CLONING AND CHARACTERIZATION OF AN ENDOGLUCANASE GENE ENCODING THERMOTOLERANT CELLULOLYTIC ENZYME FROM *Syncephalastrum racemosum* ISOLATED FROM THAILAND

BENJAMAPORN WONGANU

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(MOLECULAR GENETICS AND GENETIC ENGINEERING)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2006

ISBN 974-04-7009-2

Thesis
Entitled

CLONING AND CHARACTERIZATION OF AN ENDOGLUCANASE GENE ENCODING THERMOTOLERANT CELLULOLYTIC ENZYME FROM *Syncephalastrum racemosum* ISOLATED FROM THAILAND

Benjamaporn Wonganu
Miss Benjamaporn Wonganu
Candidate

Asst. Prof. Kusol Pootanakit, Ph. D.
Major-Advisor

Asst. Prof. Lily Eurwilaichitr, Ph. D.
Co-Advisor

Apinunt Udomkit
Asst. Prof. Apinunt Udomkit, Ph. D.
Co-Advisor

Wipa Chungjatupornchai
Assoc. Prof. Wipa Chungjatupornchai, Ph.D.
Co-Advisor

Prof. M.R. Jisnuson Svasti, Ph.D.
Dean
Faculty of Graduate Studies

V. Akkarapatumwong
Asst. Prof. Varaporn Akkarapatumwong, Ph.D.
Chair
Master of Science Programme in
Molecular Genetics and Genetic Engineering
Institute of Molecular Biology and Genetics

Thesis
Entitled

CLONING AND CHARACTERIZATION OF AN ENDOGLUCANASE GENE ENCODING THERMOTOLERANT CELLULOLYTIC ENZYME FROM *Syncephalastrum racemosum* ISOLATED FROM THAILAND

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Molecular Genetics and Genetic Engineering)
on
17 March 2006

Benjamaporn Wonganu
.....
Miss Benjamaporn Wonganu
Candidate

.....
Asst.Prof. Kusol Pootanakit, Ph.D.
Major-Advisor

.....
Asst. Prof. Lily Eurwilaichitr, Ph.D.
Co-Advisor

Apinunt Udomkit
.....
Asst. Prof. Apinunt Udomkit, Ph.D.
Member

.....
Asst.Prof. Chuenchit Boonchird, Ph.D.
Member

Wipa Chungjatupornchai
.....
Assoc. Prof. Wipa Chungjatupornchai, Ph.D.
Member

.....
Prof. M.R. Jisnuson Svasti, Ph. D.
Dean
Faculty of Graduate Studies
Mahidol University

C. Krittanai
.....
Asst. Prof. Chartchai Krittanai, Ph.D.
Acting Director
Institute of Molecular Biology and Genetics
Mahidol University

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisors, Asst. Prof. Kusol Pootanakit and Asst. Prof. Lily Eurwilaichitr, for their kind supervision, valuable advice, helpfulness and encouragement throughout this thesis project. I also greatly indebted to Asst. Prof. Apinunt Udomkit, Assoc. Prof. Wipa Chungjatupornchai, Asst.Prof. Chuenchit Boonchird for their discussions and constructive suggestions.

In addition, I would like to express my sincere appreciation for Dr. Witoon Tirasophon, Asst. Prof. Sutipa Tanapongpipat, Dr. Rutchadaporn Sriprang, and Dr. Verawat Champreda for their helpful comments and advices on this work. I am grateful to Miss Katewadee Boonyapakorn, Miss Supattra Treerattrakool, and Miss Suchada Saengwiman for their laboratory suggestion.

On the personal side, I would like to thank my father, my mother, my sisters and other members in my family for their love, understanding, support and encouragement. I am sincerely thankful to all members of B203 and of enzyme screening laboratory, BIOTEC for their kindness, helpfulness, encouragement and much-needed moral support.

Finally, I am extremely thankful to the National Center for Genetic Engineering and Biotechnology (BIOTEC) and Development and Promotion of Science and Technology Talent Project (DPST) for providing the scholarship for my studies and for this thesis.

Benjamaporn Wonganu

CLONING AND CHARACTERIZATION OF AN ENDOGLUCANASE GENE ENCODING THERMOTOLERANT CELLULOLYTIC ENZYME FROM *Syncephalastrum racemosum* ISOLATED FROM THAILAND

BENJAMAPORN WONGANU 4636875 MBMG/M

M.Sc. (MOLECULAR GENETICS AND GENETIC ENGINEERING)

THESIS ADVISORS:KUSOL POOTANAKIT, Ph.D., LILY EURWILAICHITR, Ph.D., APINUNT UDOMKIT, Ph.D., WIPA CHUNGJATUPORNCHAI, Ph.D.

ABSTRACT

A locally-collected filamentous fungus, *Syncephalastrum racemosum* (BCC18080), produced high level thermotolerant cellulases. Specifically, over 50% cellulases activities remained after incubation at 80°C for an hour. Since one of the cellulolytic enzymes, endoglucanase, is a major enzyme produced by the fungus; therefore, in this study, it is of interest to clone and characterize the endoglucanase gene from BCC18080. First, optimal condition to induce cellulase was determined. Then, BCC18080 total RNA was isolated; degenerate RT-PCR and 3'- and 5'-RACEs were employed to obtain the full-length endoglucanase gene (1,020 nucleotides encoding 340 amino acids). NCBI-BLAST search showed BCC18080 endoglucanase belonged to glycosyl hydrolase family 45. The signal peptide of this gene was as predicted by SignalP and later confirmed by N-terminal sequencing, suggesting that amino acid residues 1 to 32 was the signal peptide. Recombinant clones with and without the leader sequence were expressed in *P. pastoris*. Expression of all integrants demonstrated that the active BCC18080 endoglucanase gene was successfully produced and secreted as a 55 kDa and a 30 kDa proteins. N-terminal sequencing suggested that the 55 kDa band was the mature protein while the 30 kDa band was the truncated protein of the larger band. Since the size of the BCC18080 endoglucanase (55 kDa) was larger than the calculated molecular mass, it may contain sugar moieties. Glycoprotein analysis showed that the 55 kDa protein was glycosylated; but not in the N-linked position, while the smaller protein was not glycosylated. All recombinant proteins showed optimal temperature at 70°C and optimal pH at 5-6. They also showed activity against carboxylmethylcellulose and retained more than 50% activity for 4 hours at 70°C. This suggested that thermostable endoglucanase was successfully produced from *P. pastoris* transformants. In addition, high K_{cat} and low apparent K_m of these recombinant proteins indicated good properties of this enzyme against its substrate.

KEY WORDS: *Syncephalastrum racemosum* / ENDOGLUCANASE/ THERMOTOLERANT ENZYME / *Pichia pastoris*

126 pp. ISBN 974-04-7009-2

การโคลนและการศึกษาลักษณะของยีนเอนโดกลูคาเนสจากเชื้อราทนร้อน *Syncephalastrum racemosum* ที่แยกได้ในประเทศไทย (CLONING AND CHARACTERIZATION OF AN ENDOGLUCANASE GENE ENCODING THERMOTOLERANT CELLULOLYTIC ENZYME FROM *Syncephalastrum racemosum* ISOLATED FROM THAILAND)

เบญจมาภรณ์ วงศ์อนุ 4636875 MBMG/M

วท.ม. (อณูพันธุศาสตร์และพันธุวิศวกรรมศาสตร์)

คณะกรรมการควบคุมวิทยานิพนธ์ : กุศล ภูธนกิจ, Ph.D., ลิลี่ เอื้อวิไลจิตร, Ph.D., อภินันท์ อุดมกิจ, Ph.D., วิภา จึงจตุพรชัย, Ph.D.

บทคัดย่อ

เชื้อรา *Syncephalastrum racemosum* (BCC18080) ที่เก็บได้ในประเทศไทย สามารถผลิตเอนไซม์เซลลูเลสที่มีคุณสมบัติทนความร้อนได้ถึง 80 องศาเซลเซียส ดังนั้นงานวิจัยนี้จึงมุ่งเน้นจะทำการโคลนและศึกษาลักษณะของยีนเซลลูเลสจากเชื้อราตัวนี้โดยเฉพาะอย่างยิ่งยีนเอนโดกลูคาเนส ซึ่งเป็นเอนไซม์หลักในกลุ่มเอนไซม์ที่ย่อยเซลลูโลส งานวิจัยนี้จึงได้ทำการโคลน cDNA ของเอนไซม์เอนโดกลูคาเนสจากเชื้อ BCC18080 และศึกษาลักษณะของเอนไซม์ ในขั้นแรกเชื้อ BCC18080 ได้ถูกเหนี่ยวนำให้เกิดการสร้างเอนไซม์แล้วได้ถูกนำมาสกัด total RNA เพื่อนำมาสังเคราะห์ cDNA ด้วยวิธีการ RT-PCR, 3'- และ 5'- RACEs ผลการศึกษาพบว่า cDNA ของเอนไซม์เอนโดกลูคาเนสให้ลำดับนิวคลีโอไทด์ 1,020 ตัว ซึ่งถอดรหัสเป็นลำดับอะมิโนได้ 340 ตัว เมื่อเปรียบเทียบในฐานข้อมูลพบว่า cDNA ของเอนไซม์เอนโดกลูคาเนส จัดอยู่ใน glycosyl hydrolase กลุ่มที่ 45 ตำแหน่งอะมิโนที่ 1 ถึง 33 เป็นตำแหน่งของ signal peptide ผลการแสดงออกของ cDNA ของเอนไซม์เอนโดกลูคาเนสใน *P. pastoris* พบว่าเอนไซม์สามารถผลิตและหลั่งออกนอกเซลล์ในทั้งแบบที่มีและไม่มี leader sequence ในรูปของโปรตีน 2 ขนาดคือ 55 kDa และ 30 kDa ซึ่งผลวิเคราะห์ลำดับอะมิโนในส่วนปลาย N พบว่าโปรตีนขนาด 55 kDa เป็นโปรตีนสายสมบูรณ์ของเอนโดกลูคาเนส ในขณะที่ 30 kDa เป็นโปรตีนที่ถูกย่อยไปบางส่วน นอกจากเอนไซม์ที่โคลนได้จะมีความจำเพาะเฉพาะกับซับสเตรสของเอนโดกลูคาเนสแล้ว เอนไซม์ยังสามารถทนต่อความร้อนได้ถึง 70 องศาเป็นเวลา 4 ชั่วโมงโดยที่ความสามารถในการทำงานยังคงอยู่มากกว่า 50 เปอร์เซ็นต์ แสดงให้เห็นว่า *P. pastoris* ที่มียีนที่สร้างเอนโดกลูคาเนสจาก BCC18080 สามารถผลิตเอนไซม์เอนโดกลูคาเนสที่สามารถทนต่อความร้อนสูงได้สำเร็จ มากไปกว่านั้นค่า K_{cat} ที่สูงและค่า K_m ที่ต่ำ เป็นค่าที่แสดงให้เห็นว่าเอนไซม์เอนโดกลูคาเนสจากเชื้อราที่ศึกษาสามารถจับและย่อยกับซับสเตรสที่จำเพาะเจาะจงกับเอนไซม์ได้ดี

126 หน้า ISBN 974-04-7009-2

CONTENTS

CONTENTS (CONT.)

CONTENTS (CONT.)

CONTENTS (CONT.)

CONTENTS (CONT.)

CONTENTS (CONT.)

CONTENTS (CONT.)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (CONT.)

LIST OF FIGURES (CONT.)

LIST OF ABBREVIATIONS

A_{260}	=	absorbance at 260 nonometers
A_{280}	=	absorbance at 280 nonometers
AOX	=	alcohol oxidase gene
ATP	=	adenosine triphosphate
Arg	=	arginine
Asn	=	asparagine
BCC18080	=	*Syncephalastrum racemosum* (BCC18080)
BSA	=	bovine serum albumin
BMGY	=	buffered minimal glyceral complex medium
BMMY	=	buffered minimum methanol complex medium
bp	=	base pair(s)
CTAB	=	cetyl trimethyl ammonium bromide
^{o}C	=	degree Celcius
cDNA	=	complementary DNA
dATP	=	deoxyadenosine-5'-triphosphate
dCTP	=	deoxycytosine-5'-triphosphate
dGTP	=	deoxyguanosine-5'-triphosphate
dTTP	=	deoxythymidine-5'-triphosphate
dNTP	=	dATP, dCTP, dGTP, dTTP
DMSO	=	dimethyl sulfoxide
DNA	=	deoxyribonucleic acid
DTT	=	dithiothreitol
E. coli	=	*Eschericia coli*
EDTA	=	ethylene diamine tetraacetic acid
g	=	gram
xg	=	centrifugal force
h	=	hour (s)

LIST OF ABBREVIATIONS (CONT.)

IPTG	=	isopropyl-β-D-thiogalactopyranoside
kb	=	kilobase(s)
kDa	=	kilodalton(s)
KEX2	=	killer expression2 protein
LB	=	Luria-Bertani (medium)
M	=	molar
mM	=	milimolar
mg	=	milligram(s)
mg/ml	=	milligrams/milliliter
min	=	minute(s)
min	=	minute (s)
ml	=	milliliter(s)
mRNA	=	messenger RNA
µg	=	microgram(s)
µl	=	microliter(s)
µM	=	micromolar
MW	=	molecular weight
ng	=	nanogram
OD	=	optical density
PAGE	=	polyacrylamide gel electrophoresis
PCR	=	polymerase chain reaction
pmol	=	picomole
PNGase F	=	N-Glycosidase F
PVDF	=	polyvinylidene difluoride

LIST OF ABBREVIATIONS (CONT.)

RACE	=	rapid amplification of cDNA ends
RNA	=	ribonucleic acid
RNase	=	ribonuclease
RT-PCR	=	reverse transcription-polymerase chain reaction
SDS	=	sodium dodecyl sulfate
Sec	=	second
T_m	=	melting temperature
TEMED	=	N,N,N',N'-tetramethyl-ethylenediamine
Tris-HCl	=	tris-(hydroxylmethyl)-aminoethane hydrochloric
U	=	Unit (s)
UV	=	utraviolet
V	=	volt
v/v	=	volume/volume
w/v	=	weight/volume
X-gal	=	5-bromo-4-chloro-3-indolyl-β-D-galactoside
YEPD	=	yeast extract peptone dextrose

CHAPTER I

INTRODUCTION

1.1 *Syncephalastrum racemosum* (BCC18080)

S . racemosum (BCC18080) is a filamentous fungus that is commonly found in soil, and in animal feces in tropical and subtropical area (Figure 1). It belongs in:

Kingdom : Fungi
Phylum : Zygomycota
Class : Zygomycetes
Order : Mucorales
Family : Syncephalastraceae
Genus : Syncephalastrum
Species : racemosum

In culture, *S. racemosum* grows rapidly, producing either low-growing or tall (0.5 to 1.5 cm) erect mycelia terminating in club-shaped within 1 week. The cultures are hyaline, with surface coloration varying from nearly white to various shades of green, olive, and gray to almost black (1). *S. racemosum* reproduces asexually by non-motile spores that are arranged in a row inside a cylindrical sac called a merosporangium. The entire rows of non-motile spores are dispersed by wind (2). Sexual reproduction of this fungus is by conjugation, the fusion of gametangia resulted in the formation of zygospore (3).

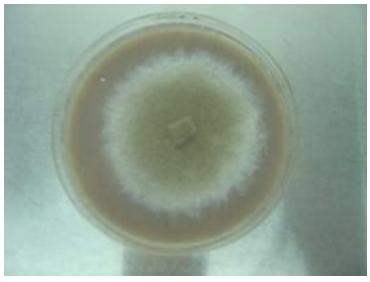


Figure 1. Filamentous fungus, *Syncephalastrum racemosum* (BCC18080)

1.2 Structure of cellulose

Cellulose consists of linear β-1, 4-linked D-glucopyranose chains in the chair configuration (Figure 2). In nature, up to 250 individual cellulose molecules are self-assembled by hydrogen bonds into crystalline structures called microfibrils (4). In addition, associated hemicelluloses relate to the aggregation of cellulose (5). Since the orientation of all adjacent cellulose molecules are not arranged in the same dimension, cellulose fibers in nature are not purely crystalline--there are crystalline and amorphous regions. In addition to these regions, cellulose fibers contain various factors that cause its surface area to form kinks or twists in the microfibrils and the formation of micropore, large pits or capillaries (6, 7, 8, 9). The effect of heterogeneous structure within microfibrils is that not only small molecules such as water can penetrate into cellulose structure but the larger molecules including cellulytic enzymes can also do (10, 11).

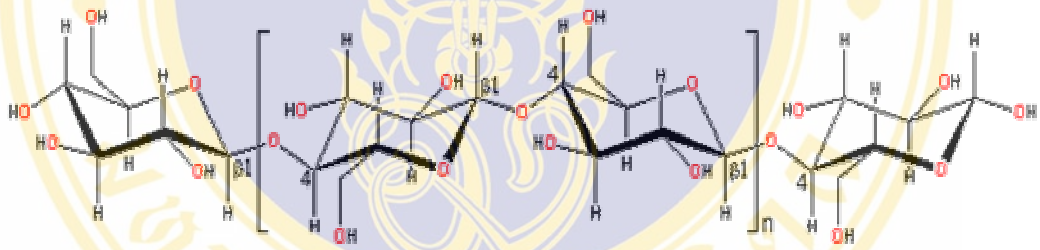


Figure 2. Conformational structure of cellulose (9).

1.3 Cellulases

1.3.1 Fungal cellulase system

Hydrolytic components of cellulase system were classified based on their mode of catalytic action and on structure in catalytic domains (12). There are: (i) endoglucanases or 1, 4-β-D-glucanohydrolases (ii) exoglucanase, including 1, 4-β-D-glucan glucanohydrolases also known as cellodextrinases and 1, 4-β-D-glucan cellobiohydrolases also known as cellobiohydrolases (iii) β-glucosidase or β-glucoside glucohydrolase (13).

Endoglucanases are inactive against crystalline cellulose. They randomly cut at internal β-glycosidic linkages of internal amorphous structure in the

cellulose polysaccharide chain, giving new chain ends and various lengths of oligosaccharides as products.

Exoglucanses act on the reducing or non-reducing ends of the polysaccharide chain, generating either glucose or cellobiose. They can act on the crystalline structure of cellulose, producing cellobiose and decreasing the degree of polymerization (DP) (14). Exoglucanases decrease DP slowly when compare to endoglucanase action (15).

β-glucosidases complete the process of cellulose hydrolysis by hydrolysing soluble cellodextrins and cellobiose and removing glucose from the non-reducing ends. Unlike the exoglucanase, β-glucosidase activity will increase when DP value decreases (16).

1.3.2 Determination of cellulase activity

1.3.2.1 Biochemical assay methods

Since three small groups of enzymes are classified as celllases (endoglucanase, exoglucanase, and β-glucosidase), varying assay methods were developed to distinguish one from the others. For endoglucanases, carboxylmethylcellulose was widely used as specific substrate. Because pH value and ionic strength of the solvent affect the viscosity of its solution (17) that change from solution to gel at temperature lower than 37^{o}C (18), hydroxyethycellulose therefore becomes more suitable as one of specific substrate for endoglucanase because it is a neutral compound (19). Normally, the amount of reducing sugar, represented as glucose, is measured as the enzyme activity by using the dinitrosalicylic acid method (20), the Somogyi-Nelson method (21), the phenol-sulfuric method (22) or the ferricyanide method, which has the greatest sensitivity (23). For exoglucanases, the direct method to determine the enzyme activity requires purified enzyme since no specific substrate is available for this group of enzymes. Usually, cotton is the best substrate (24) but phosphoric acid swollen cellulose or micro-crystalline cellulose such as Avicel is also suitable (25). For β-glucosidase, its activity is determined by using cellobiose or p-nitrophenyl-β-glucoside as a substrate (26, 27).

1.3.2.2 Practical Assay Methods

For determination of the cellulolytic capability of cellulase present in culture filtrate or a commercial crude cellulase enzyme, saccharification of filter

paper is the standard method. Filter paper activity is measured as the amount of reducing sugar formed (24). In addition, enzyme activity can also be determined rapidly by using Avicel linked with Remozol brilliant blue R. The activity is detected from the formation of soluble dyed product (28).

1.3.3 Molecular genetics of fungal cellulases

Several fungi have been reported to produce cellulases (29), such as: aerobic fungi *Trichoderma viride*, *Trichoderma reesei, Penicillium pinophilum, Sporotrichum pulverulentrum, Fusarium solani, Talaromyces emersonii* and *Trichoderma koningii* (27, 30, 31, 32, 33, 34, 35); thermophilic aerobic fungi *Sporotrichum thermophile*, *Thermoascus aurantiacus, Chaetonium thermophile* and *Humicula insolens* (36, 37, 38, 39, 40); mesophilic anaerobic fungi *Neocallimastrix frontalis, Piromonas communis* and *Sphaeromonas communis* (41, 42, 43). To improve the properties of secreted cellulases for industrial application, gene technology has been employed (44). The cellulase system of *T. reesei* has been the most widely studied as a model of fungal cellulotic system. *T. reesei* produces two exoglucanases (CBHI and CBHII), five endoglucanases (EGI, EGII, EGIII, EGIV and EGV) and two β-glucosidase (BGLI and BGLII (45, 46, 47). Since most thermophilic fungi can produce thermostable cellulases, they are increasingly being studied. For instance, the thermophilic fungus *H. insolens* cellulase system is wildely studied and is shown to be homologous to the *T. reesei* producing two cellobiohydrolases (CBHI and CBHII) and five endoglucanases (EGI, EGII, EGIII, EGIV and EGV (40).

1.3.4 Application of fungal cellulases

Since fungi produce large amounts of cellulases, there is now a world-wide interest in using fungal cellulases in various industries such as food, animal feed, the paper and pulp, textile and chemical industries (48, 49, 50, 51).

In textile industry, cellulases are used to modify the surface and properties of cellulosic fibres by removing microfibril, fuzz and loss fibres from the surface. Moreover, they are used to remove excess dye from the denim fabric and to increase the softness and color brightness of cotton fabrics (51, 52). In paper and pulp industry, celulases are used in pulp refinement and in paper recycling by removing ink from recycled paper (53). In food industry, cellulase are used to extract oil from seed, removal of external soybean coat in fermented soybean production and isolate protein

from soybean and coconut (52). In animal feed industry, cellulases are used to pretreat the lignocellulosic material in feed. This process improves the digestibility of monogastric animals (51). In chemical industry, cellulases are used in production of ethanol and other commodity products from the cellulosic biomass (52).

1.3.5 Fungal thermotolerant cellulases

Most of thermophilic or mesophilic microorganisms produce thermostable cellulases (48). Since fungi produce large amount of extracellular cellulases, fungal cellulases are widely studied and have been utilized for industrial applications. Most thermostable cellulases produced from fungi have optimal activity between 50 and 80 °C (54).

For industrial applications, themostable cellulases are ideal because they can withstand high temperature without being denatured. For example, the thermostable cellulase produced from *Aspergillus niger*. This enzyme has been widely applied in various industries such as food, textile, and paper-making. (55, 56).

Because of its advantage, thermostable cellulases have been wildely studied with the aim to improve their properties. Expression of thermostable cellulases has been reported in both bacteria and yeast. Since some evidence showed that glycosylation effect thermostability of enzyme, for instance, deglycosylation in *Humicula insolens* cellulases lower their stability (57). Therefore, most thermostable cellulases are expressed in yeast because besides providing high expression level, this system also adds glycosylation to heterologous protein.

1.4 Glycosyl hydrolase families

Since carbohydrate in oligo- and polysaccharide forms play central roles in many of biological process such as structure, food storage and utilization, viral invasion and highly selective cellular signaling events, glycoside hydrolases hydrolyzing these carbohydrate forms are important. Cellulases are one of the glycosyl hydrolases that catalyse polysaccharide form of carbohydrate, cellulose. In 1991, the classification of the glycosyl hydrolase into families was based on anino acid sequence similarity (58). This classification has been updated several times (59, 60). At the lastest update, the glycosyl hydrolase families are grouped into 86 families from more than 5,000 glycosyl hydrolases (13). This classification is based on sequence, 3D structure,

substrate specificity, and reaction mechanisms. Of the 86 families, 13 families (families 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 61 and 74) have cellulase activity (13, 61). Some families contain enzymes with diffirent substrate specificity, for example enzymes belonged to family 5 have cellulase, xylannase, and mannanase activities. At the same time, some families contain enzymes with the same substrate specificity but come from different organisms, for example, family 9 contains cellulases of bacteria, fungi, plant, and animal. In contrast, family 7 contains only fungal hydrolases whereas family 8 contains only bacterial hydrolases (13).

1.5 *Pichia pastoris*

1.5.1 General characteristics of *P. pastoris*

P. pastoris is the methylotrophic yeast that can utilize methanol as a sole source of carbon and energy (62). The metabolic pathway of methanol utilization involves several unique enzymes (63). The first step is the oxidation of methanol to formaldehyde and hydrogen peroxide catalyzed by alcohol oxidase (AOX). Since hydrogen peroxide is toxic to the cells, methanol metabolism is processed within specialized organelle, called peroxisome, where hydrogen peroxide is degraded to oxygen and water (64, 65).

There are two genes that encode for alcohol oxidases in *P. pastoris*: *AOX1* and *AOX2*. *AOX1* is responsible for a vast majority of alcohol oxidase activity in the cell. Expression of the *AOX1* gene is regulated by two mechanisms: a repression/derepression and induction mechanism (66, 67, 68). The methanol is important to induce high expression level, normally $\geq$ 30% of total soluble protein in the methanol- grown cells at growth-limiting rates (69).

1.5.2 Advantages of using *P. pastoris* as an expression system

Because of several advantages over other expression system, the methylotrophic yeast, *P. pastoris,* has become a highly successful system for the production of various heterologous proteins.

In post-translation modification, *P. pastoris* expression system has potential to perform many of the post-translation modifications in the higher eukaryotic system such as processing of signal sequence, folding, disulfide bridge formation, lipid addition and O- and N- linked glycosylation (68). Both *P. pastoris*

and *Saccharomyces cerevisiae* are capable of adding O- and N- linked carbohydrates to the secreted proteins. However, this is a problem in *S. cerevisiae* as hyperglycosylation frequently occured. In contrast, the problem of hyperglycosylation in secreted heterologous proteins does not occur when *P. pastoris* have been used (70, 71, 72). For example; In the invertase, N-linked oligosaccharide chain of invertase from *P. pastoris* compose of 30 mannose units while N-linked oligosaccharide chain of invertase from *S. cerevisiae* consists of 50-150 mannose residue (71, 73).

At the expression level, *P. pastoris* is able to express high level of heterologous proteins. Compared to *S. cerevisiae* expression level, *P. pastoris* can produce up to 30% of total cellular protein for intracellular expression or up to 80% of total secreted protein for extracellular expression. Whereas protein yields from *S. cerevisiae* can reach only 1-5% of total protein (74, 75).

1.5.3 *Pichia* expression system

Expression of any foreign proteins in *P. pastoris* requires three general steps: (i) the insertion of gene into an expreesion vector, (ii) transformation of the expression vector into the *P. pastoris* genome by any of the available techniques such as electroporation, a spheroplast generation method, or whole cell method involving lithium chroride and polyethylene glycol (76, 77), and (iii) determination of potential expression strains for protein production (65). For successful production of heterologous proteins, detailed information on expression vector, *P. pastoris* strains, gene dosage, and promoter element are important.

1.5.2.1 Expression vector

Most expression vectors designed for heterologous protein expression consist of a 0.9 kb fragment of the *P. pastoris* 5' AOX1 promoter sequences and short AOX1-derived fragment of transcriptional termination sequence (78). Between these fragments, there is a multiple cloning sites for insertion of foreign gene of interest. The alcohol oxidase open reading frame (ORF) contain generally long 5' untranslated region (116 nucleotides). In insertion process, the first ATG of the heterologous coding sequence is inserted closest to the *AOX1* ATG (67). For secretion of foreign protein, in-frame fusion of the gene of the gene of interest is constructed with the vector containing the secretion signal sequence. The secretion signals of *P. pastoris* acid phosephatase (*PHO1*) or *S. cerevisiae* α-mating factor (α-

MF) are available (65, 79). In addition, all expression vectors contain selectable markers for regulation of transformant growth. *HIS4* genes from both *P. pastoris* and *S. cerevisiae*, *ARG4* from *S. cerevisiae*, and the *Sh ble* gene from *Streptoalloteichus hindustanus* which confers resistance to the bleomycin-related drug zeocin are selectable marker genes described for molecular genetic manipulation of yeast (76, 80, 81).

1.5.2.2 *P. pastoris* strains

Suitable *P. pastoris* expression strains are essential for expression of heterologous protein. Three types of host strains which are grouped based on their ability to use methanol are available. Methanol utilization ability resulted from deletion in one or both *AOX* genes. GS115 is a wild-type host strain (Mut^+, methanol utilization plus phenotype). It contains either *AOX*1 or *AOX*2 genes and can grow on methanol at wild-type rate. KM71 is a strain which has a partial deletion in the chromosomal *AOX*1 gene and replacement with the *S.cerevisiae ARG4* gene (82). As a result, *AOX* production for methanol metabolism must rely on the weaker *AOX*2 and can grow slowly on methanol induction (Mut^s, methanol utilization slow phenotype). Sometime, strains with *AOX* mutation produce foreign protein better than wild type strains (71, 83, 84). The third host MC100-3 is a host strain with both *AOX* genes deleted. It is unable to grow on methanol (Mut^-, methanol utilization minus phenotype). In addition, there are several protease-deficient strains such as SMD1163, SMD1165, and SMD1168 which have the ability to reduce proteins degradation which are rapidly degraded by proteases from fermentor cultures (85, 86).

1.5.2.3 Gene dosage

The effect of gene copy number on gene expression is different in each situation. In some cases, one copy is sufficient while in other cases 2 or more copies are needed. Some studies indicated that an increase in copy number has effect on gene expression. For example, the expression level of heterologous aprotinin increases when gene copy number increase to five copies (73, 87).

1.5.2.4 Promoter elements

Expression of heterologous proteins in methylotrophic yeast, *P. pastoris*, requires strong inducible promoter elements. Although the *AOX*1 promoter has been widely used to express numerous foreign genes, however in some cases this

promoter is not suitable. For example, methanol utilization is not suitable for the production of food products since methane, a petroleum-related compound, is one source of methanol (65). Therefore, alternative promoters that are not induced by methanol have been adapted to use in *P. pastoris* expression system. *P. pastoris GAP*, *FLD1* and *YPT1* promoters are examples of alternative promoters. The *P. pastoris* glyceroldehyde 3-phosphate dehydrogenase (*GAP*) gene promoter is a strong constitutive promoter that requires glucose to induce expression process (88). The *FLD1* (glutathione-dependent formaldehyde dehydrogenase) promoter needs methylated amines as nitrogen sources and methanol as a carbon source for its induction (89). The *YPT1* promoter provides a low but constitutive expression level in the media containing various carbon sources such as glucose, methanol, or mannitol (90).

1.6 Secretory pathway

P. pastoris can produce heterologous intracellular or extracellular proteins. For secretory proteins, they are modified in many steps through a secretory pathway before they are secreted to the extracellular medium. First, they enter the endoplasmic reticulum (ER). Proteins in ER are folded and some modification processes such as glycosylation, disulfide bridge formation, phosphorylation, and subunit assembly initiated. Next, proteins leave the ER and go to the Golgi compartment by packing in transport vesicles. Further glycosylation and peptide processing can take place in this part. Finally, proteins are packed again in secretory vesicles and are directed to the plasma membrane where they are secreted (Figure 3) (91, 92).

1.6.1 Secretion and signal peptide

Heterologous proteins require the signal sequence to direct them to secretory pathway. Although there are several different secretion signal sequences including the native signal peptides, the *S. cerevisiae* α-factor prepro peptide has been successfully used for heterologous protein expression.

In this study, prepro sequence of *S. cerevisiae* alpha mating factor (MFα-1) was used as a signal peptide for secretion process. This signal sequence contains an open reading frame encoding 165 amino acids that contain 22 amino acids of a hydrophobic part which act as signal peptide, approximately 60 amino acids of three

consensus N-linked glycosylation sites (Asn-X-Thr), and a region of four tandem repeats of α-factor. Each repeat is preceded by spacer sequences of six to eight amino acids (variations of Lys-Arg-Glu-Ala-Asp-Ala-Glu-Ala). This site is hypothesized to contain proteolytic processing signal (79, 93). The removal of the pre sigmal occurred in the ER membrane by signal peptidase. Kex2 endopeptidase cleaves between Arg-Lys of pro leader sequence region, and then the process is followed rapidly by cleavage of Glu-Ala repeat by Ste13 (dipeptidyl aminopeptidase) protein leaving mature signal sequence (Figure 4). Normally, this signal peptide is removed from the foreign protein before exiting from the yeast cell (94). Although some studies showed that *S. cerevisiae* α-factor prepro signal sequence is a better secretion signal for expression in *P. pastoris*, it is unsuitable in some cases. The secretion efficiency is not determined by only a signal sequence but also depends on the structural nature of the recombinant protein (95).

1.6.2 Glycosylation

P. pastoris is capable of adding both N-(asparagines) and O-(serine or threonine) linked carbohydrates to secreted proteins (96). However, the numbers and types of sugar units added by mammalian cells are different from those added by lower eukaryotes such as *P. pastoris*. For instance, in O-linked glycosylation, mammalian cells add several of sugars including N-acetylgalactosamine, galactose (Gal), and sialic acid (NeuAc) while *P. pastoris* adds only mannose (Man) residues (65). Although the information involving O-linked glycosylation in *P. pastoris* is limited there are some reports that show O-linked glycosylated proteins in *P. pastoris* such as the *Aspergillus awamori* glucoamylase catalytic domain (97), human IGF-I (85), barley α-amylase 1 and 2 (98), and human single-chain urokinase-type plasminogen activator (99).

In all eukaryotes, an oligosaccharide unit $Man_8GlcNAc_2$ (where GlcNAc = *N*-acetylglucosamine and Man = mannose) is added to asparagines in N-linked site. However, the pattern of glycosylation at this site is different between higher eukaryote such as human and lower eukaryote such as *P. pastoris*. The pattern in human is $Man_{5-6}GlcNAc_2$ whereas $Man_{8-9}GlcNAc_2$ is the typical outer chain pattern occurred in *P. pastoris* (95). Although hyperglycosylation is not a frequent situation in *P. pastoris*, some foreign proteins secreted by *P. pastoris* appear to be

hyperglycosylated such as equine herpesviral 1 glycoprotein D (100), equine lactoferrin (101).

1.7 Computer programs for data analysis

1.7.1 NCBI-BLAST

Basic Local Alignment Search Tool (BLAST) is a method for rapid searching of sequence database similarity. This tool is available in http://www.ncbi.nlm.nih.gov/BLAST/. The query sequences used to submit via the BLAST Web pages can be either DNA or protein sequences. In NCBI-BLAST, the pairwise alignment is performed to find similarity between the query sequence and other sequences in the database. BLAST can interpret results into two statistical scores, a bit score and expect value (E value). The bit score indicates the degree of similarity in the alignment between the query sequence and the compared sequence in the database. This score is calculated from the number of similarity or identity residues and any gaps introduced to sequence alignment. If a bit score is high, it shows more similarity. Whereas the E-value shows as the indication of the statistical significance of a given pairwise alignment, it represents the more significant than the bit score if the E-value has a lower value (102, 103).

There are different BLAST variants available types of BLAST in NCBI-BLAST program. Each type of BLAST is selected depended on query sequence and expected results.

(1). Standard nucleotide-nucleotide BLAST (blastn) compares a nucleotide query to NCBI nucleotide database.

(2). Standard protein-protein BLAST (blastp) compares a protein query sequence to NCBI protein sequence database.

(3). Nucleotide query-protein db (blastx) compares a nucleotide sequence to NCBI protein sequence database. The query sequence is translated in six frames before each frame compares against protein database.

(4). Protein query-translated db (tbastn) compares a protein query sequence to NCBI nucleotide sequence database that are translated in all reading frames.

(5). Nucleotide query-translated db (tbastx) compares all translated query sequence in six frames against the six-frame translations of NCBI nucleotide sequence database.

1.7.2 Sequence alignment programs

The multiple sequence alignment programs are essential tools in molecular biology. The two well-known programs used in the alignment are CLUSTAL W and CLUSTAL X. Both programs are freely available and portable to run on a variety of computer and operating systems. CLUSTAL W was developed in 1994 to improve the sensitivity of progressive multiple sequence alignment through sequence weighing, positive-specific gap penalties and weight matrix choice. The program can build up a multiple alignment by using a series of pairwise alignment. First, the most closely related sequences are aligned in order to group these sequences. Then divergent sequences are added gradually to align together by keeping the location of the gaps found in the early alignment. Relationships between sequences derived from the alignment program can be visualized in phylogenetic tree by using Neighbour-Joining method (104). To focus on providing a moderm window–type interface, CLUSTAL X has progressed as the new system that is easier to use. CLUSTAL X is a new graphical interface to the CLUSTAL W program. Although alignments produced by CLUSTAL X are the same as those produced by CLUSTAL W, this tool provides all options required for multiple sequence alignment including the ability to cut and paste sequence to change the order of alignment, selection of a subset of the sequence to be realigned and selection of sub-range of the alignment to be realigned and inserted back into the original alignment. Moreover, this program can allow the user to highlight low-score region in the alignment (105, 106).

1.7.3 SWISS-MODEL

SWISS-MODEL is a server that automatically provides comparative modeling of three dimensional (3D) protein structures from submitted protein sequence. The server is available at http://swissmodel.expasy.org. This server requires only a protein sequence as the input data. Then the server will automatically select suitable templates, align sequences, and build protein structure model. Finally, the models including the detail report are sent back via email. The program DeepView

(Swiss-PdbViewer) and WebLab Viewer Pro are the programs for visualization, analysis and manipulation of protein structure (107).

1.7.4 SignalP Program

Signal peptide is the N-terminal part of the amino acid chain in the secreted protein. It comprises three regions: a positively charged n-region, followed by a hydrophobic h-region and a neutral but polar c-region. The cleavage site is found at the small and neutral position (108). The location of the signal peptide and the cleavage site can be predicted using available prediction method via WWW server. SignalP is the computational method for prediction of N-terminal signal peptides and the cleavage site based on neural network and hidden Markov model algorithms. The trained networks provide two different scores between zero and one for each amino acid residues in the protein chain. The output result from signalP shows three different scores, C, S and Y. C-score is interpreted as the cleavage site score. It estimates the probability of the position being the first in the mature protein. S-score shows as an estimate of the probability of the position belonging to the signal peptide. If the S-score value is greater than 0.5, it can predict the sequence in that position to be a signal peptide. Since C-score may appear several peaks and the S-score may fluctuate, the true cleavage site position will correspond to the maximal Y-score. Y-score is a derivative of the C-score combined with the S-score. In Figure 6 (A), the output of secreted protein shows the values of the C-, S- and Y-scores. The S-score is more than 0.5 in the position of predicted signal peptides. The C-score has one sharp peak that corresponds to the maximal Y-score. In contrast, a typical non-secreted protein has low values in all C-, S- and Y-scores as shown in Figure 5 (B) (109). SignalP is available at server http://www.cbs.dtu.dk/services/SigmalP/ (110).

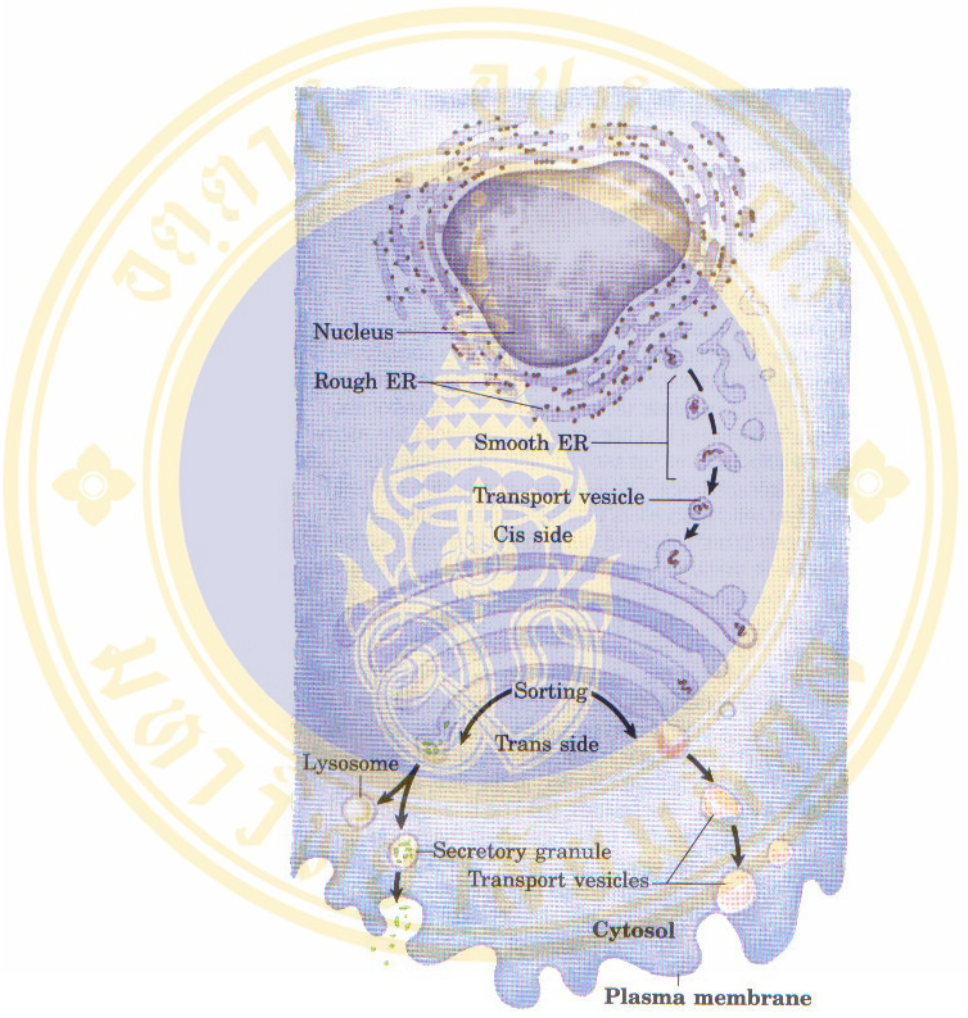


Figure 3. The secretory pathway of secreted proteins

This figure shows the pathway of proteins destined for lysosomes, the plasma membrane, or secretion. Proteins are moved from ER to the Golgi complex by transport vesicles. Modification processes occur here before proteins are sent to their final destinations (92).

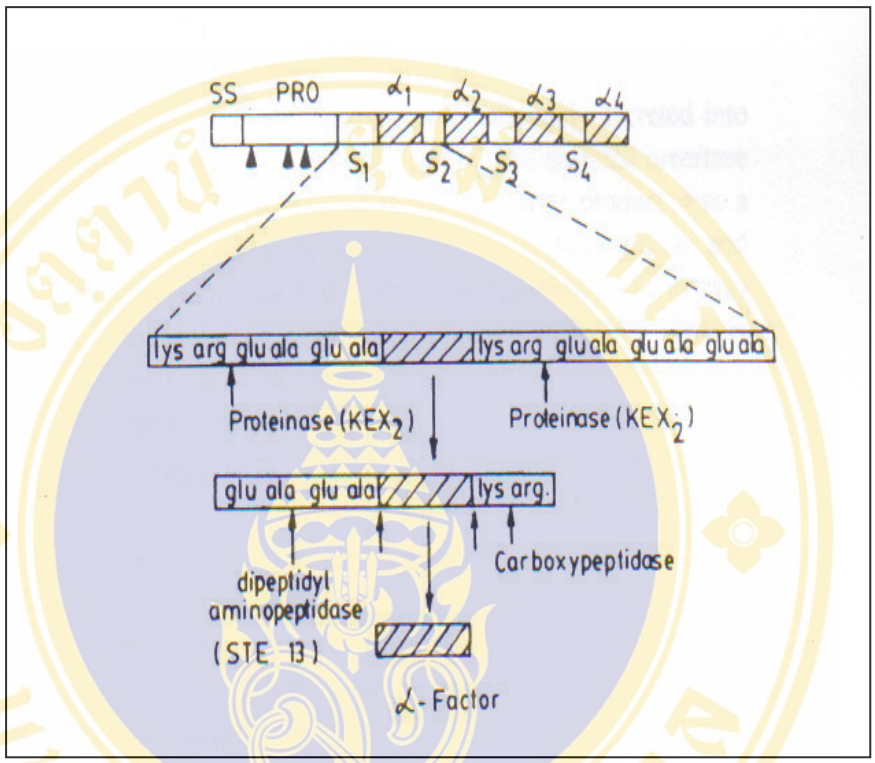


Figure 4. Processing of the *Saccharomyces cerevisiae* prepro-α-factor alpha mating factor (MFα-1)

To generate the mature α-factor peptide, prepro-α-factor are cleaved by KEX_2 and STE13 as indicated sites by the arrows. SS represents signal sequence. Pro is pro segment. Glycosylation sits are indicated by dark triangle (▲). S_1 , S_2 , S_3 , and S_4 show spacer (take from: Heslot and Gaillardin, 1992) (79).

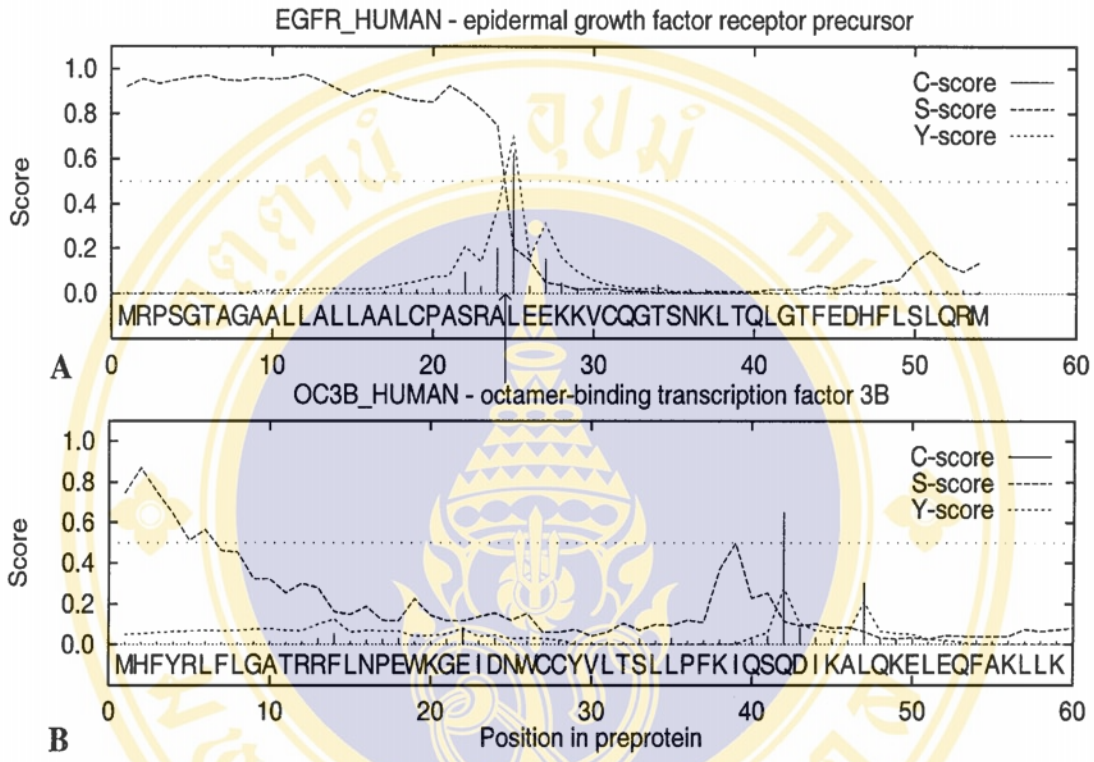


Figure 5. The network outputs from SignalP of secreted and non-secreted protein

The value of the C-, S- and Y-scores are shown for each position in the sequence. (A) Successfully predicted signal peptides of secreted protein show that the C-, S- and Y-scores are high for the position of signal peptide. After the cleavage site as marked with an arrow, the first position in the mature protein is shown. (B) For non-secreted protein, all three scores are very low throughout the sequence. (taken from Nielsen, 1997) (109).

CHAPTER II
OBJECTIVES

The preliminary screening results of more than 200 thermotolerant fungi from the BIOTEC Molecular and Enzyme Screening Laboratory demonstrated that *Syncephalastrum racemosum* (BCC18080) produced the highest level of thermotolerant cellulases. Specifically, over 50% of cellulases activity remained after incubation at 80^{o}C for an hour. One of the cellulolytic enzymes of interest is endoglucanase as it is the major cellulolytic enzyme produced in fungi, responsible for the degradation of cellulose. Therefore, my M.Sc. thesis is focused on (i) cloning and characterization of genes encoding endoglucanase from BCC18080 isolated from Thailand and (ii) expression of the active endoglucanase in the *Pichia pastoris* system.

CHAPTER III
MATERIALS

3.1 Fungal strain

Syncephalastrum racemosum (BCC18080), a filamentous fungus, was obtained from the BIOTEC Culture Collection (BCC), BIOTEC, Thailand.

3.2 Bacterial strain

Escherichaia coli, DH5α [*sup*E 44 Δ*lac*U169, (Φ80 *lac*Z ΔM15) *hsd*R17 *rac* A1 *end* A1 *gyr*A96 *thi*-1 *rel*A1], was used as a host for plasmid propagation.

3.3 Yeast strain

Pichia pastoris, KM71 (*arg*4, *his*4, *aox*1::*ARG*4), was used as a host cell for yeast expression of fungus endoglucanase gene.

3.4 Plasmid vectors

4.1 pGEM®-T easy vector (Promega, USA) is a vector for all cloning steps. A physical pGEM®-T Easy vector map was shown in Figure 6.

4.2 pPICZαA vector (Invitrogen, USA) was employed as a vector for expression of *P. pastoris* integrants containing an endoglucanase gene without leader sequence). A physical pPICZαA vector map was shown in Figure 7.

4.3 pPICZB vector (Invitrogen, USA) was utilized as a vector for expression of *P. pastoris* integrants containing an endoglucanase gene with leader sequence). A physical pPICZB vector map was shown in Figure 8.

3.5 Synthetic oligonucleotides

All nucleotide primers (Table 1) were purchased from either Bioservice Unit (BSU), BIOTEC, Thailand or PROLIGO, Singapore.

3.6 Culture media

3.6.1 Fungal culture medium

S. racemosum (BCC18080) was grown at 30°C in 5% wheat bran broth. For agar plates, 1.5% (w/v) of bacteriology agar was added.

3.6.2 Bacterial culture medium (For transformants containing pGEM®-T easy vector)

E. coli, DH5α, was grown in LB medium containing 1 % (w/v) peptone (Bio Basic Inc.), 0.5 % (w/v) yeast extract (Bio Basic Inc.) and 1% (w/v) NaCl. For LB plate agar, 2% (w/v) bacteriology agar was added. The *E.coli* transformants were grown in LB containing 100 µg/ml ampicillin (Sigma, USA). X-gal/ IPTG agar plates, 20 µl of a stock solution of 20 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) in dimethylformamide and 4 µl of a solution of isopropylthio-β-D-galactoside (IPTG), were used for selecting the transformants.

3.6.3 Bacterial culture medium (For transformant containing pPICZαA vector and pPICZB vector)

E. coli, DH5α, was grown in low-salt LB medium containing 1 % (w/v) peptone (Bio Basic Inc.), 0.5 % (w/v) yeast extract (Bio Basic Inc.) and 0.5% (w/v) NaCl. For LB plate agar, 1.5% (w/v) bacteriology agar was added. The *E. coli* transformants were grown in low-salt LB containing 25 ug/ml Zeocin™ as selectable marker.

3.6.4 Yeast culture medium

3.6.4.1 Culture and selective media

P. pastoris (strain KM71) was grown in a rich medium YEPD containing 2 % (w/v) peptone (Bio Basic Inc.), 1 % (w/v) yeast extract (Bio Basic Inc.) and 2% glucose (Sigma, USA). For YEPD plate agar, 2 %(w/v) bacteriology agar was added. The *P. pastoris* transformants were grown under selective condition in a rich medium YEPD containing 100 ug/ml Zeocin™.

3.6.4.2 Expression medium

Buffered minimal glycerol complex medium (BMGY) consists of 1% (w/v) yeast extract (Bio Basic Inc.), 2% (w/v) peptone (Bio Basic Inc.), 100 mM potassium phosphate pH 6.0, 1.34% (w/v) YNB, 4x 10^{-5}% biotin (w/v) and 1% (v/v)glycerol.

Buffered minimal methanol complex medium (BMMY) consists of 1% (w/v) yeast extract (Bio Basic Inc.), 2% (w/v) peptone (Bio Basic Inc.), 100 mM potassium phosphate pH 6.0, 1.34%(w/v) YNB, 4x 10^{-5}% biotin (w/v) and 3% (v/v) methanol.

3.7 Standard cellulase and cellulose

Cellulase from *Tricoderma reesei* used as a reference enzyme was purchased from Boehringer Mannheim, Germany. AZCL-HE cellulose was purchase from Megazyme, Ireland.

3.8 Chemicals

Zeocin™ was purchased from either Invitrogen, USA and other chemicals and solvents (analytical grade) were purchased from Gibco BRL,USA; Fluka, Switzerland; Sigma, USA or Merck, Germany.

3.9 Enzymes

Restriction enzymes and other modification enzymes were purchased from New England Biolabs, USA; Boehringer Mannheim, Germany; Promega, USA; Gibco BRL, USA; Stratagene, USA or Sigma, USA.

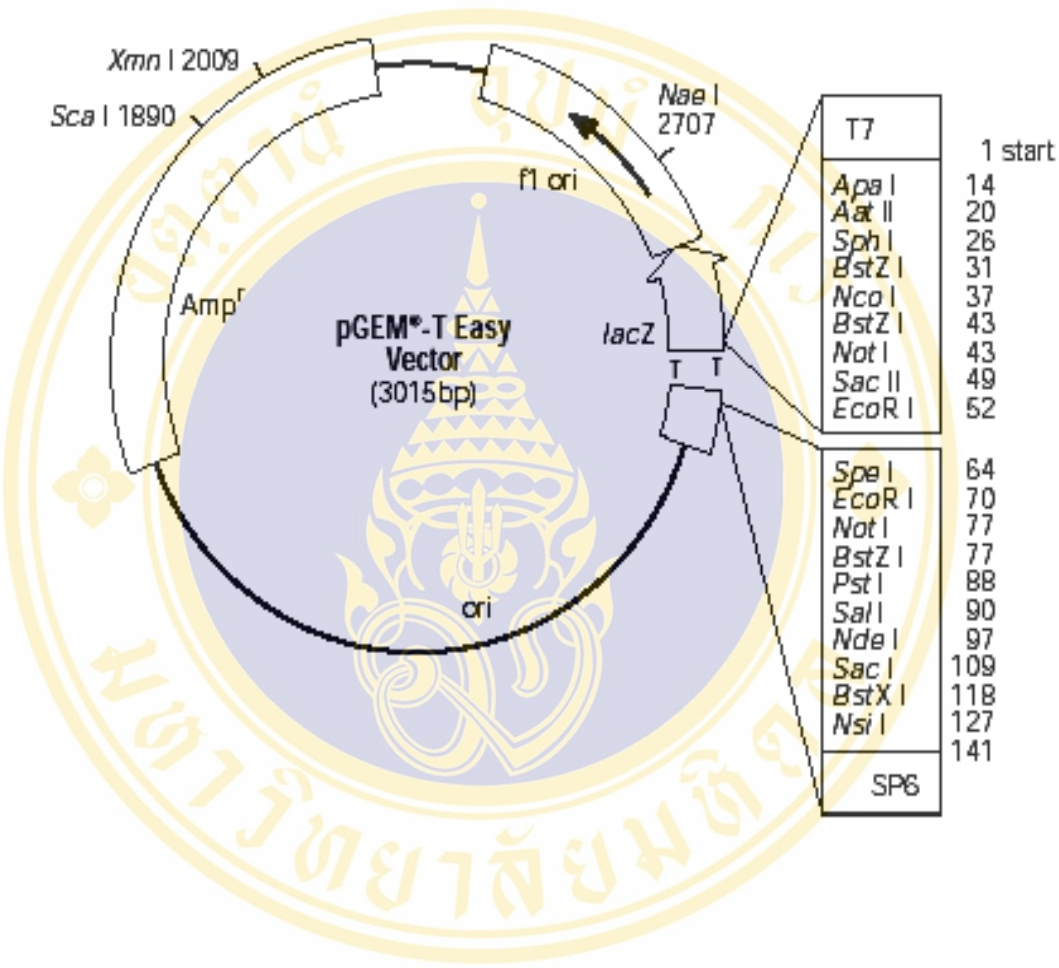


Figure 6. A physical map of pGEM®-T Easy Vector (Taken from: A pGEM®-T Easy Vector technical manual No.042)

The figure illustrates a pGEM®-T Easy Vector containing *lacZ* gene, multiple cloning sites, origin of replication of *E.coli* and amplicilin resistant marker gene.

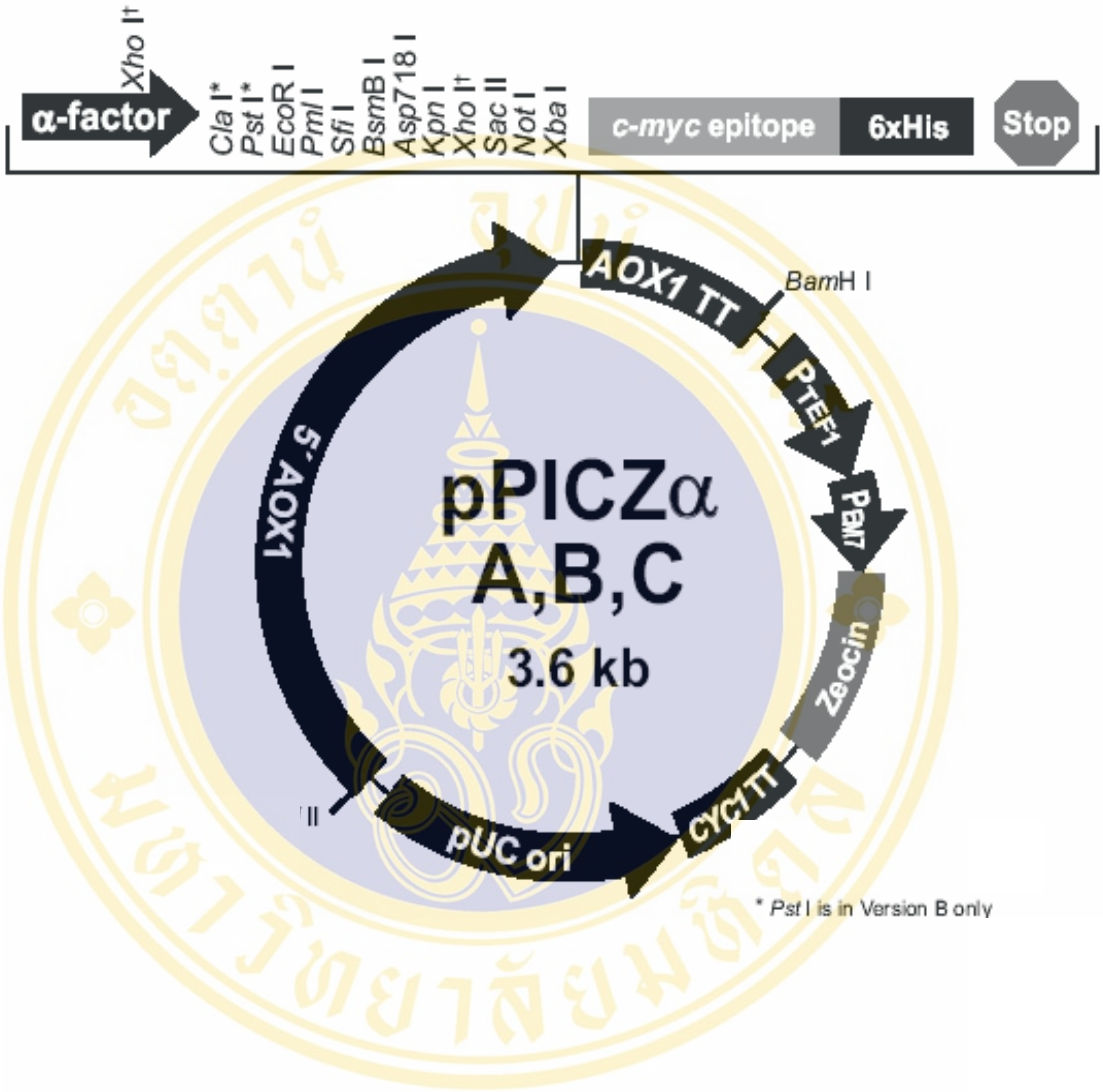


Figure 7. A physical map of pPICZαA vector (Taken from Invitrogen's instruction manual version E 010203 No.V195-20)

The figure illustrates plasmid pPICZαA containing 5'AOX1 promoter, *Sacharomyces cerevisiae* α-factor secretion signal, multiple cloning sites, AOX1 transcription termination (TT), Zeocin™ resistant gene and origin of replication of *E.coli.*

Figure 8. A physical map of pPICZB vector (Taken from Invitrogen's instruction manual version E 010203 No.V195-20)

The figure illustrates pPICZB containing 5'AOX1 promoter, multiple cloning sites, AOX1 transcription termination (TT), Zeocin^{TM} resistant gene and origin of replication of *E.coli.*

Table 1. Synthetic nucleotide primers used in each experiment. For degenerate primers, the following abbreviations are used (R= A, G; Y= T, C; N= A, T, C, G)

Experiment and Primer name	Sequence (5'→3')	Size (bp)	Tm (°C)
Partial cDNA amplification			
EndoF1	CARTGYGGNGGNAARAATTGG	21	57.8
EndoF2	TGYAAYGAYAAYCARCCNTGG	21	55.9
EndoR1	RTTRTCNGCRTTYTTRAACCA	21	51.9
EndoR2	YTTNCGNGARCANCCNGTYTT	21	56.8
Reverse transcription			
PM1 oligo(dT)	CCGGAATTCAAGCTTCTAGAGG ATCCTTTTTTTTTTTTTTT	42	70
Adapter primer			
PM2	CCGGAATTCAAGCTTCTAGAGG ATCC	26	60.5
3'-end cDNA amplification			
3'RaceEndo	GAGGGCAAGAAGATGGTCGT	20	53.1
5'-end cDNA amplification			
5'RaceEndo1	ACGACCATCTTCTTGCCCTC	20	53.1
5'RaceEndo2Nested	CCGCTGATGCTTGCAGCAGC	20	61.5
***P. pastoris* expression**			
expressEndoF1	GCGAATTCATGATCCTTCAC CGCACA	26	66.4
expressEndoF2	GCCTCGAGAAAAGAGCCGA TTGCAGC	26	67.5
expressEndoR	GCTCTAGATTAGGAGGTGCG TTCGCA	26	63

CHAPTER IV
METHODS

4.1 Cellulase activity assay using AZCL-HE cellulose

0.1% (w/v) AZCL-HE cellulose (Megazyme, Ireland) was soaked in 95% ethanol for 30 min before it was mixed with the prewarmed 1.5 % (w/v) bacteriology agar in 0.1 M phosphate buffer pH 5.8. A volume of 20 µl mixture was poured into a sterile plate. When the agar was solidified, the wells for loading samples were created by puncturing in the agar plate using a sterile 5 mm diameter tube. A volume of 20 µl of sample was loaded into each well. Then, the AZCL-HE cellulose plate was incubated overnight at 30 ^{o}C. The cellulase activity was determined from the size of the blue zone produced around the well by measuring the length from the edge of the well to the edge of the blue zone in mm.

4.2 Optimization of cellulase production

A stock culture *S. racemosum* BCC18080 (obtained from the Molecular and Enzyme Screening Laboratory, BIOTEC) was inoculated in 5% wheat bran broth and grown at 30^{o}C with 250 x g continuous shaking. To determine the optimal condition for cellulase production, a volume of 500 µl of culture supernatant was collected everyday for 8 days and subjected to the cellulase activity assay as mentioned in 4.1.

4.3 RNA isolation

4.3.1 Total RNA isolation by using TRI REAGENT

Total RNA was isolated from *S. racemosum* BCC18080 mycelia after culturing in the wheat bran broth (5%w/v) for 5 days. The mycelia were separated from the broth by filtration through a filter cloth. A gram of mycelia was frozen in liquid nitrogen and ground with mortar and pestle until a floury consistency was obtained. Then, it was homogenized by using a hand-held homogenizer (Fisher, USA) in a volume of 10 ml of TRI-Reagent (Molecular Research Center, USA) as described by the

manufacture. Specifically after homogenization, the homogenate was transferred into a 50 ml tube and stored at room temperature for 5 min. Protein and polysaccharide were removed by adding 2 ml of chloroform, followed by vigorously shaking for 15 s and stored the mixture at room temperature for 10 min. The mixture was centrifuged at 10,000x g for 20 min at 4 ^{o}C. Then, the aqueous phase was transferred to a fresh tube. The RNA was precipitated by mixing with 5 ml of isopropanol and stored at room temperature for 10 min. The precipitated RNA was collected by centrifugation at 10,000x g for 20 min at 4 ^{o}C. After the supernatant was removed, the RNA pellet was washed once with 75% ethanol and centrifuged at 7,500x g for 5 min at 4 ^{o}C. The RNA pellet was briefly air-dried for 5 min and resuspended in DEPC-treated water.The isolated total RNA was either stored at -80 ^{o}C or used immediately.

4.3.2 Determination of RNA concentration and purity

Both RNA quantity and quality were determined by using a spectrophotometer (Hitachi, Japan) at absorbance of 260 nm and 280 nm. The RNA concentration was calculated from the absorbance value measured at 260 nm by using the following equation:

$$\text{RNA concentration } (\mu g/\mu l) = \frac{A_{260} \times (\text{dilution factor}) \times 40}{1000}$$

The absorbance ratio of A_{260}/A_{280} was used to determine the RNA purity. A ratio of high purity of RNA should be between 1.8-2.0.

4.3.3 RNA electrophoresis

The quality and integrity of RNA was determined using gel electrophoresis in 1% agarose gel containing 3.5% formaldehyde. RNA agarose gel was prepared by boiling 0.4 g of agarose in 34 ml of DEPC-treated water and cooling down to 60 ^{o}C. Subsequently, 4 ml of 10X MOPS buffer [200 mM of morpholinopropanesulfonic acid (MOPS), 50mM sodium acetate, and 10 mM Na_2EDTA] and 2.2 ml of 37% formaldehyde were added, mixed and poured quickly.

The RNA sample preparation was performed by mixing 2 µg of RNA with 2 µl of RNA loadind buffer (50% glycerol, 0.4% bromophenol blue, 1 mM EDTA, 1 mg/ml EtBr) and 6 µl of sample buffer (10 ml formamide, 3.5 ml of 37% formaldehyde, 2 ml of 10X MOPS).Then the RNA mixture was heated at 65 ^{o}C for 10 min and quickly cooled on ice for 5 min.

RNA agarose gel electrophoresis was carried out in the electrophoresis tank containing 1X MOPS buffer. The RNA mixture was loaded into the slots of the submerged gel and run at 120 volts for 5 min to allow the RNA enter into the gel then at 80 volts until the dye front is about 2/3 rd of the gel length. The integrity of RNA was analysed and photographed under the UV light.

4.3.4 RNA purification

The RNA sample was purified by using RNeasy RNA purification kit (QIAGEN, Germany) to remove any color remaining which might interfere PCR reacrion from the sample. The process was performed as described by the manufacture. Briefly, the RNA sample was adjusted to a volume of 100 µl with RNase-free water. A volume of 350 µl of buffer RLT was added and followed by 250 µl of absolute ethanol. The sample mixture (700 µl) was applied to an RNeasy mini column (QIAGEN, Germany) placed in a 2 ml collection tube. A volume of 500 µl of buffer RPE was added onto the column and the column was centrifuged at 10,000x g. The flow-through was discarded and 500 µl of buffer RPE was added. The column was centrifuged and the flow-through was discarded. The RNeasy silica gel membrane was dried by centrifugation at 13,000x g for 2 min. The RNeasy column is transferred to a new 1.5 ml collection tube and RNA was eluted by 30 µl RNase- free water.

4.4 Amplification of *S.racemosum* BCC18080 endoglucanase partial sequence by RT-PCR

4.4.1 First stranded cDNA synthesis

To generate first strand cDNA from *S.racemosum* BCC18080, total RNA from *S. racemosum* BCC18080 mycelium was isolated at the time of maximal cellulase activity and used as template. In this step, the first stand cDNA was synthesized in a 20 µl volume, containing 1X RT-buffer, 10 mM each of dNTPs (dATP, dCTP, dGTP and dTTP), 200 ng of random hexamer, 1µg of total RNA, 20 U RNase OUT^{TM} Recombinant Ribonuclease Inhibitor (Fermentas, Lithunia) and 200 U SuperscriptIII reverse transcriptase (Invitrogen, USA). The reaction mixture was incubated at 42°C for 1 hr. Then, the reaction was terminated by heating at 70°C for 15 min. After the first strand cDNAs were successfully synthesized, they were kept at -20°C until used.

4.4.2 Degenerate primer design

The degenerate oligonucleotide primers were designed based on the conserved region of the endoglucanase gene of the *Rhizopus oryzae RCE1, 2* and *3* (Q8J1L2, Q8J1L10, Q 8J1L1, respectively). They are:

(1.) EndoF1 (QCGGKNW): 5'-CARTGYGGNGGNAARAATTGG-3'

(2.) EndoF2 (CNDNQPW): 5'-TGYAAYGAYAAYCARCCNTGG-3'

(3.) EndoR1 (WFKNADN): 5'- RTTRTCNGCRTTYTTRAACCA-3'

(4.) EndoR2 (KTGCSRK): 5'-YTTNCGNGARCANCCNGTYTT-3'

All primers were synthesized from BioService Unit (BIOTEC).

4.4.3 PCR amplification

After reverse transcription, first strand cDNAs were used as a template for PCR amplification. Two rounds of PCR amplification (semi-nested PCR and nested PCR) were employed using degenerate primers. The first round of PCR reaction was performed in a total volume of 50 µl mixture containing 20 µM of EndoF1 and EndoR2, 1X Promega buffer, 2.5 mM each of dNTPs, 2 mM $MgCl_2$, 4 µl cDNA and 1 U Taq polymerase (IMBG, Mahidol University). Amplification cycles were composed of one cycle of 5 min at 94 °C, 35 cycles of 94°C for 30 s, 40°C for 30 s and 72°C for 45 s, then 10 min at 72°C. The second round of PCR was continued using 4 µl of the previous PCR products using EndoF2 and EndoR1 as primers for nested PCR and using EndoF2 and EndoR2 as primers for semi-nested PCR. The second round was amplified using the same PCR conditions as the first round. The PCR products were analysed on 1% agarose gel electrophoresis.

4.4.4 DNA electrophoresis

For gel electrophoresis, the agarose gel was prepared by melting 1% (w/v) of agarose in 1X TBE (Tris-borate buffer, 89 mM boric acid, 89 mM Tris-HCL, 2 mM EDTA, pH 8.0). The gel was poured into an electrophoretic tray and allowed to set at room temperature. 1X TBE was used as an electrophoretic buffer. The digested DNA or PCR products were mixed with 1X loading dye (25% glucerol, 60 mM EDTA and 0.25% bromophenol blue) and loaded into the prepared gel and run at a constant voltage of 100 volts for 1.5 hr. The gel was stained in 2.5 µg/ml of ethidium bromide solution for 5 min and destained in water 15-30 min. The DNA was visualized under the UV light (Gel Doc model 1000, Bio-Rad, USA) and photograghed.

4.5 Cloning of BCC18080 endoglucanase gene

4.5.1 Competent cell preparation

E.coli (DH5α) was inoculated into 250 ml SOB [2% peptone (Difco), 0.5% yeast extract (Difco), 0.05% NaCl] and grown at 18 °C with shaking at 250 x g until OD_{600} reached 0.6. The cell culture was transferred into five steriled 50 ml polypropylene centrifuge tubes which were then chilled on ice for 10 min. After centrifugation at 3,000 x g for 7 min at 4°C, the supernatant was discarded. The pellet was resuspended in 80 ml ice cold TB buffer (10 mM PIPES, 55 mM $MnCl_2$, 15 mM $CaCl_2$, and 250 mM KCl). Then, the tubes were chilled on ice for 10 min and then were pooled together into one tube. Cells were harvested by centrifuging at 3,000 x g for 7 min at 4°C. The cell pellet was resuspended in 20 ml ice cold TB buffer. Dimethyl sulfoxide was added to 7% final concentration and competent cells were stored at -80°C (111).

4.5.2 DNA purification by QIAquick gel extraction kit (QIAGEN)

The DNA band of expected size was excised from the agarose gel under the UV light. QIAGEN gel extraction kit was used to purify PCR product. As described in the menufacture‘s instruction manual, after weighing the gel slice in a tube, three volumes of buffer QG were added to one volume of the gel. The tube was incubated at 50°C for 10 min or until the gel slice was completely dissolved by vortexing the tube every 2-3 min during incubation. The color of the mixture should be yellow to ensure that the pH was higher than 7.5 which aim high efficiency of DNA binding to a QIAquick column. Then, one gel volume of isopropanol was added and the solution was gently mixed. To bind DNA, the solution was applied to the QIAquick spin column which was placed in a provided 2 ml collection tube. After centrifugation for 1 min, the flow-through was discarded and the QIAquick column was placed back in the same tube. Next, 0.5 ml buffer QG was added to the QIAquick column to remove all traces of agarose and the column was centrifuged for 1 min. 0.75 ml buffer PE was added to the QIAquick column to wash DNA and the column was left at room temperature for 2-5 min. After centrifugation for 1 min, the flow-through was discarded. The QIAquick column was centrifuged for an additional 1 min at 13,000 x g. Then, the QIAquick column was placed into a clean 1.5 ml microcentrifuge tube. To elute DNA, 30 µl of elution buffer was added to the center of the QIAquick membrane and

the column was left at room temperature for 5 min. The tube was then centrifuged as above. Then, 1-2 µl of concentrated DNA was analyzed by gel electrophoresis in order to determine the purity and concentration of DNA.

4.5.3 DNA ligation

The purified DNA was directly ligated to the pGEM-T Easy vector (Promega, USA) in a molar ratio of 1:3 (vector: insert). The amount of DNA used in ligation reaction was estimated using the following equation.

$$\text{ng of an insert} = \frac{\text{ng of vector x kb size of insert x (insert: vector molar ratio)}}{\text{kb size of vector}}$$

The ligation mixture was performed in a final volume of 10 µl of reaction mixture containing 25 ng of pGEM-T Easy vector, 1X rapid ligation buffer for T4 DNA ligase [30 mM Tris-HCl (pH7.8), 10 mM $MgCl_2$, 10 mM EDTA, 1 mM ATP and 5% polyethyleneglycol] and 2.5 unite of high concentration T4 DNA ligase (Promega, USA). The ligation mixture was mixed and incubated overnight at 16 °C.

4.5.4 Transformation of competent *E.coli* DH5α cells

A volume of 100 µl of *E. coli* competent cells was mixed gently with 5 ng of ligated product. The mixture was left on ice for 30 min. The cell mixture was heat shocked at 42°C for 90 s and then immediately placed back on ice for 30 min. A volume of 900 µl of LB medium [1 % (w/v) peptone (Bio Basic Inc.), 0.5 % (w/v) yeast extract (Bio Basic Inc.) and 1% (w/v) NaCl] was added. Next, the mixture was incubated at 37 °C for 1 hr. with constant shaking. Then, the cells were spread on LB agar plate containing 100 µg/ml of ampicillin, 4 µl of 2% IPTG (isopropylthio-β-D-galactoside) and 20 µl of 20% X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside)(IPTG and X-gal were spread over the surface of an LB plate). The agar plate was incubated at 37 °C for 12-16 hr.

4.5.5 Master plate preparation and selection for recombinant clones screening

Blue and white colonies were observed after incubation at 37 °C for 12-16 hr. About twenty white clones were picked and spotted (10-15 spots per clone) on LB agar plate containing ampicillin called a master plate. The master plate was incubated at 37 °C for 12-16 hr. Recombinant clones were screened using simplified rapid size

screening method. Specifically, one colony per clone was picked by toothpick from the master plate and lysed in 30 µl of pre-warm lysis buffer [5mM EDTA, 10 % (w/v) sucrose, 0.25%(w/v) SDS, 100 mM NaOH, 60 mM KCl and 0.05% (w/v) bromophenol blue] and were incubated at 37 °C for 5 min and then placed on ice for 5 min. After centrifugation at 13,000 x g for 5 min, 20 µl of supernatant was analysed on 0.8% agarose gel electrophoresis. Those clones containing a larger size plasmid than the original pGEMT-Easy vector alone were selected for plasmid isolation.

4.5.6 Plasmid DNA extraction using CTAB method

For this method, a selected clone was inoculated into 3 ml of LB broth containing 100 µg/ml amplicillin and incubated with vigorous shaking at 37 °C for 12-16 hr. To collect the cell pellet, the cell suspension in a microcentrifuge tube was centrifuged at 10,000 x g for 10 s. The cell pellet was resuspended in 200 µl of STET buffer [8% (w/v) sucrose, 1% triton X-100, 50 mM EDTA and 50 mM Tris pH 8.0] by vortexing vigorously. Then, the cell suspension was added with 25 µl of 10 mg/ml lysozyme, incubated at room temperature for 10 min and placed in a boiling water bath for 45 sec. Next, the lysed cells were centrifuged at 12,000 x g for 15 min, cell debris was removed by using a sterile toothpick. Next, 20 µl of 5% CTAB (cetyl trimethyl ammonium bromide) was added. The tube was mixed by inverting the tube and incubating at room temperature for 30 min. The mixture was centrifuged at 12,000 x g for 5 min. Then the supernatant was removed and the DNA pellet was resuspended in 300 µl of 1.2 M NaCl. A volume of 5 µl of RNaseA (10 mg/ml) was added and the mixture was incubated at 37°C for 30 min. Next, a volume of 300 µl (an equal volume of aqueous phase) of chloroform was added and the mixture was centrifuged at 12,000 x g for 5 min. The upper aqueous phase was transferred to a fresh microtube. To precipitate plasmid DNA, approximately 600 µl (2 volumes of aqueous phase) of absolute ethanol was added. After incubation at room temperature for 5 min, the precipitated plasmid DNA was centrifuged at 12,000 x g for 15 min and then the supernatant was discarded. The DNA pellet was washed twice in 300 µl of 70 % (v/v) ethanol and air dried. A completely dried plasmid DNA pellet was resuspended in 20-30 µl of sterile distilled water. The quality and concentration of plasmid DNA was analyzed on 1% agarose gel electrophoresis.

4.5.7 Plasmid DNA extraction using QIAprep Spin Miniprep kit (QIAGEN)

A volume of 3 ml of the overnight culture was collected by centrifugation at 10,000 x g for 10 sec. The pelleted bacterial cells were resuspended in 250 µl of Buffer P1, and then 250 µl of Buffer P2 was added. After gently inverting the tube 4-6 times to mix, a 350 µl of Buffer N3 was added and the tube was inverted immediately 4-6 times and then centrifuged at 13,000 x g for 10 min in a table-top microcentrifuge. The supernatant was transferred to the QIAprep spin column inserted in a collection tube. The QIAprep spin column was centrifuged at 13,000 x g for 1 min. The flow-through was discarded. Then, The QIAprep spin column was washed with 750 µl of Buffer PE and centrifuged at 13,000 x g for 1 min. After the flow-through was discarded, the column was centrifuged for an additional 1 min to remove residual wash buffer. The QIAprep column was placed in a clean 1.5 ml microcentrifuge tube and the DNA was eluted by adding 30 µl of Buffer EB (10mM Tris-Cl, pH 8.5), left standing for 1 min and followed by centrifuging at 13,000 x g for 1 min. The quality and concentration of plasmid DNA was analyzed on 1% agarose gel electrophoresis.

4.5.8 Restriction enzyme digestion

After the recombinant plasmid was purified, approximately 1 µl of plasmid was added to the reaction mixture containing 1X buffer H [90 mM Tris-HCl (pH 7.5), 50mM NaCl and 10mM $MgCl_2$] and 2-4 units of *Eco*RI restriction endonuclease enzyme (Promega, USA) in the final volume of 10 µl. The mixture was incubated at 37°C for 3 hr and then the size of insert was determined using an agarose gel electrophoresis.

4.5.9 DNA sequencing

The recombinant clones containing cDNA of interest were further analyzed by DNA sequencing (Macrogen, Korea).

4.5.10 Sequence analysis

To translate the obtained DNA sequences into a deduced amino acid sequence, the Bioedit Sequence Alignment Edit program was used. The obtained DNA sequences from different clones were aligned together by Clustal W program. Either the nucleotide or the deduced amino acid sequence was searched in GenBank database in the NCBI website (http://www.ncbi.nlm.nih.gov) and compared to other related sequences by using Clustal W.

4.6 3' RACE method for amplification of 3' end of BCC18080 endoglucanase gene

To obtain the 3'end of cDNA of a BCC18080 endoglucanase, 3' rapid amplification of cDNA end (3'RACE) was employed (Figure 10A). A forward gene specific primer, 3'RaceEndo was designed from sequence of a partial endoglucanase sequence previously obtained. The first strand cDNA was constructed from 1 µg of total RNA from BCC18080 mycelium using PM1, oligo-dT-adaptor primer, in the presence of Superscript III reverse transcriptase (Invitrogen, Canada) as described in method 4.4.1. Then, the PCR amplification was performed using 3'RaceEndo, a gene specific primer, and PM2, an adaptor primer in a total volume of 50 µl mixture containing 20 µM of 3'RaceEndo and PM2, 1X DyNazyme buffer containing 1.5 mM $MgCl_2$, 2.5 mM each of dNTPs, 4 µl cDNA and 1 U DyNazyme EXT^{TM} DNA polymerase (Finnzyme, Finland). PCR profile of the 3'end cDNA amplification is shown in figure 10B. The PCR product was analyzed on 1% agarose gel electrophoresis. Next, it was purified, cloned and sequenced as described in protocol 4.5.2-4.5.9. The 3'end cDNA sequence was analyzed as described in 4.5.10.

4.7 5' RACE method for amplification of 5'end of BCC18080 endoglucanase gene

4.7.1 First stranded cDNA synthesis by partial heat denaturation method

The 5' end of cDNA of BCC18080 endoglucanase was obtained by 5' rapid amplification of cDNA end (5'RACE) method (Figure 11A). However, in order to prevent a strong RNA secondary structure problem, the first-strand cDNAs were generated using partial heat denaturation reverse transcription method (112). 3 µg of total RNA from BCC18080 mycelium was reverse transcribed by using 10 µM of 5'RaceEndo1, a gene specific primer. The first-stranded cDNAs were synthesized by Superscript III reverse transcriptase (Invitrogen, Canada) as described in method 4.4.1. Then, the partial heat denaturation method was performed. Specifically, the mixture was heated at 85°C for 3 min and immediately put on ice water slurry for 2 min. The reverse transcription was continued by adding 1 µl (200 unit) of SuperscriptIII reverse transcriptase and incubated at 42 °C for 60 min. Reverse transcriptase was heat-inactivated at 70°C for 15 min. A volume of 2.5 µl of 5 M NaOH was added to a

mixture to remove the RNA template. Then it was incubated at 55°C for 30 min and neutralized with 72 µl of 1% acetic acid.

4.7.2 QIAquick PCR purification

QIAquick PCR purification kit (QIAGEN) was used to purify the first stranded cDNA from dNTPs, ATP, salts and primer before poly-A tailing. The purification process was performed according to manufacturer's instruction. Briefly, to the cDNA solution five volumes of PB buffer was added and then the mixture was applied to a QIAquick spin column placed on a 2 ml collection tube. The column was centrifuged at 13,000 x g at room temperature for 1 min. The flow-through was discarded. Next, 0.75 ml of PE buffer was added before following with centrifugation at 13,000 x g for 1 min. After the flow-through was discarded, the column was centrifuged for an addition 1 min at 13,000 x g. The column was placed into a new microcentrifuge tube. Then at least 30 µl of sterile distilled water was added into the center of membrane to elute the first-strand cDNAs. The eluted cDNAs were collected in microcentrifuge tube by centrifuging at 13,000 x g for 1 min after the column was standed at room temperature for 1 min.

4.7.3 cDNAs precipitation of first strand

In order to increase the quantity of first strand cDNA, DNA precipitation was performed. The cDNA solution was adjusted with 3 M NaOAC, pH 5.2 to final concentration of 0.3 M NaOAC. Three volumes of absoluted ethanol were then added. After the mixture was kept at -20°C for overnight, it was centrifuged at 14,000 x g, 4°C for 20 min. Then, ethanol was removed carefully by pipetting. The first strand cDNA was washed twice with 2 volumes of 70% cold ethanol. After centrifugation at 14,000 x g for 10 min, first strand cDNA was eluted with 10 µl sterile distilled water.

4.7.4 Poly-A tailing of first stranded cDNA

Poly-A tail was added to the 3'end of the first-strand cDNAs (which corresponded to the 5'-end of the mRNA) in order to allow oligo(dT) primer to bind. In the final volume of 15 µl reaction, 10 µl of precipitated first-strand cDNAs were added to the reaction mixture containing 1.5 µl of 25 mM dATP, 1X TdT buffer [100 mM cacodylate buffer (pH 6.8), 1 mM $CoCl_2$, and 0.1 mM DTT], and 0.5 µl of 30 unit/µl terminal deoxynucleotidyl transferase (TdT) (Promega, USA). Then, the mixture was

incubated at 30°C for 1 h. The TdT was inactivated by incubation at 70°C for 10 min. The poly-A tailed first stranded cDNA was stored at -20°C until used.

4.7.5 Synthesis of 5' end cDNA by PCR amplification

Two gene-specific primers, 5'RaceEndo1 and 5'RaceEndo2Nested**,** were designed from the obtained partial sequence of BCC18080 endoglucanase in order to amplify the 5' end cDNA. Two rounds of PCR amplification were performed using two different primer pairs (Figure 11A). The first PCR was performed with 5'RaceEndo1, a gene specific primer and PM1, an oligo-dT-adapter primer. Then, the first PCR product was used as a template for the second PCR performed with 5'RaceEndo2Nested, a gene specific primer and PM2, an adaptor primer. Both PCR reactions were done in the presence of 4 µl of cDNAs template, 20 µM of each primer, 1X DyNaZyme buffer containing 1.5 mM of $MgCl_2$, 2.5 mM each of dNTP's and 1U of DyNaZyme EXT^{TM} DNA polymerase (Finnzyme, Finland) in a final volume of 50 µl. The first round of PCR amplification was carried out following PCR profile as shown in Figure 11B. The nested PCR was performed in the same condition, except in the annealing step, the temperature was changed to 54 °C (Figure 11B). The PCR product was analyzed on 1% agarose gel electrophoresis. Next, it was purified, cloned and sequenced as described in protocol 2.4.2-2.4.9. The 5'end of cDNA sequence was combined with the previously identified sequence of an endoglucanase and analyzed as described in 2.4.10.

4.7.6 Prediction of BCC18080 leader sequence

After obtaining the 3'end of cDNA, all partial sequences were combined and the predicted full-length of BCC18080 endoglucanase was analyzed. A leader sequence of the obtained endoglucanase was predicted by using SignalP 3.0 prediction program (http://www.cbs.dtu.dk/services/SigmalP/). The information from this prediction was used to designed three gene specific primers corresponding to the 5' and 3' end, expressEndoF1, expressEndoF2 and expressEndoR, which were used for amplification of the full-length endoglucanase cDNA with or without leader sequence (Table 1).

4.8 Construction of endoglucanase in *P. pastoris* expression vector

4.8.1 Amplification of the full-length endoglucanase cDNA with or without its own leader sequence

The full-length endoglucanase gene with or without its own leader sequence were constructed by using PCR. ExpressEndoF1 and expressEndoR primers were used for amplifying endoglucanase gene with its own leader sequence while expressEndoF2 and expressEndoR primers were used for amplifying endoglucanase gene without leader sequence. First, the first-strand cDNAs were synthesized from total RNA isolated from the BCC18080 mycelium using SuperscriptIII reverse transcriptase (Invitrogen, Canada). The reaction conditions were as recommended by the supplier as described in method 2.3.1. The PCR reaction was performed in a 50 µl mixture that contained 10 µM of each expressEndoF or expressEndoF2 and 10µM of expressEndoR, 1X DyNazyme buffer containing 1.5 mM $MgCl_2$, 2.5 mM each of dNTP's, 4 µl of cDNA and 1U of DyNaZyme EXT^{TM} DNA polymerase (Finnzyme, Finland). Amplification cycles for both constructs were as follow: one cycle of 5 min at 94°C, 35 cycles with 30 sec at 94°C, 45 sec at 58°C and 1 min at 72°C, then 10 min at 72°C. The PCR product was analyzed on 1% agarose gel electrophoresis and purified using QIAGEN gel extraction kit.

4.8.2 Preparation of the recombinant plasmids

To subclone the full-length endoglucanase gene with or without its own leader sequence into the *P. pastoris* vectors, two expression vectors, pPICZB and pPICZαA were linearized. The linearized pPICZB vector was prepared by double digestion with *Eco*RI and *Xba*I (Promega, USA) while the pPICZαA was double digested with *Xba*I and *Xho*I (Promega, USA). A digestion reaction contained approximately 3 µg of pPICZB or pPICZαA, 0.75 unit of restriction enzymes (Promega, USA), 1X reaction buffer and sterile distilled water to give a total volume of 40 µl. The digested products were purified using QIAGEN gel extraction kit after the digestion had been completed. Then, DNA ligation and transformation were performed as described. The transformants were spreaded on low salt LB agar plate [1% peptone (Bio Basic Inc.), 0.5% yeast extract (Bio Basic Inc.), 0.5% NaCl and 1.5% agar] containing 25 µg/ml of Zeocin^{TM}(Invitrogen).

4.9 Transformation of *P. pastoris*

4.9.1 Preparation of *P. pastoris* competent cells

A single colony of *P. pastoris* KM71 was grown in 50 ml of YEPD at 30°C with shaking until OD_{600} reached 1.3-1.5. Then, the cells were harvested and washed with 10 ml of cold sterile water. After centrifugation at 4,000 x g for 8 min at 4 °C, the cells were resuspended in 250 µl of ice-cold 1 M sorbitol. A volume of 40 µl of the competent cells was used for transformation reaction.

4.9.2 Preparation of linearized recombinant plasmids

The recombinant plasmids were linearized by *Dra*I digestion and then were precipitated with 60 µl of 95% (v/v) ethanol in the presence of 3 µl of 3 M NaOAc and 1.5 µl of yeast tRNA. Then the mixture was incubated at -80 °C for 15 min and centrifuged at 10,000 x g at 4 °C for 10 min. The pellet was washed with 70% ethanol and air dried. The dried pellet was resuspended in 5 µl of sterile distilled water.

4.9.3 Transformation of *P. pastoris* competent cells by electroporation

The mixture containing 40 µl of *P. pastoris* competent cells and 5 µl of linearized recombinant plasmid was transferred to an ice-cold 0.2 cm electrocuvette (Hybaid, USA) and left on ice for 5 min. The cells were pulsed by using Bio-Rad Gene Pulser (Bio-Rad, USA) with following condition: 1.5 kV, 25 µF and 200 Ω (time constant should be 4.5). Then, 1 ml of ice-cold sorbitol was added to the cuvette and the solution was transferred to a steriled microcentrifuge tube. The tube was incubated at 30°C for 1 h without shaking. Next, 1 ml of Yeast Extract Peptone Dextrose Medium (YEPD) was added and the mixture was incubated at 30°C with a constant shaking for 2 h. After centrifugation at 3,000 x g at 4°C for 5 min, the supernatant was discarded and all cells were spreaded on YEPD agar plate containing 100 µg/ml Zeocin™. The cells were grown at 30°C for 2-3 days until colonies formed.

4.9.4 Total DNA isolation from *P. pastoris*

Both the transformant and the wild-type and recombinant *P. pastoris* strains were grown at 30°C until OD_{600} reached 5-6 in YEPD or in containing 100 µg/ml Zeocin™, respectively. Approximately 4 ml of cell culture was centrifuged at 3,000 x g for 5 min to collect cells. The cell pellet was washed once with 1 ml of sterile distilled water, then resuspended in 200 µl of SCED buffer, pH 7.5 (1M sorbitol, 10 mM sodium

citrate, pH 7.5, 10 mM sodium citrate, pH 7.5, 10 mM EDTA, 10 mM DTT). Then, a volume of 5 µl of lyticase (25 U/µl) (Sigma, German) was added and the mixture was incubated at 30°C for 3-4 h. Next, 100 µl of 2 % (w/v) SDS was added, mixed gently and chilled on ice for 5 min. Then, a volume of 150 µl of 5 M potassium acetrate pH 8.9 was added to the solution. After centrifugation at 13,000 x g at 4°C for 5 min, the supernatant was transferred into a new microcentrifuge tube, 1 ml of 95% ethanol was added and the mixture was incubated at 4°C overnight, centrifuged at 10,000 x g at 4°C for 20 min, then the supernatant was removed and the pellet was resuspended gently in 500 µl of TE buffer (10 mM Tris-HCl, pH 7.4 and 1 mM EDTA pH 8.0). The DNA solution was incubated with 10 µl of 5 mg/ml RNaseA at 37°C for 1 h. An equal volume of phenol, pH 8 was added; the solution was mixed gently and then centrifuged at 13,000 x g for 5 min. The nucleic acid in an aqueous phase was transferred into a new microcentrifuge tube, an equal volume of chloroform:isoamyl alcohol (24:1) was added. The mixture was mixed gently and centrifuged at 13,000 x g for 5 min. Next, the nucleic acid in an aqueous phase was transferred into a new microcentrifuge tube**.** Then, a half volume of 7.5 M ammonium acetrate, pH 7.5 and two volumes of 95% ethanol were added. The tube was placed at -80°C for 10 min. After centrifugation at 10,000 x g at 4°C for 20 min, the cell pellet was washed once with 70 % (v/v) ethanol. The DNA was resuspended in 30µl of TE and stored at -20°C until used.

4.9.5 PCR analysis of *P. pastoris* integrants

To screen *P. pastoris* for recombinant plasmid, the PCR reaction was performed; in a total volume of 50 µl mixture containing 1x Promega buffer, 2.5 mM each of dNTPs, 2 mM $MgCl_2$, 10 pmole of each primer (5' *AOX*1 forward primer and 3'*AOX*1 reverse primer), 50-100 ng of genomic DNA and 1 U Taq polymerase (IMBG, Mahidol University). The reaction was operated in an automated thermal cycler GeneAmp PCR system model 2400 (Perkin Elmer Cetus, USA). Amplification cycles were composed of one cycle of 5 min at 94 °C, 30 cycles of 30 s at 94°C, 30 s at 50°C and 1 min at 72°C, then 7 min at 72°C. A 5 µl of aliquot of the PCR products were analyzed on 1% (w/v) agarose gel electrophoresis.

4.10 Expression of recombinant endoglucanase in *P. pastoris*

A single colony of *P. pastoris* recombinants was inoculated into 2 ml of YEPD containing 100 µg/ml Zeocin™ and grown at 30°C with shaking at 250 x g for 48 h. The cell culture was transferred to 5 ml of Buffered Glyceral-complex Medium (BMGY) medium and grown in the same condition as mentioned above until the culture reached an OD_{600} of 5-6. To induce expression of recombinant protein, the cell pellet was harvested by centrifuging at 3,000 x g for 5 min at room temperature. After the supernatant was discarded, the cell pellet was resuspended in Buffered Methanol-complex Medium (BMMY) medium using 1/5 volume of the original culture. The cell suspension was placed in a 20 ml of a glass tube. Absolute methanol was added to a final concentration of 3% very 24 h to maintain induction. The culture was collected at 0, 24, 48, 72, 96, 120, 144 h by centrifuging at the maximal speed in a bench-top microcentrifuge for 2-3 min at room temperature. The supernatant containing secreted protein was transferred to a fresh microcentrifuge tube. The secreted protein were analyzed by using SDS-PAGE and also tested for the endoglucanase activity.

4.11 Characterization of BCC18080 endoglucanase

4.11.1 Protein electrophoresis

4.11.1.1 Sample preparation

Before protein samples were loaded into SDS-polyacrylamide gel, then samples were mixed with 4X protein sample buffer [60 mM Tris-HCL (pH 7.5), 2% (w/v) SDS, 10% glycerol, 0.025% (w/v) Bromophenol blue, and 100 mM DTT], then boiled for 10 min, immediately kept on ice and centrifuged at 13,000 x g for 5 min.

4.11.1.2 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Stacking and separating gel for SDS-PAGE were set up as described in Table 2. The protein were electrophorised in protein running buffer [25 mM Tris-HCL, pH 8.2, 192 mM glycine and 0.1% (w/v) SDS] using a constant voltage of 100 volts until the dye front reached the bottom of the gel. Then, the gels were stained in staining buffer [45% (v/v) methanol, 10% (v/v) acetic acid and 0.1% (w/v) coomassie blue R-250] for 1 h. The excess stain was washed overnight with mild shaking in destaining buffer [50% (v/v) methanol and 10% (v/v) acetic acid]. The gel was dried

by placing between two wetted cellophane papers, which were clamped together. All air-bubbles, which can lead to gel cracking, were removed. The gel was dried at room temperature.

4.11.2 Glycosylation analysis

4.11.2.1 Glycoprotein sugar analysis by GelCode® glycoprotein staining kit

Glycoprotein sugar moieties in polyacrylamide gel were detected using GelCode® glycoprotein staining kit (Pierce Biotechnology, USA). After electrophoresis, the gel was fixed by immersing in 100 ml of 50% (v/v) methanol for 30 min. Then the gel was washed twice with 100 ml of 3% (v/v) acetic acid for 10 min. The gel was transfer to 25 ml of Oxidizing Solution for 25 min. Any glycols present in glycoprotein are oxidized to aldehydes. Next, the gel was immersed in 25 ml of GelCode® Glycoprotein Staining Reagent for 15 min and 25 ml of Reducing Solution for 5 min. The gel was washed extensively with 3% (v/v) acetic acid and then once with ultrapure water. Glycoproteins will appear as magenta bands with light pink background.

4.11.2.2 N-linked glycosylation analysis by PNGase F Digestion kit

In order to determine if the recombinant endoglucanase contain N-linked glycosylation, it was digested with N-Glycosidase F (PNGase F) that hydrolyzes nearly all types of N-glycan from glycoprotein. The digestion process was performed according to manufacturer's instruction. Specifically, the protein was denatured by boiling for 10 min in 1X Glycoprotein denaturing buffer [0.5% (w/v) SDS and 1% β-mecaptoethanol]. Then, the denatured protein samples were incubated in 1X G7 buffer [50 mM sodium phosphate, pH 7.5], 1% Nonidet P-40 and PNGase F (New England BioLabs) at 37°C at various time to determine the optimal period. Finally, the digested proteins were analyzed in 12% SDS-PAGE.

4.11.2.3 Concentration of the secreted protein by using ultrafiltration (100)

The centriprep® centrifugal filter devices (Amicon® Bioseparation) were used following manufacturer's instruction to concentrate the secreted protein in the culture supernatant. The device acts by forcing the liquid in a protein solution through

a membrane which retains the protein of interest. The protein solution was added to the sample container upto the fill line on the side of the container and the filtrated collector was inserted into a sample container. The centriprep was centrifuged at 13,000 x g for 20 min and this step was repeated until the expected concentration was obtained. The filtrate was discarded. The concentrated protein was transferred to a 1.5 ml eppendorf tube using a pipette.

4.11.3 Protein purification by Gel Filtration Chromatography

After the supernatant was concentrated, it was applied onto a superdex^{TM} 75 10/300 GL column (Amersham Bioscience, USA) equilibrated with 50 mM phosphate buffer pH 7. The enzyme was eluted with 50 mM phosphate buffer pH 7. The active fraction was determined using AZCL-HE cellulose. The purity of recombinant protein was analysed by SDS-PAGE combined with Coomassie blue staining.

4.11.4 Protein concentration determination

Protein concentrations were determined based on the Bradford method [113] using Bio-Rad protein assay kit. Protein standard (BSA) was diluted in distilled water to concentration of 2.0, 4.0, 6.0, 8.0 and 10.0 µg/ml. A volume of 160 µl of the protein standard and 40 µl of Bio-Rad dye reagent were mixed and placed into a microtitre plate. The mixture using distilled water instead of the protein was used as a blank. The reactions were incubated at room temperature for 5 min. The absorbance at 595 nm was measured from protein standard or samples using microplate spectrophotometer Spectra Max190 (Molecular Devices, USA). The concentration of protein samples were calculated by referring to the concentration of protein standard.

4.11.5 Assay for enzyme activity using DNS method

The carboxymethylcellulase (CMCase) activity was assayed by measuring the reducing sugars liberated, as D-glucose equivalents, using the dinitrosalicylic acid assay (DNS method) [114]. One unit (U) of specific activity was defined as the amount of enzyme releasing 1 µmole of reducing sugar per min.

The effect of temperature was examined under standard conditions (50 mM sodium phosphate pH 5.8) by varying the temperature (30-80°C). In order to determine the optimal pH, the CMCase activity was measured under several different

pH using 50mM sodium citrate (pH 3 to 5), sodium acetate (pH 4 to 6), sodium phosphate (pH 5 to 8) and Tris-HCL (pH 8-10).

The hydrolysis activity using CMC (Fluka), Avicel (Merck), Xylan(Sigma) and Cellobiose (Fluka) as substrates was assayed using reaction mixtures containing approximately 0.5 µg of protein and 10 mg of substrate in 1 ml of 50 mM sodium acetate (pH 6). The reaction was incubated for 10 min at 70°

In order to determine the remaining enzyme activity at 70 °C, the purified enzyme was incubated at 70°C for 2 h and 4 h in 50 mM acetrate buffer (pH 6), then the activity was determine using DNS method.

The hydrolysis kinetic of CMC by endoglucanase at 70 °C in acetate buffer pH 6 was also determined. Substrate concentrations from 2.5-20 mg ml^{-1} were used. Activity rates were measured at eight different substrate concentrations in triplicate. The amount of reducing sugars was determined using DNS method. The apparent kinetic constants K_m, V_{max} and K_{cat} were calculated using the equation for enzyme kinetics.

4.11.6 N-terminal peptide sequencing

4.11.7.1 Sample preparation for sequencing

The protein samples were separated in 12% SDS-PAGE as described in 4.10.1. Then, thioglycolic acid [0.015% (v/v)] was added in the upper chamber to migrate ahead of the protein and scavenge free radicals. After electrophoresis was completed, The gel was soaked in CAPS buffer [10 mM CAPS (pH 11), 10% methanol] for 5 min to remove electrophoresis buffer and salt detergent.

4.11.7.2 Blotting protein sample to PVDF membrane

In order to transfer samples from acrylamide gel onto membrane, the Mini Trans-Blot® electrophoresis transfer cell (Bio-Rad) was used. Polyvinylidene difluoride (PVDF) membrane (Bio-Rad) was pre-wet in methanol and equilibrated in CAPS buffer, the transblotting sandwich was formed and electroblotted was performed at constant voltage of 50 volts, at 4°C for 100 min.

After blotting, the membrane was removed and rinsed with large amount of HPLC water. Then the membrane was stained with 0.1% (w/v) Coomassie Blue R-250, 45% (v/v) methanol and 10% (v/v) acetic acid for 30 min and excess dye was

removed by washing in 50% (v/v) methanol until the background became clear. The membrane was rinsed with HPLC water before air-drying.

The protein bands on PVDF membrane was excised and analyzed by Biomolecular Research Facility, University of Newcastle, Australia.

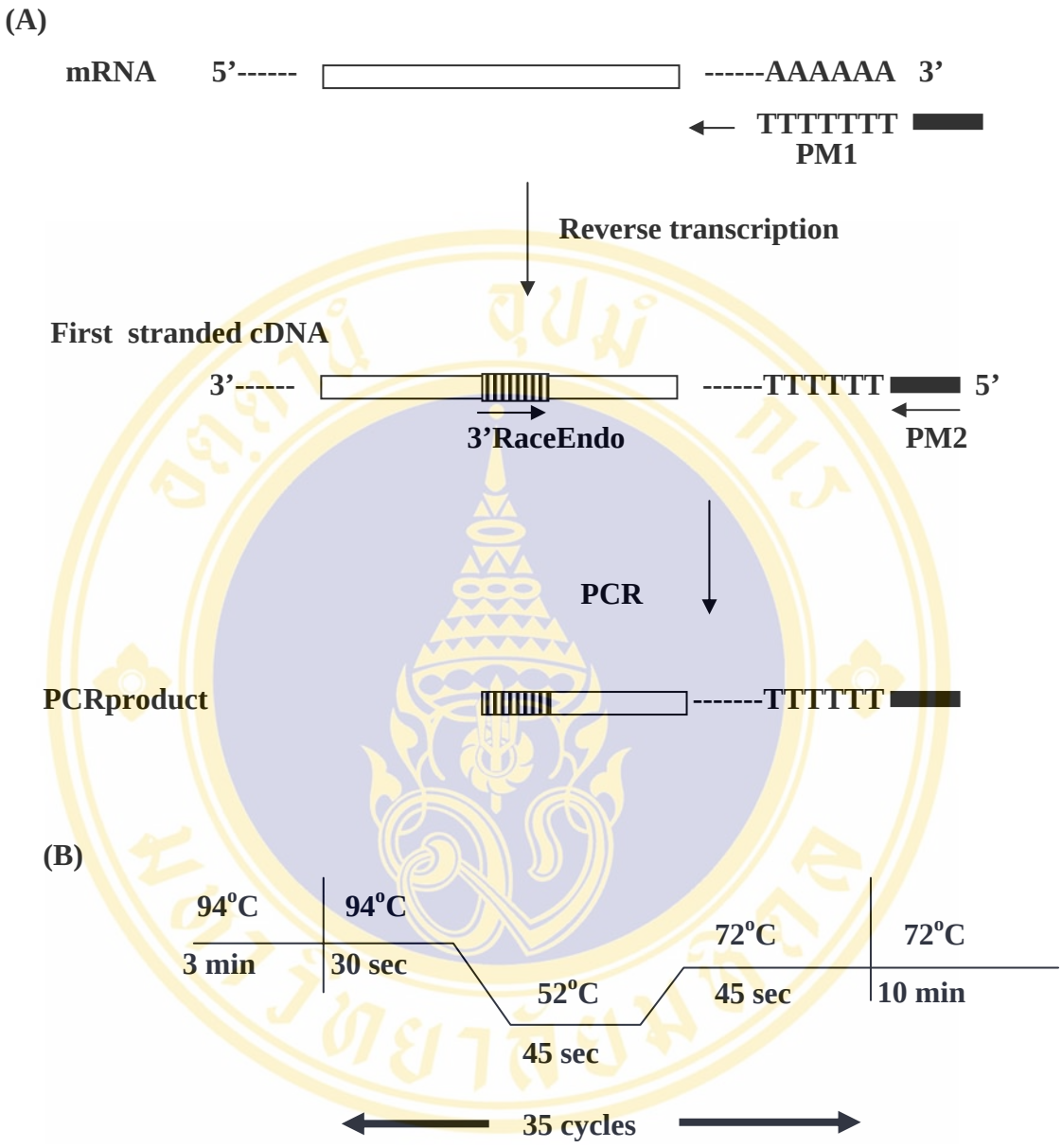


Figure 9. A schematic diagram and PCR profile of 3' end cDNA amplification (3'RACE)

(A) A schematic diagram of the 3' end cDNA amplification by 3'RACE method. The black boxes indicate synthesized nucleotides in PM2 and shade black boxes indicate nucleotide position used in designed 3'RaceEndo primer. Dotted lines represented the untranslated regions (UTR). Arrows showed position and direction of primers. **(B)** PCR profile of 3' end cDNA amplification of BCC18080 endoglucanase.

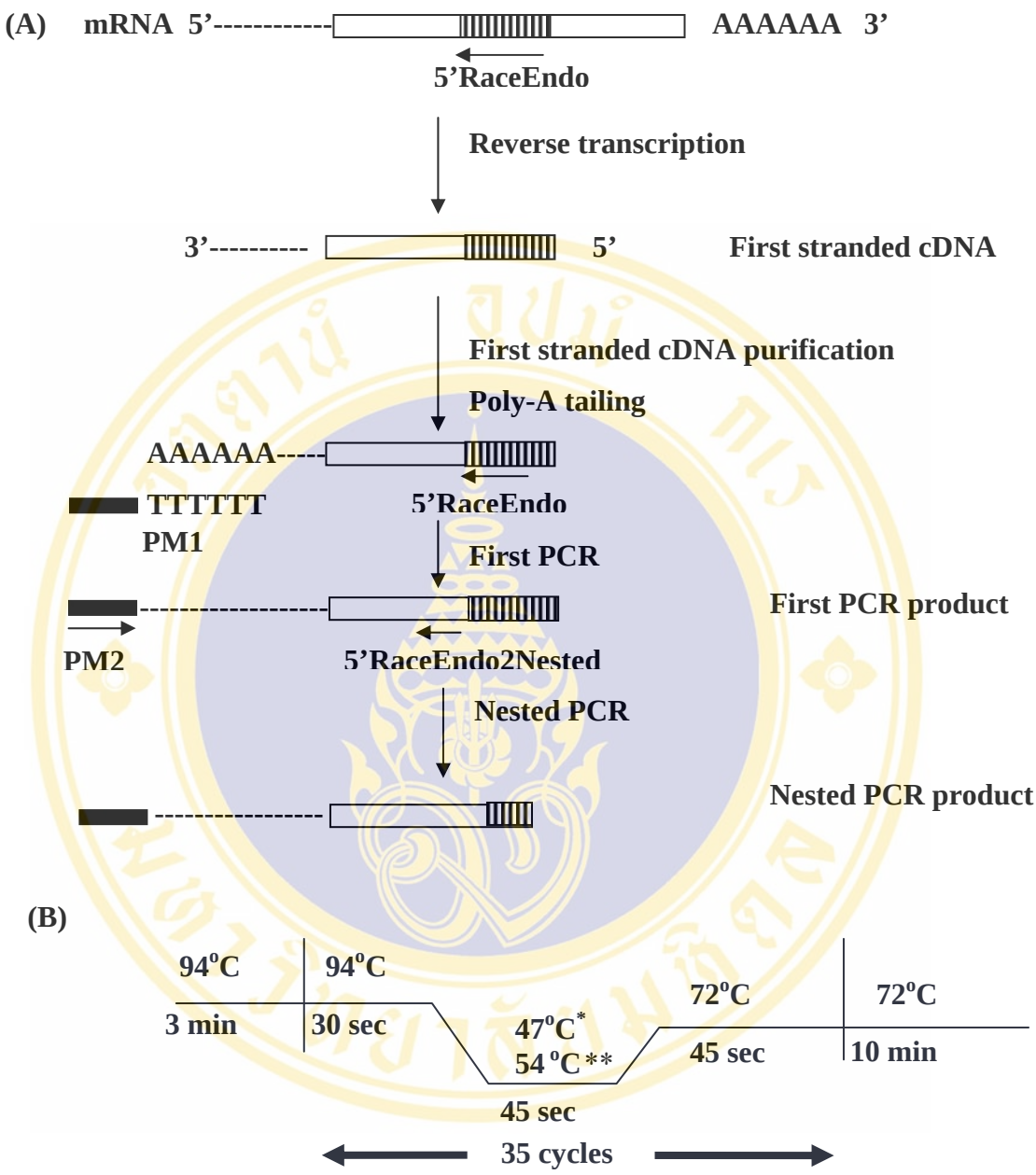


Figure 10. A schematic diagram and PCR profile of 5' end cDNA amplification (5' RACE)

(A) A schematic diagram of the 5' end cDNA amplification by 5'RACE method. A black block showed synthesized nucleotides in PM1 andPM2. Shaded black boxes indicated nucleotide position used in designed 5'RaceEndo primer and 5'RaceEndo2nested. Dotted lines represented the untranslated regions (UTR). Arrows showed position and direction of primers. **(B)** PCR profile of 5' end cDNA amplification of BCC18080 endoglucanase. Asterisks indicated the different condition of the first (*) and nested (**) PCR reactions.

Table 2. Gel preparation for SDS-PAGE for protein determination

Solution	5% of stacking SDS-PAGE Volume (ml)	12% of separating SDS-PAGE Volume (ml)
30% Acrylamide[a]	0.83	4
H_2O	3.4	3.3
1 M Tris-HCl, pH6.8	0.63	-
1.5 M Tris-HCl, pH8.8	-	2.5
10% SDS[b]	0.05	0.1
10% APS[c]	0.05	0.1
TEMED[d]	0.01	0.07
Total volume	4.97	10.07

[a]acrylamide: N,N'-methylene-bis-acrylamide 29:1

[b]sodium dodecyl sulfate

[c]ammonium persulfate

[d]N,N,N',N'-tetrametyl-ethylenediamine

CHAPTER V
RESULTS

5.1 Optimal condition for endoglucanase induction in BCC18080

In order to determine the optimal culturing condition for production of endoglucanase in BCC18080 to ensure the presence of RNA encoding the corresponding gene, the stock culture of BCC18080 was directly inoculated into the supernatant of 5% wheat bran. The cultures were grown at 250 rpm at 30°C. The supernatant was collected and tested for endoglucanase activity at various days for 8 days. The result showed that the maximal endoglucanase activity was at the fifth day with the maximal radius of the clear zone of 6 mm (Figure 11). Therefore, the optimal condition for culturing the fungus for total RNA isolation was to directly grow BCC18080 in the supernatant of 5% wheat bran with continuously shaking at 30°C, 250 rpm for 5 days.

5.2 Total RNA isolation from BCC18080 by using TRI Reagent

Total RNA was isolated from the BCC18080 using TRI REAGENT at time of maximal endoglucanase activity. RNA concentration and purity was determined by spectrophotometry at 260 and 280 nm and the quality of RNA was investigated by formaldehyde agarose gel electrophoresis. The result suggested that RNA sample showed an absorbance ratio of A_{260}/A_{280} in the range 1.8-2.0, indicating that RNA sample was pure. The result presented a predominant RNA band in size approximately 3.6 and 1.9 kb corresponding to 28S and 18S rRNA, respectively (Figure 12).

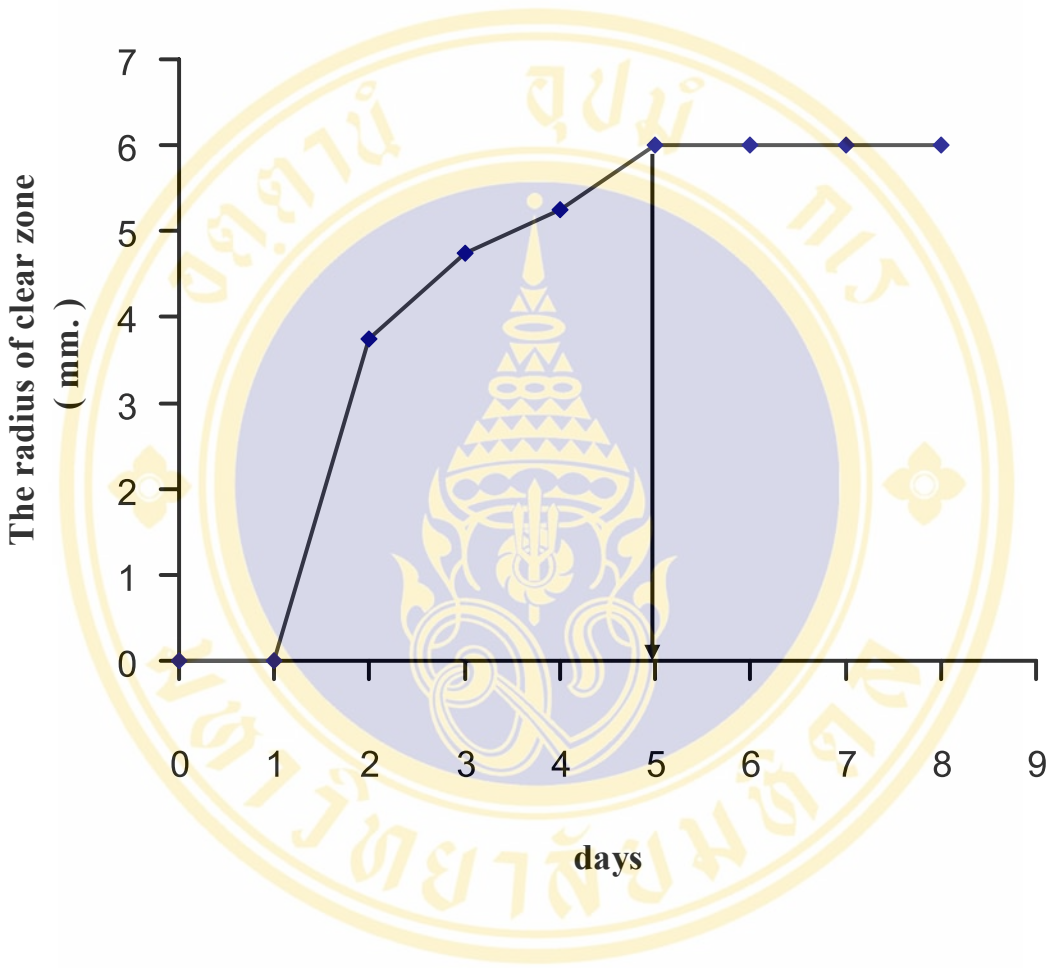


Figure 11. Optimal condition for endoglucanase induction in BCC18080

BCC18080 was directly inoculated into 5% wheat bran. The supernatant was collected at various days for 8 days. To test for the endoglucanase activity, 20 µl of supernatant was added onto 0.1% AZCL-HE cellulose plate. The radius of the clear zone indicated the endoglucanase activity. The maximal endoglucanase activity was at fifth day with the maximal radius of the clear zone at 6 mm as shown by the arrow.

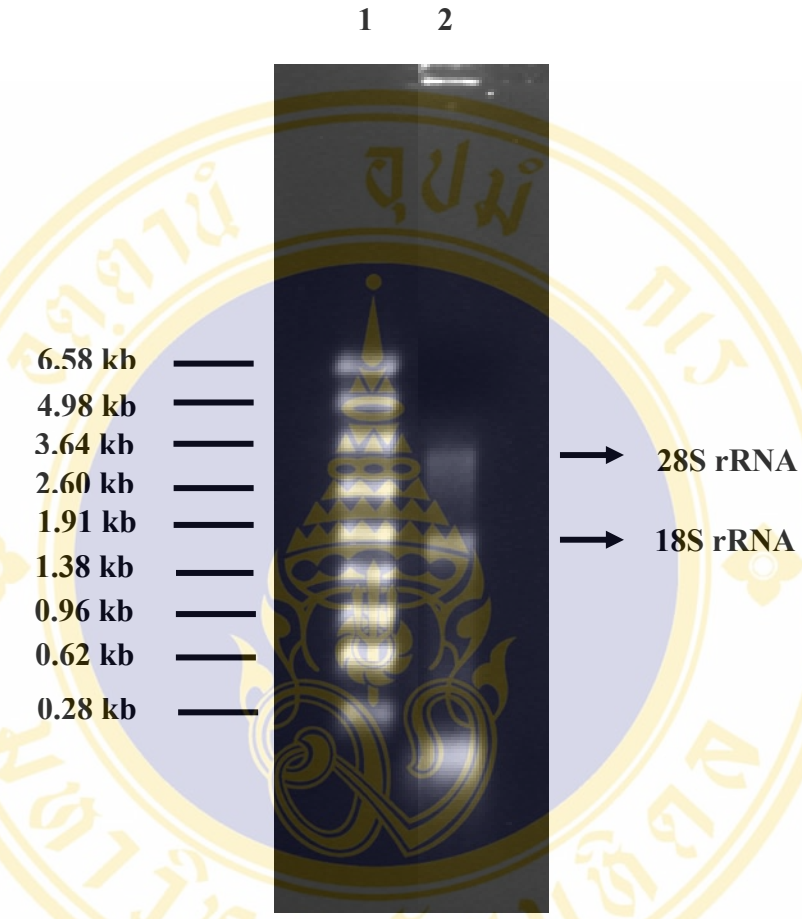


Figure 12. Total RNA isolated from BCC18080 mycelium

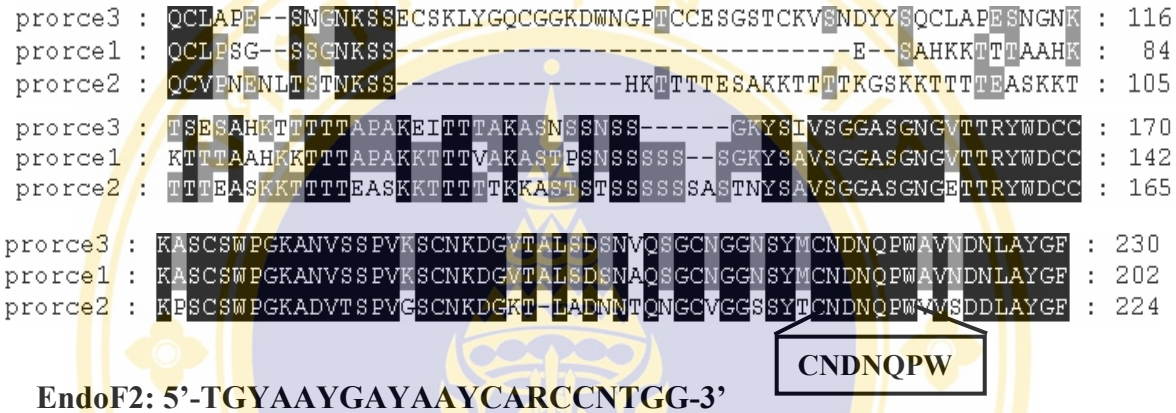
RNA was size-fractionated on 1% formaldehyde agarose gel electrophoresis. Lane 1 is 2 µg RNA molecular weight marker. Lane 2 is RNA from BCC18080. The size of major band is approximately 1.9 kb corresponding to 18S rRNA while a band of approximately 3.6 kb is corresponding to 28S rRNA.

5.3 Cloning of partial sequence of BCC18080 endoglucanase gene by RT-PCR

The degenerate primers (EndoF1, EndoF2, EndoR1 and EndoR2) were designed based on the conserved amino acid sequence of the endoglucanase from the *Rhizopus oryzae*; *RCE1, 2* and *3* (accession number Q8J1L2, Q8J1L10, Q 8J1L1, respectively) (Figure 13). The result from RT-PCR using the degenerate primers, EndoF2 and EndoR1, showed approximately a 390 bp product (Figure 14). It was purified and cloned into pGEM-T Easy vector. Next, DNA sequence analysis using NCBI-BLAST revealed that the 390 bp PCR product showed approximately 70% sequence identity to the endo-β-D-glucanase gene *RCE*1, *RCE*2, and *RCE*3 of *Rhizopus oryzae*. The partial endoglucanase contained 339 nucleotides and encoded 113 amino acids.

The alignment of the deduce amino acid sequence of BCC18080 endoglucanase with other fungal endo-β-D-glucanase genes using ClustalX program revealed a 72% amino acid identity to *Rhizopus oryzae* RCE 1, a 71% amino acid identity to *Mucor circinelloides*, a 71% identity to *Rhizopus oryzae* RCE 2, a 71% identity to *Rhizopus oryzae* RCE 3 and a 69% amino acid identity to *Phycomyces nitens* (Figure 15).

EndoF1: 5’-CARTGYGGNGGNAARAATTGG- 3’



EndoF2: 5’-TGYAAYGAYAAYCARCCNTGG-3’

EndoR1: 5’-RTTRTCNGCRTTYTTRAACCA-3’

EndoR2: 5’-YTTNCGNGARCANCCNGTYTT-3’

Figure 13. Degenerate primer design

Four degenerate primers were designed based on the conserved amino acid sequence region of *Rhizopus oryzae RCE 1, 2, 3* (accession number Q8J1L2, Q8J1L10 and Q8J161, respectively).

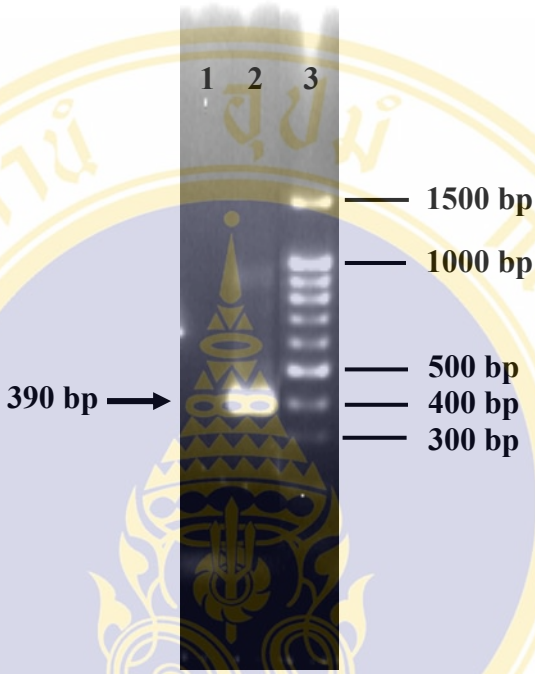


Figure 14. The RT-PCR product of BCC18080 endoglucanase

RT-PCR product was size fractionated on 1% agarose gel electrophoresis. Lane 1 is a negative control performed with distilled water. Lane 2 is the nested PCR product using EndoF2 and EndoR1 primers. The size of the nested-PCR product is approximately 390 bp. Lane 3 is a 100 bp ladder.

Figure 15. An alignment of the deduced amino acid sequence of the partial endoglucanase cDNA from BCC18080

The deduced amino acid sequence of partial endoglucanase cDNA was aligned using ClustalX with other fungal endoglucanases: *Rhizopus oryzae* endo-β-D-1, 4-glucanase genes, *RCE1* (Q8J1L2), *RCE 2* (Q8J1L10), *RCE 3* (Q 8J1L1), *Mucor circinelloides* endo-β-D-1, 4-glucanase gene (BAD95809), *Phycomyces niten* endo-β-D-1, 4-glucanase gene (BAD77808). The darkest shading indicates the most conserved region among all sequences compared.

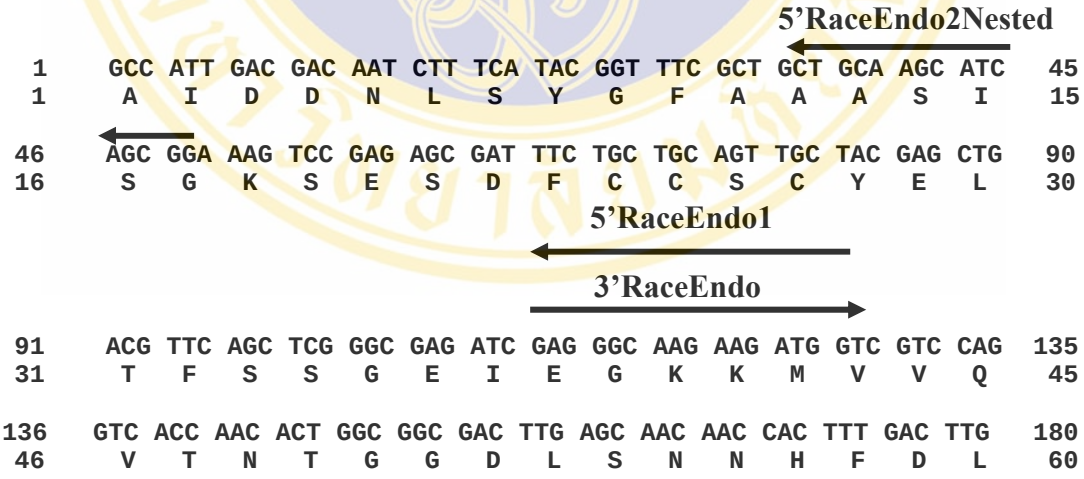


Figure 16. Nucleotide and deduced amino acid sequence of partial endoglucanase cDNA from BCC18080

The partial endoglucanase sequence from BCC18080 is 339 nucleotides encoding 113 amino acids. Data shows only positions used in specific primer design. The specific primers for the 5' RACE (5'RaceEndo1 and 5'RaceEndo2Nested) and 3' RACE (3'RaceEndo) were designed and indicated by the arrows.

5.4 Amplification of the 3'end of BCC18080 endoglucanase gene by 3'RACE method

In order to obtain the 3'end cDNA of BCC18080 endoglucanase gene, 3' RACE was performed using a gene specific primer as indicated in Figure 16. The result showed a PCR product of approximately 550 bp PCR product (Figure 17) which was purified and cloned into pGEM-T easy vector. After sequencing, the result revealed that the length of 3'end is 522 nucleotides. Translation of amino acid sequence showed 174 amino acids. A stop codon was found at the amino acid position 108th. Therefore, 3'end cDNA is only 324 nucleotides long encoding 108 amino acids (Figure 18). A polyadenylation site, AAAAAA, at 49 bp upstream of the major poly (A) sites was also found. [Although the sequence of AATAAA, a polyadenylation site, is required for mRNA 3'end formation in ~ 50% of the yeast gene, some yeast genes such as *ADH2* do not require this sequence. The mutation in this sequence do not affect in polyadenylation process (115, 116)]. The 3'end sequence analysis using NCBI-Blast revealed 74% amino acid identity to *Mucor circinelloides*, a 76% amino acid identity to *Phycomyces nitens*, a 72% amino acid identity to *Rhizopus oryzae RCE* 1, a 71% identity to *Rhizopus oryzae RCE* 2 and a 71% identity to *Rhizopus oryzae RCE* 3.

5.5 Amplification of the 5'end of BCC18080 endoglucanase gene by 5'RACE method with partial heat denaturation

Since RNA secondary structures such as hairpins or stem-loop structure may inhibit the synthesis of full-length 5'-cDNA, partial heat denaturation step during reverse transcription (112) was performed at 85°C to remove these RNA secondary structures before 5' RACE method. The result revealed a PCR product of approximately 750 bp (Figure 19) which then was purified and cloned into pGEM-T easy vector. After sequencing, the result showed the sequence of 642 nucleotides that encode 214 amino acids. Further sequence analysis showed that the first methionine residue was right next to the PM2 primer, this indicated that this methionine showed not be the starting methionine residue since an eukaryotic gene should contain 5' untranslated region (Figure 20). The second methionine, at position 13, may be the staring methionine instead. Nonetheless, the specific primer for full-length gene amplification was designed from the first methionine found in the gene.

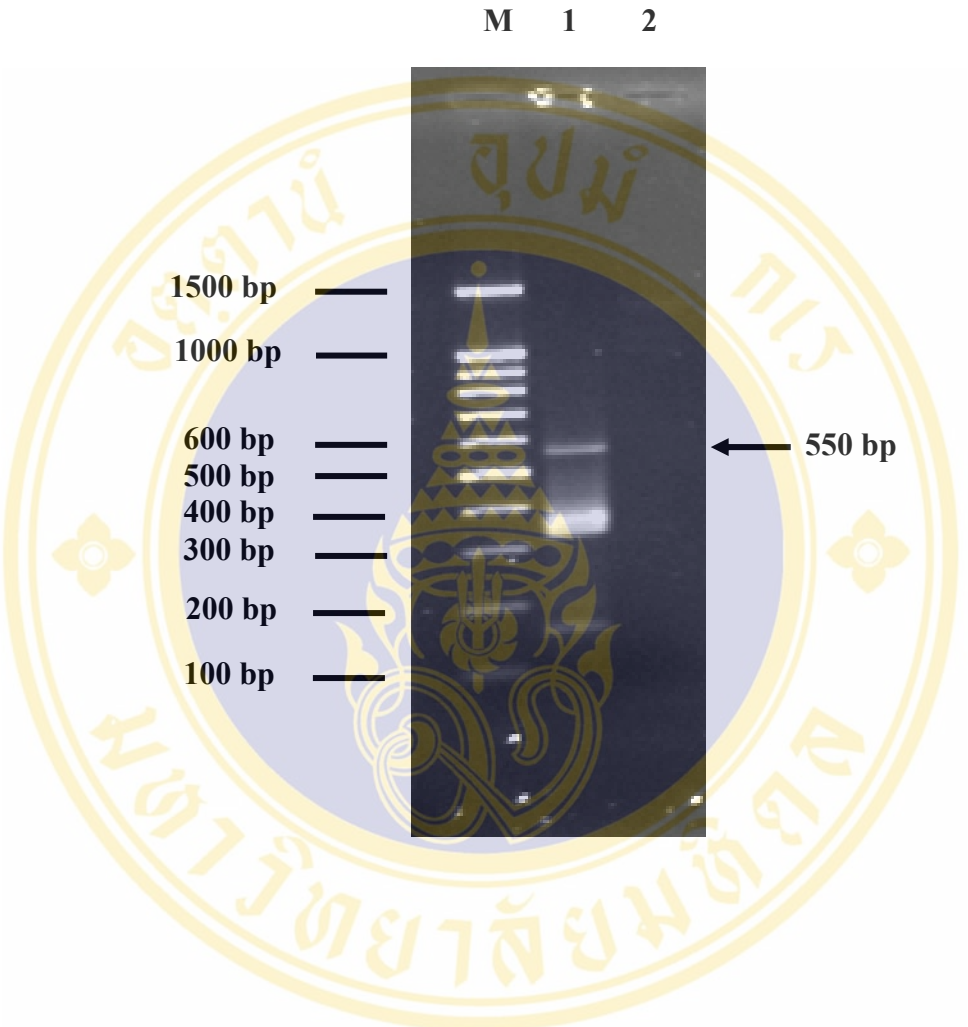


Figure 17. The 3'RACE product of BCC18080 endoglucanase gene

The 3' RACE was performed with 3'RaceEndo, a gene specific primer, and PM2, an adapter primer. Lane M is a 100 bp ladder DNA marker. Lane 1 is a 3' RACE product. The size of PCR product is approximately 550 bp. Lane 2 is a negative control.

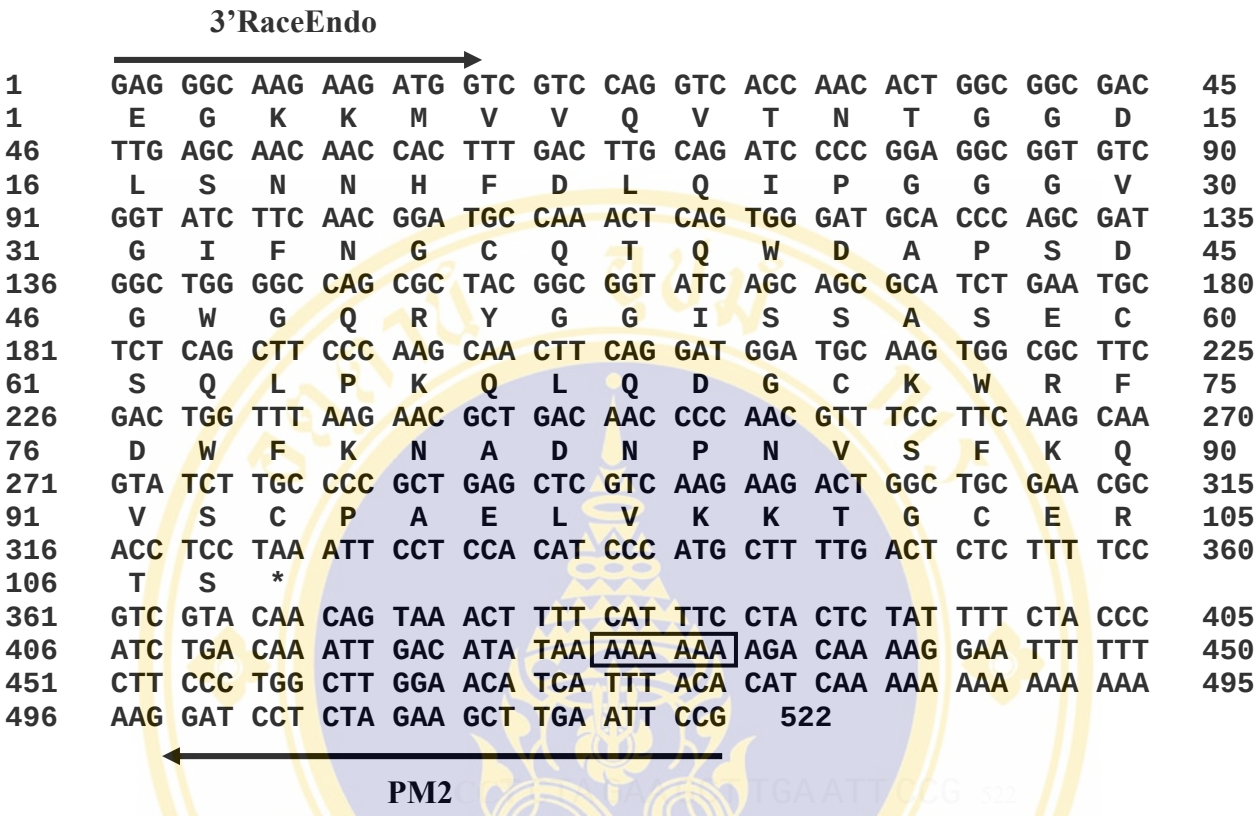


Figure18. Nucleotide and amino acid sequences of 3' end cDNA of endoglucanase from BCC18080.

The 3'end cDNA was obtained using 3'RaceEndo and PM2 primers. The 3' cDNA shows 522 nucleotides that encoded 174 amino acids. The stop codon (*) is indicated at 108th position so the 3'cDNA reveals only 324 nucleotides. The black box shows polyadenylation signal 49 bp upstream of the poly (A) sites.

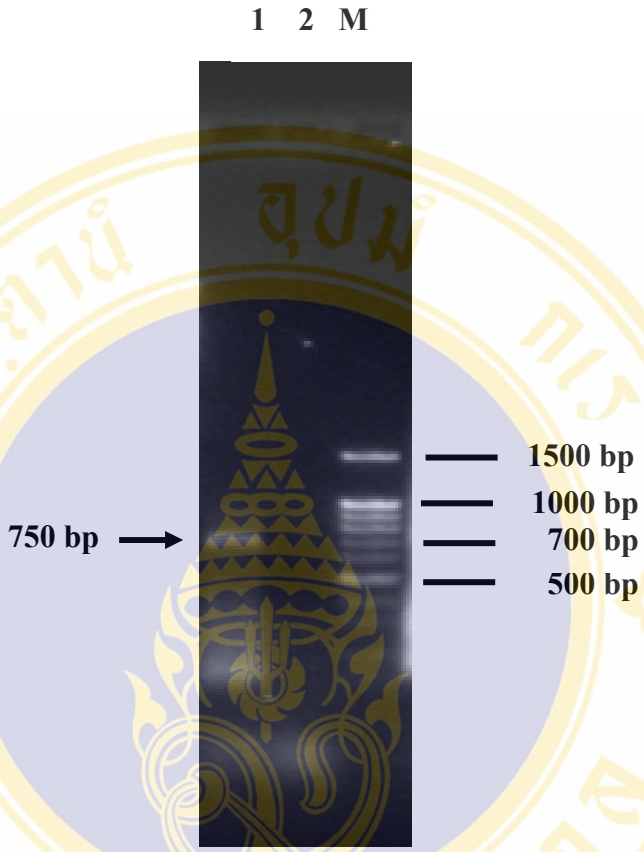


Figure 19. The 5'RACE product of BCC18080 endoglucanase gene

The 5' RACE was performed using 5'RaceEndo1, 5'RaceEndonested, gene specific primers, and PM2, an adapter primer. Lane M is a 100 bp ladder DNA marker. Lane 1 is a 5' RACE product with the size of 750 bp. Lane 2 is a negative control.

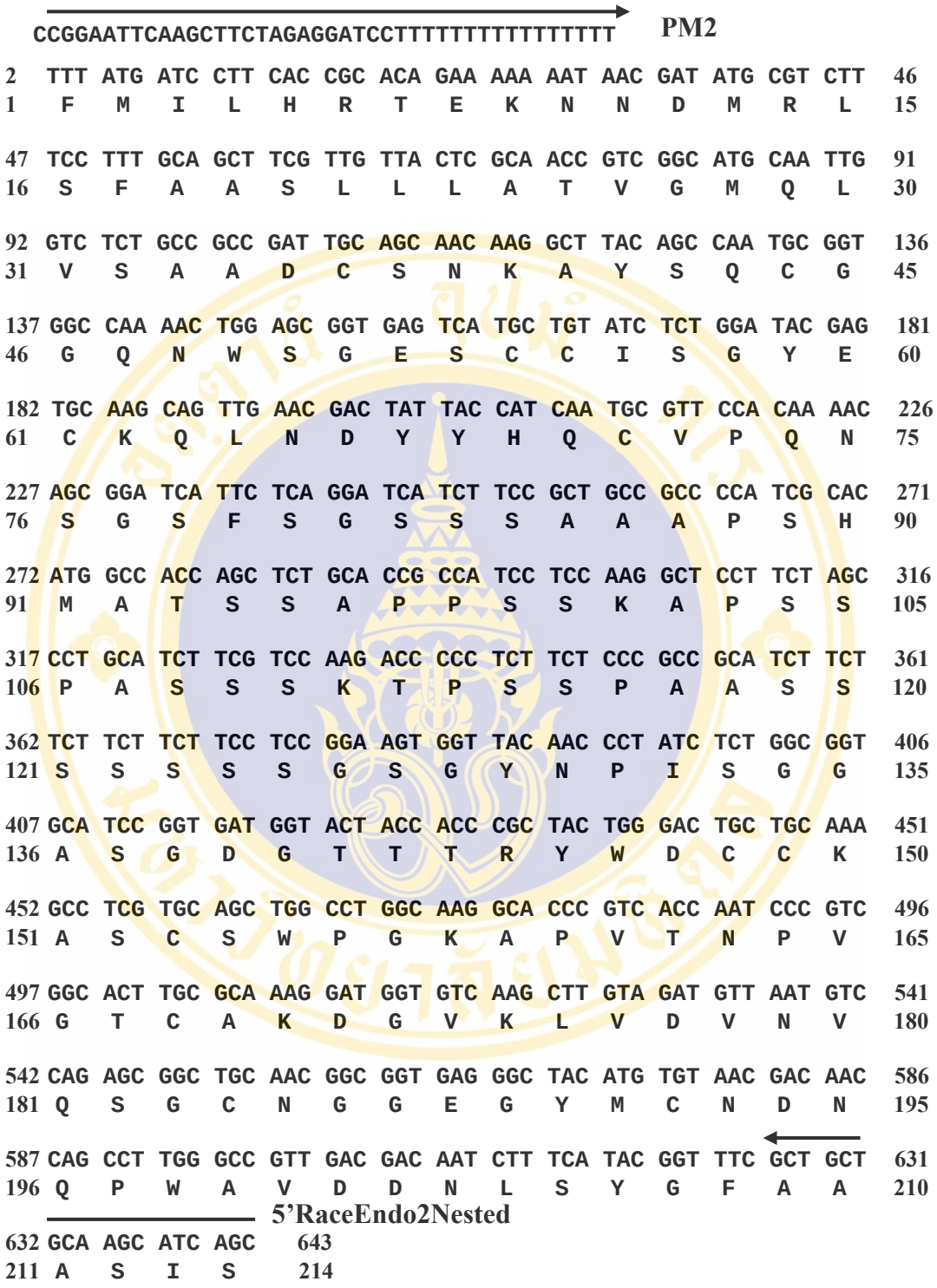


Figure 20. Nucleotide and amino acid sequences of 5' end cDNA of endoglucanase from BCC18080.

The 5'end cDNA was obtained using PM2 and 5'RaceEndo2Nested primer. The 5' cDNA reveals 642 nucleotides that encoded 214 amino acids.

5.6 Amplification of the full-length endoglucanase gene with and without its own signal sequence and construction of recombinant endoglucanase in *P. pastoris* expression vectors

The SignalP 3.0 prediction program was used to discriminate between signal peptides and non-signal peptides regions. From this prediction, the result suggested that residues 1 to 32 are signal peptides (Figure 21). In order to investigate if the fungal signal sequence functions in *P. pastoris*, the full-length gene with its own signal sequence together with the full-length gene without its own signal sequence but with the α-factor signal sequence, so called the mature gene, were amplified. ExpressEndoF1and expressEndoR specific primers for amplification of full-length gene with signal sequence and expressEndoF and expressEndoR specific primers for amplification of the mature gene were designed based on the sequence of full-length gene. The result revealed a product of 1 kb obtained from the PCR using expressEndoF1 and expressEndoR while a band of 950 bp obtained by using expressEndoF2 and expressEndoR (Figure 22). PCR fragments were purified and cloned into pGEM-T Easy vector. DNA sequencing of the inserts were performed, and the results showed that three out of five clones showed the same sequence (other two clones are different in some amino acids). One of these clones was randomly selected for further analysis. The obtained full-length gene contained 1,020 nucleotides that encoded 340 amino acids while the mature gene contained 921 nucleotides that encoded 307 amino acids (Figure 23). Sequence analysis using NCBI-BLAST showed that obtained BCC18080 endoglucanase has high sequence similarity with other fungal endoglucanases and belonged to glycosyl hydrolase family 45 (Figure 24). The alignment of the deduced amino acids of BCC18080 endoglucanase with other fungal endoglucanases (*Rhizopus oryzae*, *Mucor circinnelloides*, and *Phycomyces nitens*) showed 50-59% identities (Figure 25).

The 3D structure of BCC18080 endoglucanase was predicted from the deduced amino acid sequence using the SWISS-MODEL program. The result showed that the 3D structure of BCC18080 endoglucanase (Figure 26B) depicted six-stranded β barrels (117) which is similar to that of other endoglucanase in the glycosyl hydrolase family 45 such as endoglucanase V from *Humicola insolens* (Figure 26A). This result

confirmed that BCC18080 endoglucanase gene belonged to the glycosyl hydrolase family 45.

In order to express the BCC18080 endoglucanase gene in *P. pastoris* system, both the full-length and the mature genes were constructed in *P. pastoris* expression vector, p*PICZ*B and p*PICZ*αA, respectively (Figure 27). Since both the full-length and the mature genes obtained from PCR contained restriction sites (*Eco*RI, *Xho*I and *Xba*I), they were purified and double digested with *Eco*RI, *Xho*I or *Xba*I to facilitate in subcloning. The full-length fragment was double digested with *Eco*RI and *Xba*I before it was cloned into p*PICZ*B while the mature fragment was purified and double digested with *Xba*I and *Xho*I before it was cloned in-frame with *S.cerevisiae* α-factor secretion signal of p*PICZ*αA.

5.7 Determination of endoglucanase cDNA integration into *P. pastoris* genome

Both recombinant plasmids, a full-length endoglucanase gene subcloned into pPICZB and a mature endoglucanase gene subcloned into pPICZαA, were linerized with *Dra*I which was in 5'*AOX1* promoter. The linearized recombinant plasmids were transformed into *P. pastoris* host strain KM71 using electroporation. *P. pastoris* transformants were screened on YEPD plate containing 100 μg/ml Zeocin™. Genomic DNA of Zeocin™-resistant transformants was extracted (Method 4.9.4) to determine the intergration in *P. pastoris* genome. PCR amplification using 5'AOX and 3'AOX primers was also performed to determine the presence of the plasmid integration. The PCR product showed a band of approximately 1,400 bp as expected for the PCR fragment of the mature endoglucanase with α-factor signal sequence and a band of 1,300 bp as expected for the full-length endoglucanase with leader sequence. Whereas PCR product of 600 bp and 300 bp was obtained from *P. pastoris* transformant containing only pPICZαA and *P. pastoris* transformant containing only pPICZB, respectively (Figure 28). These results suggested that the linearized recombinant plasmids were integrated into the *P. pastoris* genome at the 5'AOX promoter.

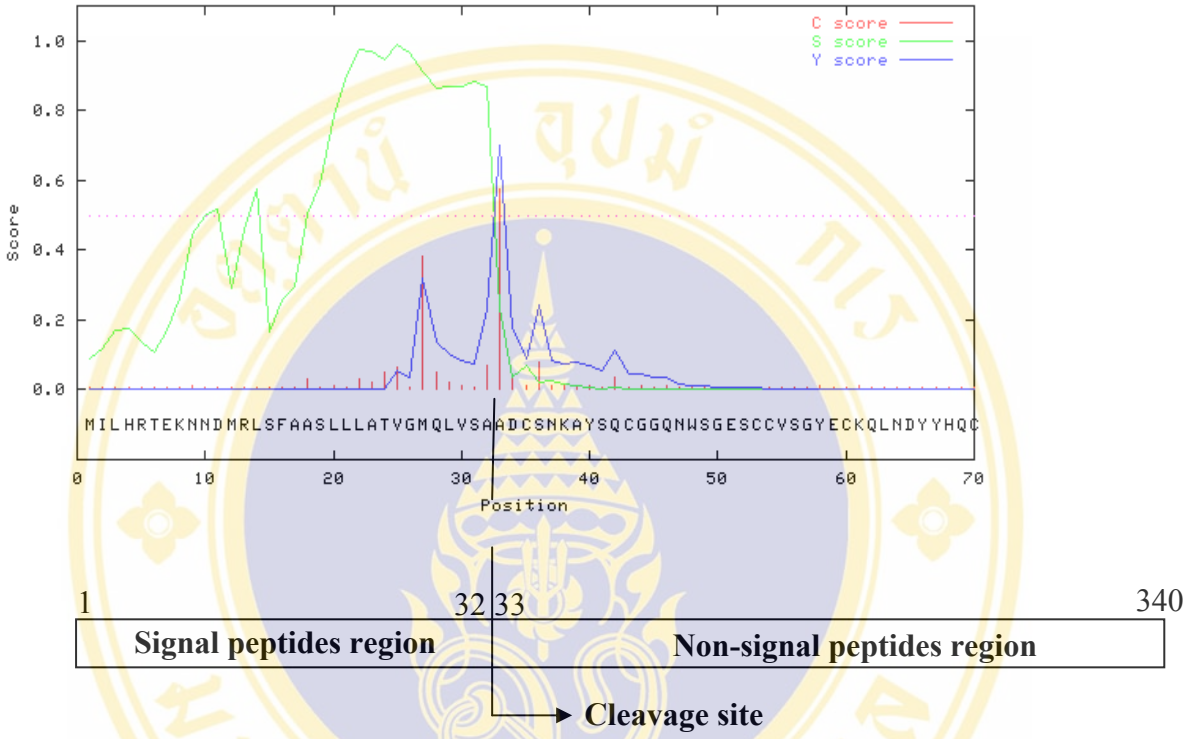


Figure 21. Position of BCC18080 endoglucanase signal peptide region and cleavage site as predicted by SignalP program.

SignalP, a signal peptide sequence prediction program, suggested that residues 1 to 32 might be the signal sequence of BCC18080 endoglucanase gene. The position of the cleavage site is between residues 32 and 33 as indicated by arrow.

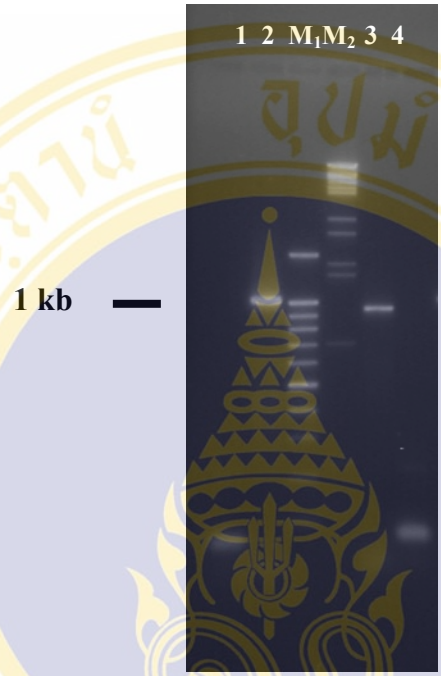
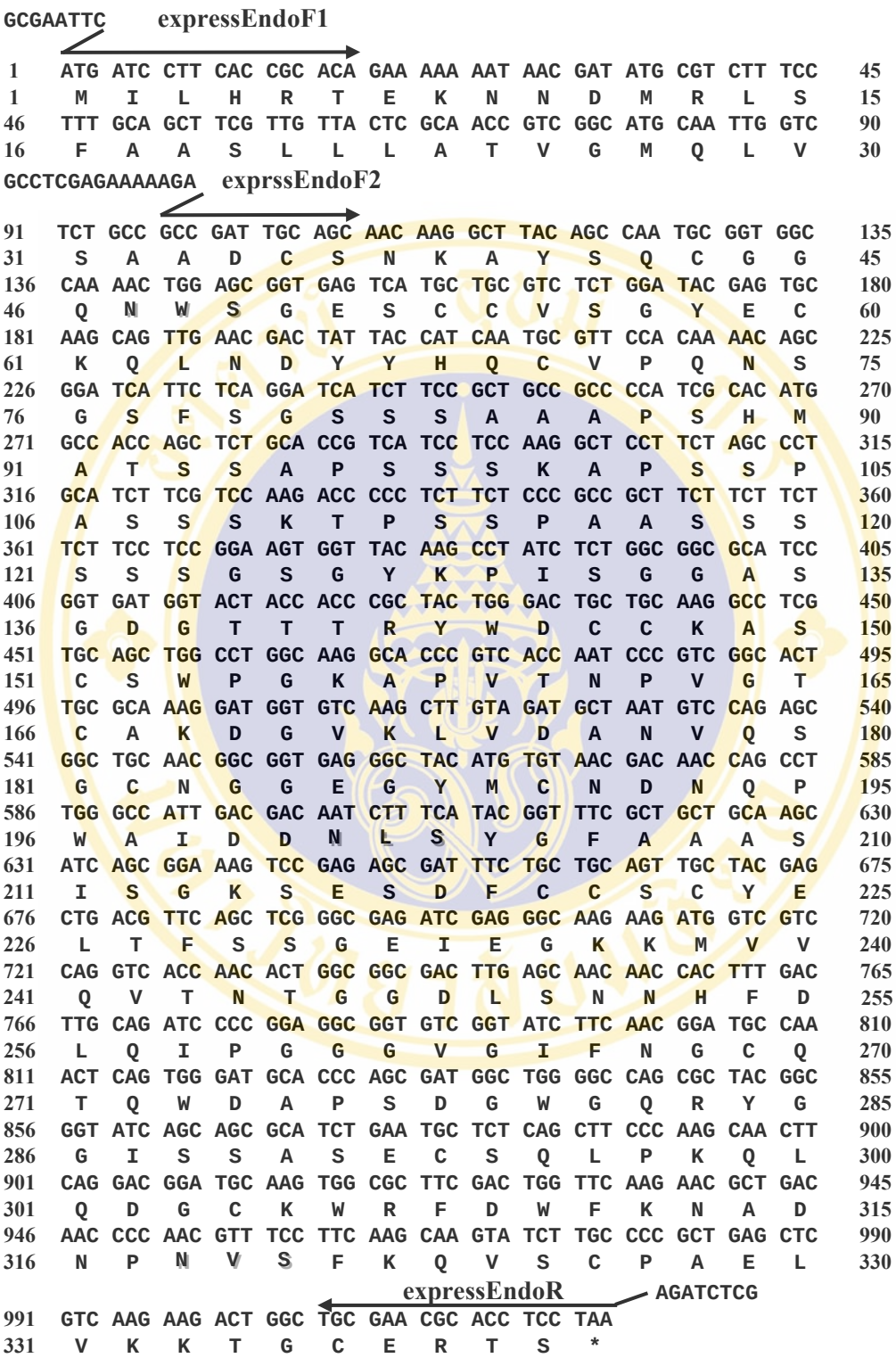


Figure 22. An agarose gel showed the PCR product of BCC18080 full-length and mature genes

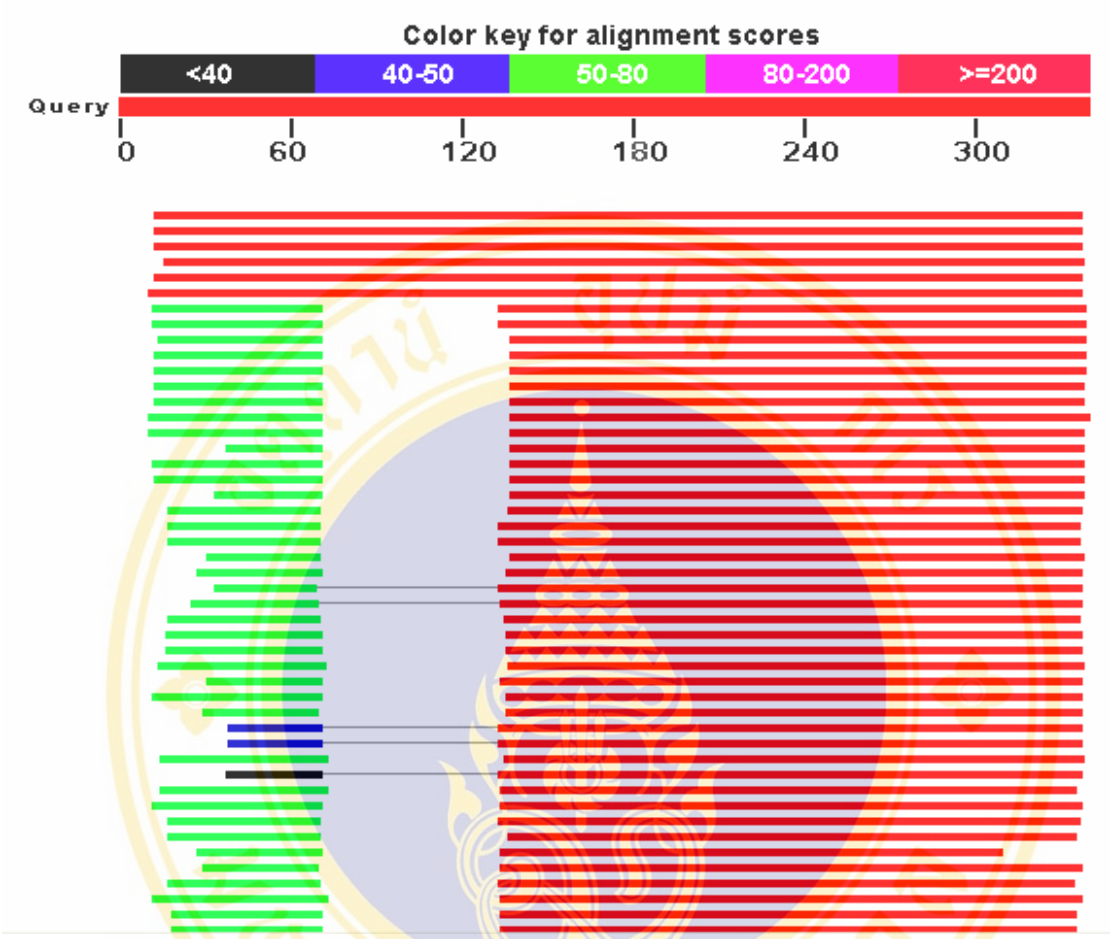
A band of approximately 1 kb containing the full-length gene is showed in the lane 2 and that of 950 bp containing the mature gene is showed in lane 3. Lanes 1 and 4 show negative controls. M_1 is a 100 bp marker and M_2 is *BstE*II-digested lambda marker.




```
1    ATG ATC CTT CAC CGC ACA GAA AAA AAT AAC GAT ATG CGT CTT TCC   45
1     M   I   L   H   R   T   E   K   N   N   D   M   R   L   S    15
46   TTT GCA GCT TCG TTG TTA CTC GCA ACC GTC GGC ATG CAA TTG GTC   90
16    F   A   A   S   L   L   L   A   T   V   G   M   Q   L   V    30
91   TCT GCC GCC GAT TGC AGC AAC AAG GCT TAC AGC CAA TGC GGT GGC   135
31    S   A   A   D   C   S   N   K   A   Y   S   Q   C   G   G    45
136  CAA AAC TGG AGC GGT GAG TCA TGC TGC GTC TCT GGA TAC GAG TGC   180
46    Q   N   W   S   G   E   S   C   C   V   S   G   Y   E   C    60
181  AAG CAG TTG AAC GAC TAT TAC CAT CAA TGC GTT CCA CAA AAC AGC   225
61    K   Q   L   N   D   Y   Y   H   Q   C   V   P   Q   N   S    75
226  GGA TCA TTC TCA GGA TCA TCT TCC GCT GCC GCC CCA TCG CAC ATG   270
76    G   S   F   S   G   S   S   S   A   A   A   P   S   H   M    90
271  GCC ACC AGC TCT GCA CCG TCA TCC TCC AAG GCT CCT TCT AGC CCT   315
91    A   T   S   S   A   P   S   S   S   K   A   P   S   S   P   105
316  GCA TCT TCG TCC AAG ACC CCC TCT TCT CCC GCC GCT TCT TCT TCT   360
106   A   S   S   S   K   T   P   S   S   P   A   A   S   S   S   120
361  TCT TCC TCC GGA AGT GGT TAC AAG CCT ATC TCT GGC GGC GCA TCC   405
121   S   S   S   G   S   G   Y   K   P   I   S   G   G   A   S   135
406  GGT GAT GGT ACT ACC ACC CGC TAC TGG GAC TGC TGC AAG GCC TCG   450
136   G   D   G   T   T   T   R   Y   W   D   C   C   K   A   S   150
451  TGC AGC TGG CCT GGC AAG GCA CCC GTC ACC AAT CCC GTC GGC ACT   495
151   C   S   W   P   G   K   A   P   V   T   N   P   V   G   T   165
496  TGC GCA AAG GAT GGT GTC AAG CTT GTA GAT GCT AAT GTC CAG AGC   540
166   C   A   K   D   G   V   K   L   V   D   A   N   V   Q   S   180
541  GGC TGC AAC GGC GGT GAG GGC TAC ATG TGT AAC GAC AAC CAG CCT   585
181   G   C   N   G   G   E   G   Y   M   C   N   D   N   Q   P   195
586  TGG GCC ATT GAC GAC AAT CTT TCA TAC GGT TTC GCT GCT GCA AGC   630
196   W   A   I   D   D   N   L   S   Y   G   F   A   A   A   S   210
631  ATC AGC GGA AAG TCC GAG AGC GAT TTC TGC TGC AGT TGC TAC GAG   675
211   I   S   G   K   S   E   S   D   F   C   C   S   C   Y   E   225
676  CTG ACG TTC AGC TCG GGC GAG ATC GAG GGC AAG AAG ATG GTC GTC   720
226   L   T   F   S   S   G   E   I   E   G   K   K   M   V   V   240
721  CAG GTC ACC AAC ACT GGC GGC GAC TTG AGC AAC AAC CAC TTT GAC   765
241   Q   V   T   N   T   G   G   D   L   S   N   N   H   F   D   255
766  TTG CAG ATC CCC GGA GGC GGT GTC GGT ATC TTC AAC GGA TGC CAA   810
256   L   Q   I   P   G   G   G   V   G   I   F   N   G   C   Q   270
811  ACT CAG TGG GAT GCA CCC AGC GAT GGC TGG GGC CAG CGC TAC GGC   855
271   T   Q   W   D   A   P   S   D   G   W   G   Q   R   Y   G   285
856  GGT ATC AGC AGC GCA TCT GAA TGC TCT CAG CTT CCC AAG CAA CTT   900
286   G   I   S   S   A   S   E   C   S   Q   L   P   K   Q   L   300
901  CAG GAC GGA TGC AAG TGG CGC TTC GAC TGG TTC AAG AAC GCT GAC   945
301   Q   D   G   C   K   W   R   F   D   W   F   K   N   A   D   315
946  AAC CCC AAC GTT TCC TTC AAG CAA GTA TCT TGC CCC GCT GAG CTC   990
316   N   P   N   V   S   F   K   Q   V   S   C   P   A   E   L   330
991  GTC AAG AAG ACT GGC TGC GAA CGC ACC TCC TAA
331   V   K   K   T   G   C   E   R   T   S   *
```

Figure 23. Nucleotide and amino acid sequences of endoglucanase from BCC18080.

After amplification using specific primers as indicated by arrows, full-length gene contained 1,020 nucleotides that encoded 340 amino acids and the mature gene contained 924 nucleotides that encoded 308 amino acids were obtained.



Sequences producing significant alignments:		species	score(Bits)	E Value
gi|27530542|dbj|BAC53956.1|	endo-beta-1,4-D-glucanase	*Rhizopus oryzae*	395	1e-108
gi|62821724|dbj|BAD95809.1|	endo-beta-D-1,4-glucanase	*Mucor circinelloides*	385	9e-106
gi|27530617|dbj|BAC53988.1|	endo-glucanase RCE3	*Rhizopus oryzae*	375	9e-103
gi|27530615|dbj|BAC53987.1|	endo-glucanase RCE2	*Rhizopus oryzae*	370	4e-101
gi|56410394|dbj|BAD77808.1|	endo-beta-D-1,4-glucanase	*Phycomyces nitens*	364	2e-99

Figure 24. NCBI- BLAST search result of BCC18080 endoglucanase gene

NCBI-BLAST result showed sequence similarity with other fungal endoglucanase genes. The different color bars indicate the score. A red bar with scale numbers showed the BCC18080 endoglucanase sequence.

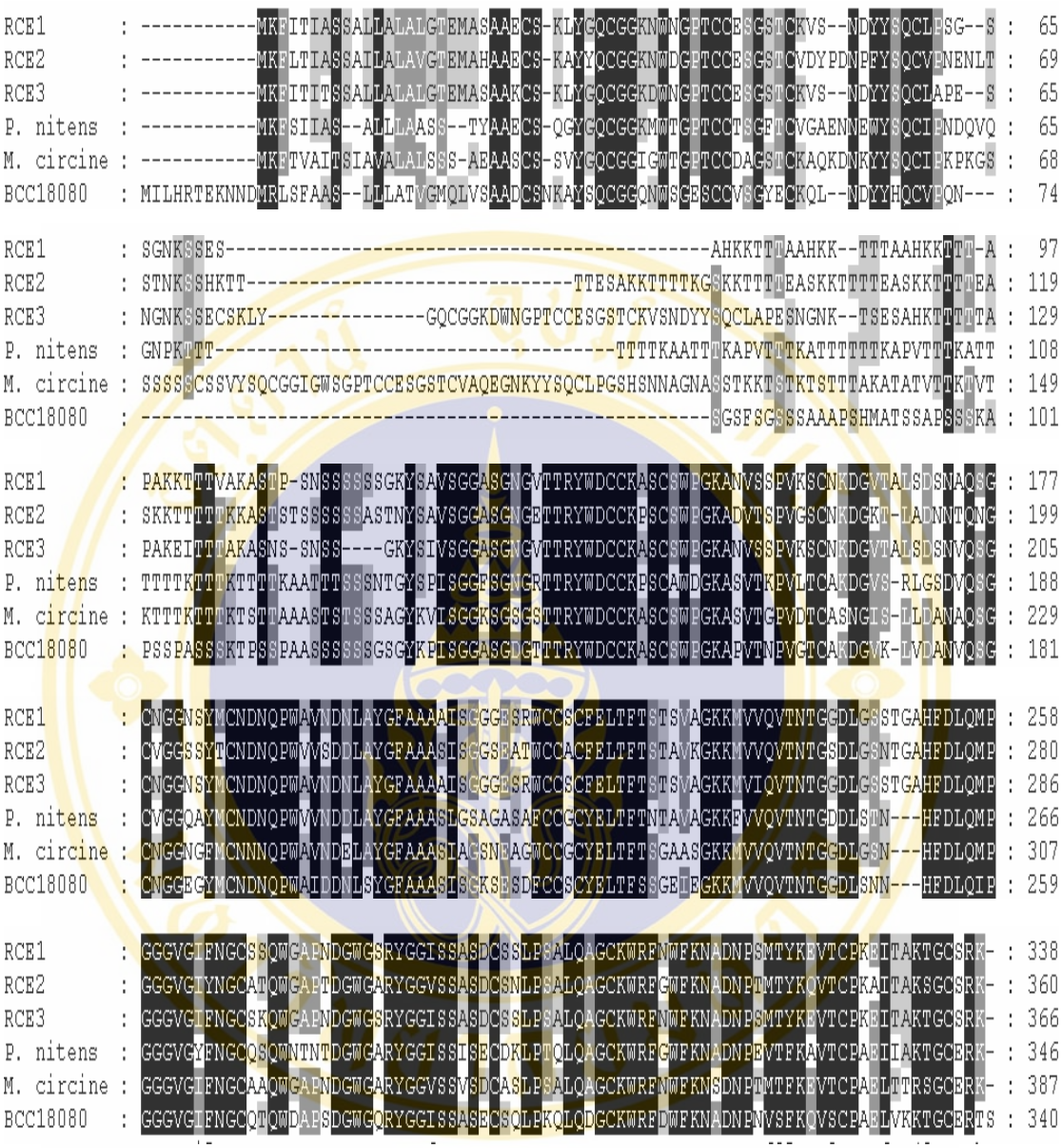
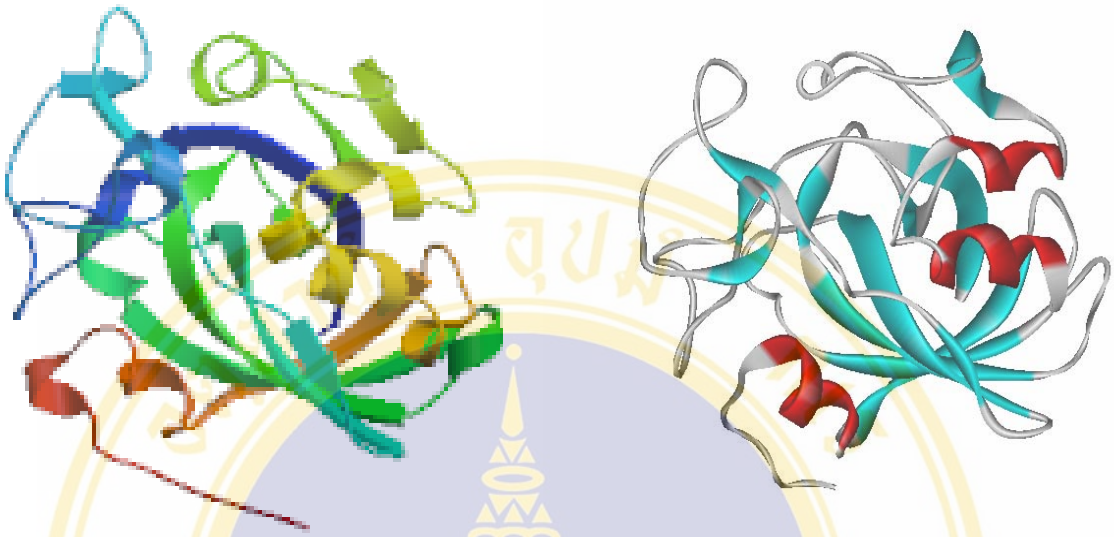


Figure 25. An alignment of the deduced amino acid sequence of endoglucanase cDNA from BCC18080 against other fungal endoglucanase genes

The deduced amino acid sequence of BCC18080 endoglucanase cDNA was aligned using ClustalX with endoglucanases of *Rhizopus oryzae RCE1, RCE 2, RCE 3, Mucor circinelloides* and *Phycomyces nitens*. The darkest shading indicates the most conserved regions among all sequences compared.



(A) **(B)**

Figure 26. 3D-structure of BCC18080 endoglucanase as predicted by the SWISS-MODEL program

The 3D-stucture of BCC18080 endoglucanase gene was predicted by using SWISS-MODEL program in www.expasy.org/swissmod/SWISS-MODEL html and WebLab viewerpro program. (A) Shows the 3D-structure of endoglucanase V belonged to glycosyl hydrolase family 45 from *Humicola insolens*. (B) Shows the predicted 3D-structure of BCC18080 endoglucanase gene.

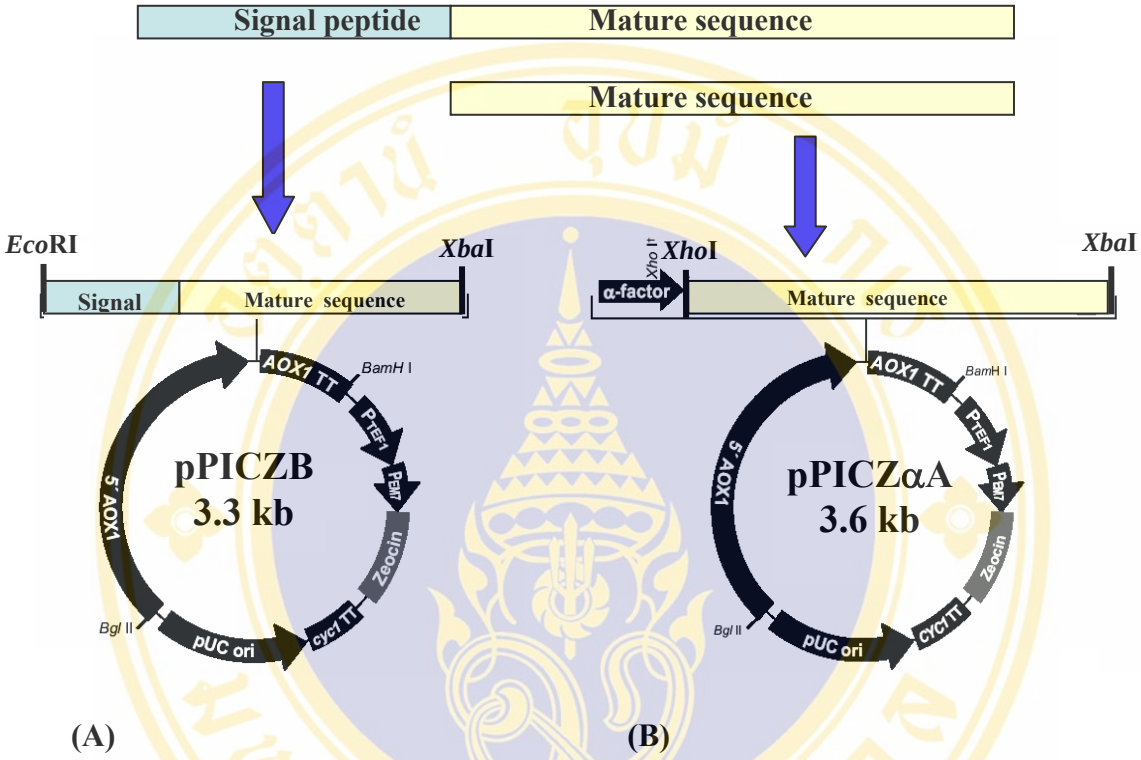


Figure 27. Construction of BCC18080 endoglucanase in *P. pastoris* expression vectors

(A) The full- length endoglucanase of BCC18080 with its own signal sequence was constructed into *p. pastoris* expression vector, pPICZB whereas (B) The mature endoglucanase gene of BCC18080 was constructed into *p. pastoris* expression vector, pPICZαA which contains *S. cerevisiae* α-factor signal sequence.

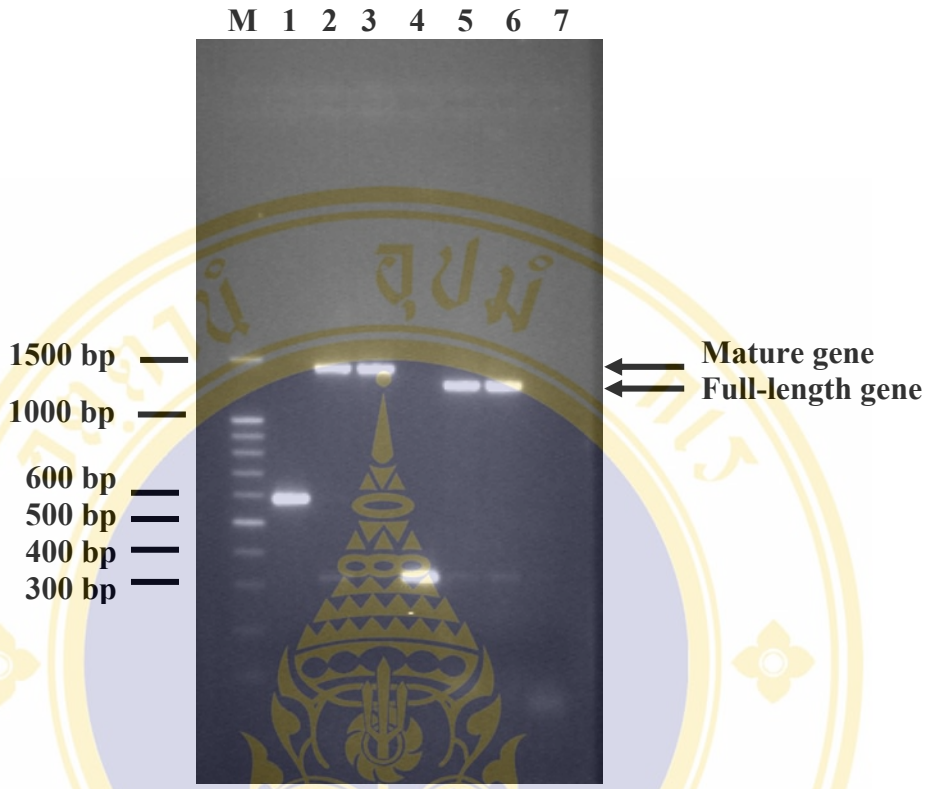


Figure 28. An agarose gel showed a result of PCR analysis to determine the integration of plasmid into the *P. pastoris* genome

PCR analysis using 5'AOX and 3'AOX primers showed that the endoglucanse gene inserted into the *P. pastoris* genome. Lane M is a 100 bp ladder DNA marker. Lane 1 is the PCR product of *P. pastoris* intergrants containing pPICZαA. Lanes 2, 3 are the PCR products of *P. pastoris* intergrant containing endoglucanase with α-factor signal sequence clone no.1 and 2, respectively. Lane 4 is the PCR product of *P. pastoris* intergrant containing pPICZB. Lane 5, 6 are the PCR products of *P. pastoris* intergrants containing endoglucanase with its own leader sequence clone no. 1 and 2, respectively. Lane 7 is a negative control using distilled water as template.

5.8 Expression of BCC18080 endoglucanase in *P. pastoris*

Two *P. pastoris* integrants, one containing the endoglucanase cDNA with its own leader sequence and another containing the endoglucanase cDNA with α-factor signal sequence, were chosen for small-scale expression as described in method 4.9. The chosen clones were grown in BMGY until OD_{600} reached 5-6 units. To induce expression, cells were pelleted and resuspended in BMMY containing 3 % (v/v) of methanol. Methanol was added every 24 h to maintain induction. The culture was collected every day for 6 days.

5.8.1 Protein analysis by using SDS-PAGE

To analyze the secreted endoglucanase, a volume of 15 µl of the culture supernatant was loaded into a 12% SDS-polyacrylamind gel. The result showed two protein bands of approximately 55 kDa and 30 kDa from both constructs and not from the *P. pastoris* without integrated genes, suggested that endoglucanase from both constructs was successfully produced and secreted from *P. pastoris*. The expected size of the secreted protein was approximately 34 kDa. Interestingly, the result revealed that both constructs gave the same protein expression patterns of both 55 kDa and 30 kDa (Figure 29A and 29B) indicating that the signal sequence of BCC18080 endoglucanase can be recognized by *P. pastoris* system which of these 2 bands is the enzyme endoglucanase will have to be confirmed by N-terminal sequencing. Next, N-terminal sequencing was carried out to determine if the cleavage of the fungal secreted sequence was as expected. But firstly, the cellulase activity assay will be performed to determine if the proteins produced from both constructs were endoglucanase.

5.8.2 Endoglucanase activity test by AZCL-HE cellulose

In order to determine the endoglucanase activity of the recombinant proteins, 20 µl of culture supernatant was applied in each well of AZCL-HE cellulose plate and the reaction was incubated at 30°C for 12 h. The blue zone indicating endoglucanase activity was observed for both transformants with its own leader or with α-factor signal sequence. The supernatant after 4 days of induction gave the largest blue zone (Figure 30A and 30B) while the blue zone was not detected from the transformant containing only pPICZB or pPICZαA, suggesting that *P. pastoris* containing endoglucanase cDNA with its own leader sequence or with α-factor signal sequence

secreted active endoglucanase protein. However, to identify the native of two bands, 55 and 30 kDa, purification and N-terminal sequencing were carried out.

5.8.3 Purification of recombinant BCC18080 endoglucanase

The secreted proteins from *P. pastoris* containing either BCC18080 endoglucanase cDNA with its own leader sequence or with α-factor signal sequence were purified. Specifically, after transformants were grown in BMMY containing 3 % (v/v) of methanol for 4 days, the cell culture was centrifuged and the supernatant was concentrated using ultrafiltration, and then applied to Superdex^{TM}75 column. The column was then eluted with 50 mM phosphate buffer pH7. The 55 kDa protein was obtained in fractions no.11 and no.12, while the 30 kDa protein was obtained in fractions no.15 and no.16. The result of purification of the secreted protein from both constructs was similar (Figure 31). A volume of 20 µl from each purified protein was applied in an AZCL-HE cellulose plate to assay for endoglucanase activity. The blue zones were detected from all purified protein fractions 11, 12, 15, and 16 (Figure 32). The results suggested that both the 55 and 30 kDa proteins possed the endoglucanase activity.

5.9 N-terminal sequencing

To confirm that both the 55 and the 30 kDa proteins were in fact endoglucanase, N-terminal sequence analysis of recombinant endoglucanases were performed. The result revealed that the first ten amino acids of purified 55 kDa protein from transformants containing endoglucanase cDNA either with its own leader sequence or with α-factor signal sequence were identical. They started at amino acid residues 33 of BCC18080 endoglucanase sequence which were A-D-C-S-N-K-A-Y-Y-Q (Appendix 1. and 2.). The result was in agreement with that predicted by SignalP program. Sequence of the 30 kDa protein revealed that the first ten amino acids were at amino acid residues 121 to 130 of the BCC18080 endoglucanase sequence. The residues were S-S-S-G-Y-G-Y-K-P-K (Appendix 3.). The result suggested that the 30 kDa protein is the partial fragment of BCC18080 endoglucanase.

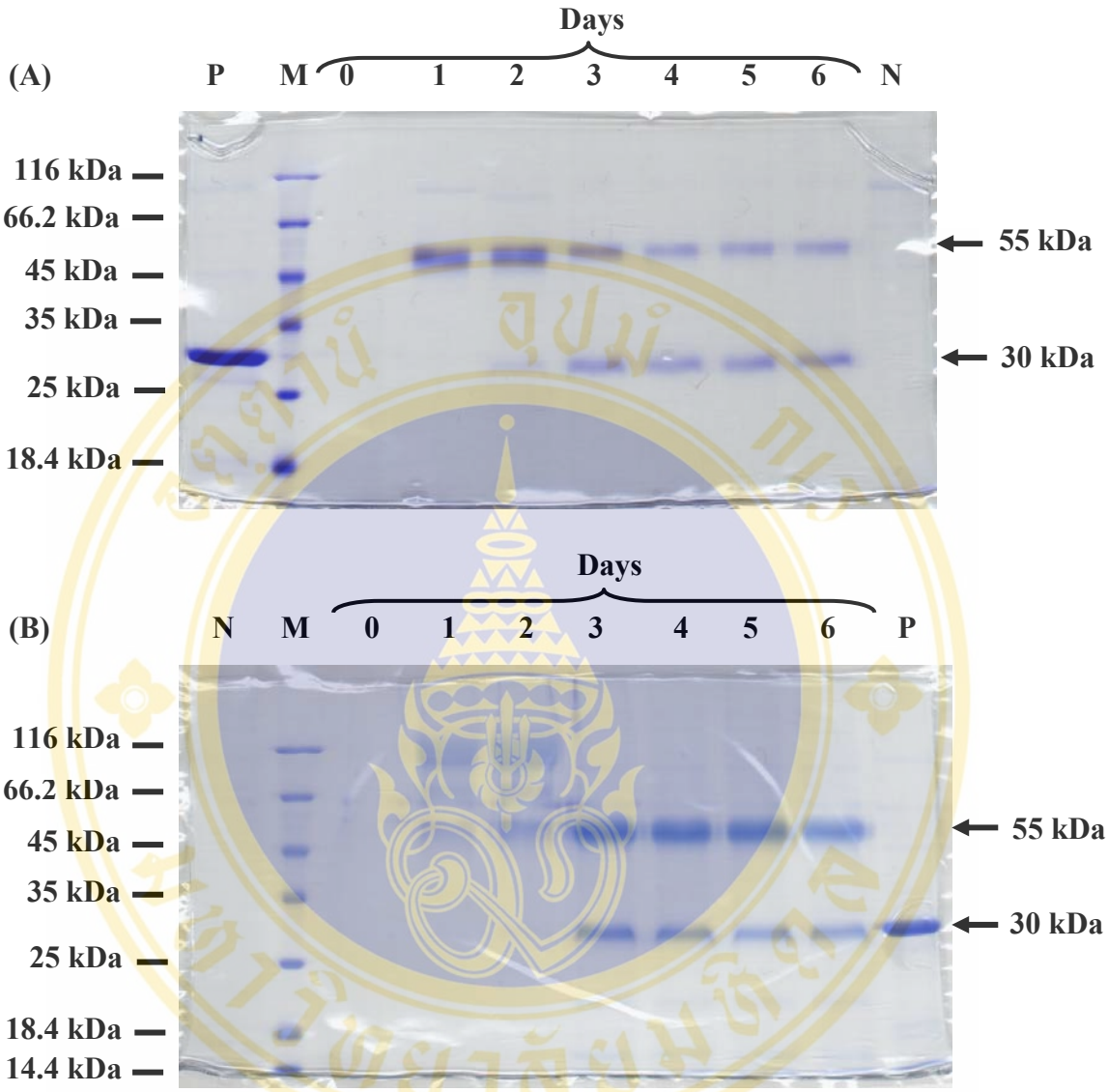
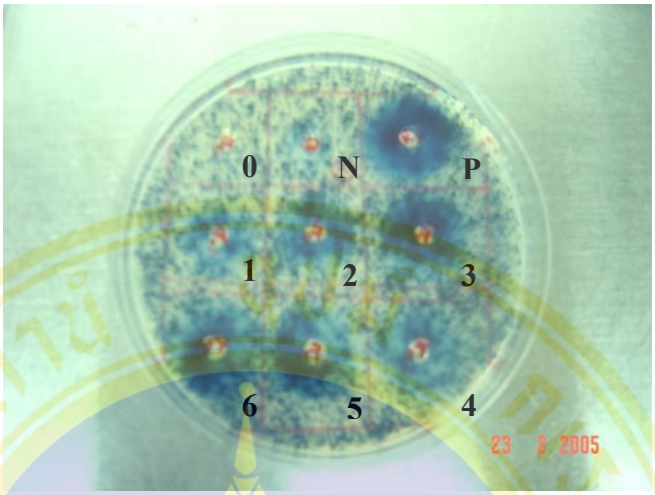


Figure 29. SDS-polyacrylamide gel electrophoresis of *P. pastoris* integrants expressing BCC18080 endoglucanase

(A) SDS-polyacrylamide gel electrophoresis of *P. pastoris* integrant clone no.1 containing BCC18080 endoglucanase gene with its own signal peptide. (B) SDS-polyacrylamide gel electrophoresis of *P. pastoris* integrant clone no.1 containing BCC18080 endoglucanase with α-factor signal sequence. These clones were grown in BMMY containing 3% methanol. The culture supernatant was collected at 0, 1, 2, 3, 4, 5 and 6 days. 15 µl of culture supernatant was loaded into each lane (lane 0-6). Lane P is the culture supernatant of xylanase gene loaded as a positive control. M is a protein marker (Fermentas). Lane N is culture supernatant of *P. pastoris* integrant containing only pPICZαA (A) or pPICZB (B) after 4 days of induction.

(A)



(B)

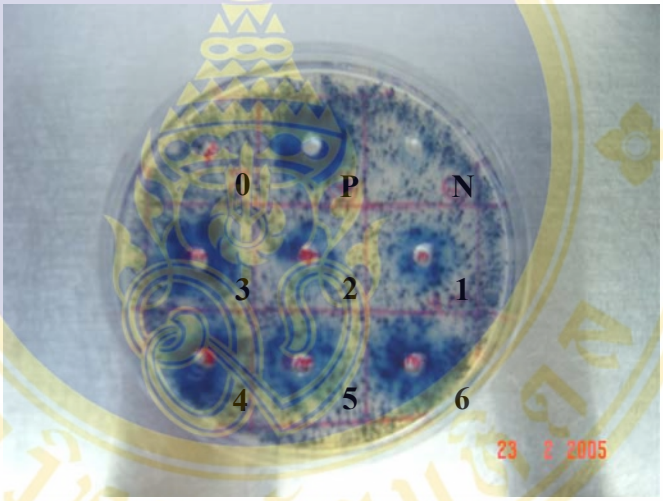


Figure 30. Endoglucanase activity as assay on AZCL-HE cellulose plates

The culure supernatant from *P. pastoris* integrants containing BCC18080 endoglucanase with its own leader sequence (A) or with α-factor signal sequence (B) after induction for 0-6 days were collected. A 20 µl of culture supernatant was applied in each well in AZCL-HE cellulose plate to assay the endoglucanase activity. The blue zone was observed after incubating at 30°C for 12 h. Wells 0-6 is *P .pastoris* integrant containing BCC18080 endoglucanase with its own leader sequence (A) or with α-factor signal sequence (B) after induction for 0-6 days, respectively. Well N is *P. pastoris* integrant containing only pPICZB (A) or pPICZαA (B). Well P is *Trichoderma reesei* cellulose (Boehringer Mannheim Gmbh, Germany) (0.126 units) used as a positive control.

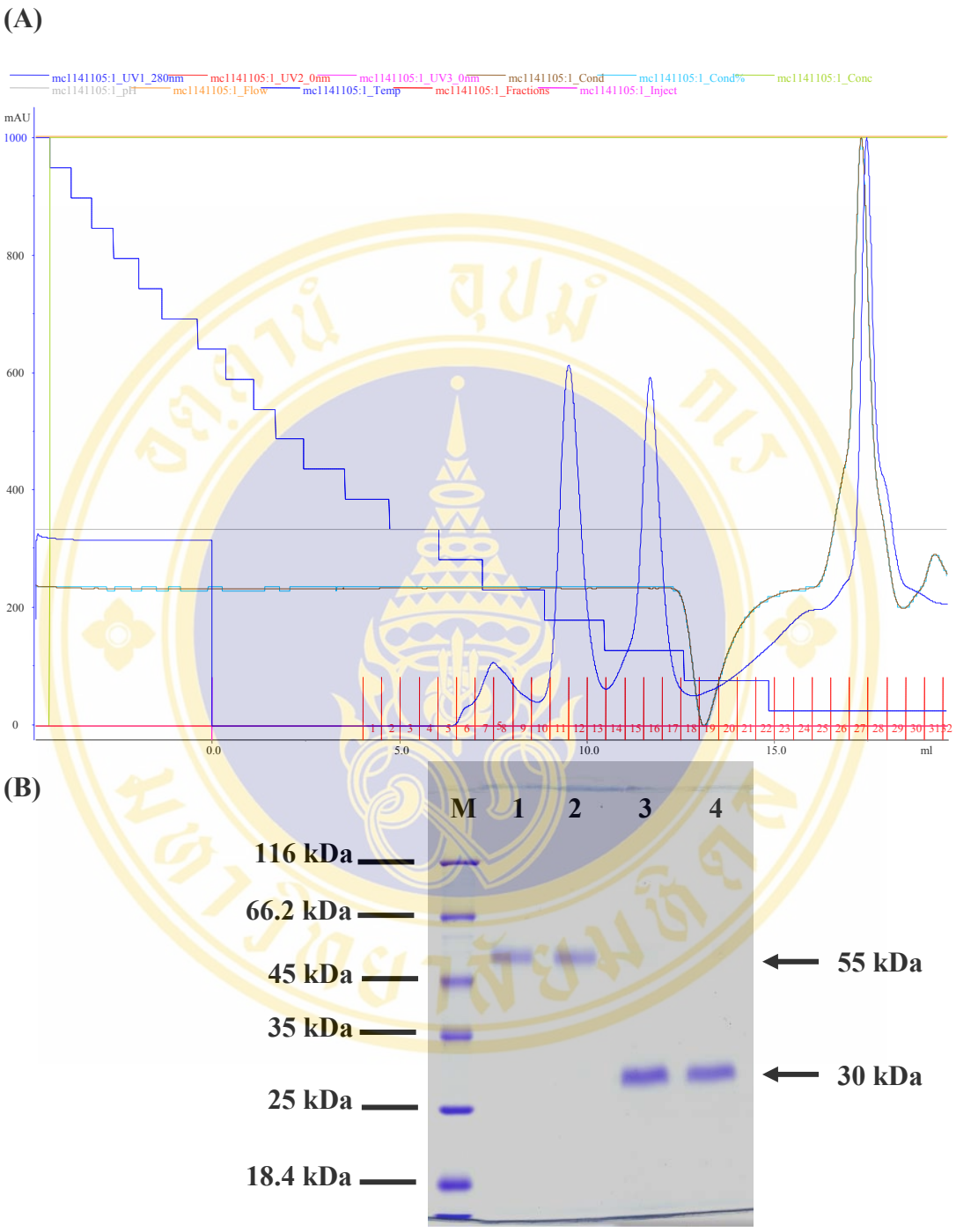


Figure 31. Purification of the secreted proteins by gel filtration

Two secreted proteins with the sizes of 55 kDa and 30 kDa were separated by size using Superdex^{TM} 75 column. After the column was eluted with 50 mM phosphate buffer pH 7 the chromatogram (A) showed that the 55 kDa protein was obtained in fractions no.11 and no.12 (Lane 1 and 2, respectively) while the 30 kDa protein was obtained in the fractions no.15 and no.16 (Lane 3 and 4, respectively).

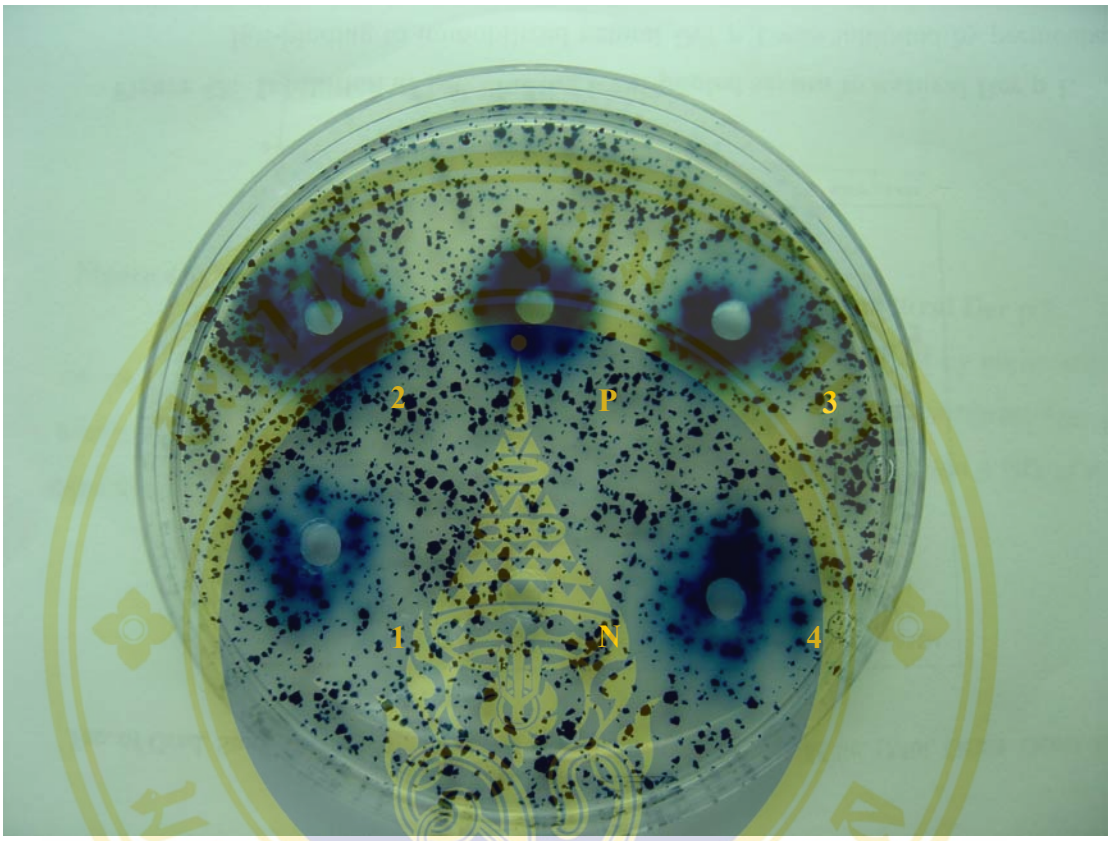


Figure 32. Endoglucanase activity assay of purified recombinants using AZCL-HE cellulose plate.

A volume of 20 µl of both recombinant purified proteins (55 kDa and 30 kDa) from *P. pastoris* integrant clone no.1 containing BCC18080 endoglucanase with its own leader sequence or with α-factor signal sequence was applied into each well of AZCL-HE cellulose plate. The blue zone was detected from all wells except well N, which was a negative control containing only 50 mM phosphate buffer pH 7. Well 1 and 2 were purified 55 kDa and 30 kDa, respectively from *P. pastoris* integrant clone no.1 containing BCC18080 endoglucanase with its own leader sequence. Well 3 and 4 were purified 55 kDa and 30 kDa, respectively from *P. pastoris* integrant clone no.1 containing BCC18080 endoglucanase with α-factor signal sequence. Well P was 0.126 units of *Trichoderma reesei* cellulase (Boehringer Mannheim Gmbh, Germany) used as positive control.

5.10 Glycoprotein analysis

5.10.1 Glycoprotein staining of the recombinant protein

Since the sizes of the BCC18080 endoglucanase protein bands were larger than the calculated molecular mass from its deduced amino sequence, one possibility was that the recombinant endoglucanase produced from *P. pastoris* was glycosylated. To investigate this, the GelCode ® Glycoprotein Staining kit which was used to detect glycoprotein sugar moieties in polyacrylamide gel was employed. The result revealed that the purified band of 55 kDa (lane 1) gave a magenta stained while the purified band of 30 kDa did not (lane 2). A glycoprotein used as a positive control in lane P (Figure 33) showed a magenta band with light pink background indicating this method was susccessful. Lane N which was a negative control was not stained. The result suggested that the 55 kDa secreted protein produced from *P. pastoris* integrant containing endoglucanase gene was glycosylated.

5.10.2 N-linked deglycosylation

Next, to determine if the recombinant glycosylated 55 kDa endoglucanase band was N-linked, sequence analysis was performed to check for potential N-linked glycosylation site. The result showed that the deduced amino acid of BCC18080 endoglucanase contained potential N-liked glycosylation sites (N-X-S/T) at the positions 47, 201 and 318 (Figure 34). To investigate this, N-Glycosidase F known as PNGaseF which is an enzyme that cleaved between GlcNAc and asparagines residues of high mannose and complex oligosaccharide from N-linked glycoprotein was employed. An N-linked glycosylated protein, RNaseB, which was a 21.5 kDa glycoprotein (Sigma, USA) was used as a positive control in this experiment. After treatment with PNGaseF, the result showed that there was no change in the mobility of the recombinant proteins from both *P. pastoris* integrants containing endoglucanase cDNA with its own leader sequence or with α-factor signal sequence while the band of the RNaseB shifted down to approximately 16.4 kDa (Figure 35). This suggested that the BCC18080 endoglucanase recombinant protein was not N-linked glycosylated.

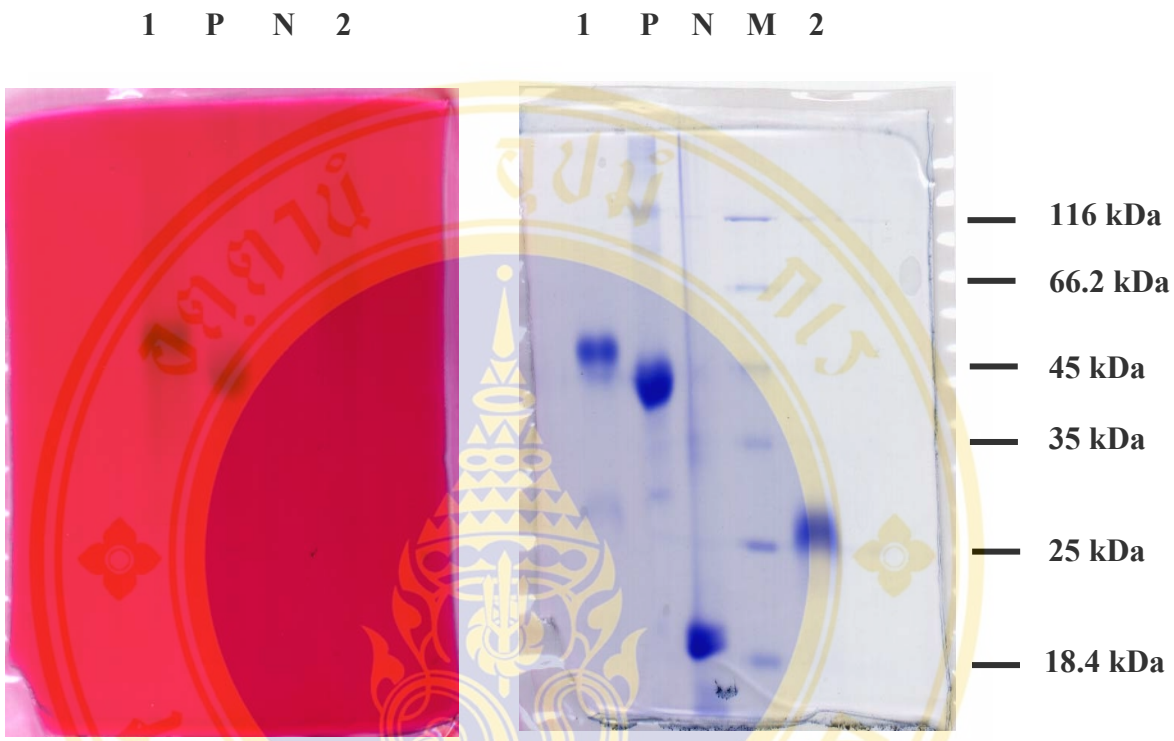


Figure 33. Glycoprotein staining analysis of the recombinant endoglucanase expressed in *P. pastoris*

(A) Half of SDS-polyacrylamide gel stained with GelCode® Glycoprotein Staining kit (B) another half of SDS-polyacrylamide gel stained with coomassie blue R-250. Lane 1 is the purified 55 kDa protein expressed in *P. pastoris* containing BCC18080 endoglucanase gene. Lane P is the horseradish peroxidase used as a positive control. Lane 2 is the purified 30 kDa protein expressed in *P. pastoris* containing BCC18080 endoglucanase gene. Lanes N and M are the soybean trypsin inhibitor used as a negative control and protein molecular weight marker (Fermentas), respectively.

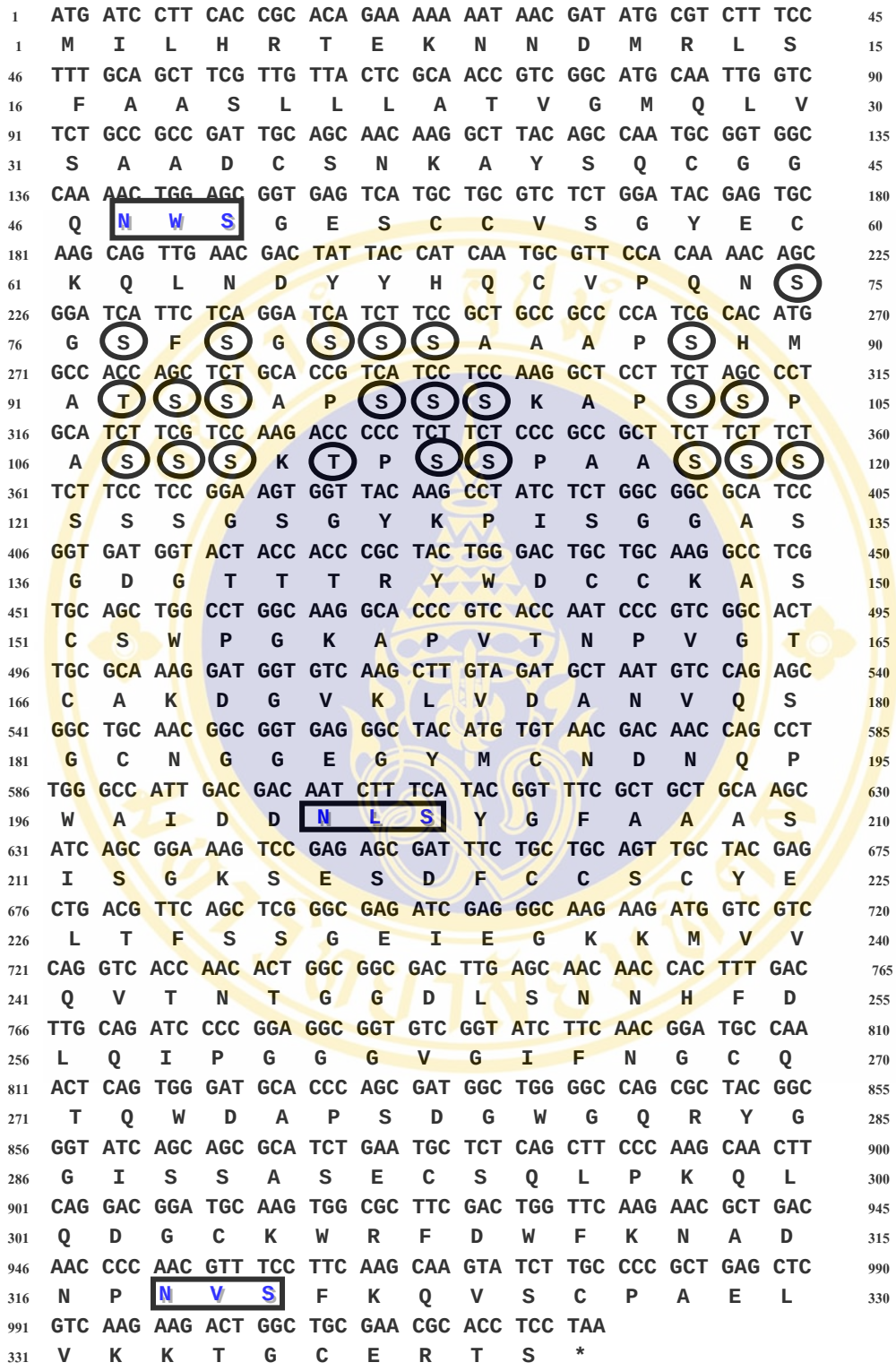


Figure 34. The predicted N- and O-liked glycosylation sites in the BCC18080 endoglucanase gene

The BCC18080 endoglucanase sequence contained a single open reading frame encoding a deduced 340-amino acid**.** Potential N-linked glycosylation sites (N-X-S/T) were found at positions 47, 201 and 318 as indicated in black boxes. Potential O-linked glycosylation sites at the linker region are indicated in the black circles.

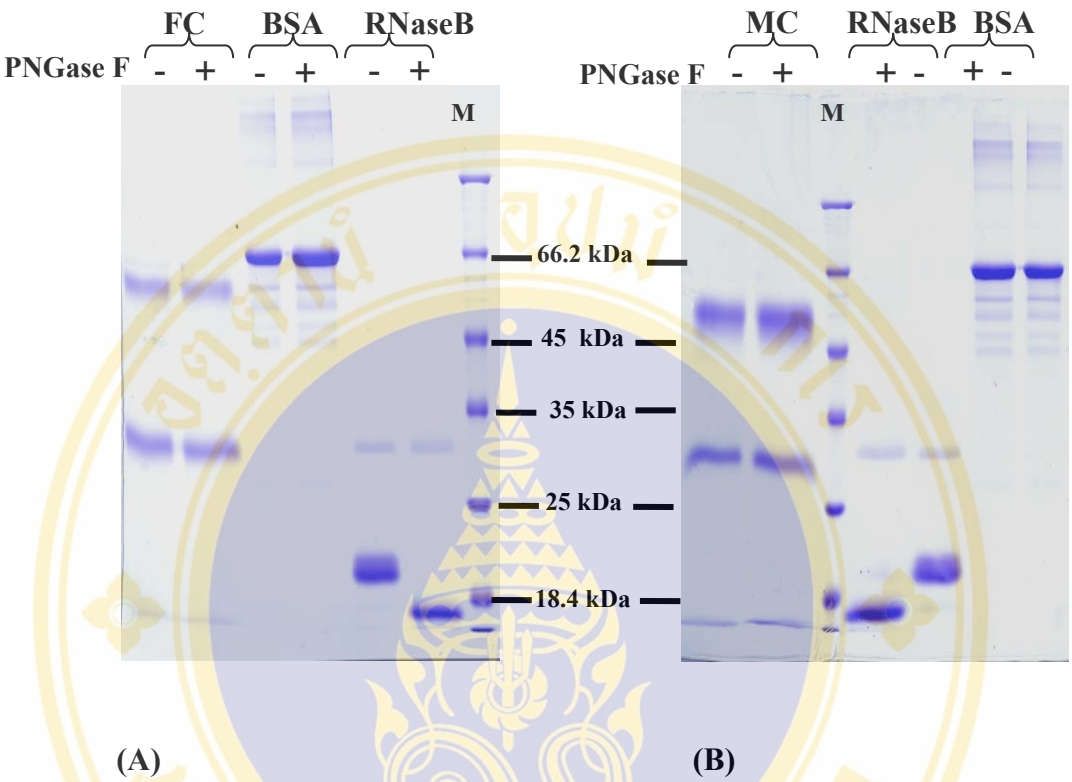


Figure 35. N-linked deglycosylation assay of recombinant endoglucanase expressed in *P. pastoris*

(A) N-linked deglycosylation of the recombinant proteins from *P. pastoris* integrant containing BCC18080 endoglucanase with its own leader sequence and (B) N-linked deglycosylation of the recombinant proteins from *P. pastoris* integrant containing BCC18080 endoglucanase with α-factor signal sequence. Lanes FC and MC are the recombinant proteins with (+) and without (-) N-glycosidase F. RNaseB with (+) and without (-) N-glycosidase F are used as positive controls. BSA with (+) and without (-) N-glycosidase F are used as negative control. Lane M is a protein marker (Fermentus). + represents the reaction with N-glycosidase F while – represents the reaction without N-glycosidase F.

5.11 Characterization of the recombinant BCC18080 endoglucanase

5.11.1 Specific activity of the recombinant endoglucanase

Substrate specificity of purified recombinant BCC18080 endoglucanases (55 kDa and 30 kDa protein from *P. pastoris* integrant containing BCC18080 endoglucanase with its own signal sequence or with α-factor signal sequence) was determined using carboxylmethylcellulose (substrate for endoglucanase), Avicel (substrate for avicelase), birchwood xylan and cellobiose. The reaction was performed by incubating the reaction mixture containing 10 mg of each substrate in 1 ml of 50 mM sodium acetate buffer (pH 6) at 70°C for 10 min. Both the purified 55 kDa and 30 kDa recombinant proteins from *P. pastoris* integrants containing BCC18080 endoglucanase with its own signal sequence or with α-factor signal sequence showed endoglucanase specific activity. The results revealed that all purified recombinant proteins showed activity against carboxylmethylcellulose (CMC) but not against Avicel, birchwood xylan and cellobiose (Table 3). The results suggested that both constructs expressed functional endoglucanase and that both the larger (55 kDa) and the smaller (30 kDa) proteins possed the endoglucanase activity.

5.11.2 Determination of the optimal temperature and pH of the recombinant endoglucanase

To determine the optimal temperature of the recombinant enzyme, the enzyme activity was measured at various temperatures from 30°C to 80°C at pH 5.8. The optimal temperature of all purified recombinant proteins was 70°C (Figure 36A) which was similar to that of the native endoglucanase from BCC18080 (Figure 36B).

The optimal pH of the recombinant enzyme was determined as follow: 50 mM citrate buffer (pH 3-5), 50 mM sodium acetate buffer (pH 4-6), 50 mM sodium phosphate buffer (pH 5-8) and 50 mM Tris-HCl buffer (pH 8-10). The optimal pH of the 55 kDa purified proteins from *P. pastoris* integrant containing BCC18080 endoglucanase with its own signal sequence or with α-factor signal sequence was at pH 6 in 50 mM sodium acetate buffer (Figure 37A). Whereas the optimal pH of 30 kDa purified proteins from *P. pastoris* integrant containing BCC18080 endoglucanase with its own signal sequence or with α-factor signal sequence was at pH 5 in 50 mM citrate buffer (Figure 37B), which was similar to that of native endoglucanase from BCC18080 (Figure 38).

5.11.3 Thermostability of recombinant endoglucanase

The stability of enzyme at high temperature was determined by incubating the recombinant enzyme at 70^{o}C for 2 h and 4 h in 50 mM sodium acetate buffer (pH 6) and the assay for the enzyme activity as described in the Method 4.11.5. The result showed that more than 50% of the endoglucanase activity was detected after incubation for 4 h (Table 4).

5.11.4 Enzyme kinetics

The hydrolysis kinetic of carboxylmethylcellulose (CMC) by recombinant endoglucanase in 50 mM sodium acetate buffer (pH 6) at 70^{o}C was determined. The result of hyperbolic curve V_{max}, K_m and K_{cat} were calculated using the equation for enzyme kinetics commonly referred to as the Henri-Michaelis-Menten equation (118). The apparent K_m value of the purified endoglucanase (55 kDa) from *P. pastoris* integrant containing BCC18080 endoglucanase with its own signal sequence or with α-factor signal sequence was lowered than that of the purified 30 kDa endoglucanase from *P. pastoris* integrant containing BCC18080 endoglucanase with its own signal sequence or with α-factor signal sequence (Table 5). This indicated that the larger protein had better affinity for CMC substrate. In addition, the K_{cat} value of the 55 kDa protein was higher, suggesting that the enzymatic reaction of the larger protein proceeded faster indicating it was more efficient than the smaller protein.

Table 3.

Analysis of substrate specificity of recombinant endoglucanase expressed in *P. pastoris*

	Hydrolysis activity (U/mg protein)			
	MCU	**MCL**	**FCU**	**FCL**
CMC	108.66	87.33	103	86.66
Avicel	-	-	-	-
Cellobiose	-	-	-	-
Xylan	-	-	-	-

Substrate specificity of the recombinant endoglucanases expressed in *P. pastoris* was determined by using various substrates in 50 mM sodium acetate buffer pH 6 under 70°C. MCU and MCL are 55 kDa and 30 kDa respectively protein from *P. pastoris* integrant containing BCC18080 endoglucanase with α-factor signal sequence. FCU and FCL are 55 and 30 kDa proteins,respectively from *P. pastoris* integrant containing BCC18080 endoglucanase its own signal sequence. " - " represents non-detectable level of enzyme.

Table 4.

Thermostability of the recombinant endoglucanase tested at 70°C

	% remaining activity			
	MCU	**MCL**	**FCU**	**FCL**
Optimal temperature (70°C)	100	100	100	100
Thermostability				
2 h incubation	75.35	75.32	73.01	80.97
4 h incubation	59.29	55.20	52.24	56.44

Thermostability of the recombinant endoglucanase was determined at 70°C using carboxylmethycellulose as substrate in in 50 mM sodium acetate buffer (pH 6) for 2 h and 4 h. Percentage of the remaining activity of an enzyme was calculated.

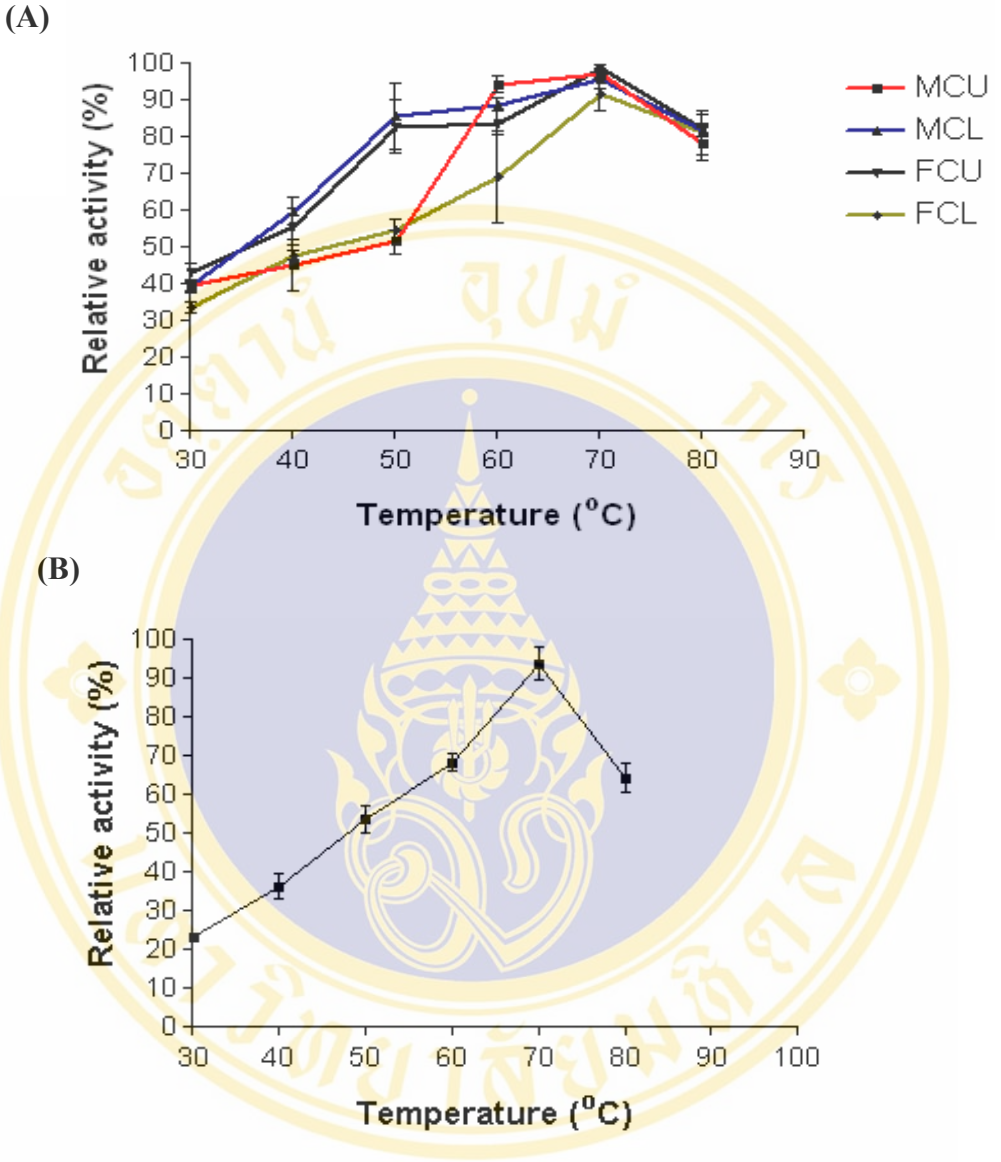


Figure 36. Optimal temperature of purified recombinant proteins and the native BCC18080 endoglucanase

The optimal temperature was determined at various temperatures from 30°C to 80°C in phosphate buffer pH 5.8 against 1% carboxylmethylcellulose for 10 min. The optimal temperature of purified recombinant proteins (A) and the crude enzyme containing BCC18080 endoglucanase (B) were 70°C. MCU and MCL represented 55 kDa and 30 kDa protein respectively from *P. pastoris* integrant containing BCC18080 endoglucanase with α-factor signal sequence. FCU and FCL represented 55 and 30 kDA protein respectively from *P. pastoris* integrant containing BCC18080 endoglucanase with its own signal sequence.

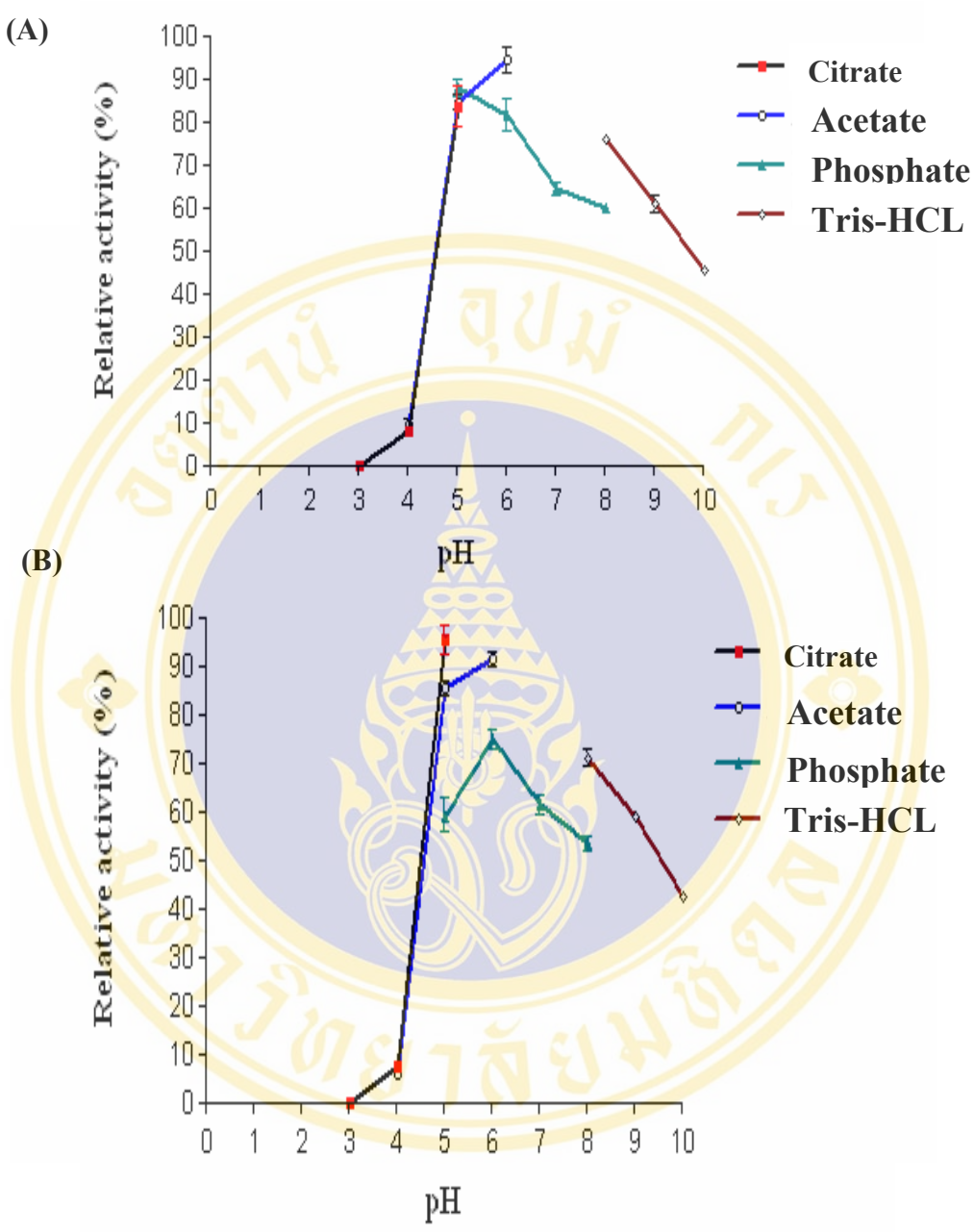


Figure 37. Optimal pH of 55 kDa (A) and 30 kDa (B) protein from *P. pastoris* integrant containing BCC18080 endoglucanase with its own signal sequence or with α-factor signal sequence

To determine the optimal pH of the purified recombinant proteins, the enzyme activity was determined at 70°C in various buffers: 50 mM citrate buffer (pH 3-5), 50 mM sodium acetate buffer (pH 4-6), 50mM sodium phosphate buffer (pH 5-8) and 50mM Tris-HCl buffer (pH 8-10). Substrate used in this study was CMC.

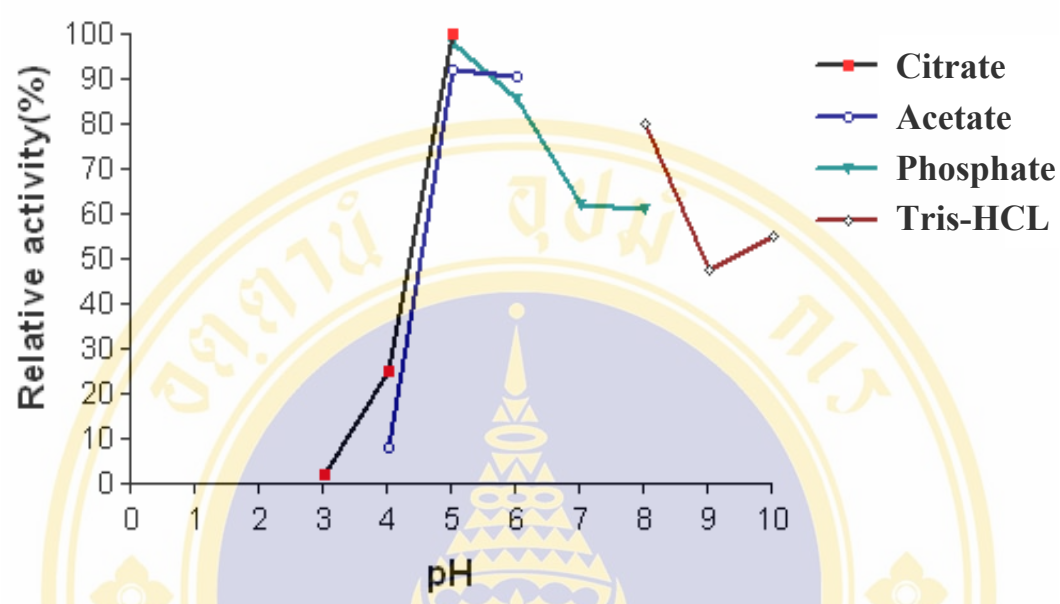


Figure 38. Optimal pH of BCC18080 endoglucanase

The enzyme activity of the BCC18080 endoglucanase was determined at 70°C in various buffer: 50 mM citrate buffer (pH 3-5), 50 mM sodium acetate buffer (pH 4-6), 50mM sodium phosphate buffer (pH 5-8) and 50mM Tris-HCl buffer (pH 8-10). CMC was used as a substrate in this study.

Table 5.

Apparent kinetic constants of the purified recombinant enzyme on carboxylmethylcellulose (CMC)

	V_{max} (µmol/ min per mg protein)	K_m (gl^{-1})	K_{cat} (s^{-1})
MCU	284.2	19.07	161.05
MCL	307.5	25.69	123
FCU	252.7	18.99	143.2
FCL	307.5	24.43	123

After recombinant proteins were incubated in various concentrations of carboxylmethylcellulose (CMC) in 50 mM sodium acetate buffer (pH 6) at 70°C for 10 min, the enzyme activity was determined. V_{max}, K_m and K_{cat} were calculated using the the Henri-Michaelis-Menten equation from the obtained hyperbolic curve. MCU and MCL represented 55 kDa and 30 kDa respectively protein from *P. pastoris* integrant containing BCC18080 endoglucanase with α-factor signal sequence. FCU and FCL represented 55 kDa and 30 kDa protein respectively from *P. pastoris* integrant containing BCC18080 endoglucanase with its own signal sequence.

CHAPTER VI
DISCUSSION

6.1 Efficiency of endoglucanase production after using 5% wheat bran broth as an inducer

Several substrates can be used to induce cellulase production in fungi such as cellulose, wheat bran, rice straw, agro- waste, cellulosic waste, and saw dust (119). Wheat bran is the most commonly used as inducer in cellulase production, since it is conveniently available and affordable. In an addition, this substrate has been found to be a suitable material for cellulase production in *Aspergillus oryzae* (120).

In this thesis, successful endoglucanase production was performed by growing BCC18080 directly on 5% wheat bran broth with continuously shaking at 30°C, 250 rpm for 8 days. The maximal endoglucanase activity was first detected on the fifth day. Moreover, the growing BCC18080 can be easily separated from the medium containing insoluble form of wheat bran.

6.2 Molecular cloning of the full-length endoglucanase gene with or without its own signal sequence

In order to obtain the full-length BCC18080 endoglucanase gene, RT-PCR and 5'-and 3'- RACEs were performed. The partial endoglucanase sequence was achieved by RT-PCR using the four degenerate primers designed based on the conserved amino acid sequence of various endoglucanase genes from members of the phylum *Zygomycota*. The 3' and 5' end of cDNA were obtained by RACE method. Since the expected size of 5'end cDNA is quite long, partial heat denaturation method was performed during the reverse transcription step to avoid RNA secondary structure formation such as hairpins and stem loop formation which often occurs in cDNA synthesis (112). After using this method, 5'end cDNA was successfully obtained and subsequently, the predicted full-length endoglucanase gene was obtained. The NCBI-BLAST search revealed that the amino acid sequence deduced from BCC18080

endoglucanase was homologous to the deduced amino acid sequences of *Rhizopus oryzae RCE1*, *RCE2*, and *RCE3* which consist of N-terminal cellulose binding domains (CBDs), linker domains, and catalytic domains belonging to glycosyl hydrolase family 45 (121). In addition, the 3D structure of BCC18080 endoglucanase was predicted using its deduced amino acid sequence via SWISS-MODEL program. The result revealed that the 3D structure of BCC18080 endoglucanase depicted six-stranded β barrels (117). The structural pattern of BCC18080 endoglucanase showed homology pattern with the catalytic domain of endoglucanase V from *Humicula insolens* which belongs in the family 45 glycosyl hydrolase families (122). Therefore, the catalytic domain of BCC18080 endoglucanase was considered to belong to family 45.

Sequence analysis suggested that the second methionine should be the start codon since the gene should contain 5' untranslated region (Figure 20). In addition, the alignment of the full-length BCC18080 endoglucanase against endoglucanase genes from fungi in the phylum zygomycota showed that the start codons of other endoglucanase genes was aligned in the same position with the second methionine of BCC18080 endoglucanase gene (Figure 25).

Since several fungal enzyme sequences contain the native signal sequence at N-terminal segment, most fungi are capable of secreting these enzymes in the medium. Here, the native signal sequence was determined by computational method, SignalP 3.0, to predict the position of N-terminal signal peptides and the cleavage site (110). The result suggested that the amino acid residues 1 to 32 were signal peptides and the cleavage site is between residues 32 and 33 from the first methionine found in amino acid sequence. N-terminal sequencing was performed and confirmed the exact the cleavage site which was between residues 32 and 33.

In order to investigate if the native signal sequence can function in *P. pastoris* the full-length gene with its own signal sequence together with the full-length gene with α-factor signal sequence, so called the mature gene, were amplified. The full-length endoglucanase with its own signal sequence has an open reading frame of 1,020 nucleotides encoding a polypeptide of 340 amino acids with a calculated mass of 37 kDa while the mature gene contained 921 nucleotides that encoded 307 amino acids with a calculated mass of 34 kDa.

6.3 Expression of BCC18080 endoglucanase in *P. pastoris*

Expression of endoglucanase genes have been reported in several host organisms. For instance, endoglucanase from *Humicula grisea* var *thermodia* was expressed in *Aspergillus oryzae* (123), endoglucanase from *Aspergillus niger* was expressed in *Saccharomyces cerevisiae* (55), endoglucanase from *Macrophomina phaseolina* was expressed in *Escherichica coli* (124), or endoglucanase from *Streptomyces viridosporus* expressed in *P. pastoris* (125). In this study, *P. pastoris* was used as the expression host for the production of the heterologous proteins since this system can perform many of the post-translational modifications which are found in higher eukaryotes such as processing of signal sequences, folding, disulfide bridge formation, certain types of lipid addition, and O- and N-linked glycosylations.

The BCC18080 full-length endoglucanase gene with its own leader peptide was subcloned into *P. pastoris* expression vector, pPICZB; while, the mature gene was subcloned into *P. pastoris* expression vector, pPICZαA, which contained the *S. cerevisiae* MFα-1 signal sequence. After both recombinant plasmids were linearized and then transformed into *P. pastoris*, the integration of endoglucanase cDNA into *P. pastoris* was determined by screening for zeocin-resistant and PCR amplification using 5'AOX and 3'AOX primers. Then, *P. pastoris* integrants containing either the full-length endoglucanase sequence with its own signal sequence or the mature endoglucanase sequence with α-factor signal sequence were expressed in a small-scale. To induce expression, the cells were grown in BMMY containing 3% methanol at every time of induction (1-6 days). The culture supernatant was collected every day for 6 days.

SDS-PAGE analysis of the recombinant proteins showed that both constructs gave the same protein expression patterns of both 55 kDa and 30 kDa whereas these bands were not found from the *P. pastoris* without integrated gene. These results indicated that the fungal signal sequence of BCC18080 can function in *P. pastoris*. In addition, the secreted proteins of both constructs showed activity against AZCL-HE cellulose. The results suggested that *P. pastoris* transformants containing endoglucanase cDNA with its own signal sequence or without but with α-factor signal sequence can produce active endoglucanase.

Since the expected size of the secreted protein was approximately 34 kDa, N-terminal sequencing of both the 55 kDa and the 30 kDa was performed. Before sequencing, two protein bands were purified by Superdex^{TM} 75 column. The 55 kDa protein was separated into fractions no.11 and 12 while the protein of 30 kDa was obtained in fractions no.15 and 16. The N-terminal sequencing results revealed that the first ten amino acid of purified 55 kDa was the mature protein whereas the first ten amino acids of 30 kDa band corresponded to the internal sequence of the BCC18080 endoglucanase, suggesting that 30 kDa protein band is the truncated protein which contained only some parts of linker domain and the catalytic domain. Nonetheless, both the purified 55 kDa and 30 kDa proteins showed endoglucanase activity against AZCL-HE cellulose. All endoglucanase genes from *Rhizopus oryzae* belonging to glycosyl hydrolase family 45 consisted of three distinct domains: a cellulose binding domain at their N-terminus, linker domain, and catalytic domain (121). A similar arrangement has been found in the BCC18080 endoglucanase. The deduced amino acid sequence from BCC18080 endoglucanase gene comprised a cellulose binding domain at residues 33 to 71 in its N-terminus, linker domain at residues 72 to 136, and the catalytic domain at residues 137 to 340. The N-terminal sequence of the 30 kDa protein revealed the first amino acid started at serine 121 (Figure 39). There is a report about linker region of many cellulases which suggests that the interdomain linker region of many cellulases is sensitive to proteolytic digestion; thus, enzymes with and without cellulose binding domains have been found in the culture supernatants of several cellulolytic organisms (126). As the 30 kDa protein contained the catalytic domain at C-terminus, it also exhibited endoglucanase activity against AZCL-HE cellulose.

6.4 Glycoprotein analysis

Since BCC18080 endoglucanase protein bands observed from SDS-polyacrylamide showed molecular weights much higher than the calculated molecular weights, they may be glycosylated by *P. pastoris*. Therefore glycosylation was determined in recombinant protein bands, 55 and 30 kDa, by the GelCode ® Glycoprotein Staining kit. The result showed that only the purified 55 kDa band showed the magenta stain while the purified 30 kDa was not. This suggested that 55 kDa secreted protein produced from *P. pastoris* was glycosylated. One possibility was

that the recombinant endoglucanase might be N-linked glycosylated. From the amino acid sequence of BCC18080 endoglucanase, potential N-linked glycosylation sites (N-X-S/T) at positions 47, 201 and 318 were found. To determine N-linked glycosylation, the recombinant proteins were treated with PNGaseF which is an enzyme that cleaved between GlcNAc and asparagines residues of high mannose and complex oligosaccharide from N-linked glycoprotein. The result showed that there was no change in the mobility of the recombinant proteins. This suggested that the 55 kDa protein was not N-linked glycosylated as predicted (Figure 34). Another glycosylation type, O-linked glycosylation, may occur. Although evidence for O-glycosylation of recombinant protein expressed in *P. pastoris* is limited, several proteins such as human lysosomal α-mannosidase, human single-chain urokinase-type plastminogen activator protein and mouse gelatinase B showed O-linked when expressed in *P. pastoris* (127). The purified 55 kDa protein might be glycosylated with O-linked oligosaccharide since it contained relatively long linker peptides which are often heavily glycosylated in many serine and threonine sites in the linker region (127). This evidence should be confirmed by a lectin binding assay against O-linked protein, *Galanthus nivalis* lectin.

6.5 Characterization of recombinant BCC18080 endoglucanase secreted from *P. pastoris*

The activities of the purified 55 kDa and 30 kDa proteins of both constructs against carboxylmethylcellulose (CMC), Avicel, birchwood xylan and cellobiose was determined for substrate specificities. The results showed that both the purified 55 kDa and the 30 kDa recombinant proteins secreted from *P. pastoris* integrants containing BCC18080 endoglucanase with its own signal sequence or without but with the α-factor signal sequence showed activity toward CMC but no activity against other substrates. Since, CMC is a specific substrate for endoglucanase, the result suggested that the purified 55 kDa and the 30 kDa proteins have endoglucanase activity. This activity against CMC of recombinant BCC18080 endoglucanase was similar to the activity of *Rhizopus oryzae* and *Humicula grisea* var *thermoidea* (121, 125).

The purified 55 kDa recombinant protein from *P. pastoris* integrants containing BCC18080 endoglucanase with its own signal sequence and without but with the α-factor signal sequence showed endoglucanase activities of 103, and 108.66 U/mg protein; whereas, the purified 30 kDa recombinant protein from *P. pastoris* integrants containing BCC18080 endoglucanase with its own signal sequence and without but with the α-factor signal sequence showed endoglucanase activities of 86.66, and 87.33 U/mg protein (Table 3). However, the endoglucanase activity of the purified recombinant proteins from both constructs was not significantly different. Combining this with the optimal pH and temperature results, this suggested that both constructs might produce the same 55 kDa and the same 30 kDa proteins. Moreover, N-terminal sequencing results of the 55 kDa from both constructs confirmed that both constructs in fact produced the same protein.

Most of the known thermostable cellulases are produced by thermophilic and mesophilic fungi (54, 55). Although BCC18080 is not a thermophilic fungus, it produced thermostable endoglucanase since more than 50% endoglucanase activity remained after incubation for 4 hours. The endoglucanase of all purified recombinant proteins from both constructs was not significant different although the 55 kDa protein was glycosylated. This is somewhat unexpected since N-linked glycosylation is essential for thermal- and pH-stability (O-linked glycosylation is needed for secretion and activity) (128).

The K_{cat} value in enzyme kinetic determination showed enzyme reaction capacity; whereas, the apparent K_m value represents the substrate concentration at which half of enzyme active sites in the sample are filled. If the K_m value of an enzyme is low, it shows better affinity for the substrate (118). The purified endoglucanase 55 kDa had a lower K_m value than that of the 30 kDa purified endoglucanase. The result suggested a better affinity of the purified 55 kDa protein for the cellulase substrates. This is in fact agreed with the sequence data as this protein contained a cellulose binding domain (CBD). The evidence for the presence of this domain which lowered the apparent K_m has been found in the study cellulases from *Humicula insolens*. The result shows that the cellulases without CBD domain has give higher apparent K_m value than those containing CBD (40). Moreover, the purified endoglucanase of the 55 kDa protein has higher K_{cat} value than that of the

purified 30 kDa endoglucanase. This suggested that the 55 kDa protein has more capacity to carry out the reaction when compare with the 30 kDa endoglucanase.

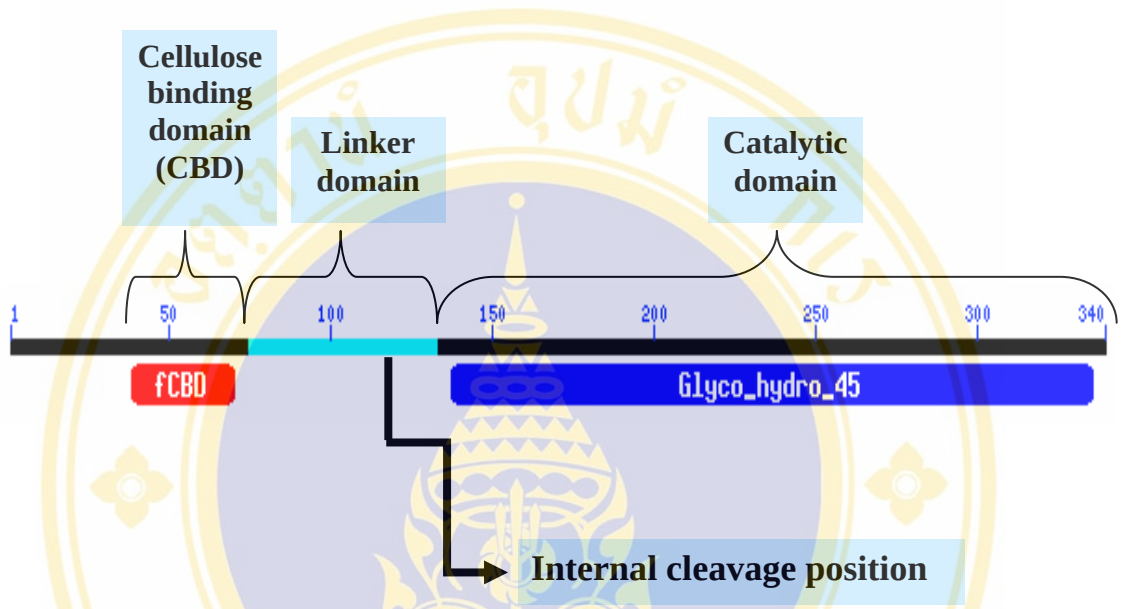


Figure 39. The domains arrangement of BCC18080 endoglucanase

The NCBI-BLAST result revealed that amino acid sequence deduced from BCC18080 endoglucanase which was homologous to endoglucanase sequences of *Rhizopus oryzae* consists of N-terminal cellulose binding domain (CBD) at protein residue 33 to 71 in its N-terminus, linker domain at protein residue 72 to 136, and catalytic domain belonging to glycosyl hydrolase family 45 at protein residue 137 to 340. From N-terminal sequence of 30 kDa protein, it revealed that the first amino acid started from residue 121 as indicated by the arrow.

CHAPTER VII
CONCLUSIONS

1. BCC18080 endoglucanase was produced with maximum activity on the fifth day when it was grown in 5% wheat bran with continuously shaking at 30°C, 250 rpm.
2. The full-length BCC18080 endoglucanase contained 1,020 nucleotides encoding a protein of 340 amino acids with a calculated mass of 37 kDa while the mature gene contained 924 nucleotides that encoded 308 amino acids with a calculated mass of 34 kDa.
3. Based on amino acid sequence similarity. the BCC18080 endoglucanase cDNA was classified into the glycosyl hydrolase family 45 which.
4. The predicted 3D structure of BCC18080 endoglucanase cDNA depicted six-stranded β barrels. This structure showed homology with other endlglucanases belonging to the glycosyl hydrolase family 45.
5. The active BCC18080 endoglucanase protein was successfully expressed in *P. pastoris* and secreted into the culture medium by using either α-factor signal sequence or its own signal sequence.
6. The signal peptide sequence of BCC18080 endoglucanase is from amino acid residues 1 to 32 and the cleavage site is between residues 32 and 33.
7. The secreted proteins of *P. pastoris* transformants containing BCC18080 endoglucanase showed protein expression patterns of 55 kDa and 30 kDa. Both proteins showed endoglucanase activity.
8. From N-terminal sequencing, the 55 kDa proteins from both constructs are of the same protein. From N-terminal sequencing, the 30 kDa protein came from internal sequence the cleavage in the linker region of the BCC18080 endoglucanase.
9. The 55 kDa protein was glycosylated; but was not N-linked glycosylated. The 30 kDa protein was not glycosylated.
10. The purified proteins of 55 kDa and 30 kDa showed substrate specificity against carboxylmethylcellulose (CMC).

11. The optimal temperature of all purified recombinant proteins was 70^{o}C which was similar to that of the endoglucanase from BCC18080.

12. The optimal pH of the 55 kDa purified proteins from *P. pastoris* integrant was at pH 6 in 50 mM sodium acetate buffer whereas the optimal pH of the 30 kDa purified protein was at pH 5 in 50 mM citrate buffer.

13. BCC18080 endoglucanase is a thermostable endoglucanase since it retained more than 50% endoglucanase activity after incubation for 4 hours.

14. The purified 55 kDa endoglucanase had lower K_m value than that of the purified 30 kDa endoglucanase, suggesting a better affinity for the substrate.

15. The purified 55 kDa endoglucanase has a higher K_{cat} than that of purified 30 kDa endoglucanase, suggesting that the 55 kDa protein can carry out the enzyme reaction faster when compare to the 30 kDa protein.

REFERENCES

1. Ribes JA, Vanover-Sams CL, Baker DJ. Zygomycetes in human disease. Clin Microbiol Rev 2000;13(2):236-301.
2. Ingold CT, Zoberi MH. The asexual apparatus of Mucorales in relation to spore liberation. Transactions of the British Mycological Society 1963;46(1):115-134.
3. Webster J. Zygomycotina. In: Introduction of fungi. second ed. USA: Cambridge University press; 1980.
4. Vries P, Visser J. Aspergillus enzymes involved in degradation of plant cell wall polysaccharides. Microbiology and Molecular Biology Reviews 2001;65:497-522.
5. Atalla RH, Hackney JM, Uhlin I, Thompson NS. Hemicelluloses as structure regulators in the aggregation of native cellulose. Int J Biol Macromol 1993;15(2):109-12.
6. Blouin FA, Martin LF, Rowland SP. Gel permeation properties of cellulose.III. Measurement of pore structure of unmodified and of mercerized cottons in fibrous from. Textile Res.J. 1970;40:809-813.
7. Cowling EB. Physical and chemical constraints in the hydrolysis of cellulose and lignocellulosic materials. Biotechnol Bioeng Symp 1975(5):163-81.
8. Fan LT, Lee YH, Beardmore DH. Mechanism of the enzymatic hydrolysis of cellulose: effect of major structural features of cellulose on enzymatic hydrolysis. Biotechnol Bioeng 1980;22:177-199.
9. Marchessault RH, Sundararajan. PR. Cellulose. NewYork: Academic Press, Inc.; 1993.
10. Stone JE, Scallan. AM. A structural model for the cell wall of water-swollen wood pulp fibres based on their accessibility to macromolecules. Cellulose Chem. Technol. 1968;3:343-358.

11. Stone JE, Scallan AM, Donefer E, Ahlgren E. Digestibility as a simple function of a molecule of similar size to a cellulase enzyme. Adv. Chem. Ser. 1969;95:219-241.
12. Henrissat B, Teeri TT, Warren RA. A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. FEBS Lett 1998;425(2):352-4.
13. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 2002;66(3):506-77, table of contents.
14. Teeri TT. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends Biotechnol 1997;15:160-167.
15. Kleman-Leyer K, Siika-Aho M, Teeri T, Kirk T. The Cellulases Endoglucanase I and Cellobiohydrolase II of Trichoderma reesei Act Synergistically To Solubilize Native Cotton Cellulose but Not To Decrease Its Molecular Size. Appl. Environ. Microbiol 1996;8(62):2883-2887.
16. McHale A, Coughlan MP. The cellulolytic system of Talaromyces emersonii Purification and characterization of the extracellular and intracellular β-glucosidases. Biochimica et Biophysica Acta 1981;662(1):152-159.
17. Iwasaki T, Tokuyasu K, Funatsu M. Determination of Cellulase Activity Employing Glycol Cellulase as a Substrate. J Biochem (Tokyo) 1964;55:30.
18. Li LH, Flora RM, King KW. Individual roles of cellulase components derived from Trichoderma viride. Arch Biochem Biophys 1965;111(2):439-47.
19. Child JJ, Eveleigh DE, Sieben AS. Determination of cellulase activity using hydroxyethylcellulose as substrate. Can J Biochem 1973;51(1):39-43.
20. Miller GL, Blum R, Glennon WE, Burton AL. Measurement of carboxymethylcellulase activity. Anal.Biochem 1960;2:127-132.
21. Nelson N. A Photometric Adaptation of the Somogyi Method for the Determination of Glucose. J. Biol. Chem 1944;153:375-380.
22. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. Colorimetric method for determination of sugars and related substances. Anal. Chem 1956;28:350-356.

23. Wood TM, McCrae SI. The purification and properties of the C1 component of Trichoderma koningii cellulase. Biochem. J. 1972;128:1183-1192.
24. Mandel M, Weber J. Exogluconase activity by microorganisms. Adv. Chem. Ser. 1969;95:391-414.
25. Enari,T.M. Markkanen P. Production of cellulolytic enzymes by fungi. In: T.K.Ghose, A.Fiechter, N.Blakebrough, editors. Advances in biochemical engineering: Springer-verlag; 1977. p. 12.
26. Selby K, Maitland C. The Cellulase of Trichoderma viride. Biochem J 1967;104:716.
27. Wood T. Cellulolytic enzyme system of. Trichoderma. koningii. Biochem J 1968;109:217-227.
28. Leisolab M, Linkob M. Determination of the solubilizing activity of a cellulase complex with dyed substrates. Analytical Biochemistry 1976;70(2):592-599.
29. Wood ,T.M. Mechanisms of cellulose degradation by enzymes from aerobic and anarobic and anaerobic fungi. In: M.P.Coughlan, editor. enzyme systems for lignocellulose degradation. New york: Elsevier applied science; 1989.
30. Ogawa K, Toyama N. Resolution of the *Trichoderma viride* cellulolytic complex. J. Ferment. Technol 1972;45:671-680.
31. Gong CS, Ladisch MR, Tsao TG. Biosynthesis, purification and mode of action cellulases of *Trichoderma reesei*. Adv. Chem. Ser. 1979;181:261-287.
32. Wood TM, McCrae SI. The cellulase of Penicilium pinophilum: Synergism between enzyme component in solubilizing cellulose with special reference to the involvement of two immunologically distinct cellulobiohydrolases. Biochem J 1986;234:93-99.
33. Eriksson KE. Enzyme mechanisms involved in cellulose hydrolysis by the rot fungus *Sporotrichum pulverulentum*. Biotechnol Bioeng 1978;70:317-332.
34. Wood TM. The cellulase of Fusarium solani. Resolution of the enzyme complex. Biochem J 1969;115:457-464.
35. Folan MA, Coughlan MP. The cellulase complex in the culture filtrate of the themophilic fungus, *Talaromyces emersonii*. Int J. Biochem 1979;10:505-510.

36. Bhat KM, Gaikwad JS, Maheshwari R. Purification and characterization of an extracellular B-glucosidase from the thermophilic fungus *Sporotrichum thermophile* and its influence on cellulase activity. J. Gen. Microbiol 1993;139:2825-2832.

37. Bhat KM, Maheshwari R. *Sporotrichum thermophile:* Growth, cellulose degradation and cellulase activity. Appl. Environ. Microbiol 1987;53:2175-2182.

38. Khandke KM, Vithayathil PJ, Murthy SK. Purification of xylanase, B-glucanase, endocellulase, and exocellulase from a thermophilic fungus, *Thermoascus aurantiacus* A-131. Arch Biochem Biophys 1989;274:491-500.

39. Ganju RK, Murthy SK, Vithayathil PJ. Purification and functional characteristic of an endocellulase from *Chaetomium thermophile* var. *coprophile*. Carbohydr. Res. 1990;197:245-255.

40. Schulein M. Enzymatic properties of cellulases from *Humicola insolens*. J Biochem 1997;57:71-81.

41. Wood TM, Wilson CA, McCrae SI, Joblin KN. A highly active extracellular cellulase from the anaerobic rumen fungus *Neocallimastrix frontalis*. FEBS. Microbiol. Lett 1986;34:37-40.

42. Wood TM, Wilson CA. Studies on the capacity of the cellulase of the anaerobic rumen fungus Piromonas communis P to degrade hydrogen bond-ordered cellulose. Appl Microbiol Biotechnol 1995;43(3):572-8.

43. Phillips MW, Gordon GL. Growth characteristics on cellobiose of three different anaerobic fungi isolated from the ovine rumen. Appl Environ Microbiol 1989;55(7):1695-702.

44. Knowles JKC, Teeri TT, Penttila M, Saloheimo M, Harkki A, Lehtovaara P. Molecular Genetics of Fungal Cellulases. In: M.P.Coughlan, editor. Enzyme systems for linocellulose degradation: Elsevier applied science; 1989. p. 51-56.

45. Kubicek CP, Penttila ME. Regulation of production of plant polysaccharide degrading enzymes by Trichoderma. In: Harman GE, Kubicek CP, editors. richoderma and Gliocladium. London: Taylor & Francis Ltd; 1998. p. 49-72.

46. Nogawa M, Goto M, Okada H, Morikawa Y. L-Sorbose induces cellulase gene trancription in the cellulolytic fungus *Trichoderma reesei*. Curr. Genet. 2001;38:329-334.

47. Takashima S, Nakamura A, Hidaka M, Masaki H, Uozumi T. Molecular cloning and expression of the novel fungal beta-glucosidase genes from *Humicola grisea* and *Trichoderma reesei*. J. Biochem 1999;125:728-736.

48. Bhat MK, Bhat S. Cellulose degrading enzymes and their potential industrial applications. Biotechnol Adv 1997;15(3-4):583-620.

49. Himmel ME, Ruth MF, Wyman CE. cellulase for commodity products from cellulosic biomass. Curr. Opin. Biotechnol 1999;10:358-364.

50. Beguin P, Aubert JP. The biological degradation of cellulose. FEBS Microbiol. Rev. 1993;13:25-58.

51. Mandel M. Application of cellulases. Biochem. Soc. Trans. 1985;13:414-415.

52. Valjamae P. The Kinetics of cellulose enzymatic hydrolysis. Uppsala: Uppsala University; 2002.

53. Dienes D, Egyhazi A, Reczey K. Treatment of recycled fiber with *Trichoderma* cellulase. Industrial Crop and products 2004;20:11-21.

54. Maheswari R, Bharadwaj G, Bhat MK. Thermophilic Fungi: their physiology and enzyme. Microbiol Mol Biol Rev 2000;64(3):461-488.

55. Hong J, Tamaki H, Akiba S, Yamamoto K, Kumagai H. Cloning of a gane encoding a highly stable endo-B-1,4-Glucanase from *Aspergillus niger* and its expression in yeast. Journal of Bioscience and Bioengineering 2001;92(5):434-441.

56. Blain JA. Industrial enzyme production. In: Smith JE, editor. The filamentous fungi. London: Edward Arnold; 1975.

57. Hayashida S, Yoshioka H. The role of carbohydrate moiety on thermostability of cellulases from Humicola insolens YH-8. Agric. Biol. Chem 1979;44(3):481-487.

58. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 1991;280 (Pt 2):309-16.

59. Henrissat B, Bairoch A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 1993;293 (Pt 3):781-8.

60. Henrissat B, Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. Biochem J 1996;316 (Pt 2):695-6.

61. Henrissat B. A new cellulase family. Mol Microbiol 1997;23(4):848-9.

62. Ogata K, Nishikawa H, Ohsugi M. A yeast capable of utilizing methanol. Agric. Biol. Chem 1969;33:1519-1520.

63. Veenhuis M, van Dijken JP, Harder W. The significance of peroxisomes in the methabolism of one-carbon compounds in yeast. Adv. Microb. Physiol 1983;24:1-82.

64. Waterham HR, Russell KA, Vries YD, Cregg JM. Perixisomal targeting import, and assembly by alcohol oxidase in *Pichia pastoris*. J.Cell.Biol 1977;139(6):1419-1431.

65. Cereghino JL, Cregg JM. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS. Microbiol.Rev 2000;24:45-66.

66. Tschopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR. Expression of *LacZ* gene from two methanol- regulated promoters in *Pichia pastoris*. Nucleic. Acid. Res 1987;15:3859-3876.

67. Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. Mol. Cell. Biol 1985;5:1111-1121.

68. Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol. Cell. Biol 1989;9:1316-1323.

69. Couderc R, Baratti J. Oxidation of methanol by yeast *Pichia pastoris*: Purification and propoties of alcohol oxidase. Agric. Biol. Chem 1980;44:2279-2289.

70. Grinna LS, Tschopp JF. Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast *Pichia pastoris*. Yeast 1989;5:107-115.

71. Tschopp JF, Svelow G, Kosson R, Craig W, Grinna LS. High-level secretion of glycosylated invertase in the methyllotrophic yeast, *Pichia pastoris*. Bio/Technology 1987;5:1305-1308.

72. Kukuruzinska MA, Bergh MLE, Jackson BJ. Protein glycosylation in yeast. Ann. Rev. Biochem 1987;59:915-944.

73. Sreekrishna K, Kropp KE. *Pichia pastoris*. In: Wolf K, editor. Nonconventional Yeasts in Biotechnology: Springer-Verlag Berlin Heidelberg; 1996.

74. Clare JJ, Rayment FB, Balantine SP, Sreekrishna K, Romanos MA. High-level expression of Tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem repeat intergration of the gene. Bio/Technology 1991;9(5):455-460.

75. Seigel RS, Brierley RA. Methylotrophic yeast *Pichia pastoris* produced high-cell density fermentration with high cell yield as vehicle for recombinant protein production. Biotech. Bioeng 1989;34:403-404.

76. Cregg JM, Barringer KJ, Hessler AY, Madden KR. *Pichia pastoris* as a host system for the production of foreign protein. Mol. Cell. Biol 1985;5:3376-3385.

77. Cregg JM, Russell KA. Transformation Methods. Mol. Biol 1998;103:27-39.

78. Koutz PJ, Davis GR, Stillman CA, Barringer KJ, Cregg JM, Thill GP. Structural comparison of the *Pichia pastoris* alcohol oxidase genes. Yeast 1989;5:167-177.

79. Heslot H, Gaillardin C. Molecular biology and genetic engineering of yeasts. Florida: CRC Press; 1992.

80. Cregg JM, Madden KR. Use of site -specific recombination to regenerate selectable markers. Mol. Gen. Genet 1989;219:320-323.

81. Higgin DR, Busser K, Comiskey J, Whittier PS, Purcell TJ, Hoeffler JP. Small vector for expression based on dominant drug resistance with direct multicopy selection. Methods Mol. Biol 1998;103:41-53.

82. Cregg JM, Madden KR. Developement of yest transformation systems and construction of methanol-utilization-defective mutants of *Pichia pastoris* by gene disruption. In: Stewart GG, Russell I, Klein RD, Hiebsch RR, editors.

Biological Research on Industial Yeasts. Boca Raton, FL.: CRC Press; 1987. p. 1-18.

83. Cregg JM, Tschopp JF, Stillman CA, Siegel R, Akong M, Craig WS, et al. High level expression and efficient assembly of hepatitis B surface antigen in the methylotrophic yeast, *Pichia pastoris*. Bio/Technology 1987;5:479-485.
84. Chiruvolu V, Cregg JM, Meagher MM. Recombinant protein production in an alcohol oxidase-defective strain of *Pichia pastoris* in fed-batch fermentation. Enzyme Microb. Technol 1997;21:277-283.
85. Brierley RA. Secretion of recombinant human insulin-like growth factor I (IGF-1). Methods Mol. Biol 1998;103:149-177.
86. White CE, Hunter MJ, Meininger DP, White LR, Komives EA. Large-scale expression, purification and characterization of small fragments of thrombomodulin:the roles of the sixth domain and of methionine 388; 1995.
87. Thill GP, Davis GR, Stillman GR, Holtz C, Brierly R, Engel M, et al. Positive and negative effects of multi-copy intergrated expression vectors on protein expression in *Pichia pastoris*. In: Proc 6 th Int Symp on Genetic of Industial Microorganisms. Strasbourg; 1990. p. 477-490.
88. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM. Isolation of *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 1997;186:37-44.
89. Shen S, Sulter G, Jeffries TW, Cregg JM. A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. Gene 1998;216:93-102.
90. Sear IB, O'Connor J, Rossanase OW, Glick BS. A versatile set of vectors for constitutive and regulated gen expression in *Pichia pastoris*. Yeast 1998;14:783-790.
91. Conesa A, Punt PJ, Luijk Nv, Hondel CAMJJvd. The Secretion Pathway in Filamentous Fungi: A Biotechnological View. Fungal Genetics and Biology 2001;33:155–171.
92. Nelson DL, Cox MM. Lehninger Principles of Biochemistry. 3 ed: Worth; 2000.

93. Kurjan J, Herskowitz I. Structure of a yeast pheromone gene (MF?): A putative ?-factor precursor contains four tandem copies of mature ?-factor. Cell 1982;30(3):933-943.

94. Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, et al. Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. 1984;81:4642-4646.

95. Goochee CF, Gramer MJ, Andersen DC, Bahr JB, Rasmussen JR. The oligosaccharides of glycoproteins: bioprocess factors affecting oligosaccharide structure and their effect on glycoprotein properties. Biotechnology (NY) 1991;9:1347-1355.

96. Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr. Opin. Biotechnol 2002;13:329–332.

97. Heimo H, Palmu K, Suominen I. Expression in *Pichia pastoris* and purification of *Aspergillus awantori* glucoamylase catalytic domain. Protein Expr. Purif 1997;11:304.

98. Juge N, Andersen JS, Tull D, Roepstorff P, Svensson B. Overexpression, purification, and characterization of recombinant barley alpha-amylase 1 and 2 secreted by the methylotrophic yeast *Pichia pastoris*. Protein Expr. Purif 1996;8:204-214.

99. Tsujikawa M, Okabayashi K, Morita M, Tanabe T. Secretion of variant of human single-chain urokinase-type plasminogen activator without an N-glycosylation site in the methylotrophic yeast, *Pichia pastoris* and characterization of the secreted product. Yeast 1996;12:541-553.

100. Ruitenberg KM, Gilkerson JR, Wellington JE, Love DN, Whalley JM. Equine herpesvirus 1 glycoprotein D expressed in *Pichia pastoris* is hyperglycosylated and elicits a protective immune response in the mouse model of EHV-1 disease. Virus Res 2001;79(1-2):125-35.

101. Paramasivam M, Saravanan K, Uma K, Sharma S, Singh TP, Srinivasan A. Expression, purification, and characterization of equine lactoferrin in *Pichia pastoris*. Protein Expr Purif 2002;26(1):28-34.

102. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J. Mol. Biol 1990;215:403-410.

103. Madden T. The BLAST Sequence Analysis Tool. In; 2003. http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.610

104. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22(22):4673-80.

105. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. Trends Biochem Sci 1998;23(10):403-5.

106. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25(24):4876-82.

107. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 2003;31(13):3381-5.

108. von Heijne G. Signal sequences. The limits of variation. J Mol Biol 1985;184(1):99-105.

109. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 1997;10(1):1-6.

110. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340(4):783-95.

111. Inoue H, Nojima H, Okayana H. High efficiency transformation of *Escherichia coli* with plasmids. Gene 1990;96:23-28.

112. Huttemann M. Partial heat denaturation step during reverse transcription and PCR screening yield full-length 5'-cDNAs. Biotechniques 2002;32(4):730-736.

113. Bollag DM, Rozycki MD, Edelstein SJ. Protein Concentration Determination. In: Protein Methods: A JOHN WILEY & SONS; 1996.

114. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem 1959;31(3):426-428.

115. van Helden J, del Olmo M, Perez-Ortin JE. Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. Nucleic Acids Res 2000;28(4):1000-10.

116. Humphrey T, Birse CE, Proudfoot NJ. RNA 3' end signals of the S.pombe ura4 gene comprise a site determining and efficiency element. Embo J 1994;13(10):2441-51.

117. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol 1997;7(5):637-44.

118. Copeland RA. Kinetics of single-substrate enzyme reactions. In: Enzyme: a practical introduction to structure, mechanism, and data analysis. 2 ed. New York: John Wiley & Sons; 2000.

119. Pandey A, P.Selvakumar, R.Soccol C, Nigam P. Solid state fermentation for the production of industrial enzymes.

120. Bioprocess and product developement. In: DBT Annual report; 2004-2005.

121. Murashima K, Nishimura T, Nakamura Y, Koga J, Moriya T, Sumida N, et al. Purification and characterization of new endo-1,4-B-D-glucanase from *Rhizopus oryzae*. Enzyme Microb. Technol 2002;30:319-326.

122. Davies GJ, Tolley SP, Henrissat B, Hjort C, Schulein M. Structure of oligosaccharide-bound forms of the endoglucanase V from *Humicula insolens* at 1.9 A resolution. Biochemistry 1995;34:16210-16220.

123. Takashima S, Nakamura A, Hidaka M, Masaki H, Uozumi T. Cloning, sequencing and expression of the cellulase genes of *Humicula grisea* var. *thermoidea*. Journal of biotechnology 1996;50:137-147.

124. Wang H, Jones RW. Cloning, characterization and functional expression of an endoglucanase- encoding gene from the phytopathogenic fungus Macrophomina phaseolina. Gene 1995;158:125-128.

125. Thomas L, Crawford DL. Cloning of clustered Streptomyces viridosporus T7A lignocellulose catabolism genes encoding peroxidase and endoglucanase and their extracellular expression in Pichia pastoris. Can. J. Microbiol./Rev. can. microbiol 1998;44(4):364-372.

126. Linder M, Teeri TT. The roles and function of cellulose-binding domains. Journal of biotechnology 1997;57:15-28.

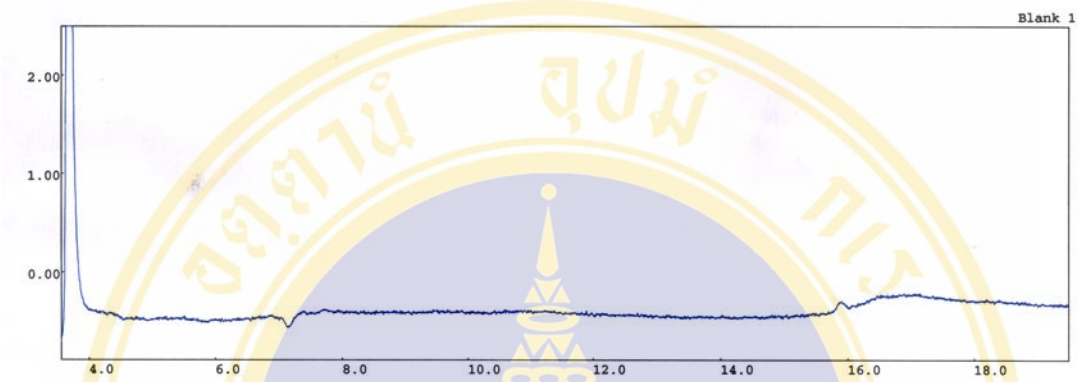
127. Bretthauer RK, Castellino FJ. review: Glycosylation of *Pichia pastoris*-derived protein. Biotechnol. Appl. Biochem 1999;30:193-200.

128. Muerivuori H, Siegler KM, Sands JA, Montenecourt BS. Regulation of cellulaes biosynthesis and secretion in fungi. Biochemical Society Transactions 1985;13:411.

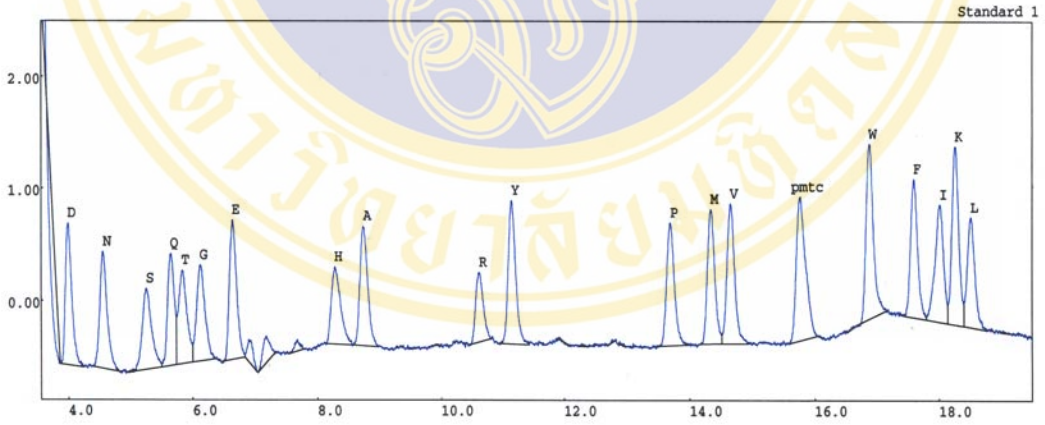
APPENDIX

Appendix 1. N-terminal sequences of purified 55 kDa protein from transformant clones containing endoglucanase cDNA with its own leader sequence

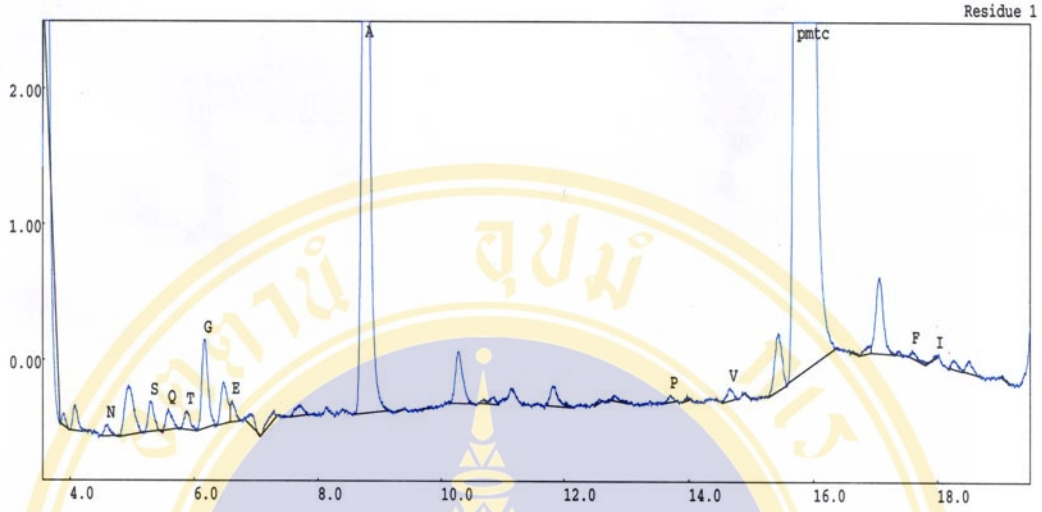
Blank



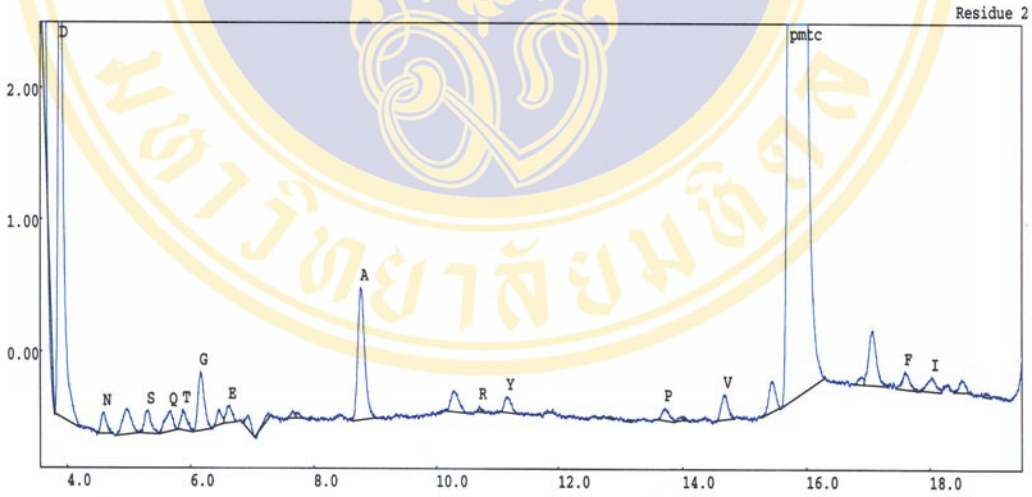
Standard



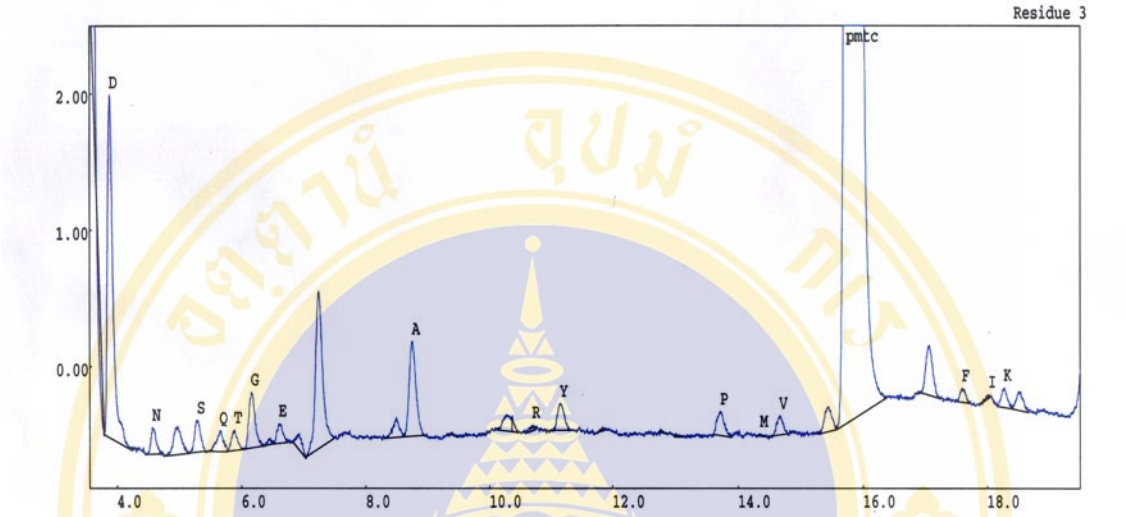
Residue 1



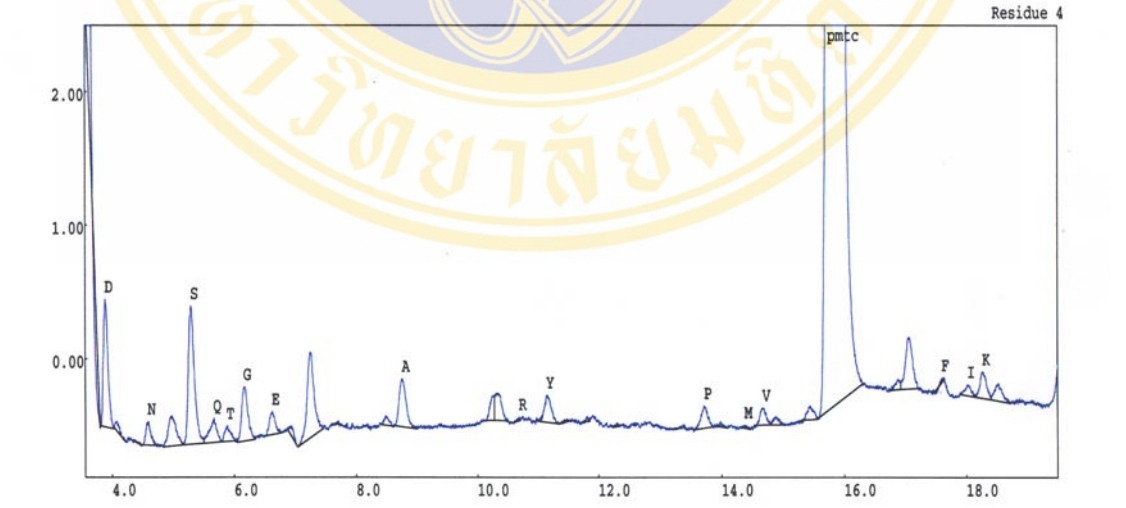
Residue 2



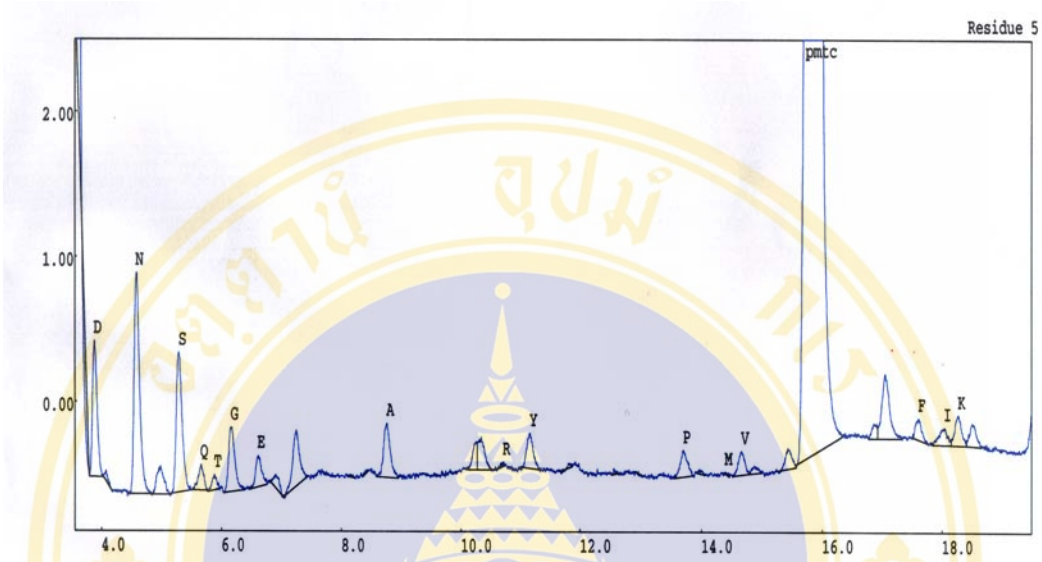
Residue 3



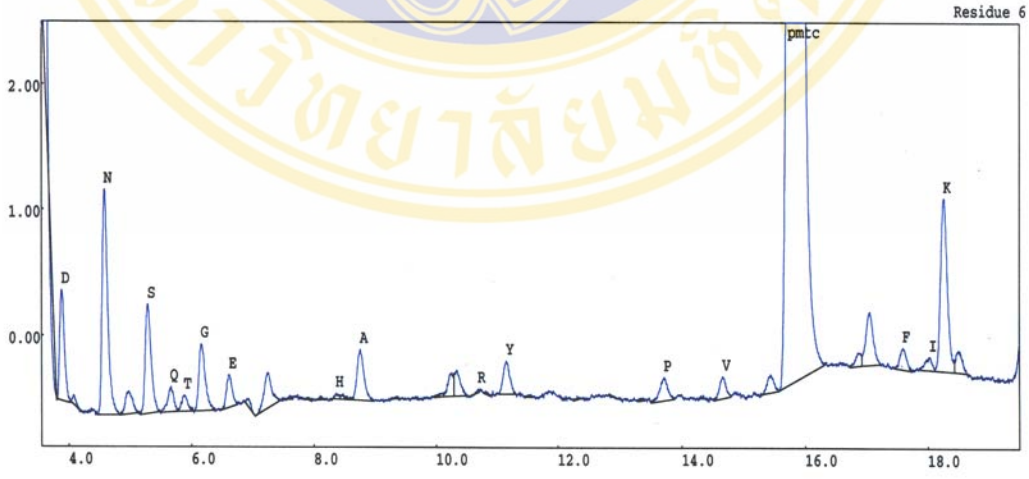
Residue 4



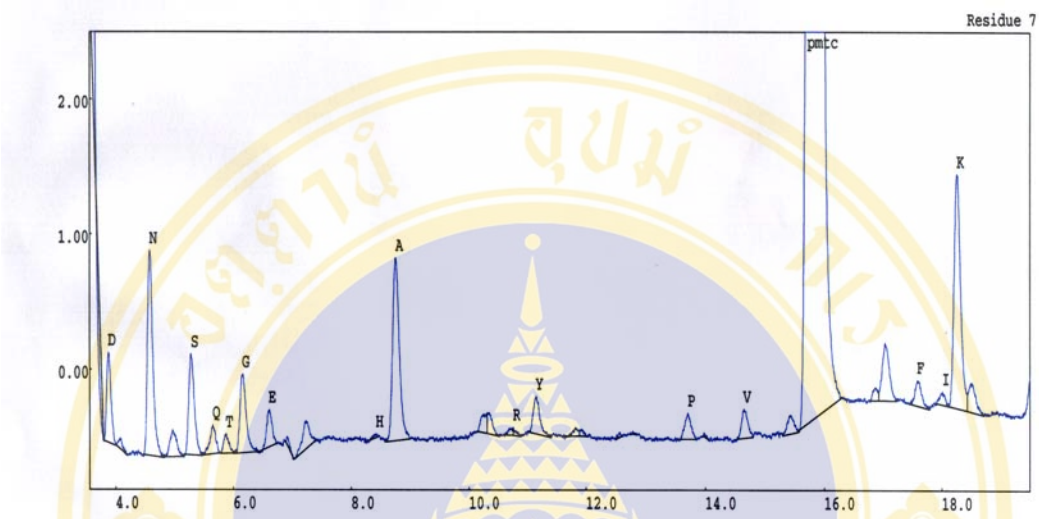
Residue 5



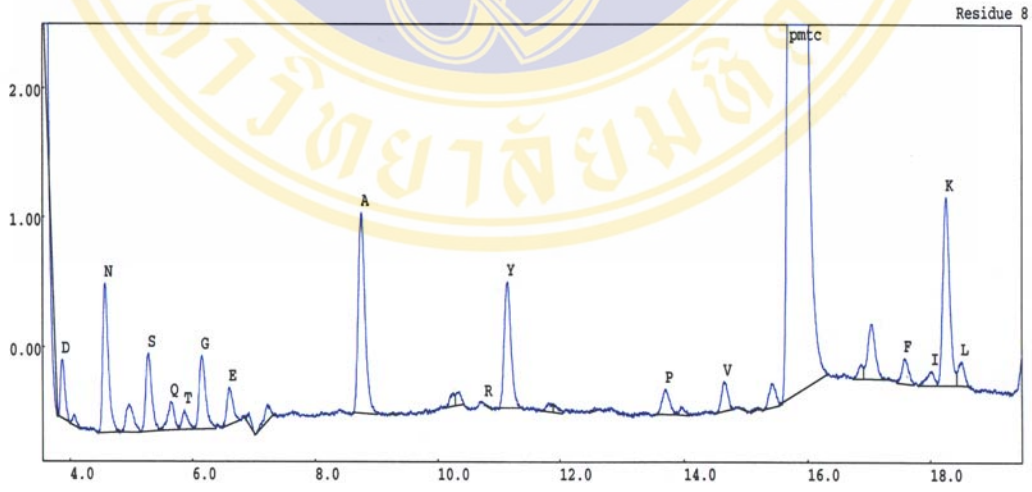
Residue 6



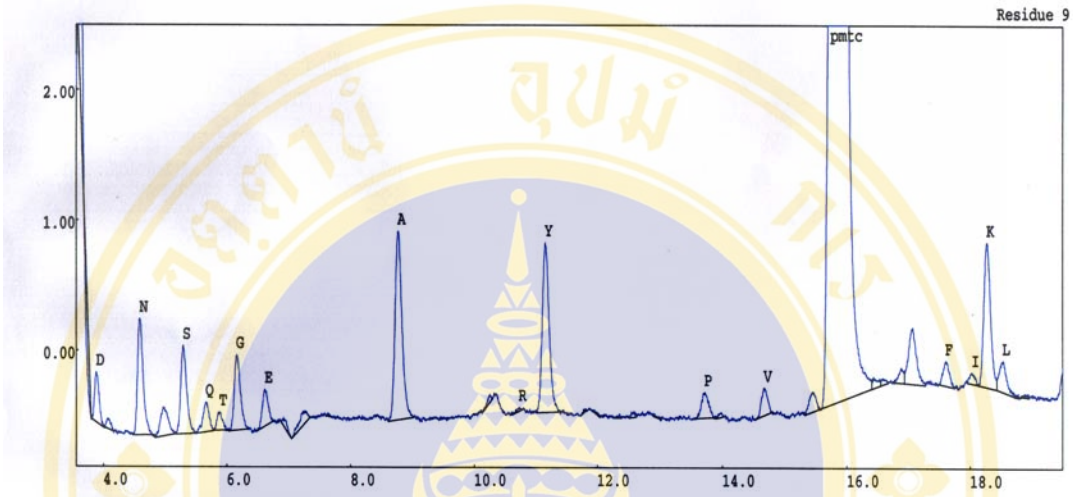
Residue 7



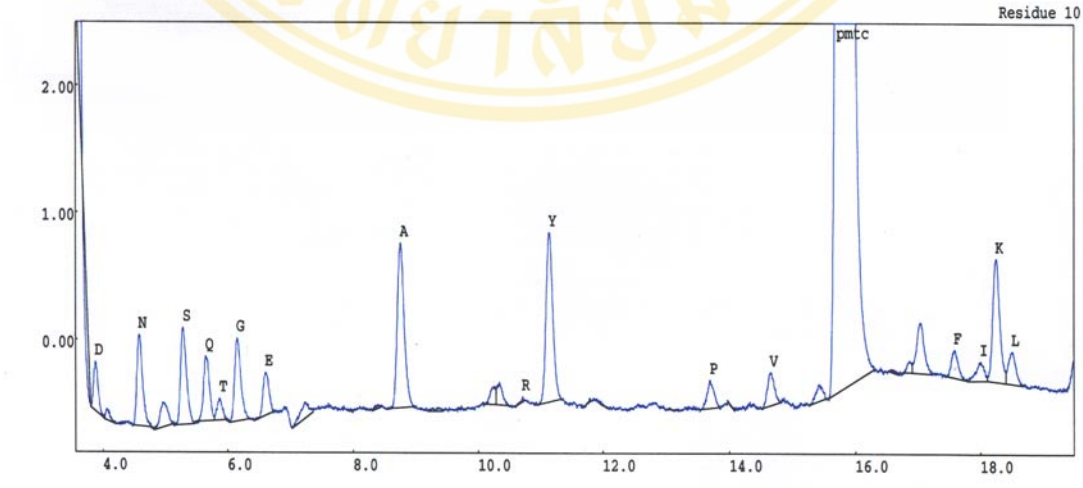
Residue 8



Residue 9

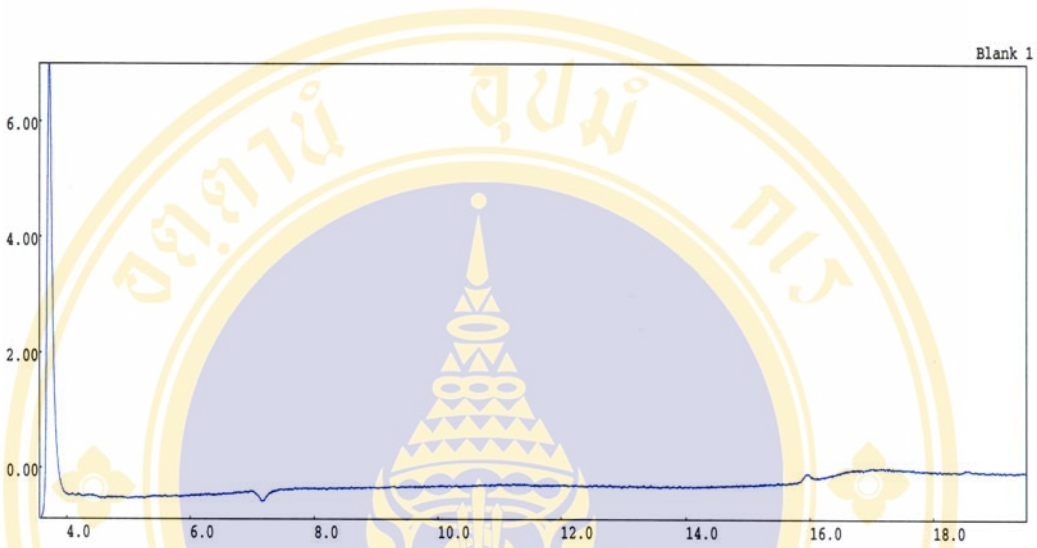


Residue 10

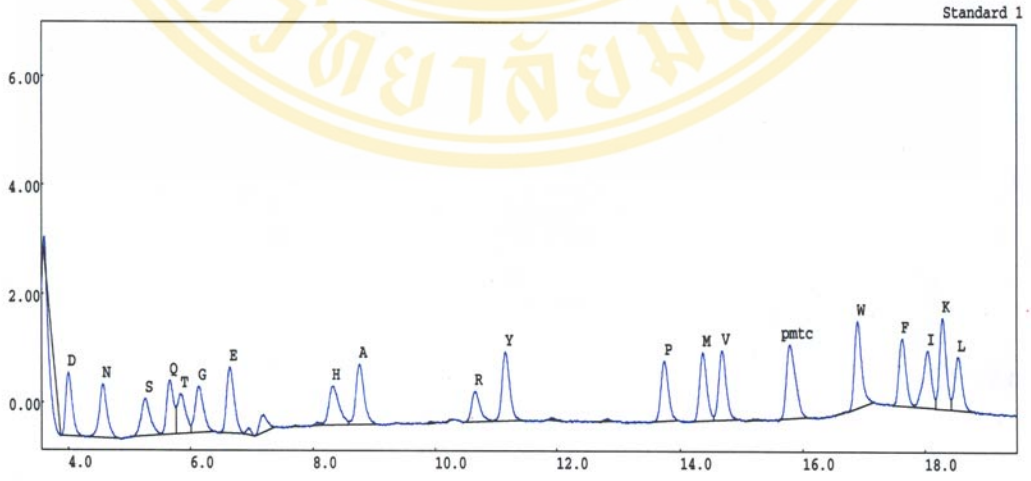


Appendix 2. N-terminal sequences of purified 55 kDa protein from transformant clones containing endoglucanase cDNA without its own leader sequence

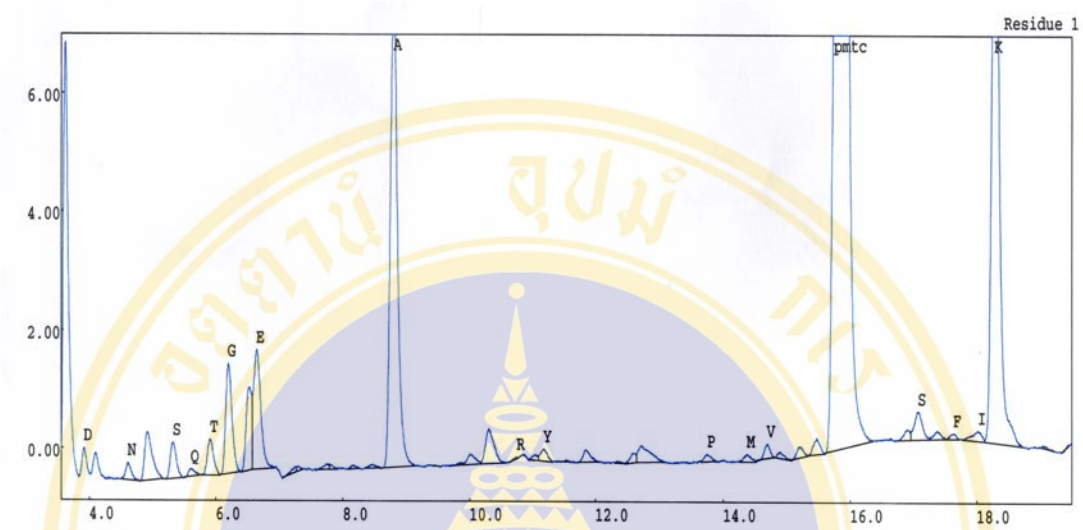
Blank



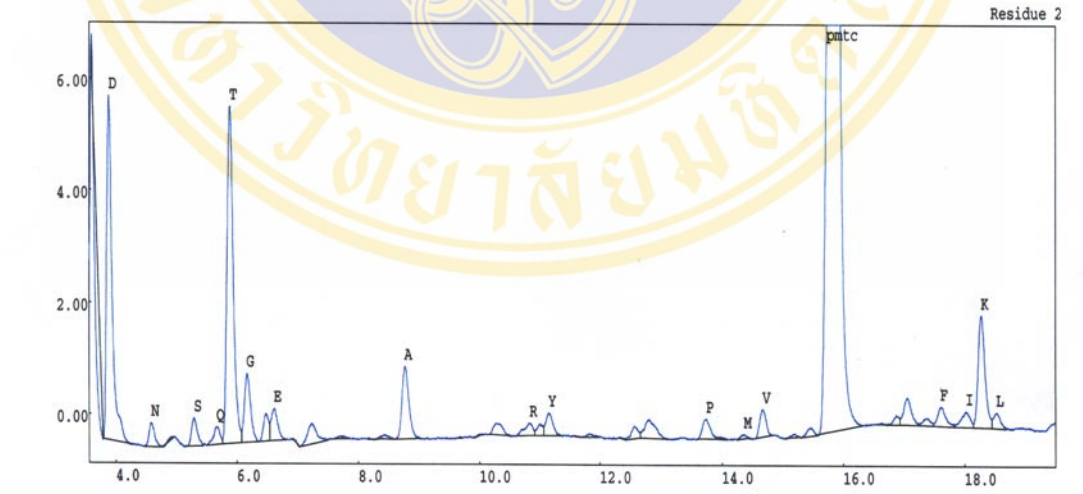
standard



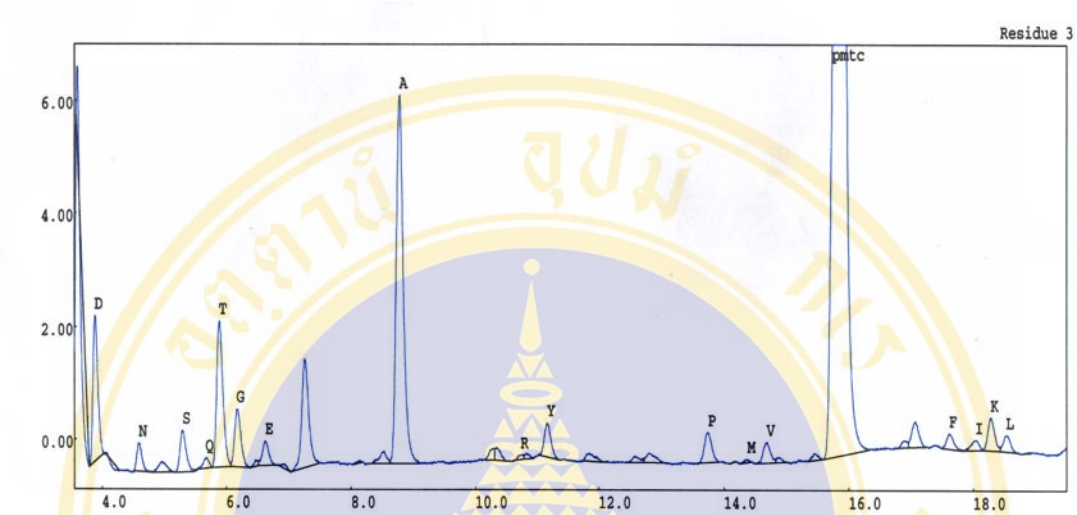
Residue 1



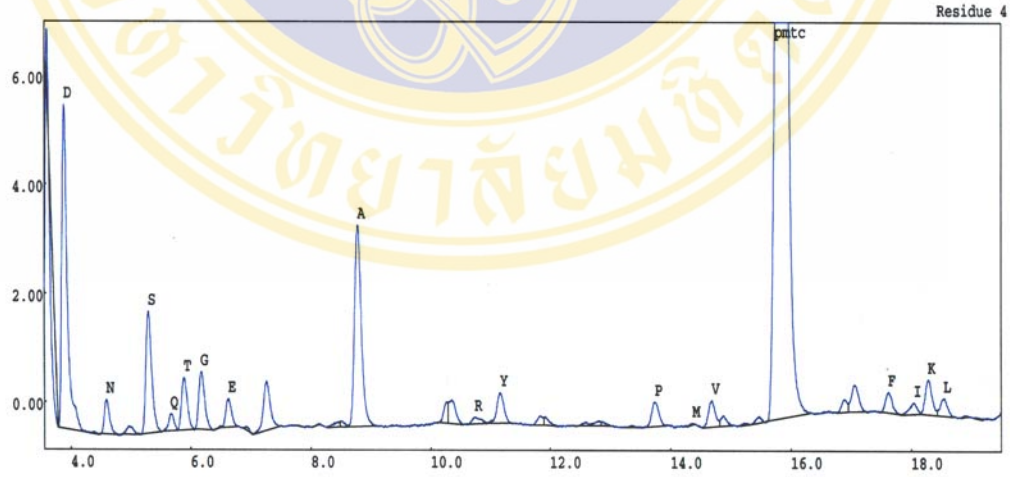
Residue 2



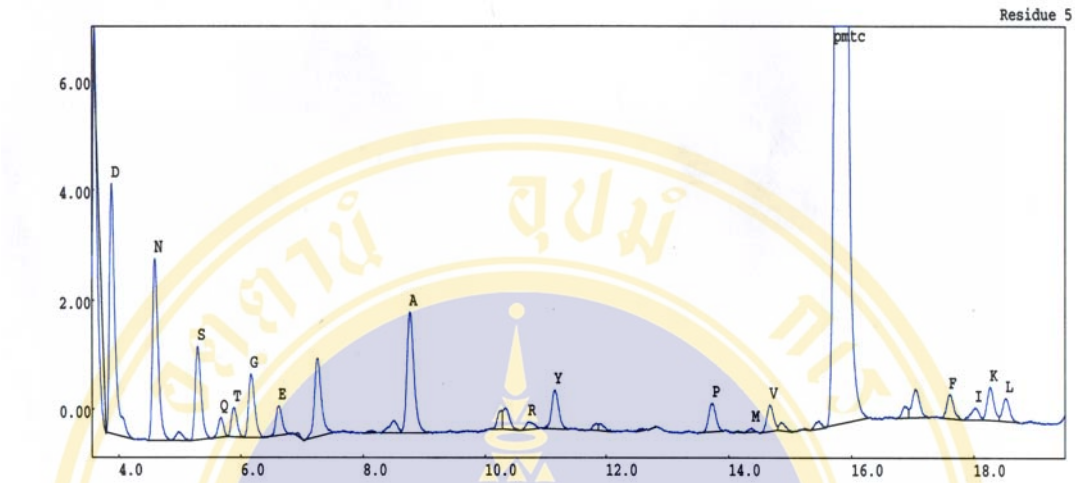
Residue 3



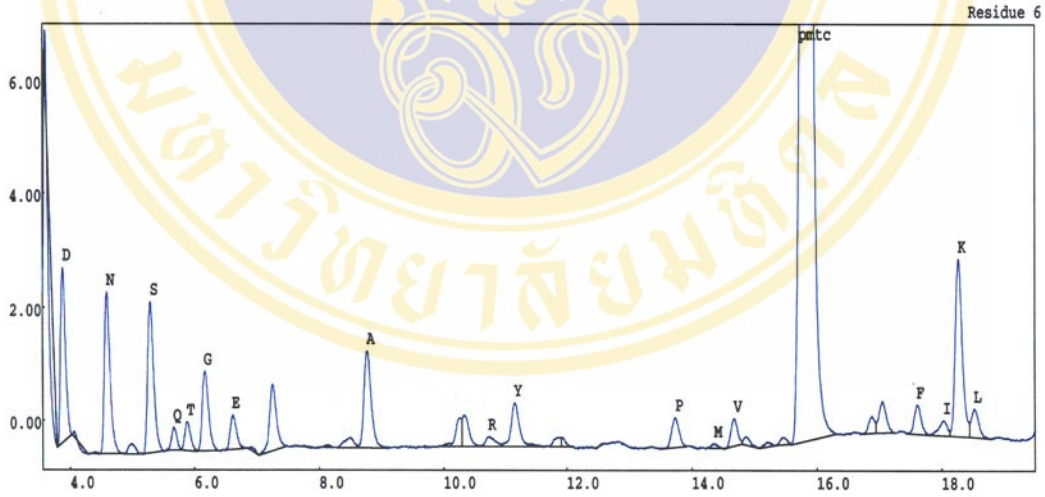
Residue 4



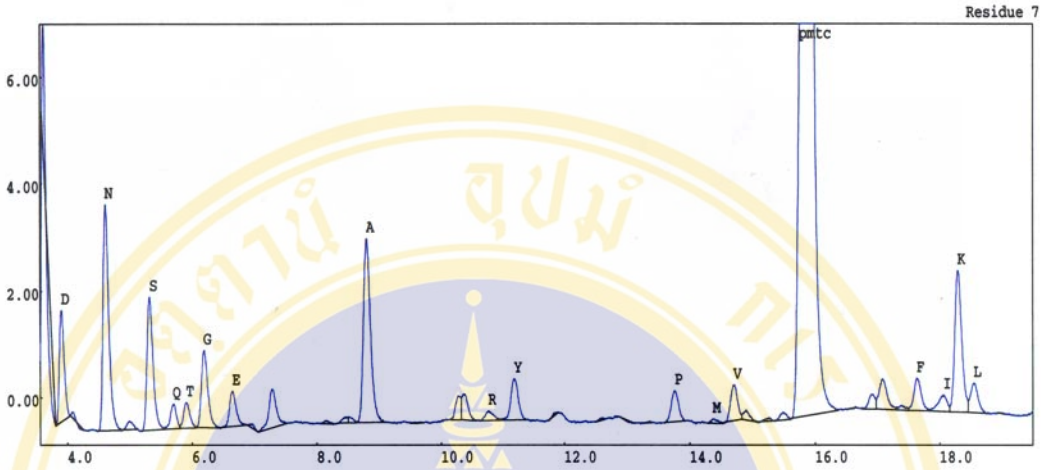
Residue 5



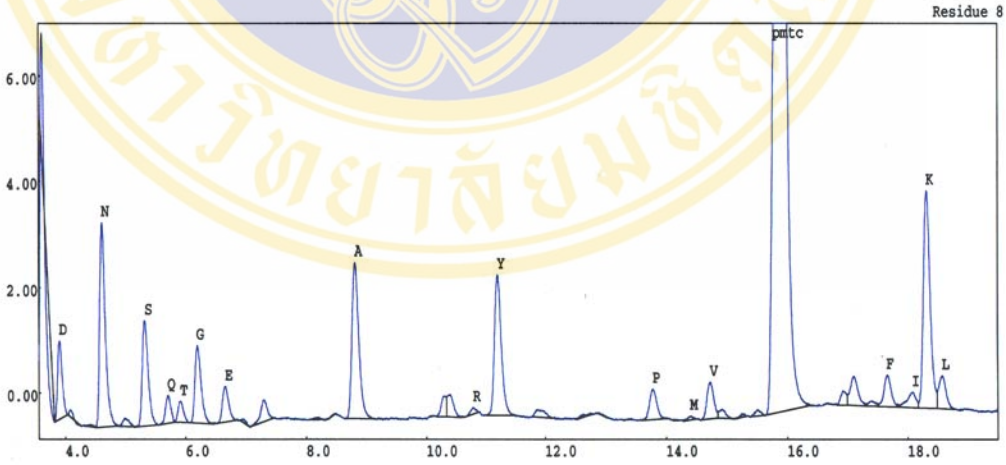
Residue 6



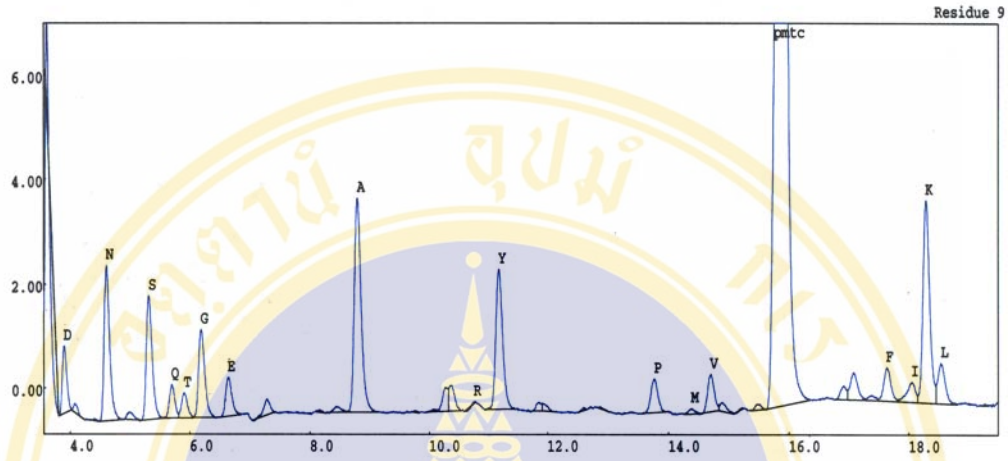
Residue 7



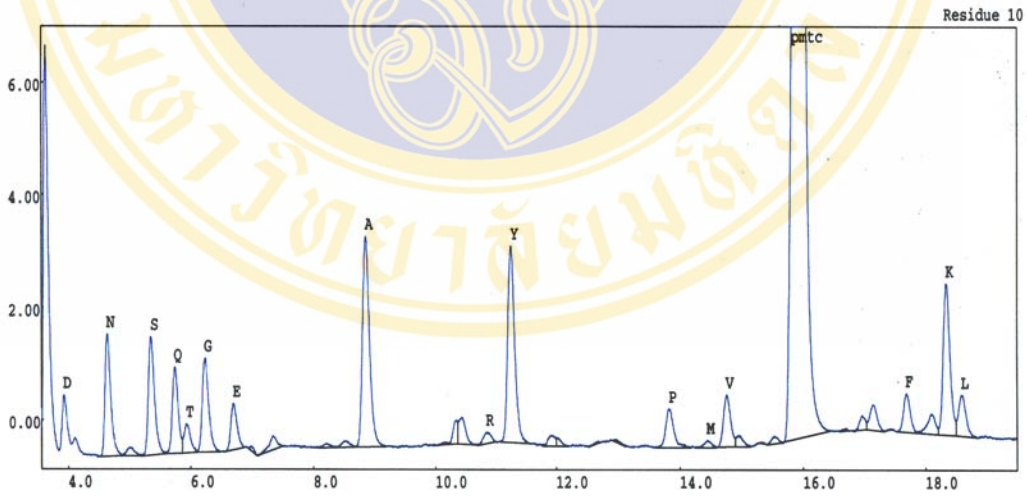
Residue 8



Residue 9

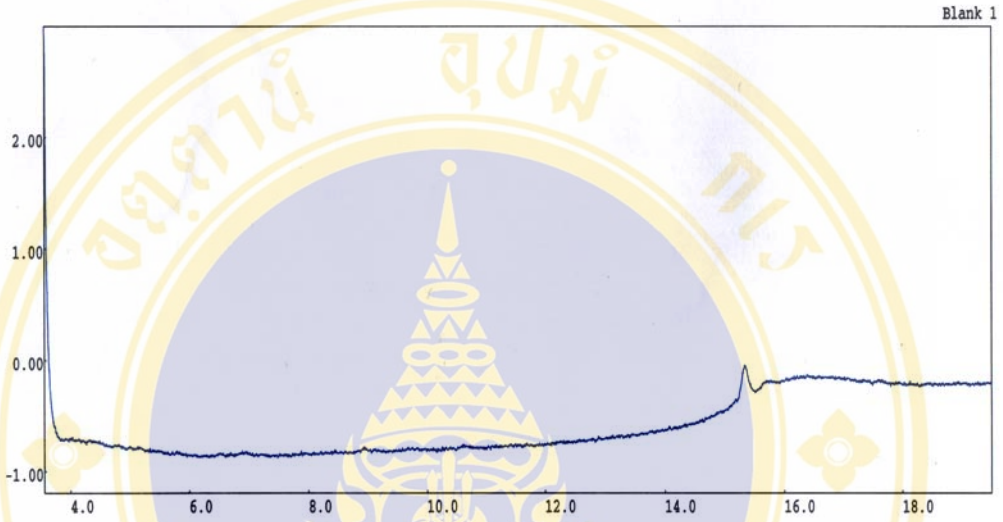


Residue 10

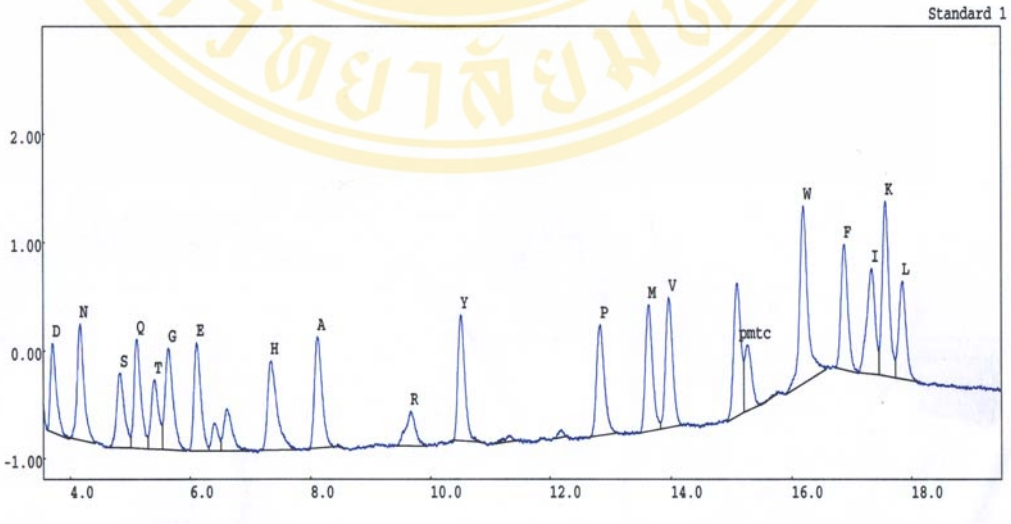


Appendix 3. N-terminal sequences of 30 kDa protein from transformant clones containing endoglucanase cDNA without its own leader sequence

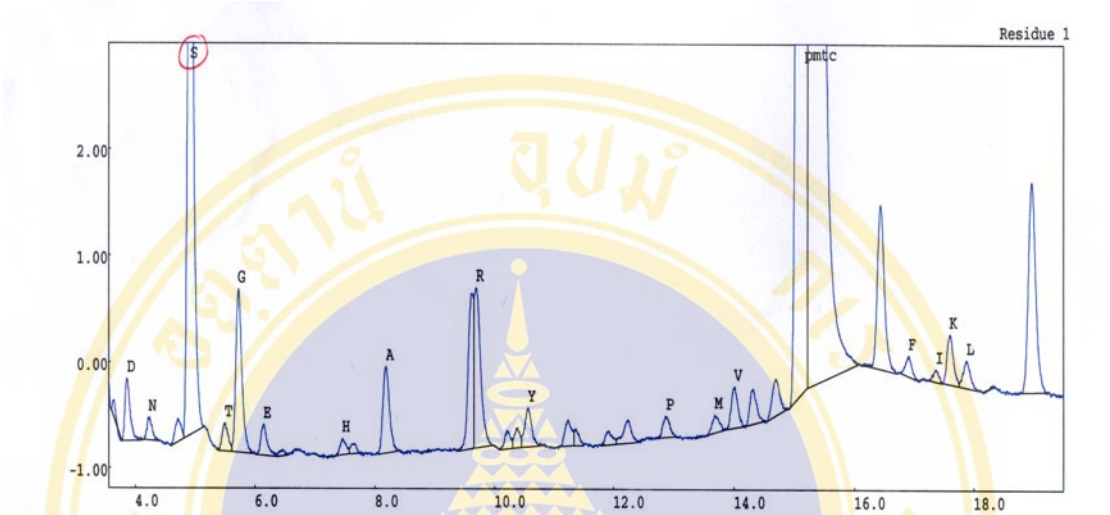
Blank



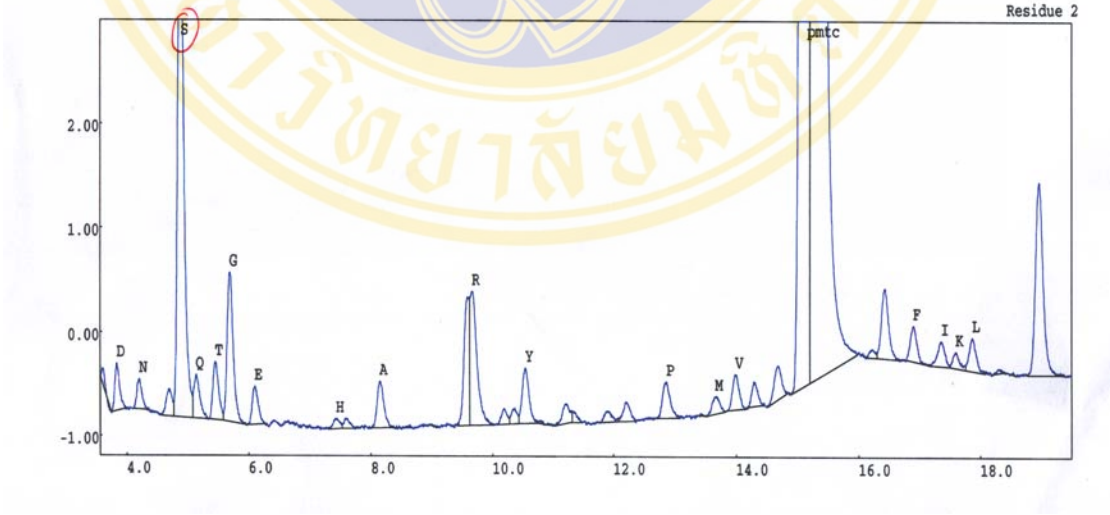
Standard



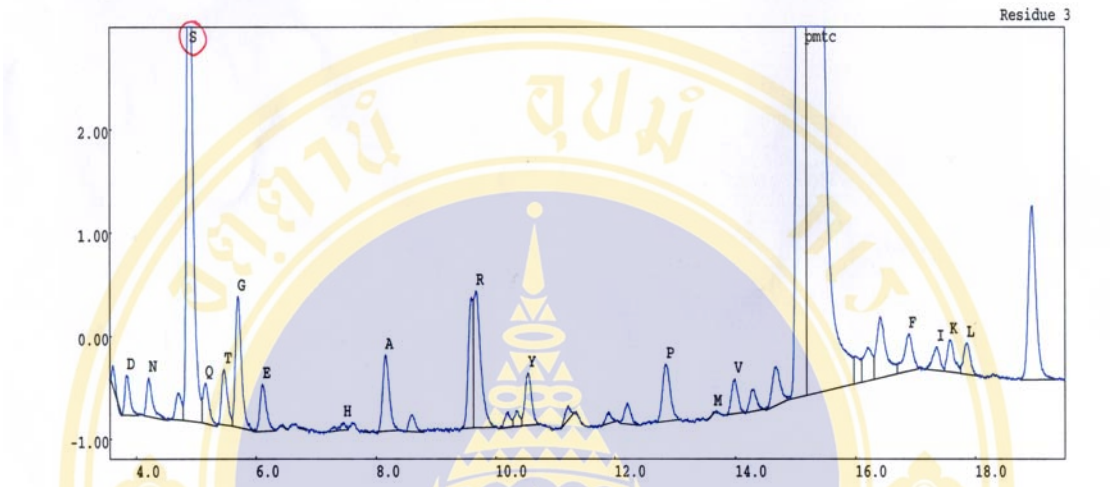
Residue 1



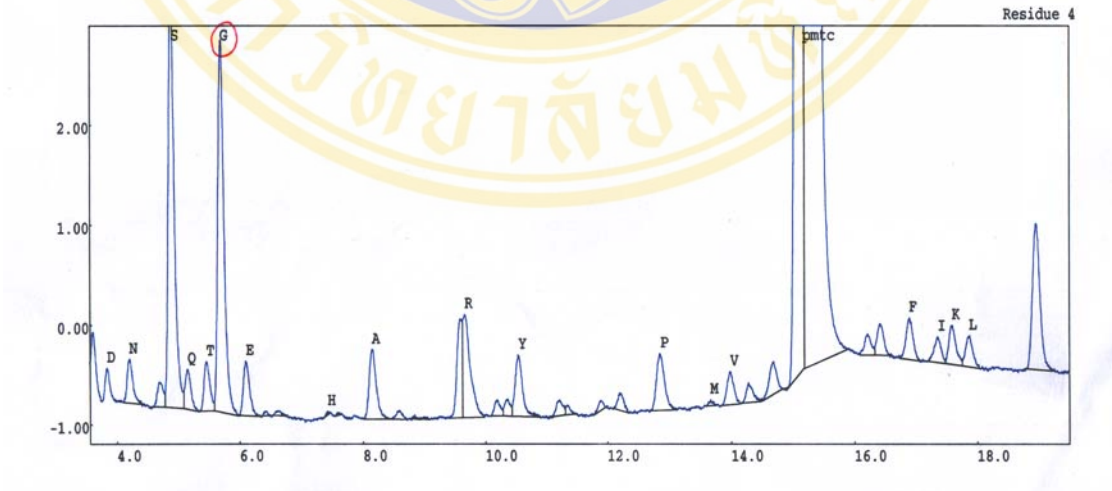
Residue 2



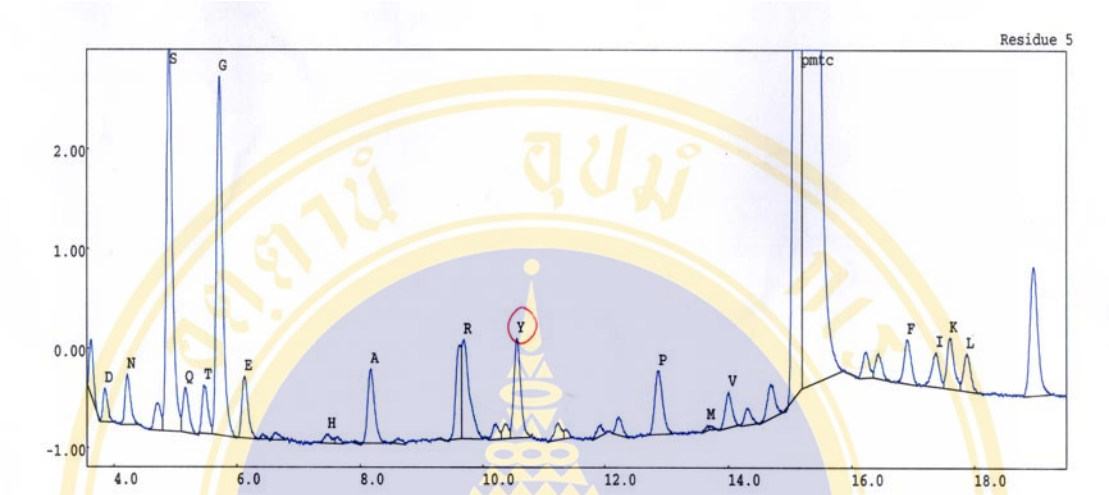
Residue 3



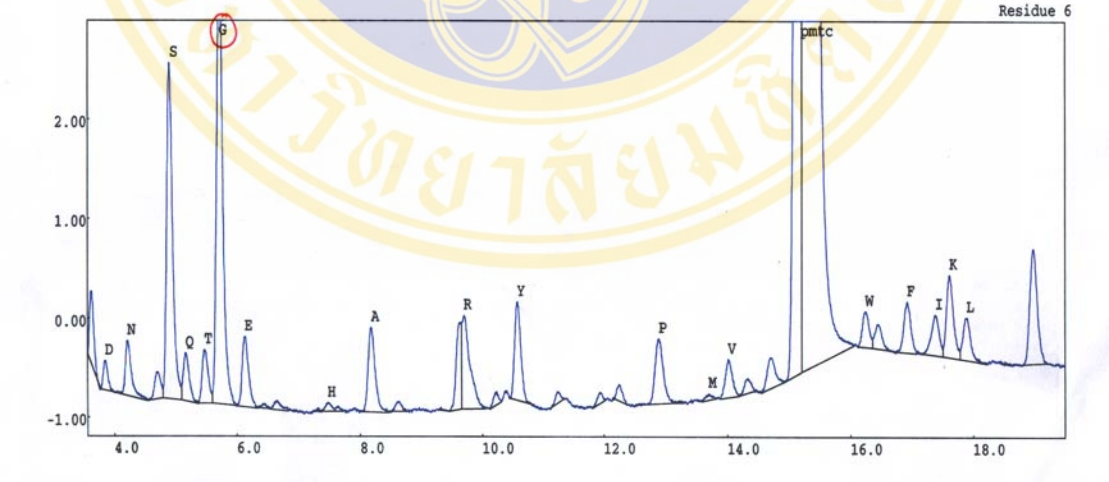
Residue 4



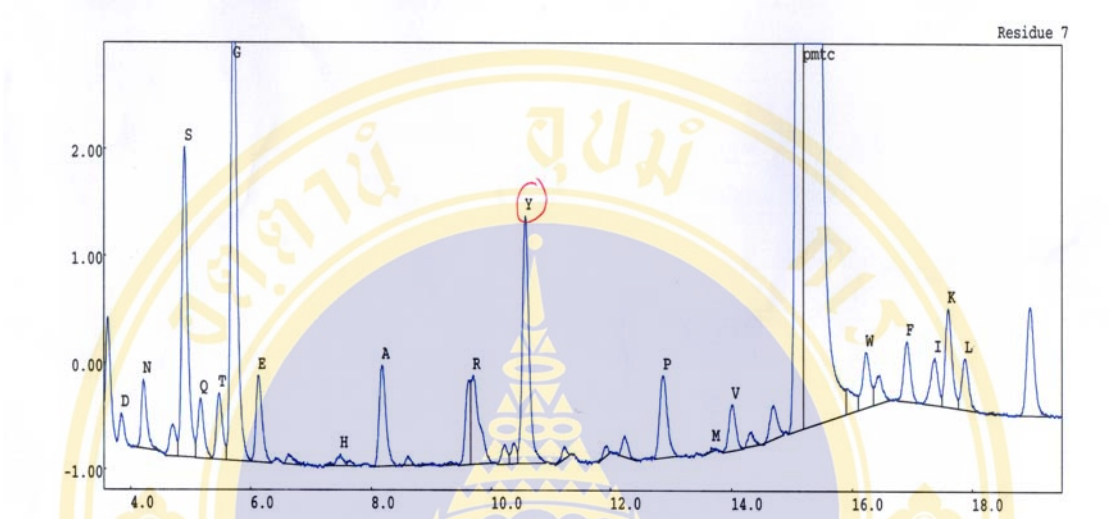
Residue 5



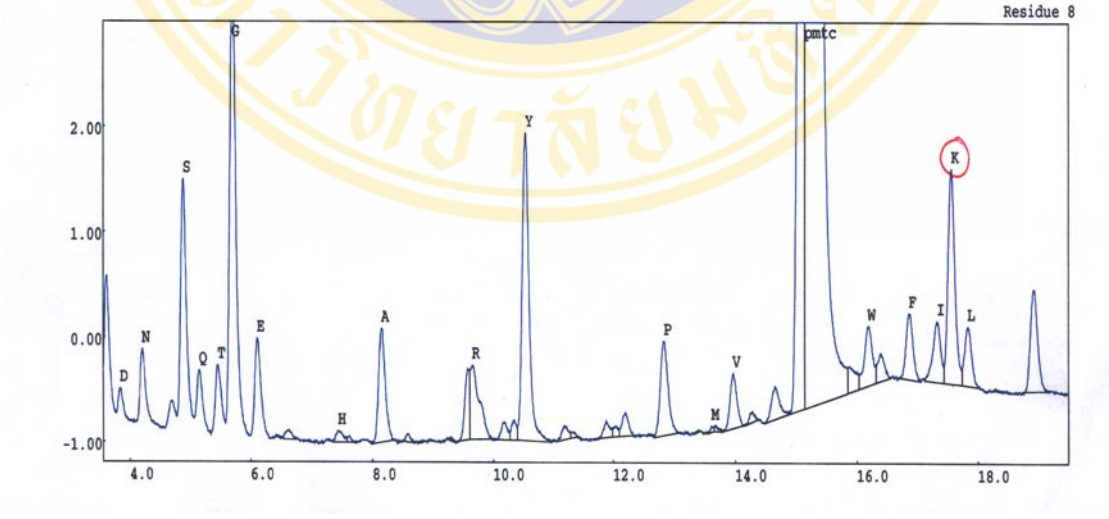
Residue 6



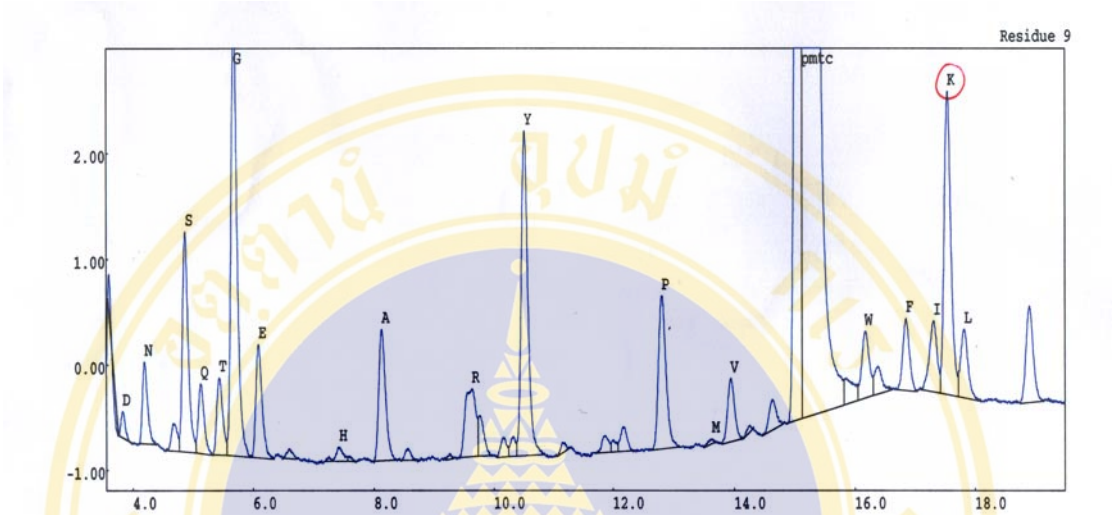
Residue 7



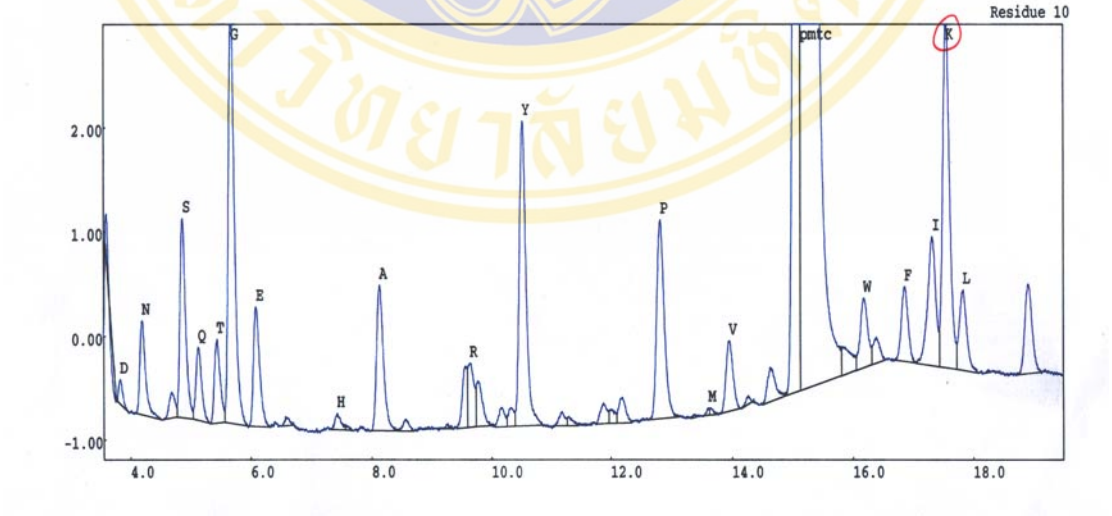
Residue 8



Residue 9



Residue 10



BIOGRAPHY

NAME	Miss Benjamaporn Wonganu
DATE OF BIRTH	27 July 1980
PLACE OF BIRTH	Roi-Et, Thailand
INSTITUTIONS ATTENDED	Khonkean University, 1999-2002: Bachelor of Science (Biology) Mahidol University, 2003-2005: Master of Science (Molecular Genetics and Genetic Engineering)
RESEARCH GRANT	Development and Promotion of Science and Technology Talent Project (DPST) National Center for Genetic Engineering and Biotechnology (BIOTEC)
HOME ADDRESS	18 Tavapiban Rd. Amphur Muang district, Roi-et, Thailand 57210 Tel. +66-(0)65304567 E-mail: fonja_benja@yahoo.com