic compounds, may reduce the efficiency of the fermentation (Nigam, 2001).

However, the rice straw should have potential to be developed for the production of ethanol by this SSCF fermentation. Further studies can be carried out on the optimal conditions of fermentation to increase the production of ethanol approximately closed to the theoretical values.

CHAPTER V

CONCLUSION AND FUTURE PERSPECTIVES

Lignocellulosic biomass is an attractive source of biorefinery applications because it is renewable, non-food raw materials, inexpensive, and environmental friendly. Due to the complexity and recalcitrant structure of the polymer, it requires collective groups of enzyme known as cellulases to work synergistically for complete degradation. The filamentous fungus *T. reesei* is an efficient colonizer of the biomass producing all the enzymes needed for efficient breakdown of crystalline cellulose to glucose, except β-glucosidases which partly mycelium-bound and presented in relatively small amount when compared to *Aspergillus* sp. Nonetheless, the utilization of these enzymes is not currently in commercial production and still requires research and development. Therefore, isolating of novel cellulases and protein engineering are now considered as big challenges to improve the activity of enzymes and their thermostability, as well as to make them resistant to extreme pH or tolerant to organic solvents for both academic purposes and industrial applications. Recently, the priority research is focusing on the development of recombinant microorganism with the objectives to improve their level of enzymatic expression, and adapt them to extreme industrial conditions for cost-effective process. *Y. lipolytica* is non-conventional yeast dedicated for powerful expression system and high-throughput screening of mutant library via directed evolution. This method uses random and combinatorial approaches which mimic the natural evolution and do not require sufficient insight into the structure – function relationships to introduce mutations. To summarize, the following can be concluded from this thesis:

1. Cloning and expression of *A. oryzae* glucan 1,3-beta-glucosidase in *P. pastoris*

P. pastoris is the host of choice for heterologous expression because of its high secretion ability with low level of endogenous protein, and the initial purity of the product in the culture medium is higher when it is induced by high concentration of methanol led to the ease of purification process. A glucan 1,3-beta-glucosidase A gene (*exgA*) from *A. oryzae* using *S. cerevisiae* signal peptide (α-factor) was favorably

expressed under the control of either constitutive or inducible promoter in *P. pastoris*. The expression level of recombinant ExgA under the *AOX*1 promoter was higher than *GAP* promoter. Purified ExgA was stable at temperature ranging from 20-45°C and over the range pH 5-8. The recombinant enzyme also exhibited the molecular weight similar to the mature peptide of the enzyme from *A. oryzae*, and tolerated to glucose inhibition with the kinetic values of K_i, K_M, V_{max}, and k_{cat} of the enzyme for pNPβG were 365 mM, 0.56 mM, 10042 μmol min^{-1} mg of $protein^{-1}$, 410 s^{-1}, respectively. The limited of ExgA activity at high temperature indicates that there is room for improvement for the ExgA by means of random mutagenesis. Moreover, the success to express of individual cellulase could be further studies by optimizing the cocktail mixture to achieve the most effective synergistic activity.

2. Comparison of *T. reesei* endoglucanase II and cellobiohydrolase II expression level in the yeast *P. pastoris* and *Y. lipolytica*

The level of recombinant enzyme expression is affected by many factors from its gene, promoters, secretion signals, hosts, through protein stability. Therefore, a numbers of recombinant host and genetic elements should be tested to optimize the expression level and low chance of failure. The sequences encoding of *T. reesei egl*II and *cbh*II were successfully cloned and expressed in *Y. lipolytica*. This is the first report that these cellulase genes were heterologously expressed in this yeast. Two different *Y. lipolytica* strains (zeta and Po1d), promoters (*POX* and *TEF*) and secretion signals (native and preproLip2) were also employed to construct recombinant yeast expression strains. Enzymatic activity of both proteins was maximized in *Y. lipolytica* Po1d strain using constitutive promoter and preoproLip2 secretion signal. Endoglucanase activity produced by optimized strain was similar to its homolog expressed in *P. pastoris*. However, when recombinant *P. pastoris* was induced by 3% (v/v) methanol, the activity was increased from the same strain by about seven-fold. In contrast, cellobiohydrolase activity expressed in *Y. lipolytica* was higher than in *P. pastoris* about 30%. Transformed *Y. lipolytica* produced up to 15 mg/L endoglucanase and 50 mg/L cellobiohydrolase within 48 h when compared to 96 h in *P. pastoris*, with the specific activity of both proteins being greater than their homologs produced by *P. pastoris*. Recombinant proteins produced by *Y. lipolytica*

were glycosylated but had a little effect on their enzymatic activity and properties. These results suggested that *Y. lipolytica* is potentially an alternative host for heterologous expression and high-throughput screening in the following experiments.

3. Directed evolution of *T. reesei* endoglucanase II to improve its thermostability in *Y. lipolytica*

In addition to overexpress EGII, the thermostability of EGII was also improved by directed evolution in *Y. lipolytica*. Two different mutagenesis strategies, random- and site-saturation mutagenesis, were used to improve the desired property. The mutant *egl*2 library was constructed under constitutive promoter, cultured in glycerol as the carbon source, and screening the mutants by measuring the ratio of OD_{540} after incubated the enzyme at 70 °C for 45 min using a micro-3,5-dinitrosalicylic acid assay. The results demonstrated that the amino acid at position 257 is responsible for thermostability. Variants T257N and T257D were the best thermostable mutants which half of its activity was lost at 70 °C within 120 min. For further studies, various recombinations of individual changes could enable a more thorough characterization and quantification of the role of co-operativity in protein stability. The developed method in this study showed that *Y. lipolytica* is potentially an excellent and attractive system for heterologous expression, and it can be adapted to obtain mutants that are advantageous for industrial applications such as activity, pH and temperature optima, stability, and substrate specificity through high-throughput screening.

4. Application of the recombinant cellulases in ethanol fermentation of rice straw

The recombinant *T. reesei* endoglucanase II (EGII) and cellobiohydrolase II (CBHII), and *A. oryzae* glucan 1,3-beta-glucosidase A (ExgA) previously produced in *P. pastoris* were solely utilized or in combination with commercial cellulase supplemented with xylanase for hydrolysis of the alkaline peroxide pretreated rice straw. The yeasts, *S. cerevisiae* and *P. stipitis* were simultaneously applied for co-fermentation at 35 °C for 5 days. From the results, the highest yield of ethanol was obtained about 0.86 g/L or 0.71 g/g substrate indicating the rice straw should have potential to be developed for the production of ethanol by this SSCF fermentation.

REFERENCES

Abdeljabbar, D. M., Song, H. J., and Link, A. J. (2012). *Trichoderma reesei* cellobiohydrolase II is associated with the outer membrane when overexpressed in *Escherichia coli*. *Biotechnology Letters*. 34: 91-96.

Adachi, T., Ito, J., Kawata, K., Kaya, M., Ishida, H., Sahara, H., Hata, Y., Ogino, C., Fukuda, H., and Kondo, A. (2008). Construction of an *Aspergillus oryzae* cell-surface display system using a putative GPI-anchored protein. *Applied Microbiology and Biotechnology*. 81: 711-719.

Agbor, V. B., Cicek, N., Sparling, R., Berlin, A., and Levin, D. B. (2011). Biomass pretreatment: Fundamentals toward application. *Biotechnology Advances*. 29: 675-685.

Araujo, E. F., Barros, E. G., Caldas, R. A., and Silva, D. O. (1983). Beta-glucosidase activity of a thermophilic cellulolytic fungus, *Humicola* sp. *Biotechnology Letters*. 5(11): 781-784.

Badieyan, S., Bevan, D. R., and Zhang, C. (2012). Study and design of stability in GH5 cellulases. *Biotechnology and Bioengineering*. 109(1): 31-44.

Balat, M., Balat. H., and Öz, C. (2008). Progress in bioethanol processing. *Progress in Energy and Combustion Science*. 34(5): 551-573.

Bankar, A. V., Kumar, A. R., and Zinjarde, S. S. (2009). Environmental and industrial applications of *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology*. 84: 847-865.

Barr, B. K., Hsieh, Y. L., Ganem, B., and Wilson, D. B. (1996). Identification of two functionally different classes of exocellulases. *Biochemistry*. 35(2): 586-592.

Barth, G., and Gaillardin, C. (1997). Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica. FEMS Microbiology Reviews*. 19: 219-237.

Bayer, E. A., Chanzyt, H., Lamed, R., and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Structural Biology*. 8: 548-557.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J. L., Molina-Jouve, C., and Nicaud, J. M. (2009). *Yarrowia lipolytica* as a model for bio-oil production. *Progress in Lipid Research*. 48: 375-387.

Berlin, A., Gilkes, N., Kilburn, D., Bura, R., Markov, A., Skomarovsky, A., Okunev, O., Gusakov, A., Maximenko, V., Gregg, D., Sinitsyn, A., and Saddler, J. (2005). Evaluation of novel fungal cellulase preparations for ability to hydrolyze softwood substrates – evidence for the role of accessory enzymes. *Enzyme and Microbial Technology*. 37: 175-184.

Bhat, M. K. (2000). Cellulases and related enzymes in biotechnology. *Biotechnology Advances*. 18: 355-383.

Bidlack, J., Malone, M., and Benson, R. (1992). Molecular structure and component integration of secondary cell walls in plants. *Proceedings of the Oklahoma Academy of Science*. 72: 51-56.

Biely, P., Vrsanska, M., and Claeyssens, M. (1991). The endo-l,4-β-glucanase I from *Trichoderma reesei* action on β-1,4-oligomers and polymers derived from D-glucose and D-xylose. *European Journal of Biochemistry*. 200: 157-163.

Blanchin-Roland, S., Otero, R. R. C., and Gaillardin, C. (1994). Two upstream activation sequences control the expression of the XPR2 gene in the yeast *Yarrowia lipolytica*. *Molecular and Cellular Biology*. 14(1): 327-338.

Boer, H., Teeri, T. T., and Koivula, A. (2000). Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters. *Biotechnology and Bioengineering*. 69: 486-494.

Boisset, C., Fraschini, C., Schülein, M., Henrissat, B., and Chanzy, H. (2000). Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A From *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. *Applied and Environmental Microbiology*. 66(4): 1444-1452.

Bollok, M., Resina, D., Valero, F., and Ferrer, P. (2009). Recent patents on the *Pichia pastoris* expression system: expanding the toolbox for recombinant protein production. *Recent Patents on Biotechnology*. 3: 192-201.

Boonvitthya, N., Bozonnet, S., Burapatana, V., O'Donohue, M. J., Chulalaksananukul, W. (2012a). Comparison of the heterologous expression of *Trichoderma reesei* endoglucanase II and cellobiohydrolase II in the yeasts *Pichia pastoris* and *Yarrowia lipolytica*. *Molecular biotechnology*. (*In press*)

Boonvitthya, N., Tanapong, P., Kanngan, P., Burapatana, V., and Chulalaksananukul, W. (2012b). Cloning and expression of *Aspergillus oryzae* glucan 1,3-beta-glucosidase A in *Pichia pastoris*. (Submitted)

Bordes, F., Fudalej, F., Dossat, V., Nicaud, J. M., and Marty, A. (2007). A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *Journal of Microbiological Methods*. 70: 493-502.

Bornscheuer, U. T., and Pohl, M. (2001). Improved biocatalysts by directed evolution and rational protein design. *Chemical Biology*. 5: 137-143.

Casaregola, S., Lépingle, A., Bon, E., Neuvéglise, C., Nguyen, H. V., Artiguenave, F., Wincker, P., and Gaillardin, C. (2000). Genomic exploration of the hemiascomycetous yeasts: 7. *Saccharomyces servazzii*. *FEBS Letters*. 487: 47-51.

Cavallius, J., Zoll, W., Chakraburtty, K., and Merrick, W. C. (1993). Characterization of yeast EF-1α: Non-conservation of post-translational modifications. *Biochimica et Biophysica Acta*. 1163: 75-80.

Centre national de la recherche scientifique, Organization. (2007). Carbohydrate-active enzymes database. [Online]. Retrieved from; http://www.cazy.org/Home.html/. [2012, January 16]

Cereghino, J. L., and Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews*. 24: 45-66.

Cereghino, G. P. L., Sunga, A. J., Cereghino, J. L., and Cregg, J. M. (2001). *Genetic Engineering*: *Expression of foreign genes in the yeast Pichia pastoris,* vol. 23. New York: Kluwer Academic/Plenum Publishers.

Chadha, B. S., Kanwar, S. S., and Garcha, H. S. (1995). Simultaneous saccharification and fermentation of rice straw into ethanol. *Acta Microbiologica et Immunologica Hungarica*. 42(1): 71–75.

Chandel, A. K., and Singh, O. V. (2011). Weedy lignocellulosic feedstock and microbial metabolic engineering: advancing the generation of 'Biofuel'. *Applied Microbiology and Biotechnology.* 89: 1289-1303.

Chandrakant, P., and Bisaria, V. S. (1998). Simultaneous bioconversion of cellulose and hemicellulose to ethanol. *Critical Reviews in Biotechnology.* 18(4): 295-331.

Chang, M. C. Y. (2007). Harnessing energy from plant biomass. *Chemical Biology*. 11: 677-684.

Chang, T., and Yao, S. (2011). Thermophilic, lignocellulolytic bacteria for ethanol production: Current state and perspectives. *Applied Microbiology and Biotechnology*. 92: 13-27.

Chen, H., Han, Y., and Xu, J. (2008). Simultaneous saccharification and fermentation of steam exploded wheat straw pretreated with alkaline peroxide. *Process Biochemistry*. 43(12): 1462-1466.

Cheng, K. K., Zhang, J. A., Ping, W. X., Ge, J. P., Zhou, Y. J., Ling, H. Z., and Xu, J. M. (2008). Sugarcane bagasse mild alkaline/oxidative pretreatment for ethanol production by alkaline recycle process. *Applied Biochemistry and Biotechnology*. 151(1): 43-50.

Choi, N. S., Kim, B. H., Park, C. S., Han, Y. J., Lee, H. W., Choi, J. H., Lee, S. G., and Song, J. J. (2009). Multiple-layer substrate zymography for detection of several enzymes in a single sodium dodecyl sulfate gel. *Analytical Biochemistry*. 386: 121-122.

Chu, B. C. H. and Lee, H. (2007). Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation. *Biotechnology Advances*. 25(5): 425-441.

Claeyssens, M., Tilbeurgh, H. V., Kamerling, J. P., Berg, J., Vrsanskai, M., and Bielyi, P. (1990). Studies of the cellulolytic system of the filamentous fungus *Trichoderma reesei* QM 9414 substrate specificity and transfer activity of endoglucanase I. *Biochemical Journal.* 270: 251-256.

Covert, S. E., Bolduc, J., and Cullen, D. (1992). Genomic organization of a cellulase gene family in *Phanerochaete chrysosporium. Current Genetics*. 22: 407-413.

Cregg, J. M., Cereghino, J. L., Shi, J., and Higgins, D. R. (2000). Recombinant protein expression in *Pichia pastoris*. *Molecular Biotechnology*. 16: 23-52.

Cregg, J. M. (2007). *Methods in Molecular Biology: Pichia Protocols*. 2nd. vol. 389. New Jersey: Humana Press, Inc.

Cregg, J. M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K., and Chappell, T. (2009). Expression in the yeast *Pichia pastoris*. *Methods in Enzymology*. 463: 169-189.

Curreli, N., Fadda, M. B., Rescigno, A., Rinaldi, A. C., Soddu, G., Sollai, F., Vaccargiu, S., Sanjust, E., and Rinaldi, A. (1997). Mild alkaline/oxidative pretreatment of wheat straw. *Process Biochemistry*. 32(8): 665–670.

Dalby, P. A. (2007). Engineering enzymes for biocatalysis. *Recent Patents on Biotechnology*. 1: 1-9.

Daly, R., and Hearn, M. T. W. (2005). Expression of heterologous proteins in *Pichia pastoris*: A useful experimental tool in protein engineering and production. *Journal of Molecular Recognition*. 18: 119-138.

Dan, S., Marton, I., Dekel, M., Bravdo, B., He, S., Withers, S. G., and Shoseyov, O. (2000). Cloning, expression, characterization, and nucleophile identification of Family 3, *Aspergillus niger* β-glucosidase. *The Journal of Biological Chemistry*. 275(7): 4973-4980.

Dashtban, M., Schraft, H., and Qin, W. (2009). Fungal bioconversion of lignocellulosic residues; opportunities and perspectives. *International Journal of Biological Sciences*. 5(6): 578-595

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*. 3(9): 853-859.

de Vries, R. P., and Visser, J. (2001). *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiology and Molecular Biology Reviews*. 65(4): 497-522.

Doi, R. H. (2008). Cellulases of mesophilic microorganisms cellulosome and noncellulosome producers. *Annals of the New York Academy of Sciences*. 1125: 267-279.

Domínguez, Á., Fermiñán, E., Sánchez, M., González, F. J., Pérez-Campo, F. M., García, S., Herrero, A. B., Vicente, A. S., Cabello, J., Prado, M., Iglesias, F. J., Choupina, A., Burguillo, F. J., Fernández-Lago, L., and López, M. C. (1998). Non-conventional yeasts as hosts for heterologous protein production. *International Microbiology*. 1: 131-142.

Dubois, B., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956). Colormetric method for determination of sugars and relative substances. *Analytical Chemistry*. 28: 350-356.

Edlund, U., Svensson, M., and Albertsson, A. (2012). Microsphere valorization of forestry derived hydrolysates. *European Polymer Journal*. 48: 372-383.

Eijsink, V. G. H., Gåseidnes, S., Borchert, T. V., Burg, B. V. D. (2005). Directed evolution of enzyme stability. *Biomolecular Engineering*. 22: 21-30.

Fan, L. T., Lee, Y., and Gharpuray, M. M. (1982). The Nature of lignocellulosics and their pretreatments for enzymatic hydrolysis. *Advances in Biochemical Engineering/Biotechnology*. 23: 157-187.

Fenel, F., Leisola, M., Jänis, J., and Turunen, O. (2004). A de novo designed N-terminal disulphide bridge stabilizes the *Trichoderma reesei* endo-1,4-xylanase II. *Journal of Biotechnology*. 108: 137-143.

Fujita, Y., Takahashi, S., Ueda, M., Tanaka, A., Hirofumi, O., Morikawa, Y., Kawaguchi, T., Arai, M., Fukuda, H., and Kondo, A. (2002). Direct and efficient production of ethanol from cellulosic material with a yeast strain displaying cellulolytic enzymes. *Applied and Environmental Microbiology*. 68(10): 5136-5141.

Fujita, Y., Ito, J., Ueda, M., Fukuda, H., and Kondo, A. (2004). Synergistic saccharification, and direct fermentation to ethanol, of amorphous cellulose by use of an engineered yeast strain codisplaying three types of cellulolytic enzyme. *Applied and Environmental Microbiology*. 70(2): 1207-1212.

Gaillardin, C., and Ribet, A. M. (1987). *LEU2* directed expression of β-galactosidase activity and phleomycin resistance in *Yarrowia lipolytica*. *Current Genetics*. 11: 369-375.

Gellissen, G., Kunze, G., Gaillardin, C., Cregg, J. M., Berardi, E., Veenhuis, M., and Klei, I. V. D. (2005). New yeast expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adeninivorans* and *Yarrowia lipolytica* – A comparison. *FEMS Yeast Research*. 5: 1079-1096.

Ghose, T. K. (1987). Measurement of cellulase activities. *Pure and Applied Chemistry*. 59 (2): 257-268

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller Jr, R. C., and Warren, R. A. J. (1991). Domains in microbial beta-1, 4-glycanases: Sequence conservation, function, and enzyme families. *Microbiological Reviews*. 55(2): 303-315.

Girio, F.M., Fonseca, C., Carvalheiro, F., Duarte, L.C., Marques, S., and Bogel-Łukasik, R. (2010). Hemicelluloses for fuel ethanol: A review. *Bioresource Technology*. 101: 4775-4800.

Goldschmidt, F. (2008). From Cellulose to Ethanol: Engineering Microorganisms to Produce Biofuel. [Online]. Retrieved from; http://www.ibp.ethz.ch/research/aquaticchemistry/teaching/archive_past_lectures/term_paper_08_09/HS08_GOLDSCHMIDT_rev_termpaper.pdf/[2011, July 25]

Gould, J. M. (1984). Alkaline peroxide delignification of agricultural residues to enhance enzymatic saccharification. *Biotechnology and Bioengineering*. 26(1): 46–52.

Gray, K. A., Zhao, L., and Emptage, M. (2006). Bioethanol. *Current Opinion in Chemical Biology*. 10(2): 141-146.

Gusakov, A. V. (2011). Alternatives to *Trichoderma reesei* in biofuel production. *Trends in Biotechnology*. 29(9): 419- 425.

Hamelinck, C. N., van Hooijdonk, G., and Faaij, A. P. C. (2005). Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. *Biomass and Bioenergy*. 28(4): 384–410.

Harrison, M. J., Nouwens, A. S., Jardine, D. R., Zachara, N. E., Gooley, A. A., Nevalainen, H., and Packer, N. H. (1998). Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of *Trichoderma reesei*. *European Journal of Biochemistry*. 256(1): 119-127.

Have, A. T., Tenberge, K. B., Benen, J. A. E., Tudzynski, P., Visser, J., and van Kan, J. A. L. (2002). *The Mycota XI: Agricultural Applications*. Verlag Berlin Heidelberg: Springer.

Hendriks, A. T. W. M., and Zeeman, G. (2009). Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technology*. 100: 10-18.

Henrissat, B., Driguez, H., Viet, C., and Schülein, M. (1985). Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Nature Biotechnology*. 3: 722-726.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal.* 280: 309-316.

Herpoël-Gimbert, I., Margeot, A., Dolla, A., Jan, G., Mollé, D., Lignon, S., Mathis, H., Sigoillot, J., Monot, F., and Asther, M. (2008). Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. *Biotechnology for Biofuels*. 1(18): 1-12.

Higgins, D. R., and Cregg, J. M. (1998). *Methods in Molecular Biology*: *Introduction to Pichia pastoris* (Pichia Protocols). vol. 103. New Jersey: Humana Press, Inc.

Hong, J., Wang, Y., Kumagai, H., and Tamaki, H. (2007). Construction of thermotolerant yeast expressing thermostable cellulase genes. *Journal of Biotechnology*. 130: 114-123.

Igarashi, K., Wada, M., Hori, R., and Samejima, M. (2006). Surface density of cellobiohydrolase on crystalline celluloses. A critical parameter to evaluate enzymatic kinetics at a solid–liquid Interface. *Federation of European Biochemical Societies Journal*. 273: 2869-2878.

Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttila, M., and Samejima, M. (2009). High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolaseion I on crystalline cellulose. *The Journal of Biological Chemistry*. 284(52): 36186–36190.

Ilmén, M., den Haan, R., Brevnova, E., McBride, J., Wiswall, E., Froehlich, A., Koivula, A., Voutilainen, S. P., Siika-Aho, M., la Grange, D. C., Thorngren, N., Ahlgren, S., Mellon, M., Deleault, K., Rajgarhia, V., van Zyl, W. H., and Penttilä, M. (2011). High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. *Biotechnology for Biofuels.* 4(30): 1-15.

Ito, J., Fujita, Y., Ueda, M., Fukuda, H., and Kondo, A. (2004). Improvement of cellulose-degrading ability of a yeast strain displaying *Trichoderma reesei* endoglucanase II by recombination of cellulose-binding domains. *Biotechnology Progress*. 20: 688-691.

Kadam, K. L., Forrest, L. H., and Jacobson, W. A. (2000). Rice straw as a lignocellulosic resource: collection, processing, transportation, and environmental aspects. *Biomass and Bioenergy*. 18(5): 369-389.

Karimi, K., Emtiazi, G., and Taherzadeh, M. J. (2006). Ethanol production from diluted-acid pretreated rice straw by simultaneous saccharification and fermentation with *Mucor indicus*, *Rhizopus oryzae*, and *Saccharomyces cerevisiae*. *Enzyme and Microbial Technology*. 40(1): 138-144.

Karlsson, J., Siika-aho, M., Tenkanen, M., and Tjerneld, F. (2002). Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of *Trichoderma reesei*. *Journal of Biotechnology*. 99: 63-78.

Karmakar, M., and Ray, R. R. (2011). Current trends in research and application of microbial cellulases. *Research Journal of Microbiology*. 6(1): 41-53.

Kaya, M., Ito, J., Kotaka, A., Matsumura, K., Bando, H., Sahara, H., Ogino, C., Shibasaki, S., Kuroda, K., Ueda, M., Kondo, A., and Hata, Y. (2008). Isoflavone aglycones production from isoflavone glycosides by display of *β*-glucosidase from *Aspergillus oryzae* on yeast cell surface. *Applied Microbiology and Biotechnology*. 79:51-60.

Kerscher, S., Dröse, S., Zwicker, K., Zickermann, V., and Brandt, U. (2002). *Yarrowia lipolytica*, a yeast genetic system to study mitochondrial complex I. *Biochimica et Biophysica Acta*. 1555: 83-91.

Kim, Y. S., Jung, H. C., and Pan, J. G. (2000). Bacterial cell surface display of an enzyme library for selective screening of improved cellulase variants. *Applied and Environmental Microbiology*. 66(2): 788-793.

Kipper, K., Väljamäe, P., and Johansson, G. (2005). Processive action of cellobiohydrolase Cel7A from *Trichoderma reesei* is revealed as 'burst' kinetics on fluorescent polymeric model substrates. *Biochemical Journal*. 385: 527-535.

Klemm, D., Heublein, B., Fink, H., and Bohn, A. (2005). Cellulose: Fascinating biopolymer and sustainable raw material. *Angewandte Chemie International Edition*. (44): 3358-3393.

Kleywegt, G. J., Zou, J. Y., Divne, C., Davies, G. J., Sinning, I., Ståhlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T. T., and Jones, T. A. (1997). The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *Journal of Molecular Biology*. 272: 383-397.

Knowles, J., Lehtovaara, P., Penttilä, M., Teeri, T. Harkki, A. and Salovuori, I. (1987). The Cellulase Genes of *Trichoderma*. *Antonie van Leeuwenhoek*. 53(5): 335-341.

Krishna, S. H. and Chowdary, G. V. (2000). Optimization of simultaneous saccharification and fermentation for the production of ethanol from lignocellulosic biomass. *Journal of Agricultural and Food Chemistry*. 48(5): 1971-1976.

Kubicek, C. P. (1992). The cellulase proteins of *Trichoderma reesei*: Structure, multiplicity, mode of action and regulation of formation. *Advances in Biochemical Engineering/Biotechnology*. 45: 1-27.

Kuhad, R. C., Singh, A., and Eriksson, K. L. (1997). Microorganisms and enzymes involved in the degradation of plant fiber cell walls. *Advances in Biochemical Engineering/Biotechnology*. 57: 46-125.

Kumar, R., Singh, S., and Singh, O. V. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of Industrial Microbiology and Biotechnology*. 35: 377-391.

Kumar, P., Barrett, D. M., Delwiche, M. J., and Stroeve, P. (2009). Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Industrial and Engineering Chemistry Research*. 48(8): 3713-3729.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227:680-685.

Langston, J., Sheehy, N., and Xu, F. (2006). Substrate specificity of *Aspergillus oryzae* family 3 β-glucosidase. *Biochimica et Biophysica Acta*. 1764: 972-978.

Laplace, J. M., Delgenes, J. P., Moletta, R., and Navarro, J. M. (1993). Cofermentation of glucose and xylose to ethanol by a respiratory-deficient mutant of *Saccharomyces cerevisiae* co-cultivated with a xylose-fermenting yeast. *Journal of Fermentation and Bioengineering*. 75(3): 207-212.

Larsson, P. T., Wickholm, K., and Iversen, T. (1997). A CP/MAS 13C NMR investigation of molecular ordering in celluloses. *Carbohydrate Research.* 302: 19-25.

Laureano-Perez, L., Teymouri, F., Alizadeh, H., and Dale, B. E. (2005). Understanding factors that limit enzymatic hydrolysis of biomass characterization of pretreated corn stover. *Applied Biochemistry and Biotechnology.* 121-124: 1081-1099.

le Dall, M. T., Nicaud, J. M., and Gaillardin, C. (1994). Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Current Genetics*. 26: 38-44.

Lee, H. J., and Brown Jr, R. M. (1997). A comparative structural characterization of two cellobiohydrolases from *Trichoderma reesei*: A high resolution electron microscopy study. *Journal of Biotechnology*. 57: 127-136.

Lee, T. M., Farrow, M. F., Arnold, F. H., and Mayo, S. L. (2011). A structural study of Hypocrea jecorina Cel5A. *Protein Science.* 20(11): 1935-1940.

Lever, M. (1972). A new reaction for colorimetric determination of carbohydrates next term. *Analytical Biochemistry*. 47: 273-279.

Li, P., Anumanthan, A., Gao, X., Ilangovan, K., Suzara, V. V., Düzgüneş, N., and Renugopalakrishnan, V. (2007). Expression of recombinant proteins in *Pichia Pastoris*. *Applied Biochemistry and Biotechnology*. 142: 105-124.

Li, W. F., Zhou, X. X., and Lu, P. (2005). Structural features of thermozymes. *Biotechnology Advances*. 23: 271-281.

Linder, M., and Teeri, T. T. (1996). The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and a high exchange rate on crystalline cellulose. *Proceedings of the National Academy of Sciences.* 93: 12251-12255.

Liu, D., Zhang, R., Yang, X., Xu, Y., Tang, Z., Tian, W., and Shen, Q. (2011a). Expression, purification and characterization of two thermostable endoglucanases cloned from a lignocellulosic decomposing fungi *Aspergillus fumigatus* Z5 isolated from compost. *Protein Expression and Purification*. 79: 176-186.

Liu, S., Chen, W., Liu, G., and Xing, M. (2011b). Enhanced secreting expression and improved properties of a recombinant alkaline endoglucanase cloned in *Escherichia coli. Journal of Industrial Microbiology and Biotechnology.* 38: 855-861.

Lloyd, T. A., and Wyman, C. E. (2005). Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. *Bioresource Technology*. 96: 1967-1977.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002). Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*. 66(3): 506-577.

Lynd, L. R., van Zyl, W. H., McBride, J. E., and Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: An update. *Biotechnology*. 16: 577-583.

Macauley-Patrick, S., Fazenda, M. L., McNeil, B., and Harvey, L. M. (2005). Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*. 22: 249-270.

Madzak. C., Blanchin-Roland, S., Cordero, O. R. R., and Gaillardin, C. (1999). Functional analysis of upstream regulating regions from the *Yarrowia lipolytica* XPR2 promoter. *Microbiology*. 145: 75-87.

Madzak, C., Tréton, B., and Blanchin-Roland, S. (2000). Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *Journal of Molecular Microbiology and Biotechnology*. 2(2): 207-216.

Madzak, C., Gaillardin, C., and Beckerich, J. M. (2004). Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: A review. *Journal of Biotechnology*. 109: 63-81.

Maheshwari, R., Bharadwaj, G., and Bhat, M. K. (2000). Thermophilic fungi: Their physiology and enzymes. *Microbiology and Molecular Biology Reviews*. 64(3): 461-488.

Maki, M., Leung, K. T., and Qin, W. (2009). The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International Journal of Biological Sciences*. 5(5): 500-516.

Makino, T., Skretas, G., and Georgiou, G. (2011). Strain engineering for improved expression of recombinant proteins in bacteria. *Microbial Cell Factories*. 10(32): 1-10.

Mandels, M., and Weber, J. (1969). The production of cellulase. In cellulases and their applications. *Advances in Chemistry Series*. 95: 391-413.

Mansfield, S. D., Mooney, C., and Saddler, J. N. (1999). Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnology Progress.* 15: 804-816.

Masárová, J., Mislovicová, D., Gemeiner, P., and Michalková, E. (2001). Stability enhancement of *Escherichia coli* penicillin G acylase by glycosylation with yeast mannan. *Biotechnology and Applied Biochemistry*. 34: 127-133.

Mayer, A. F., Hellmuth, K., Schlieker, H., Lopez-Ulibarri, R., Oertel, S., Dahlems, U., Strasser, A. W. M., and van Loon, A. P. G. M. (1999). An expression system matures: a highly efficient and cost-effective process for phytase production by recombinant strains of *Hansenula polymorpha*. *Biotechnology and Bioengineering*. 63: 373–381.

Medve, J., Ståhlberg, J., and Tjerneld, F. (1994). Adsorption and synergism of cellobiohydrolase I and II of *Trichoderma reesei* during hydrolysis of microcrystalline cellulose. *Biotechnology and Bioengineering*. 44(9): 1064-1073.

Medve, J., Karlsson, J., Lee, D., and Tjerneld, F. (1998). Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: Adsorption, sugar production pattern, and synergism of the enzymes. *Biotechnology and Bioengineering*. 59(5): 621- 634.

Mega, T., and Matsushima, Y. (1979). Comparative studies of three exo-*β*-glycosidases of *Aspergillus oryzae*. *Journal of Biochemistry*. 85: 335-341.

Merino, S. T., and Cherry, J. (2007). Progress and challenges in enzyme development for biomass utilization. *Advances in Biochemical Engineering/Biotechnology*. 108: 95-120.

Messner, R., Kubicek-Pranz, E. M., Gsur, A., and Kubicek, C. P. (1991). Cellobiohydrolase II is the main conidial-bound cellulase in *Trichoderma reesei* and other *Trichoderma* strains. *Archives of Microbiology*. 155: 601-606.

Miettinen-Oinonen, A. (2004). *Trichoderma reesei strains for production of cellulase for the textile industry*. Doctoral Thesis. Faculty of Biosciences Department of Biological and Environmental Sciences Division of Genetics University of Helsinki.

Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N., and Arnold, F. H. (2000). Directed evolution study of temperature adaptation in a psychrophilic enzyme. *Journal of Molecular Biology*. 297: 1015-1026.

Mosier, N. S., Hall, P., Ladisch, C. M., and Ladisch, M. R. (1999). Reaction kinetics, molecular action, and mechanisms of cellulolytic proteins. *Advances in Biochemical Engineering/Biotechnology*. 65: 23-40.

Mtui, G. Y. S. (2009). Recent advances in pretreatment of lignocellulosic wastes and production of value added products. *African Journal of Biotechnology*. 8(8): 1398-1415.

Müller, S., Sandal, T., Kamp-Hansen, P., and Dalbøge, H. (1998). Comparison of expression systems in the Yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica* cloning of two novel promoters from *Yarrowia lipolytica. Yeast*. 14: 1267-1283.

Nakazawa, H., Okada, K., Kobayashi, R., Kubota, T., Onodera, T., Ochiai, N., Omata, N., Ogasawara, W., Okada, H., and Morikawa, Y. (2008). Characterization of the catalytic domains of *Trichoderma reesei* endoglucanase I, II, and III, expressed in *Escherichia coli*. *Applied Microbiology and Biotechnology*. 81: 681-689.

Nakazawa, H., Okada, K., Onodera, T., Ogasawara, W., Okada, H., and Morikawa, Y. (2009). Directed evolution of endoglucanase III (Cel12A) from *Trichoderma reesei*. *Applied Microbiology and Biotechnology*. 83: 649-657.

Nelson, N. (1944). A photometric adaptation of the Somogyi method for determination of glucose. *Journal of Biological Chemistry*. 153: 375-380.

Nicaud, J. M., Fabre, E., and Gaillardin, C. (1989). Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. *Current Genetics*. 16: 253-260.

Nicaud, J. M., Madzak, C., Broek, P., Gysler, C., Duboc, P., Niederberger, P., and Gaillardin, C. (2002). Protein expression and secretion in the yeast *Yarrowia lipolytica*. *FEMS Yeast Research*. 2: 371-379.

Nidetzky, B., Steiner, W., Hayn, M., and Claeyssens, M. (1994). Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: A new model for synergistic interaction. *Biochemical Journal.* 298: 705-710.

Nigam, J. N. (2001). Ethanol production from wheat straw hemicellulose hydrolysate by *Pichia stipitis*. *Journal of Biotechnology*. 87(1): 17–27.

Nutt, A., Sild, V., Pettersson, G., and Johansson, G. (1998). Progress curves a mean for functional classification of cellulases. *European Journal of Biochemistry* 258(1): 200-206.

O'Donohue, M. J., and Kneale, G. G. (1994). *Methods in Molecular Biology: DNA-Protein Interactions* (Principles and Protocols). vol 30. New Jersey: Humana Press, Inc.

Office of Agricultural Economics, Government. (2012). Basics data of agricultural economics. [Online]. Retrieved from; http://www.oae.go.th/download/download_journal/fundamation-2553.PDF. [2012,April 19]

Öhgren, K., Bengtsson, O., Gorwa-Grauslund, M. F., Galbe, M., Hahn-Hägerdal, B., and Zacchi, G. (2006). Simultaneous saccharification and co-fermentation of glucose and xylose in steam-pretreated corn stover at high fiber content with *Saccharomyces cerevisiae* TMB3400. *Journal of Biotechnology*. 126(4): 488-498.

Okada, H., Tada, K., Sekiya, T., Yokoyama, K., Takahashi, A., Tohda, H., Kumagai, H., and Morikawa, Y. (1998a). Molecular characterization and heterologous expression of the gene encoding a low-molecular-mass endoglucanase from *Trichoderma reesei* QM9414. *Applied and Environmental Microbiology*. 64(2): 555-563.

Okada, H., Sekiya, T., Yokoyama, K., Tohda, H., Kumagai, H., and Morikawa, Y. (1998b). Efficient secretion of *Trichoderma reesei* cellobiohydrolase II in *Schizosaccharomyces pombe* and characterization of its products. *Applied Microbiology and Biotechnology*. 49: 301-308.

Okada, H., Mori, K., Tada, K., Nogawa, M., and Morikawa, Y. (2000). Identification of active site carboxylic residues in *Trichoderma reesei* endoglucanase Cel12A by site-directed mutagenesis. *Journal of Molecular Catalysis B: Enzymatic*. 10: 249-255.

O'sullivan, A. C. (1997). Cellulose: the structure slowly unravels. *Cellulose*. 4: 173-207.

Otero, R. C., and Gaillardin, C. (1996). Efficient selection of hygromycin-B-resistant *Yarrowia lipolytica* transformants. *Applied Microbiology and Biotechnology*. 46: 143-148.

Peng, F., Ren, J., Xu, F., Bian, J., Peng, P., and Sun, R. (2009). Comparative study of hemicelluloses obtained by graded ethanol precipitation from sugarcane bagasse. *Journal of Agricultural and Food Chem*istry. 57: 6305-6317.

Penttilä, M. E., André, L., Lehtovaara, P., Bailey, M., Teeri, T. T., and Knowles, J. K. (1988). Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae. Gene*. 63(1): 103-112.

Peterson, R., and Nevalainen, H. (2012). *Trichoderma reesei* RUT-C30 – thirty years of strain improvement. *Microbiology*. 158: 58-68.

Pignède, G., Wang, H. J., Fudalej, F., Seman, M., Gaillardin, C., and Nicaud, J. M. (2000). Autocloning and amplification of *LIP2* in *Yarrowia lipolytica. Applied and Environmental Microbiology*. 66(8): 3283-3289.

Polizeli, M. L. T. M., Rizzatti, A. C. S., Monti, R., Terenzi, H. F., Jorge, J. A., and Amorim, D. S. (2005). Xylanases from fungi: Properties and industrial applications. *Applied Microbiology and Biotechnology*. 67: 577-591.

Qin, Y., Wei, X., Song, X., and Qu, Y. (2008a). Engineering endoglucanase II from *Trichoderma reesei* to improve the catalytic efficiency at a higher pH optimum. *Journal of Biotechnology*. 135: 190-195.

Qin, Y., Wei, X., Liu, X., Wang, T., and Qu, Y. (2008b). Purification and characterization of recombinant endoglucanase of *Trichoderma reesei* expressed in *Saccharomyces cerevisiae* with higher glycosylation and stability. *Protein Expression and Purification*. 58: 162-167.

Rabelo, S. C., Filho, R. M., and Costa, A. C. (2008). A comparison between lime and alkaline hydrogen peroxide pretreatments of sugarcane bagasse for ethanol production. *Applied Biochemistry and Biotechnology*. 144(1): 87-100.

Rabinovich, M. L., Melnik, M. S., and Bolobova, A. V. (2002). Microbial cellulases (Review). *Applied Biochemistry and Microbiology.* 38(4): 305-321.

Rastogi, G., Bhalla, A., Adhikari, A., Bischoff, K. M., Hughes, S. R., Christopher, L. P., and Sani, R. K. (2010). Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. *Bioresource Technology*. 101: 8798-8806.

Reese, E. T. (1976). History of the cellulase program at the U.S. army natick development center. *Biotechnology and Bioengineering Symposium*. 6: 9-20.

Riou, C., Salmon, J., Vallier, M., Günata, Z., and Barre, P. (1998). Purification, characterization, and substrate specificity of a novel highly glucose-tolerant β-glucosidase from *Aspergillus oryzae. Applied and Environmental Microbiology*. 64(10): 3607-3614.

Roe, J. H. (1955). The determination of sugar in blood and spinal fluid with anthrone reagent. *Journal of Biological Chemistry*. 212(1): 335-343.

Romanos, M. A., Scorer, C. A., and Clare, J. J. (1992). Foreign gene expression in yeast: A review. *Yeast*. 8: 423-488.

Romanos, M. (1995). Advances in the use of *Pichia pastoris* for high-level gene expression. *Biotechnology*. 6: 527-533.

Rooyen, R., Hahn-Hägerdal, B., la Grange, D. C., and Zyl, W. H. (2005). Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. *Journal of Biotechnology*. 120: 284-295.

Rosgaard, L., Pedersen, S., Langston, J., Akerhielm, D., Cherry, J. R., and Meyer, A. S. (2007). Evaluation of minimal *Trichoderma reesei* cellulase mixtures on differently pretreated barley straw substrates. *Biotechnology Progress*. 23(6): 1270-1276.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K., and Jones, T. A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science*. 249(4967): 380-386.

Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature*. 454: 841- 845.

Saha, B. C. (2003). Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology*. 30: 279-291.

Saha, B. C. and Cotta, M. A. (2006). Ethanol production from alkaline peroxide pretreated enzymatically saccharified wheat straw. *Biotechnology Progress*. 22: 449-453.

Saha, B. C. and Cotta, M. (2007). Enzymatic saccharification and fermentation of alkaline peroxide pretreated rice hulls to ethanol. *Enzyme and Microbial Technology*. 41(4): 528-532.

Saloheimo, A., Henrissat. B., Hoffrén, A. M., Teleman, O., and Penttilä, M. (1994). A novel, small endoglucanase gene, EGL5, from *Trichoderma reesei* isolated by expression in yeast. *Molecular Microbiology*. 13(2): 219-228.

Saloheimo, M., Lehtovaara, P., Penttika, M., Teeri, T. T., Stihlhergb, J., Johansson, G., Pettersson, G., Claeyssens, M., Tomme, P., and Knowles, J. K. C. (1988). EGIII, a new eudoglucanase from *Trichoderma reesei:* The characterization of both gene and enzyme. (Recombinant DNA; cellulase; *Trichoderma;* gene sequence; module shufIling; active site; phage λ vector). *Gene*. 63: 11-21.

Saloheimo, M., Nakari-setala, T., Tenkanen, M., And Penttila, M. (1997). cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *European Journal of Biochemistry*. 249: 584-591.

Sambrook, J., Fritch, E. F., and Maniatis, T. (1989). *Molecular cloning a laboratory manual*. 2nd ed. vols. 1, 2 and 3. New York: Cold Spring Harbor Laboratory Press.

Sánchez, C. (2009). Lignocellulosic residues: Biodegradation and bioconversion by fungi. *Biotechnology Advances*. 27: 185-194.

Sandgren, M., Shaw, A., Ropp, T. H., Wu, S., Bott, R., Cameron, A. D. Ståhlberg, J., Mitchinson, C., and Jones, T. A. (2001). The X-ray crystal structure of the *Trichoderma reesei* Family 12 Endoglucanase 3, Cel12A, at 1.9 Å resolution. *Journal of Molecular Biology*. 308: 295-310.

Schmidt, F. R. (2004). Recombinant expression systems in the pharmaceutical industry. *Applied Microbiology and Biotechnology*. 65: 363-372.

Schuster, A., and Schmoll, M. (2010). Biology and biotechnology of *Trichoderma*. *Applied Microbiology and Biotechnology*. 87: 787-799.

Sen, S., Dasu, V. V., and Mandal, B. (2007). Developments in directed evolution for improving enzyme functions. *Applied Biochemistry and Biotechnology*. 143: 212-223.

Sharrock, K. R. (1988). Cellulase assay methods: a review. *Journal of Biochemical and Biophysical Methods*. 17(2): 81-105.

Shoseyov, O., Shani, Z., and Levy, I. (2006). Carbohydrate binding modules: Biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews*. 70(2): 283-295.

Shyamala, S., Ravikumar, S., Vikramathithan, J., and Srikumar, K. (2011). Isolation, purification, and characterization of two thermostable endo-1,4-β-D-glucanase forms from *Opuntia vulgaris*. *Applied Biochemistry and Biotechnology*. 165: 1597-1610.

Sodoyer, R. (2004). Expression systems for the production of recombinant pharmaceuticals. *Biodrugs*. 18(1): 51-62.

Somogyi, M. (1952). Note on sugar determination. *Journal of Biological Chemistry*. 195: 19-25.

Srisodsuk, M., Reinikainen, T., Penttilä, M., and Teeri, T. T. (1993). Role of the interdomain linker peptideo *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. *The Journal of Biological Chemistry*. 266(26): 20756-20761.

Steffens, D. L., and Williams, J. G. K. (2007). Efficient site-directed saturation mutagenesis using degenerate oligonucleotides. *Journal of Biomolecular Techniques*. 18:147-149.

Stöcker, M. (2008). Biofuels and biomass-to-liquid fuels in the biorefinery: Catalytic conversion of lignocellulosic biomass using porous materials. *Angewandte Chemie International Edition*. 47: 9200-9211.

Sukumaran, R. K. Singhania, R. R., and Pandey, A. (2005). Microbial cellulases-production, applications and challenges. *Journal of Scientific and Industrial Research*. 64: 832-844.

Sun, R. C., Tomkinson, J., Ma, P. L., and Liang, S.F. (2000). Comparative study of hemicelluloses from rice straw by alkali and hydrogen peroxide treatments. *Carbohydrate Polymers*. 42(2): 111-122.

Suominen, P. L., Mäntylä, A. L., Karhunen, T., Hakola, S., and Nevalainen, H. (1993). High frequency one-step gene replacement in *Trichoderma reesei* II. Effects of deletions of individual cellulase genes. *Molecular and General Genetics*. 241: 523-530.

Taherzadeh, M. J., and Karimi, K. (2008). Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: a review. *International Journal of Molecular Sciences*. 9(9): 1621-1651.

Taniguchi, M., Tohma, T., Itaya, T., and Fujii, M. (1997). Ethanol production from a mixture of glucose and xylose by co-culture of *Pichia stipitis* and a respiratory-deficient mutant of *Saccharomyces cerevisiae*. *Journal of Fermentation and Bioengineering*. 83(4): 364-370.

Teeri, T. T., Lehtovaara, P., Kauppinen, S., Salovuori, I., and Knowles, J. (1987). Homologous domains in *Trichoderma reesei* cellulolytic enzymes: Gene sequence and expression of cellobiohydrolase II. (Recombinant DNA; cellulose hydrolysis; conserved domains; evolution; glycosylation; phage 1 vector; module shuffling; protein families; transcript mapping; introns). *Gene*. 51: 43-52.

Teeri, T. T. (1997). Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Tibtech*. 15: 160-167.

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*. 72: 211-222.

Tokuhiro, K., Ishida, N., Kondo, A., and Takahashi, H. (2008) Lactic fermentation of cellobiose by a yeast strain displaying *β*-glucosidase on the cell surface. *Applied Microbiology and Biotechnology*. 79: 481-488.

van Zyl, W. H., Lynd, L. R., den Haan, R., and McBride, J. E. (2007). Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Advances in Biochemical Engineering/Biotechnology*. 108: 205-235.

Vasan, P. T., Piriya, P. S., Prabhu, D. I. G., and Vennison, S. J. (2011). Cellulosic ethanol production by *Zymomonas mobilis* harboring an endoglucanase gene from *Enterobacter cloacae*. *Bioresource Technology*. 102: 2585-2589.

Viikari, L., Alapuranen, M., Puranen, T., Vehmaanper, J., and Siika-aho, M. (2007). Thermostable enzymes in lignocellulose hydrolysis. *Advances in Biochemical Engineering/Biotechnology*. 108: 121-145.

Waffenschmidt, S., and Jaenicke, L. (1987). Assay of reducing sugars in the nanomole range with 2,2'-bicinchoninate. *Analytical Biochemistry*. 165(2): 337-340.

Weil, J., Westgate, P., Kohlmann, K., and Ladisch, M. R. (1994). Cellulose pretreatments of lignocellulosic substrates. *Enzyme and Microbial Technology*. 16(11): 1002-1004.

Wood, T. M. (1988). *Biomass part A: cellulose and hemicellulose*. New York: Academic.

Wood, T. M., McCrae, S. I., and Bhat, K. M. (1989). The mechanism of fungal cellulase action synergism between enzyme components of *Penicillium pinophilum* cellulase in solubilizing hydrogen bond-ordered cellulose. *Biochemical Journal*. 260: 37-43.

Wood, T. M., and Garcia-Campayo, V. (1990). Enzymology of cellulose degradation. *Biodegradation.* 1: 147-161.

Worldwide Protein Data Bank, Organization. (2012a). Biological macromolecular resource. [Online]. Retrieved from; http://www.rcsb.org/pdb/explore/ explore.do? structureId=1EG1. [2012, January 16].

Worldwide Protein Data Bank, Organization. (2012b). Biological macromolecular resource. [Online]. Retrieved from; http://www.rcsb.org/pdb/explore/ explore.do?structureId=3QR3. [2012, January 16].

Worldwide Protein Data Bank, Organization. (2012c). Biological macromolecular resource. [Online]. Retrieved from; http://www.rcsb.org/pdb/explore/ explore.do?structureId=1H8V. [2012, January 16].

Worldwide Protein Data Bank, Organization. (2012d). Biological macromolecular resource. [Online]. Retrieved from; http://www.rcsb.org/pdb/explore/ explore.do?structureId=1CEL. [2012, January 16].

Worldwide Protein Data Bank, Organization. (2012e). Biological macromolecular resource. [Online]. Retrieved from; http://www.rcsb.org/pdb/explore/ explore.do?structureId=1CB2. [2012, January 16].

Wu, S., and Letchworth, G. J. (2004). High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. *Biotechniques*. 36: 152-154.

W.W. Norton and Company, Incorporate. (2010). Genetic Engineering and Biotechnology. [Online]. Retrieved from; http://www.wwnorton.com/college/ biology/microbiology2/ch/12/etopics.aspx. [2012, April 19]

Xuan, J. W., Fournier, P., and Gaillardin, C. (1988). Cloning of the *LYS5* gene encoding saccharopine dehydrogenase from the yeast *Yarrowia lipolytica* by target integration. *Current Genetics*. 14: 15-21.

Yang, B., Boussaid, A., Mansfield, S. D., Gregg, D. J., and Saddler, J. N. (2002). Fast and efficient alkaline peroxide treatment to enhance the enzymatic digestibility of steam-exploded softwood substrates. *Biotechnology and Bioengineering*. 77(6): 678-684.

Yang, D., Weng, H., Wang, M., Xu, W., Li, Y., and Yang, H. (2010). Cloning and expression of a novel thermostable cellulose from newly isolated *Bacillus subtilis* strain I15. *Molecular Biology Reports*. 37: 1923-1929.

Yin, J., Li, G., Rena, X., and Herrler, G. (2007). Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of Biotechnology*. 127: 335-347.

Zhang, Y. H. P., and Lynd, L. R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering*. 88(7): 797-824.

Zhang, Y. H. P., Himmel, M. E., and Mielenz, J. R. (2006). Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances*. 24: 452-481.

Zhang, Y. H. P., Hong, J., and Ye, X. (2009a). *Biofuels: Methods and Protocols* (Methods in Molecular Biology). vol. 581. New Jersey: Humana Press, Inc.

Zhang, A. L., Luo, J. X., Zhang, T. Y., Pan, Y. W., Tan, Y. H., Fu, C. Y., and Tu, F. Z. (2009b). Recent advances on the GAP promoter derived expression system of *Pichia pastoris*. *Molecular Biology Reports*. 36: 1611-1619.

Zheng, L., Baumann, U., and Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*. 32(14): 1-5.

Zhou, S., Davis, F. C., and Ingram, L. O. (2001). Gene integration and expression and extracellular secretion of *Erwinia chrysanthemi* Endoglucanase CelY (*celY*) and CelZ (*celZ*) in Ethanologenic *Klebsiella oxytoca* P2. *Applied and Environmental Microbiology*. 67(1): 6-14.

BIOGRAPHY

Mr. Nassapat Boonvitthya was born on June 16th, 1983 in Bangkok, Thailand. He got a Bacherlor's Degree of Science, Major in Genetics, Faculty of Science, Chulalongkorn University in 2005. During his undergraduate, he received first class honour and gold medal award. Moreover, he also earned excellent graduate student award from Department of Botany and excellent graduate student award from Faculty of Science (professor Tab's gold medal award). Later in 2006, he studied in Program in Biological Science, Faculty of Science, Chulalongkorn University, with received a two years-grant from Chulalongkorn University Graduate Scholarship to Commemorate the 72nd Anniversary of His Majesty King Bhumibol Adulyadej, and four years-grant from Chulalongkorn University Dutsadi Phiphat Scholarship, which is collaboration between Chulalongkorn University and PTT public company. During two years study in France, he was employed as Recherche Engineer at INSA-Toulouse in the project "Improvement of immobilization of a GH43 xylosidase" for one year under supervisor of Dr. Michael J. O'Donohue.

Outcome from this study:

Boonvitthya, N., Bozonnet, S., Burapatana, V., O'Donohue, M. J., and Chulalaksananukul, W. (2012). Comparison of the heterologous expression of *Trichoderma reesei* endoglucanase II and cellobiohydrolase II in the yeasts *Pichia pastoris* and *Yarrowia lipolytica*. (In press)

Boonvitthya, N., Tanapong, P., Kanngan, P., Burapatana, V., and Chulalaksananukul, W. (2012). Cloning and expression of *Aspergillus oryzae* glucan 1,3-beta-glucosidase A in *Pichia pastoris*. (Submitted)

Boonvitthya, N., Bozonnet, S., Burapatana, V., O'Donohue, M. J., and Chulalaksananukul, W. (2012). Directed evolution of *Trichoderma reesei* endoglucanase II to improve its thermostability in *Yarrowia lipolytica*. (Preparation)

Conference and proceeding:

Boonvitthya, N., Bozonnet, S., Burapatana, V., Chulalaksananukul, W., and O'Donohue, M. (2011). Overexpression and characterization of *Trichoderma reesei* endoglucanase II and Cellobiohydrolase II in two yeast platforms *Yarrowia lipolytica* and *Pichia pastoris*. 16th Biological Sciences Graduate Congress. National University of Singapore. Singapore, December, 12-14 (Oral presentation).

Desrousseaux, M., **Boonvitthya, N.**, Dumon, C., and O'Donohue, M. (2011). Improvement of immobilization of a GH 43 xylosidase. 10th International Symposium on Biocatalysis. Giardini, Sicily, Italy, October (Poster presentation).

Boonvitthya, N., Bozonnet, S., Burapatana, V., Chulalaksananukul, W., and O'Donohue, M. (2011). Optimization of recombinant expression of *Trichoderma reesei* endoglucanase II and cellobiohydrolase II in yeasts for directed evolution. 9th carbohydrate Bioengineering Meeting. Lisbon, Portugal, May (Poster presentation).

Boonvitthya, N., Bozonnet, S., Burapatana, V., Chulalaksananukul, W., and O'Donohue, M. (2009). Construction of recombinant yeasts expressing *Trichoderma reesei* endoglucanase II and cellobiohydrolase II for efficient cellulose degradation. INRA, France, September (Poster presentation).

Boonvitthya, N., Suwannarangsee, S., Winayanuwattikun, P., Burapatana, V., and Chulalaksananukul, W. (2008). Preparing of cellulolytic enzyme overproduction of *Trichoderma reesei* and *Aspergillus* sp. in *Pichia pastoris*. 13rd Biological Sciences Graduate Congress. National University of Singapore. Singapore, December, 15-17 (Poster presentation).

Boonvitthya, N., Winayanuwattikun, P., Piapukiew, J., and Chulalaksananukul, W. (2007). Cloning and expression of *Fusarium solani* NAN 103 lipase in *Pichia pastoris*. 12th Biological Sciences Graduate Congress. University of Malaya. Malaysia, December, 17-19 (Poster presentation).