

**CHARACTERIZATION OF RICE GLYCOSIDE
HYDROLASE FAMILY I - CLUSTER AT/OS 6 ENZYMES**



**A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Biochemistry
Suranaree University of Technology
Academic Year 2016**

การศึกษาคุณลักษณะของเอนไซม์ glycoside hydrolase family I - cluster
At/Os 6 ในข้าว



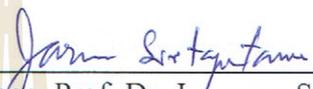
วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาชีวเคมี
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2559

CHARACTERIZATION OF RICE GLYCOSIDE HYDROLASE

FAMILY I - CLUSTER AT/OS 6 ENZYMES

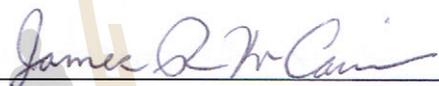
Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Master's Degree.

Thesis Examining Committee



(Assoc. Prof. Dr. Jarawan Siritapetawee)

Chairperson



(Prof. Dr. James R. Ketudat-Cairns)

Member (Thesis Advisor)



(Assoc. Prof. Dr. Mariena Ketudat-Cairns)

Member



(Dr. Sakesit Chumnarnsilpa)

Member



(Prof. Dr. Santi Maensiri)

Vice Rector for Academic Affairs
and Internationalisation



(Prof. Dr. Santi Maensiri)

Dean of Institute of Science

ลิงก์ ฐานข้อมูล : การศึกษาคุณลักษณะของเอนไซม์ glycoside hydrolase family I - cluster At/Os 6 ในข้าว (CHARACTERIZATION OF RICE GLYCOSIDE HYDROLASE FAMILY I - CLUSTER AT/OS 6 ENZYMES). อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.เจมส์ เกตุทัต-คาร์นส์, 112 หน้า.

เอนไซม์เบตาไกลูโคซิเดสของพืชส่วนใหญ่จัดอยู่ในตระกูลที่ 1 ของไกลโคไซด์ไฮโดรเลส ซึ่งประกอบไปด้วยเอนไซม์ทรานสกลูโคซิเดส อย่างไรก็ตาม เอนไซม์ไกลโคไซด์ไฮโดรเลส กลุ่มที่ 1 ยังมีความสามารถในการย้ายหมู่น้ำตาลกลูโคสให้กับสารใหม่ๆ ในพืชต่างๆ จากความคล้ายคลึงของลำดับเบสกรดอะมิโน Os5BGlu19 ของข้าวถูกจัดอยู่ในกลุ่มของไกลโคไซด์ไฮโดรเลส กลุ่มที่ 1 (GH1) At/Os cluster 6 ซึ่งมี Os9Glu31 ของข้าวที่สามารถย้ายหมู่น้ำตาลให้กับสารอื่นๆ การศึกษาการทำงานทางชีวเคมีของ Os5BGlu19 และ Os9Glu31 ที่กลายพันธุ์ เพื่อศึกษาเพิ่มเติมเกี่ยวกับการทำงานของเอนไซม์ในไกลโคไซด์ไฮโดรเลส กลุ่มที่ 1 cluster 6 cDNA ของ Os5BGlu19 ได้ตัดต่อเข้าไปในพลาสมิด pET32a และ pPZC α BNH8/eGFP เพื่อผลิตโปรตีนใน *Escherichia coli* และในยีสต์ *P. pastoris* ตามลำดับ จากการศึกษาไม่พบโปรตีนที่ละลายน้ำได้หรือกิจกรรมของเอนไซม์ที่ผลิตจาก pET32a/Os5BGlu19 ใน *E. coli* ส่วนโปรตีนที่ผลิตได้จาก *P. pastoris* พบว่า มีอยู่ในเซลล์เมื่อส่องด้วยกล้องจุลทรรศน์ ฟลูออเรสเซนส์ จึงได้ทำการตัดเบสบางส่วนที่ไม่ตรงกับ Os5BGlu19 ออกจาก cDNA Os5BGlu19 ที่ถูกตัดเบสบางส่วนออกได้ถูกผลิตโปรตีนใน *E. coli* และได้ทำการยืนยันโดย western blot ด้วย Os5BGlu19 แอนติบอดี โปรตีนถูกทำให้บริสุทธิ์ด้วยวิธี immobilized metal affinity chromatography (IMAC) เพื่อนำมาตรวจสอบกิจกรรมของเอนไซม์ Os5BGlu19 สามารถย้ายหมู่น้ำตาลไปให้ ferulic acid (FA) เพื่อสะสม 1-feruloyl- β -D-glucose (FAG) ในสารสกัดจากใบข้าว ที่ตรวจสอบโดย UPLC โดยใช้ negative ion electrospray ionization tandem mass spectrometry (LC-MSMS) โดยใช้โหมด multiple reactant monitoring (MRM)

เพื่อศึกษาการแทนที่ของกรดอะมิโน tryptophan ลำดับที่ 243 ในการย้ายหมู่น้ำตาลของ Os9BGlu31 จึงได้ทำการกลายพันธุ์ไปเป็นกรดอะมิโนซึ่งประกอบไปด้วย ไกลซีน (G) ฮีสติดีน (H) ไลซีน (K) กลูตามีน (Q) อาร์จินีน (R) เซอรีน (S) และ วาลีน (V) โดยเทคนิค site directed mutagenesis Os9BGlu31 wild type และกลายพันธุ์ สามารถผลิตโปรตีนใน *E. coli* สายพันธุ์ Origami B (DE3) และถูกทำแยกให้บริสุทธิ์โดย IMAC และ His₆ tags ถูกตัดออกโดย TEV protease

UPLC ถูกใช้ในการวิเคราะห์ปริมาณของ 4NP ที่ถูกปลดปล่อยออกมาเพื่อหากิจกรรมของเอนไซม์ Os9BGlu31 wild type และตัวกลายพันธุ์โดนใช้ตัวรับ (acceptors) 22 ตัว Os9BGlu31 กลายพันธุ์ที่ตำแหน่ง W243 (G Q และ V) พบว่า สามารถย้ายหมู่น้ำตาลไปที่ phenolic acids ฮอร์โมนพืช และ ฟลาโวนอยด์ ได้ดีกว่า wild type Os9BGlu31 กลายพันธุ์ที่ตำแหน่ง W243 สามารถสังเคราะห์สารได้หลายชนิด เช่น กลูโคไซด์ และ บิส-กลูโคไซด์ กับ กรดไฮดรอกซีฟีนอลิก และ ฟลาโวนอยด์ แสดงถึงอิทธิพลของการย้ายหมู่น้ำตาลของ Os9BGlu31 ที่ตำแหน่งกรดอะมิโน W243 ว่ามีความสำคัญในการย้ายหมู่น้ำตาลไปให้สารอื่น



LINH THUY TRAN : CHARACTERIZATION OF RICE GLYCOSIDE
HYDROLASE FAMILY I - CLUSTER AT/OS 6 ENZYMES.

THESIS ADVISOR : PROF. JAMES R. KETUDAT-CAIRNS, Ph.D. 112 PP.

GLYCOSIDE HYDROLASE/TRANSGLUCOSIDASE/ RICE/BETA-
GLUCOSIDASE/ RECOMBINANT PROTEIN EXPRESSION/MUTAGENESIS

Plant β -glucosidases mostly belong in glycoside hydrolase family 1 (GH1), which includes transglucosidases as well. Enzymes in GH1 have been reported to catalyze novel transglycosylation reactions in different plants. Based on sequence similarity, rice Os5BGlu19 is classified in GH1 phylogenetic cluster At/Os 6, which also includes rice Os9BGlu31 transglucosidase. I characterized the biochemical functions of Os5BGlu19 and Os9Glu31 mutants in order to learn more about the functions of GH1 cluster 6 enzymes. An Os5BGlu19 cDNA was inserted into pET32a and pIZCaBNH8/eGFP vectors to express in *Escherichia coli* and *P. pastoris*, respectively. However, no soluble protein or activity could be detected for protein expression from pET32a/Os5BGlu19 in *E. coli*, and protein which was expressed in *P. pastoris* was found to be localized inside the cell by fluorescence microscopy of the C-terminally fused enhanced green fluorescent protein (eGFP) tag. Therefore, Os5BGlu19 was truncated on both termini to remove extra sequence that did not match the homology model structure of Os5BGlu19 and might contain protease sites or sequence targeting the protein to the yeast vacuole. Truncated Os5BGlu19 fusion protein was expressed in *E. coli* and its expression confirmed by western blot analysis with anti-Os5BGlu19 peptide antibody. The protein was partially purified by

immobilized metal-affinity chromatography (IMAC) to identify the enzyme activity. Os5BGlu19 could transfer glucose to ferulic acid (FA) to accumulate 1-feruloyl- β -D-glucose (FAG) in reactions with rice leaf extracts, as determined by the UPLC and electrospray ionization tandem mass spectrometry (LC-MSMS) in the multiple reactant monitoring (MRM) mode.

In order to determine the influence of substitutions at Os9BGlu31 amino acid residue tryptophan 243 on transglycosylation of various acceptors, this position was mutated to glycine (G), histidine (H), lysine (K), glutamine (Q), arginine (R), serine (S) and valine (V) by site directed mutagenesis. Os9BGlu31 wild type and its mutants were successfully expressed in Origami B(DE3) and purified by IMAC, cleavage from the N-terminal thioredoxin and His₆ tags by TEV protease, and a second round of IMAC. The relative activity of Os9BGlu31 wild type and mutants with twenty-two acceptors were compared by UPLC. Some of the new W243 mutants (G, Q and V) showed higher activity than wild type in glycosylation of phenolic acids, phytohormones and flavonoids, although not as high as the previously generated W245N mutant. Several Os9BGlu31 W243 mutants could catalyze synthesis of multiple products, such as glucoside and bis-glucoside with hydroxyphenolic acids and flavonoids. Thus, the Os9BGlu31 W243 position has profound influence on its transglycosylation activity.

School of Chemistry

Academic Year 2016

Student's signature _____

Advisor's signature _____

AKNOWLEDEMENTS

I would like to express my special thanks of my gratitude to my thesis advisor Prof. Dr. James R. Ketudat-Cairns who gave me the opportunity to study toward my master's degree in Biochemistry, for his valuable supervision, patient guidance and enthusiastic encouragement.

I am very grateful to Assoc. Prof Dr. Mariena Ketudat-Cairns for supporting and providing knowledge on my thesis work, and for spiriting during her English camp II, as well. I sincerely thank Assoc. Prof Dr. Jaruwat Siritapetawee and Dr. Sakesit Chumnarnsilpa for biochemistry subject lectures, advice and providing helpful comments. I am grateful to all SUT biochemistry faculty for their instruction. I also thank my lab members and all friends in the School of Chemistry, Suranaree University of Technology for their help.

For research funding, I am grateful to the National Research University Project of the Commission on Higher Education of Thailand and Suranaree University of Technology and the Center for Biomolecular Structure, Function and Application, Institute of Science, Suranaree University of Technology.

Finally, my deepest gratitude will go to all of my family members for understanding, supporting and cheering me throughout my study in Thailand.

Linh Thuy Tran

CONTENTS

	Page
ABSTRACT IN THAI	I
ABSTRACT IN IN ENGLISH	III
ACKNOWLEDGEMENTS	V
CONTENTS.....	VI
LIST OF TABLES	X
LIST OF FIGURES	XI
LIST OF ABBREVIATIONS AND SYMBOLS	XIV
CHAPTER	
I INTRODUCTION	1
1.1 General introduction	1
1.2 Research objectives	3
II LITERATURE REVIEW	4
2.1 Carbohydrates	4
2.2 Glycoconjugates.....	5
2.2.1 Glycoside hydrolases	6
2.2.2 Glycosyltransferases and transglycosidases.....	8
2.3 Overview of β -Glucosidase	9
2.4 Plant glycoside hydrolase family 1	11
2.5 Rice β -Glucosidases	12
2.6 Plant transglucosidase	14

CONTENTS (Continued)

	Page
2.6 Recombinant protein expression systems.....	18
2.8 Site directed mutagenesis and directed evolution	22
2.9 Secondary metabolites in plant.....	25
2.10 Detection of metabolites by liquid chromatography and mass spectrometry	29
III MATERIALS AND METHODS	32
3.1 General materials	32
3.1.1 Chemicals and reagents.....	32
3.1.2 Plasmids and bacterial strains	35
3.1.3 Oligonucleotide primers.....	35
3.2 General Methods	37
3.2.1 Preparation of <i>E. coli</i> strains DH5 α , XL1-Blue, BL 21(DE3) and Origami B(DE3) competent cells.....	37
3.2.2 Preparation of <i>P. pastoris</i> SMD1168H competent cells	37
3.2.3 Transformation of plasmids into <i>E. coli</i> competent cells.....	38
3.2.4 Transformation of plasmids into <i>P. pastoris</i> competent cells.....	38
3.2.5 Plasmid isolation by alkaline lysis method	39
3.2.6 QIAGEN plasmid miniprep	40
3.2.7 Agarose gel electrophoresis for DNA	41
3.2.8 Purification of DNA bands from gels	42
3.2.9 SDS-PAGE electrophoresis	42

CONTENTS (Continued)

	Page
3.2.10 Determination of protein concentration	43
3.3 Cloning and expression of Os5BGlu19	44
3.3.1 Cloning of optimized Os5BGlu19 cDNA into pET32a vector	44
3.3.2 Cloning of optimized Os5BGlu19 cDNA into pPICZ α BNH8/eGFP	44
3.3.3 Expression of Os5BGlu19 in <i>E. coli</i> cells.....	45
3.3.4 Expression of pPICZ α BNH8/Os5BGlu19/eGFP in <i>P. pastoris</i>	46
3.4 Cloning, expression, purification of truncated Os5BGlu19.....	48
3.5 Immunoblotting	49
3.5.1 Antibody production	49
3.5.2 Western blotting	50
3.6 Identification of Os5BGlu19 substrates from rice leaf extraction of KDML105 rice	51
3.7 Assay of Os5BGlu19 activity	51
3.8 UPLC electrospray triple quadrupole tandem mass spectrometry	52
3.9 Substrates purification by LH20 column.....	53
3.10 Characterization of relative activities of Os9BGlu31 transglucosidase wild type and its mutants	53
3.10.1 Construction of pET32a/DEST/TEV/Os9BGlu31	53
3.10.2 Site directed mutagenesis of Os9BGlu31	53
3.10.3 Expression and purification of Os9BGlu31 and its mutants	56

CONTENTS (Continued)

	Page
3.10.4 Relative activities of Os9BGlu31 wild type and mutants	57
IV RESULTS AND DISCUSSION	58
4.1 Cloning and expression of Os5BGlu19	58
4.1.1 Cloning and expression of pET32a/Os5BGlu19 in <i>E. coli</i>	58
4.1.2 Cloning and expression of pPICZ α BNH8/eGFP/Os5BGlu19 in <i>P. pastoris</i>	64
4.2 Cloning, expression and purification of Truncated Os5BGlu19	68
4.3 Identification of Os5BGlu19 substrates from rice leaf extract.....	72
4.4 Donor substrate purification and identification by LH20 column	78
4.5 Characterization of relative activities of Os9BGlu31 transglucosidase wild type and its mutants.....	81
4.5.2 Expression and purification of Os9BGlu31 wild type and its mutants.....	81
4.5.2 Relative activities of Os9BGlu31 wild type and its mutants with various acceptors.....	83
V CONCLUSION	92
REFERENCES	95
APPENDIX.....	109
CURRICULUM VITAE.....	112

LIST OF TABLES

Table	Page
2.1 <i>E.coli</i> strains that were be used in this study (adapted from the Novagen pET system manual)	21
2.2 Representative secondary metabolite classes.....	27
3.1 Chemical reagents and sources	32
3.2 Oligonucleotide primers for Os5BGlu19 cloning	36
3.3 Oligonucleotide primers for site-directed mutagenesis of Os9BGlu31	36
3.4 Appropriate antibiotics for each plasmid and <i>E. coli</i> strain.....	40
3.5 Cycling parameters for amplification of Os5BGlu19 including <i>Xba</i> I site.....	45
3.6 Cycling parameters for Os9BGlu31 amplification	55
3.7 Cycling parameters for mutation of pET32a/Os9BGlu31 by the QuikChange® site-directed mutagenesis method.....	56
4.1 Concentrations of 1-O-feruloyl- β -D-glucose ester (FAG) in reactions of Os5BGlu19 with rice extracts and ferulic acid (FA) determined by UPLC-QQQ-MSMS	77

LIST OF FIGURES

Figure	Page
2.1	The hydrolysis of a glycoside catalyzed by a glycoside hydrolase (GH) 6
2.2	Proposed inverting (a) and retaining (b) mechanisms..... 8
2.3	Phylogenetic tree of predicted protein sequences of rice and Arabidopsis glycoside hydrolase family 1 genes 17
2.4	Function of glycosynthase is generated by mutation of a glycoside hydrolase 23
2.5	The synthesis of a thioglycoside 24
2.6	Schematic view of the phenylpropanoid biosynthesis 28
2.7	General structure of major flavonoids 29
4.1	Agarose gel electrophoresis analysis of cloning optimized Os5BGlu19 into pET32a..... 59
4.2	Amino acid sequence alignment of predicted rice β -glucosidase Os5BGlu19, rice Os3BGlu6, rice Os9BGlu31 transglucosidase and acyl-glucose–dependent glucosyltransferases from the petals of carnation and delphinium (DcAA5GT and DgAA7GT) 62
4.3	SDS-PAGE analysis and western blot of Os5BGlu19 expression 63
4.4	Agarose gel electrophoresis analysis of pPICZ α BNH8/Os5BGlu19/eGFP double restriction enzyme digests 65
4.5	Fluorescence measurement of clones from twenty colonies that were expressed in BMMY with 1% methanol induction for 4 days..... 66

LIST OF FIGURES (Continued)

Figure	Page
4.6	Detection of Os5BGlu19-eGFP in <i>P. pastoris</i> cells by fluorescence and confocal microscopy and by western blot analysis of Os5BGlu19 after pPICZ α B/Os5BGlu19/eGFP was expressed in <i>P. pastoris</i> 67
4.7	Western blot analysis of Os5BGlu19 fractions from attempted purification by IMAC 67
4.8	A homology model of Os5BGlu19 based on the X-ray crystal structure of the Os3BGlu6 β -glucosidase covalent intermediate with 2-fluoro -alpha-D-glucoside (PDB code 3GNR)..... 69
4.9	Amino acid sequence alignment and construction of TrOs5BGlu19 70
4.10	SDS-PAGE and western blot analysis of expression of TrOs5BGlu19 in <i>E. coli</i> induced with different IPTG concentrations 71
4.11	TrOs5BGlu19 purification fraction analysis..... 72
4.12	Expression profile of Os5BGlu19 in different organs and tissues..... 74
4.13	The standard curve of specific fragment ion response for FAG concentration at m/z 175 from the parent ion at 355 m/z 75
4.14	Chromatograms and specific fragment ion spectra from UPLC triple quadrupole tandem mass spectrometry in MRM mode chromatograms 76
4.15	FAG ion abundances of reactions containing one of 11 elution fractions from LH20 column chromatography determined by UPLC-QQQ-MS 79
4.16	FAG ester peak ion abundances in assays of fractions extracted from TLC determined by UPLC triple quadrupole-MSMS 80

LIST OF FIGURES (Continued)

Figure	Page
4.17	83
Coomassie blue SDS PAGE gel of Os9BGlu31 W243V throughout purification.....	
4.18	86
Relative rates of Os9BGlu31 enzyme and W243 mutants with various acceptors.	
4.19	87
Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants for transfer of glucose to gibberellin GA ₄ as phytohormone acceptor.....	
4.20	88
Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants with 4NPGlc and chloramphenicol	
4.21	89
Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants transfer of glucose to 4-hydroxybenzoic acid.....	
4.22	90
Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243G mutants transfer of glucose to vanillic acid	
4.23	91
Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants transfer of glucose to luteolin.....	

LIST OF ABBREVIATIONS AND SYMBOLS

(n/μ/m)g	(nano, micro, milli) Gram
(μ/m)l	(micro, milli) Liter
(μ/m)M	(micro, milli) Molar
°C	Degree (s) Celcius
(k)bp	(kilo) Base pair DNA
cDNA	Complementary deoxynucleic acid
DNA	Deoxynucleic acid
DNaseI	Deoxyribonuclease I
dNTPs	Deoxynucleoside triphosphates
ESI	Electrospray ionization
Glc	Glucose
h	Hour
IPTG	Isopropyl β-D-thioglucoopyranoside
kDa	Kilo Dalton
LB	Luria-Bertani lysogeny broth
min	Minute
MW	Molecular weight
MS	Mass spectrometry
nm	Nano meter
4NP	4-nitrophenyl

LIST OF ABBREVIATIONS AND SYMBOLS (Continued)

OD	Optical density
RNA	Ribonucleic acid
RNase	Ribonuclease
Rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
SDS-PAGE	Polyacrylamide gel electrophoresis with SDS
QQQ	Triple quadrupole
TEMED	Tetramethylenediamine
Tris	Tris-(hydroxymethyl)-aminoethane
UPLC	Ultra-high performance liquid chromatography
v/v	Volume per volume
w/v	Weight per volume

CHAPTER I

INTRODUCTION

1.1 General introduction

Glycoside hydrolase (GH) family 1 (GH1) is primarily considered to be a family of β -glucosidases, although it also includes enzymes with other specificities. Beta-glucosidases (E.C. 3.2.1.21) are GH enzymes that play vital roles in a wide range of living organisms by releasing terminal glucose residues from glucoconjugates (Ketudat Cairns et al., 2012). They have been classified into glycoside hydrolase families GH1, GH2, GH3, GH5, GH9, GH30 and GH116 based on their protein sequences (Cantarel et al., 2009). β -Glucosidases and related enzymes have different specificities, depending on their roles in animals, plants and microorganisms, and these different activities lead to a range of applications. These include lignification, catabolism of cell wall oligosaccharides, defense, phytohormone conjugate activation, and scent release in plants, plant-microbe and insect interactions, biomass conversion in microorganisms, and breakdown of glycolipids and exogenous glucosides in animals.

In addition to hydrolysis, many β -glucosidases have high transglycosylation activity (Opassiri et al., 2003, 2004), and related enzymes have recently been found to act as transglucosidases (TGs), which transfer glucose from one glucoconjugate to another with little hydrolysis (Matsuba et al., 2010; Luang et al., 2013). The GH1 family has recently been recognized to contain TGs that function in production of glucoconjugates in plants. Previous plant transglucosidase studies found that an acyl-

glucose, 1-O- β -D-vanillyl-glucose (VG), acts as the donor molecule in the glucosyl transfer reaction at the 5 position of the anthocyanin cyanidin 3-glucoside in carnations. VG also acts as the sugar donor in the 7-O-glucosylation reaction of cyanidin 3-glucoside in delphinium petals (Matsuba et al., 2010). In *Arabidopsis*, sinapoyl-glucose acts as a bis-functional donor in sinapoylation and the glucosylation of the *p*-coumaroyl moiety on the anthocyanin A11 (Miyahara et al., 2013). Moreover, GH1-anthocyanin glucosyltransferase (actually a TG) and the SCPL acyltransferase also contribute to further glycosylation and acylation of anthocyanin pigments in delphinium. Particularly, 4-hydroxybenzoyl-glucose acts as a zwitter donor in acylation and glucosylation in modification of anthocyanin in delphinium sepals, to generate the compounds required for coloration of blue flowers (Nishizaki et al., 2014).

Many natural products are conjugated with various sugars and organic acids, which is an important part of their detoxification and storage mechanism. The attachment of glycosyl moieties forms additional derivatives of secondary products that increases the diversity of the metabolome and affects biological functions (Sumner et al., 2007). The measurement of lower abundance secondary metabolites has been achieved by chromatography and mass spectrometry. Plant extracts containing a large number of analytes could be separated by chromatography followed by analysis by high resolution mass spectrometry (Hill et al., 2009).

Rice Os9BGlu31 is a GH1 transglycosidase that demonstrated the highest activity with 4-hydroxycinnamic acid β -D-glucose ester donors, such as feruloyl-glucose, 4-coumaroyl-glucose, and sinapoyl-glucose, while the free acids of these compounds acted as the best acceptors (Luang et al., 2013). As noted above, these acyl glucoses are known to serve as donors in acyl and glucosyl transfer reactions in the

vacuole. Additionally, the Os9BGlu31 W243N mutant was reported to produce multiple products in reactions with phenolic acceptors, such as ferulic, 4-hydroxybenzoic, 4-coumaric, caffeic, sinapic and vanillic acids (Komvongsa et al., 2015). The products were identified by liquid chromatography-linked electrospray ionization tandem mass spectrometry (ESI-MS/MS).

In this study, we characterized the ability of rice Os5BGlu19 expressed in *Escherichia coli* to transfer glucose to ferulic acid, which was chosen as an acceptor, based on its high rate of glycosylation by Os9BGlu31 in our previous work. Furthermore, additional mutations of the W243 residue, which resides in the aglycone/acceptor binding site of Os9BGlu31 were made and characterized for their effects on glucosyl transfer to different acceptors. The glucoconjugate compounds were produced by transglucosylation with glucose donor and specific compounds of interest in order to clarify the possibilities for modify the acceptor specificity of Os9BGlu31 and possible applications to produce glucoconjugates.

1.2 Research objectives

1. To clone and express Os5BGlu19 β -glucosidase, optimize for suitable expression conditions and purify the protein.
2. To identify substrates for Os5BGlu19, as a means of assessing its biochemical and biological function.
3. To investigate the relative transglycosylation activity of Os9BGlu31 wild type and its W243 mutants on a range of possible acceptors, to see how different amino acids at this position affect substrate specificity.

CHAPTER II

LITERATURE REVIEW

2.1 Carbohydrates

Carbohydrates are a major class of naturally occurring organic compounds, the monomers of which consist of a hydrate of carbon with the overall formula $C_n(H_2O)_n$ where n is equal to or greater than three. Among the well-known carbohydrates are various sugars, starches, and cellulose, all of which are important for the maintenance of life in both plants and animals (Robyt et al., 1998). Although the structures of many carbohydrates appear to be quite complex, the chemistry of these substances usually involves only two functional groups - ketone or aldehyde carbonyls and alcohol hydroxyls (Matthews et al., 1999).

Carbohydrates are classified into three major classes on the basis of complexity and behavior on hydrolysis, monosaccharides, oligosaccharides, and polysaccharides. As noted above, the monosaccharides, which are the smallest carbohydrates, have a general chemical formula of an unmodified monosaccharide of $(CH_2O)_n$. The aldehyde or ketone group of a straight-chain monosaccharide will react reversibly with a hydroxyl group on a different carbon atom to form a hemiacetal or hemiketal. A heterocyclic ring is formed with an oxygen bridge between two carbon atoms. A 5-atom-ring is called a furanose and a 6-atom-ring is called a pyranose form (Pigman et al., 1972). In monosaccharides, the reference carbon for determining D- or L- chirality is the highest-numbered asymmetric carbon atom in the monosaccharide carbon chain,

and the anomeric prefix (α or β) relates the configuration at the anomeric (or glycosidic) center relative to that of the reference carbon atom (Allen et al., 1992). Polymerization of monosaccharides (two to about twenty residues) generates oligosaccharides connected by glycosidic bonds, which can be hydrolyzed by enzymes or acid to give the constituent monosaccharide units. A linear oligosaccharide has both a reducing and a non-reducing end. The reducing end of an oligosaccharide is the monosaccharide residue with a hemiacetal or hemiketal functionality, while the non-reducing end is the monosaccharide residue in acetal or ketal form (Pigman et al., 1972). Polysaccharides are larger polymers and can be grouped by the kinds of monosaccharides of which they are composed. They range in structure from linear to highly branches. Examples include storage polysaccharides, such as starch and glycogen, and structural polysaccharides, such as cellulose and chitin (Matthews et al., 1999).

2.2 Glycoconjugates

Compounds consisting of carbohydrates covalently linked to other types of chemical moieties are classified as glycoconjugates. The major groups of glycoconjugates are glycosides, including glycoproteins, glycopeptides, peptidoglycans, glycolipids, and lipopolysaccharides, and acyl sugar esters (Allen et al., 1992).

Glycoconjugates are formed in a process termed glycosylation. This ubiquitous mechanism can regulate or promote the functionality of various compounds (Lodish et al., 2000). The diversity of carbohydrate structure is generated by the formation and breakdown of glycosidic bonds mediated by glycosyltransferases (GT), glycoside hydrolases (GH), transglycosidases (TG), glycan phosphorylases and polysaccharide

lysases (Lairson et al., 2004). These enzymes are listed in the CAZy (Carbohydrate-Active EnZymes) database, where enzymes are catalogued by sequence, structure and mechanistic similarities (Cantarel et al., 2009).

2.2.1 Glycoside hydrolases

Glycoside hydrolases (EC 3.2.1) are enzymes that catalyze the hydrolysis at the anomeric carbon of glycosides, leading to formation of a hemiacetal or hemiketal and the corresponding free aglycon, as shown in Figure 2.1 (Henrissat et al., 1991).

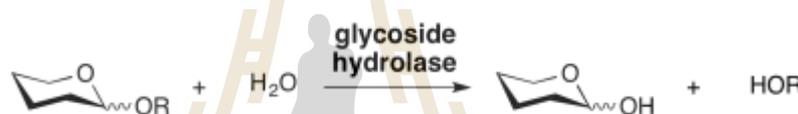


Figure 2.1 The hydrolysis of a glycoside catalyzed by a glycoside hydrolase (GH).

The glycoside hydrolases are divided according to the stereochemical outcome of the hydrolysis reaction into retaining and inverting enzymes. The biological functions of glycoside hydrolases are implicated in many essential processes of life, such as hydrolysis of structural or storage polysaccharides, protecting against pathogens, turnover of cell surface carbohydrate, etc. (Henrissat et al., 2000). The simplest classification of glycoside hydrolases is that based on their reaction and substrate specificities, i.e. the Enzyme Commission (EC) number, but classification by amino acid sequence similarity gives more insight into catalytic mechanism (Henrissat et al., 1991). Members of a sequence-related family will have similar folds, and this can direct the choice of appropriate search models for molecular replacement and opens up

the potential for homology modelling of related sequences. Currently, 145 sequence-based families of glycoside hydrolases are known and they are available on the CAZY website (www.cazy.org) that provides a regularly updated sequence-based classification, which allows reliable prediction of mechanism, active site residues and possible substrates. (Cantarel et al., 1999). The largest of the glycoside hydrolase clans is the GH-A clan, in which the proton donor and the nucleophile are found on β -strands 4 and 7, respectively, of the $(\beta/\alpha)_8$ barrel of the catalytic domain (Jenkins et al., 1995). It includes GH families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113, and 128, which contain enzymes that possess different substrate specificities.

Enzymatic hydrolysis of the glycosidic bond takes place through general acid catalysis requiring two critical residues: a proton donor and a nucleophile/base. Catalysis by glycoside hydrolases is performed with retention or inversion of the anomeric carbon configuration, depending on the mechanism utilized (Figure 2.2). In inverting GHs (Figure 2.2a), the catalytic acid residue donates a proton to the anomeric carbon while the catalytic base residue removes a proton from a water molecule, increasing its nucleophilicity to facilitate its attack on the anomeric center on the opposite side to the departing bond (Koshland et al., 1953). In contrast, retaining GHs (Figure 2.2b) perform their catalysis with two steps, glycosylation and deglycosylation, respectively. In the first (glycosylation) step, the catalytic acid/base facilitates departure of the leaving group by donating a proton to the glycosyl oxygen atom, while the nucleophile forms an enzyme-glycon covalent intermediate. In the second (deglycosylation) step, the process is reversed with a water molecule attacking with basic support from the catalytic acid/base to displace the catalytic nucleophile from the glycon.

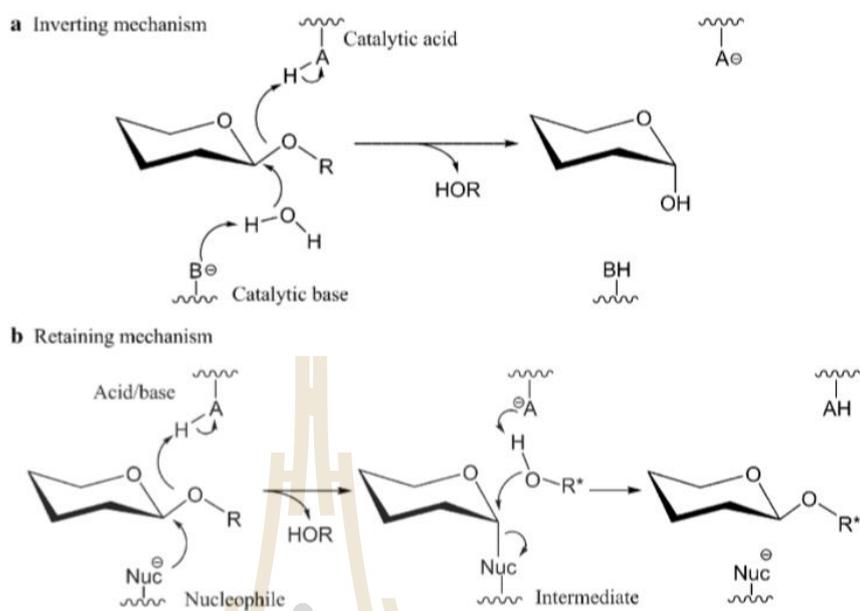


Figure 2.2 Proposed inverting (a) and retaining (b) mechanisms. AH represents a catalytic acid residue; B⁻, a catalytic base residue; Nuc, a catalytic nucleophile residue; A⁻, a catalytic acid/base in basic form. R, an aglycon moiety; and HOR*, an exogenous nucleophile, often a water molecule (Koshland et al., 1953).

2.2.2 Glycosyltransferases and transglycosidases

The biosynthesis of glycans is primarily determined by the glycosyltransferases that assemble monosaccharide moieties into linear and branched glycan chains. They have in common the ability to catalyze a group-transfer reaction in which the monosaccharide moiety of a simple nucleotide sugar donor substrate (e.g., UDP-Gal, GDP-Fuc, or CMP-Sia) is transferred to the acceptor substrate (Crout et al, 1998). The cloning and sequencing of more than 500 genomes has shown that glycosyltransferases are a very common enzyme type, representing 1–2% of the

genome. More than 30,000 glycosyltransferase sequences are known across all kingdoms, and they comprise approximately 103 glycosyltransferase families defined by primary sequence analysis (<http://www.cazy.org/GlycosylTransferases.html>).

In contrast, transglycosidases (TGs) are classified as enzymes that transfer a sugar from a donor other than a nucleotide phosphate or phospholipid glycoside to an acceptor (Lairson and Withers, 2004). TGs utilize the same mechanism as various retaining glycoside hydrolases. Reaction of the nucleophile of a retaining glycoside hydrolase with a substrate gives a glycosyl-enzyme intermediate that can be cleaved either by water to give the hydrolysis product, or by another acceptor (often another carbohydrate alcohol), to give a transglycosylation product that is a new glycoside or oligosaccharide (Sinnott et al., 1990).

2.3 Overview of β -Glucosidases

Beta-glucosidase (EC 3.2.1.21) is one among the earlier discovered and widely studied enzymes, due to its universal distribution and well defined activity, wide variety of substrates and simple nature of enzyme assay (Shewale et al., 1982). The important roles of β -glucosidases in various biological and biotechnological processes include hydrolysis of isoflavone glucosides, the production of fuel ethanol from agricultural residues, the release of aromatic compounds from flavorless precursors, etc. (Lieberman et al., 2007).

There are several β -glucosidases found in mammals, including the GH1 enzymes lactase-phloridzin hydrolase (LPH) and cytoplasmic β -glucosidase, the GH30 enzyme human acid β -glucosidase (GBA1) and the GH116 bile acid β -glucosidase (GBA2) (Ketudat Cairn and Esen, 2010). GBA1 and GBA2 are both β -glucosidases

that cleave glucosylceramide (GlcCer) to glucose and ceramide (Matern et al., 2001). Mutations in the *GBA1* gene cause Gaucher disease, a severe lysosomal storage disorder characterized by accumulation of GlcCer in tissue macrophages (Hruska et al., 2008), while *GBA2* defects cause hereditary spastic paraplegia (Martin et al., 2013) and autosomal recessive congenital ataxia with spasticity (Hammer et al., 2013). Although cytoplasmic β -glucosidase has low activity toward glucosylceramide, it appears to mainly function on exogenous plant glucosides. The intestinal hydrolase LPH functions in food digestion and has both β -galactosidase activity for digestion of lactose and β -glucosidase activity toward exogenous glucosides (Ketudat Cairns and Esen, 2010).

In insects, the *Drosophila melanogaster* genome contains only one GH1 gene, but other insects have adapted glycosides and glycoside hydrolases from the plants on which they feed for protection and digestive purposes. Myrosinases which are a type of thio- β -glucosidase found in certain insects and plants, hydrolyze non-toxic glucosinolates to release toxic defense compounds. The insect β -glucosidases play roles in glycolipid and dietary glycoside degradation, as well as in cellular signaling (Marana SR et al., 2001; Zagrobelny M et al., 2008).

Plant β -glucosidases are involved in beta-glucan synthesis during cell wall development, pigment metabolism, fruit ripening, and defense mechanisms (Ketudat-Cairns et al., 2015). In cellulolytic microorganisms, beta-glucosidase is involved in cellulase induction (due to its transglycosylation activities) and cellulose hydrolysis (Chen et al., 2007). Fungal β -glucosidases are a part of the cellulose degrading enzymes that divide cellulose into two glucose molecules to complete the glucan-breakdown work of endoglucanases and cellobiohydrolases. The action of β -glucosidases can protect these enzymes from the product inhibition effect of cellobiose. Therefore, the

application of β -glucosidases in the conversion of high-cellulose-content biomass to fermentable sugars for the production of fuel ethanol is an intensively studied area. Hydrolysis of soybean isoflavone glycosides is an important application of β -glucosidase in industries. Isoflavones are known to prevent certain cancers, lowers risks of cardiovascular diseases and improve bone health, but it has been revealed that the biological effects are mainly due to the aglycon form of isoflavones rather than their glycosylated form, because intestine absorbs aglycon forms faster (Izumi et al., 2000).

2.4 Plant glycoside hydrolase family 1

Plant β -glucosidases are found mainly in GH families 1 and 3, although putative GH5 and GH116 β -glucosidases are also found in plants (Ketudat Cairns et al., 2015). The functions of these enzymes are thought to include roles in defense, cell wall catabolism and lignification, signaling, and plant secondary metabolism. The plant GH3 family includes enzymes which turnover cell wall components and release toxic plant compounds from their glycosides for defense purposes. The barley HvExoI isoenzyme, which belongs to the GH3 family, exhibited hydrolytic activity with β -linked glucopolysaccharide (β -glucan), oligosaccharide and aryl glucoside substrates, and produced a variety of β -linked oligosaccharide products through transglucosylation activities (Hrmova et al., 1998; Luang et al., 2010).

Glycoside hydrolase family 1 (GH1) contains β -glucosidases with diverse specificities. Most plant GH1 enzymes have been categorized into a closely related subfamily based on their high sequence similarity, yet they still display a wide range of activities (Opassiri et al., 2006). Plant GH1 glycoside hydrolases often show high specificity for their aglycon or saccharide leaving groups, with substrates, which

include oligosaccharides, cyanogenic glucosides, phytohormone glycoconjugates, flavonoid and isoflavonoid glycosides, the monoterpene indole alkaloid precursor strictosidine, and benzoxanoids (Luang et al., 2010). Plant GH1 members also include myrosinases (thio- β -glucosidases) hydrolyzing the S-glycosidic bonds of plant 1-thio- β -D-glucosides (glucosinolates), (Burmeister et al., 1995), β -mannosidases, β -galactosidases, β -glucuronidases, β -fucosidases, diglycosidases like primeverosidase (Mizutani et al., 2002), furcadin hydrolase (Ahn et al., 2004) and isoflavone 7-O- β -apiosyl- β -1,6-glucosidase (Chuankhayan et al., 2005), hydroxyisourate hydrolase, which hydrolyzes an internal bond in a purine ring, rather than a glycosidic bond (Raychaudhuri et al., 2002), and glucosyl and galactosyl transferases (Moellering et al., 2010; Matsuba et al., 2010). These members catalyze their reactions with a molecular mechanism leading to overall retention of the anomeric configuration, which involves the formation and breakdown of a covalent glycosyl enzyme intermediate, as noted above. All of the enzymes display a common (β/α)₈ TIM barrel structure. Apart from plant myrosinases and animal Klotho (KL) subfamily members, all GH1 β -glucosidases contain two conserved catalytic glutamate residues located at the C-terminal end of β -strands 4 and 7 (Jenkins et al., 1995).

2.5 Rice β -Glucosidases

The relationship between rice and Arabidopsis GH1 protein sequences is described by the phylogenetic tree of predicted protein sequences shown in Figure 2.3 (Opassiri et al., 2006). There are forty Arabidopsis (*Arabidopsis thaliana*) and thirty-four rice (*Oryza sativa*) genes encoding apparently functional GH1 proteins separated into eight amino acid-sequence-based phylogenetic clusters that contain both

Arabidopsis and rice genes. In some cases, enzymes that conduct the same or similar functions are grouped together in the same phylogenetic cluster. For instance, enzymes which act on monolignol glycosides, including rice Os4BGlu14, Os4BGlu16 and Os4BGlu18, and Arabidopsis BGLU45, 46 and 47, are grouped with *Pinus contorta* coniferin/syringin β -glucosidase (PC AAC69619). Baiya et al. (2014) reported that Os4BGlu16 and Os4BGlu18 preferentially hydrolyze monolignol glucosides and discussed their possible role in lignification.

All putative rice β -glucosidase protein sequences contain the putative catalytic acid/base and nucleophilic glutamate residues, except Os4BGlu14 and Os9BGlu33, in which the acid/base glutamate is replaced with glutamine, as is seen in thioglucosidases (Opassiri, 2006). The catalytic acid/base and nucleophile consensus sequences are: W-X-T/I-F/L/I/V/S/M-N/A/L/I/D/G-E-/I/Q and V/I/L-T/S/H-E-N-G, respectively (Czjzek et al., 2000; Hoffman et al., 1999). β -Glucosidases with glutamate replaced with glutamine at the acid/base position have been shown to be effective transferases in the presence of a good leaving group aglycon and a nucleophilic acceptor (Müllegger et al., 2005). Therefore, it is possible that all of these putative rice GH1 glycosidases are active enzymes.

Currently, several rice β -glucosidases from different At/Os phylogenetic clusters have been isolated and characterized (Rouyi et al., 2014). Os3BGlu6 in At/Os 1 functions in hydrolysis of β -1,4-linked oligosaccharides, hydrophobic glycosides (Seshadri et al., 2009) and hydrolyzes GA₄ 1-O-acyl glucose ester (Hua et al., 2013). Moreover, the At/Os cluster 4 contains Os3BGlu7 and Os3BGlu8, which are β -glucosidases acting on oligosaccharides, and Os7BGlu26, which is as a β -D-mannosidase (Opassiri, 2003; Kuntothom, 2009). Os4BGlu12 β -glucosidase in At/Os

7 is also an exoglucanase, which plays a role in cell wall remodeling (Opassiri 2010). Moreover, Os4BGlu12 could release the phytohormones tuberonic acid and salicylic acid from their glycosides, suggesting function in phytohormone, as well as oligosaccharide metabolism (Wakuta et al., 2011; Hiromi et al., 2013). *In-vitro* study of rice GH1 β -glucosidases found that Os4BGlu12 and Os4BGlu13 hydrolyze tuberonic acid beta-glucoside (TAG) and the recombinant Os4BGlu13 also contributed to TAG, SAG, oligosaccharide and GA₄-GE hydrolysis in the rice plant (Wakuta et al., 2010; Hua et al., 2015).

2.6 Plant transglucosidases

The GH1 family has recently been recognized to contain TGs that function in production of glucoconjugates in plants (Moellering et al., 2010; Matsuba et al., 2010, Luang et al., 2013). The Arabidopsis At/Os8 representative, sensitive to freezing 2 (SFR2), was shown to be the chloroplast galactolipid:galactolipid galactosyltransferase (GGGT), which disproportionates the galactosyl residues of galactosyl diacyl glycerides to produce diacyl glycerol and β -linked oligogalactosyl diacyl glyceride from monogalactosyl diacyl glycerides (Moellering et al., 2010). Os11BGlu36, the rice At/Os8 member is expected to have the same function as a GGGT, although active protein has yet to be produced by recombinant methods.

Previous plant transglucosidase studies found that an acyl-glucose, 1-O- β -D-vanillyl-glucose (VG), acts as the donor molecule in the glucosyl transfer reaction at the 5 position of the anthocyanin cyanidin 3-glucoside in carnations (Matsuba et al., 2010). VG also acts as the sugar donor in the 7-O-glucosylation reaction in delphinium petals. This reaction is catalyzed by the acyl glucose-dependent anthocyanin 5-O-

glucosyltransferase (AA5GT) in carnation and by 7-O-glucosyltransferase (AA7GT) in delphinium. The Dc AA5GT and Dg AA7GT proteins were shown to be members of GH1, based on protein sequence similarity. In *Arabidopsis*, sinapoylglucose acts as a bis-functional donor by AtBGLU10 glucosyltransferase to form sinapoylation and the glucosylation of the *p*-coumaroyl moiety on the anthocyanin A11 (Miyahara et al., 2013).

Interestingly, the acyl-glucose esters mentioned above also serve as acyl donors for the vacuolar serine carboxypeptidase-like (SCPL) acyltransferases (Mugford et al., 2009) and for intracellular feruloylation of arabinoxylans (Obel et al., 2003). SCPL proteins have recently emerged as a new group of acyltransferase enzymes, which share homology with peptidases but lack protease activity and instead are able to acylate natural products (Lehfeldt, et al., 2000). Several SCPL acyltransferases have been characterized to date from dicots, including an enzyme required for the synthesis of glucose polyesters that may contribute to insect resistance in wild tomato (*Solanum pennellii*) and enzymes required for the synthesis of sinapate esters associated with UV protection in *Arabidopsis* (Mugford et al., 2009). Additionally, GH1-anthocyanin glucosyltransferase and the SCPL acyltransferase also contribute to biosynthetic pathways for anthocyanin pigments in delphinium. Particularly, *p*-hydroxybenzoyl-glucose acts as a zwitter donor in acylation and glucosylation in modification of anthocyanin in delphinium sepals, to generate the compounds required for coloration of blue flowers (Nishizaki et al., 2014).

Previously, our lab characterized Os9BGlu31, a member of the GH1 At/Os6 cluster, which also contains DcAA5GT, and DgAA7GT (Luang et al., 2013). Rice Os9BGlu31 is a glycoside hydrolase family GH1 transglycosidase that acts to transfer

glucose between phenolic acids, phytohormones, and flavonoids. This enzyme demonstrated the highest activity with phenolic acyl glucose donors, such as feruloyl-glucose, 4-coumaroyl-glucose, and sinapoyl-glucose, which are known to serve as donors in acyl and glucosyl transfer reactions in the vacuole, while the free acids of these compounds presented as the best acceptors. Os9BGlu31 and similar TG provide a means to equilibrate these plant secondary metabolites. Furthermore, Os9BGlu31 appears to transfer glucose from the unsaturated fatty acid esters 1-O-oleoyl glucose and 1-O-linoleoyl glucose, based on the build-up of these glucose esters in rice plants in which the Os9BGlu31 gene is knocked out (Komvongsa et al., 2015b). Additionally, the Os9BGlu31 W243N mutant was reported to produce multiple products in reactions with phenolic acceptors, such as ferulic, 4-hydroxybenzoic, 4-coumaric, caffeic, sinapic and vanillic acids (Komvongsa et al., 2015). The products were identified by liquid chromatography-linked electrospray ionization tandem mass spectrometry (ESI-MS/MS). Furthermore, Os9BGlu31 and the some closely related isoenzymes, such as Os1BGlu5, Os5BGlu19, Os5BGlu22, Os6BGlu25, Os9BGlu32 and Os9BGlu33, may play similar roles in rice, but the activity of other GH1 cluster At/Os6 members are yet to be determined. These enzymes also have potential to be applied to *in vitro* glycodiversification without the need for a nucleotide sugar intermediate (Gantt et al., 2011).

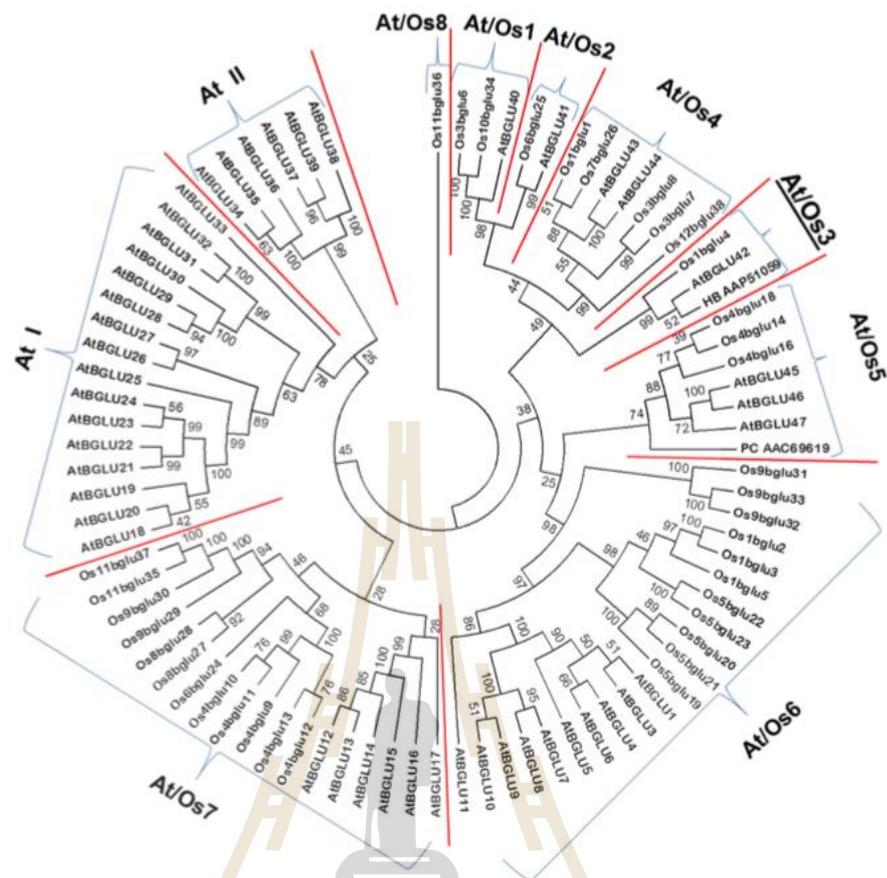


Figure 2.3 Phylogenetic tree of predicted protein sequences of rice and Arabidopsis glycoside hydrolase family 1 genes (Opassiri et al., 2006). The 7 clusters that contain both Arabidopsis and rice sequences that are clearly more closely related to each other than to other Arabidopsis or rice sequences outside the cluster are numbered 1–7, while the outgroup cluster for which the Arabidopsis orthologue is not shown is numbered 8. The figure was taken from Rouyi et al. (2014).

2.7 Recombinant protein expression systems

Use of recombinant protein has increased greatly as a technique to apply in many fields. It has contributed widely to the understanding of the regulation of gene expression via transcriptional, translational and post translational mechanisms. Many recombinant expression systems have been developed to produce high protein yield that facilitates studies of protein structure and biochemical and biological activity (Shewry et al., 1992). Among the many systems available for heterologous protein production, *E. coli* is a particularly popular system due to many useful characteristics, such as growing rapidly at high density on inexpensive substrates, and the availability of an increasingly large number of cloning and expression vectors and mutant host strains (SC Makrides et al., 1996). *E. coli* strains used in this study, as well as related strains, and their descriptions are shown in Table 2.1.

Several GH1 enzymes have been expressed in *E. coli* such as an active fusion protein with N-terminal thioredoxin and 6x histidine tags from the pET32a or its derivative pET32a/DEST expression vectors. A GH1 β -D-glucosidase from rice, BGlu1 (later given the systematic name Os3BGlu7), was expressed in *E. coli*, and characterized to have high transglucosylation activity to produce oligosaccharide products (Opassiri et al., 2003). The cDNA encoding Os4BGlu12 was used to produce recombinant enzyme, which hydrolyzed β -linked oligosaccharides and 4NP-glycosides, with the best expression from the *E. coli* Origami B(DE3) strain (Opassiri et al., 2006). Os9BGlu31 transglucosidase was also produced in the *E. coli* Origami B(DE3) strain (Luang et al., 2013). Notably, Origami(DE3), Origami B(DE3), Origami 2(DE3) and Rosetta-gami(DE3) are modified *E. coli* strains with oxidizing cytoplasm, which allows the formation of disulfide bridges in intracellularly expressed proteins.

E. coli expression vectors utilizing tags such as SUMO, maltose binding protein (MBP) and thioredoxin are designed to promote soluble expression and are commercially available. Pre-expression of sulfhydryl oxidase may markedly promote disulfide bond formation (Nguyen et al., 2011). However, the flaw of the system is that insoluble and inactive proteins are co-produced due to codon bias, rapid translation relative to protein folding, and lack of proper phosphorylation and glycosylation (Khow et al., 2012).

Yeast has several advantages over *E. coli* as an expression system. In particular, yeast can efficiently secrete proteins into the medium, facilitating protein purification and this also allows disulfide bond formation and glycosylation of proteins to occur (Shewry et al., 1992). Like bacteria, yeast are simple to cultivate on inexpensive growth media, and there is a formidable array of techniques for the manipulation of foreign genes for expression in yeast cells. *S. cerevisiae* is a well characterized eukaryotic model organism for production of heterologous proteins. Contrary to bacterial host systems, *S. cerevisiae* possesses the ability to perform eukaryote-type post-translational modifications and secretion, which leads to a reasonable cost of post-fermentation *in vitro* purification and modification (Schmidt et al., 2004). Nevertheless, *S. cerevisiae* is often limited as an expression system by low yields. Methylotrophic yeasts, such as *P. pastoris* and *Hansenula polymorpha*, retain all the advantages of *S. cerevisiae* but provide a reliable means of achieving greatly elevated yields.

P. pastoris can produce large amounts of recombinant proteins by methanol induction of the alcohol oxidase 1 (AOX) promoter, and has been reported to effectively synthesize eukaryotic post-translationally modified proteins (Cregg et al., 2000). Researchers have achieved much higher yields of plant glycoside hydrolases, such as

barley α -amylase (Juge et al., 1996) and Thai rosewood β -glucosidase (Ketudat Cairns et al., 2000) in *P. pastoris* than in *S. cerevisiae*. An efficient system for expression of barley β -D-glucan exohydrolase HvExoI in recombinant *P. pastoris* was expression from a codon-optimized cDNA in a protease-deficient strain, at 20 °C, to produce a good yield of enzyme (Luang et al., 2010). Similar success was achieved with the GH1 enzyme Os4BGlu16 utilizing the same system, allowing it to be characterized as a monolignol β -glucosidase (Baiya et al., 2014).



Table 2.1 *E. coli* strains that were used in this study (adapted from the Novagen pET system manual).

<i>E. coli</i> strain	Application description	Antibiotic resistance
BL21(DE3)	General purpose expression host	None
Origami(DE3)	Enhances disulfide bond formation in the cytoplasm by mutations in glutathione reductase (<i>gor</i>) genes	Tetracycline, Kanamycin
Origami B(DE3)	Enhances disulfide bond formation in the cytoplasm by mutations in redox genes and <i>lacZY</i> mutant may enhance and the solubility and activity of different target proteins	Tetracycline, Kanamycin
Rosetta(DE3)	Enhances the expression of eukaryotic proteins that contain codons rarely used in <i>E. coli</i> due to expression of extra t-RNA genes	Chloramphenicol
Rosetta-gami(DE3)	Combines the properties of Origami and Rosetta	Kanamycin, Tetracycline, Streptomycin, and Chloramphenicol
BL21pLysS(DE3)	Tightly controlled expression for toxic protein expression	Chloramphenicol

2.8 Site directed mutagenesis and directed evolution

Site directed mutagenesis (SDM) is an important tool to engineer the structures and functions of proteins. It can have more advantage than the classical mutations using chemical or physical agents to cause a single base change, an insertion of DNA, or a deletion, in that it is more specific. Therefore, SMD is able to probe the importance of specific residues, such as their roles in enzyme catalysis (Edelheit et al., 2009). A method based on PCR using a thermostable DNA polymerase and a plasmid vector as the template, and mutagenic oligonucleotide primers containing the desired mutation complementary to both strands of a target sequence is currently popular (Braman et al., 1996). In this method, the primers anneal to DNA template, which replicates the plasmid DNA with the mutations to produce a product containing a strand break. The mixture mutant and parental DNA plasmids are then treated with *DpnI* to remove the template methylated DNA from the newly synthesized demethylated mutant DNA and transformed into *E. coli* cells where the nick is ligated by host repair enzymes (Ishii et al., 1998).

Numerous studies have used site-directed mutagenesis to identify the catalytic acid/base and nucleophile represent as the most essential residues in an active site of a GH or GT enzyme. The mutation of these groups often does not eliminate activity, but can significantly alter the catalytic mechanism, depending heavily on the specific enzyme and on the type of amino acid substitution (Peracchi et al., 2001). Various studies have applied site-directed mutagenesis to determine catalytically important residues and produce a new enzyme class such as glycosynthases (Mackenzie et al., 1998). In the initial report of a glycosynthase, a retaining GH1 β -glucosidase was mutated at its catalytic nucleophile and lost hydrolysis activity, while transglycosidase

activity still remained. Glycosyl fluoride as a glycoside donor can transfer sugar to an acceptor in a process that inverts the anomeric stereochemistry (Figure 2.4).

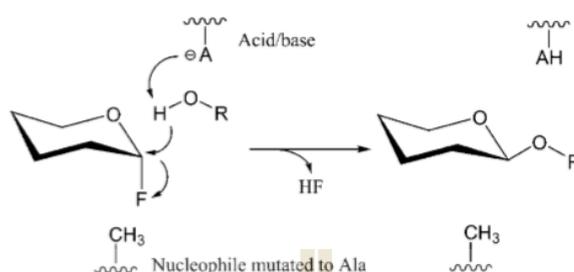


Figure 2.4 Function of glycosynthase is generated by mutation of a glycoside hydrolase. The glycosynthase catalyzes the synthesis of a glycoside from a glycosyl fluoride. (Mackenzie et al., 1998)

The mutation of general acid/base residue in retaining GHs can generate thioglycosylases, which can produce thioglycosides from activated substrates with leaving groups such as 2,4-dinitrophenyl glycosides in the presence of highly nucleophilic sugar thiols that do not require basic assistance of the base residue (Rich and Withers, 2009) (Figure 2.5a). Later, a thioglycosynthase was generated by the removal of both the nucleophile and the catalytic acid/base residue in the retaining glycosidase. This enzyme can catalyze transglycosylations with glycosyl fluorides of inverted anomeric configuration and thiosugar acceptors (Bojarova and Kren et al., 2009; Jahn et al., 2004) to synthesize thioglycosides (Figure 2.5b).

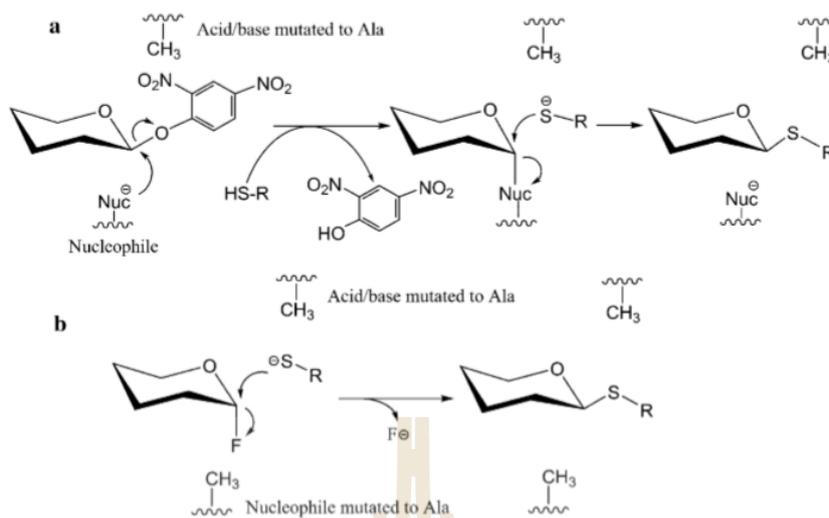


Figure 2.5 The synthesis of a thioglycoside. The reaction was catalyzed by a thioglycosylase with 2,4-dinitrophenyl glycosides and highly nucleophilic sugar thiols or by a thioglycosynthase with inverted anomeric configuration of glycosyl fluorides and thiosugar acceptors (Rich and Withers, 2009).

Moreover, improvement of enzyme activity or identification of residues involved in catalysis can be done by direct evolution that modifies the mechanism, specificity or activity (Kuchner et al., 1997). This approach was applied to produce a library of modified enzymes altered by saturation mutagenesis for each of 189 amino acid residues in a xylanase. This library was screened for increased thermostability and stability compared to the wild type enzymes (Palackal et al., 2004). A GH1 β -glycosidase from *Thermus thermophilus* was converted to β -transglycosidase by random mutation to enhance the rate and substrate specificity of the natural transglycosidase activity (Teze et al., 2013).

2.9 Secondary metabolites in plants

Organic compounds known as secondary metabolites play a common role in plant defense mechanisms against herbivores, pests, and pathogens (Fraenkel et al, 1959). Based on their biosynthetic origins, plant natural products can be divided into three major groups: terpenes (such as plant volatiles, cardiac glycosides, carotenoids and sterols), phenolics (such as phenolic acids, coumarins, lignans, stilbenes, flavonoids, tannins and lignin) and nitrogen containing compounds (such as alkaloids and glucosinolates) (Crozier et al., 2007). A representative list of secondary metabolite classes is shown in Table 2.2 (Sumner et al., 2007). Many natural products are conjugated with various sugars and organic acids, which is an important part of their detoxification and storage mechanism. The attachment of additional chemical moieties forms additional derivatives of secondary products that increases the diversity of the metabolome and affects biological functions (Sumner et al., 2007).

Phenylpropanoids are a diverse family of secondary metabolites containing an aromatic phenyl group linked to a 3-C propane side chain. The phenylpropanoid pathway results in the transformation of stilbenes, coumarins, phenylpropenes, 2-phenylchroman-containing flavonoids, and monolignols (and subsequently lignins) from *p*-coumaroyl CoA as a precursor molecule, as shown in Figure 2.6 (Stack et al., 2001). Hydroxycinnamates (HCAs) derived from *p*-coumaric acid, CoA esters or hydroxycinnamaldehydes activated with Coenzyme A or glucose in the conjugation of monosaccharides, organic acids, lipids and amines to become acyl donors for modifications of secondary metabolites (Strack et al., 2001). Sinapic, *p*-coumaric and ferulic acids are deposited during the early stages of lignification, are also integrated into lignins being involved in ligno-polysaccharide cross-links (Tobimatsu et al., 2012).

Three different monolignols, *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, are polymerized to lignin polymer, which functions in plants to provide cell wall integrity, mechanical support, water transport, and assimilation through the vascular system, and are also involved in defense mechanisms (Freudenberg et al., 1968)

Glycosylation functions to increase the size and hydrophilicity of phenylpropanoid compounds and is modulated by glycosyltransferase and hydrolase activities (Bowles et al., 2005). The GT1 family is the largest glycosyltransferase family in *Arabidopsis* and includes GT that glycosylate small molecule acceptors. Most phytohormones, derivatives of shikimic acid such as flavonoids and mevalonate (terpenoid, steroids), alkaloids and many amino acid derivatives exist as glycosides and act as acceptors for further glycosylation (Vaistij et al., 2009). For example, the main subclasses of flavonoids are the flavones, flavonols, flavan-3-ols, isoflavones, flavanones and anthocyanidins (Figure 2.7). Hydroxyl groups are usually present at the 4', 5 and 7 positions, where sugars can be attached to increase the water solubility of flavonoids, while other substituents, such as methyl groups and isopentyl units, make the flavonoids more lipophilic (Strack et al., 1992).

Table 2.2 Representative secondary metabolite classes.

Artemisinins	Hydroxycinnamic acids
Acetophenones	Isoflavonoids
Alkaloids (<i>imidazole, isoquinoline, piperidine/pyridine, purine, pyrrolizide, quinoline, quinolizidine, terepene, tropane, and tropolone alkaloids</i>) Amines	Isothiocyanates
	Lignins/Lignans
	Non protein amino acids
Anthranoids/Anthraquinones	Phenanthrenes
Anthocyanidins	Phenolics
Aristolochic acids	Phenols (<i>phloroglucinols, acylphloroglucinols, etc.</i>)
Aurones	Phenylpropanoids
Azoxylglycosides	Polyacetylenes
Benzenoids	Polyines
Coumarins	Polyketides
Cyanogenic glycosides	Steroidal and Triterpenoid Saponins
Condensed tannins	Stilbenes
Dibenzofurans	Taxols
Flavonoids (<i>flavanols, flavones, flavanones, etc.</i>)	Terepenoids (<i>hemi, mono, sesqui, di, tri, and tetra</i>)
Glucosinolates	Thiosulfinates
Hydroxybenzoic acid	Xanthones

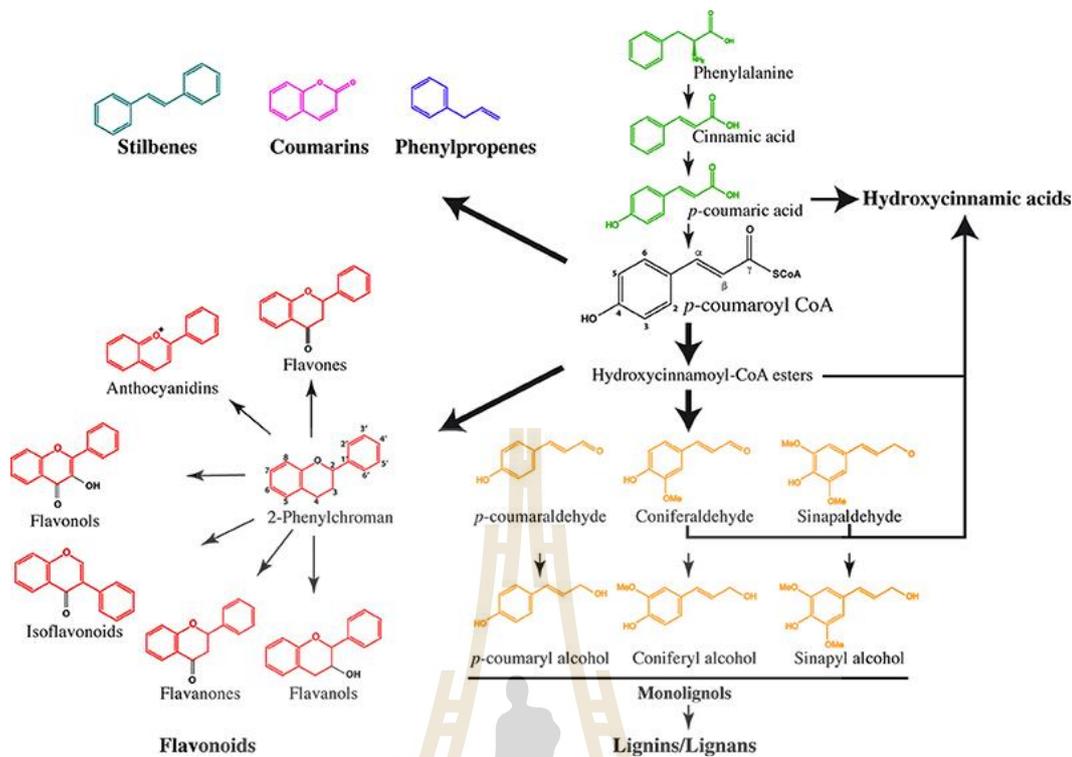


Figure 2.6 Schematic view of the phenylpropanoid biosynthesis (Le Roy et al., 2016)

The pathway begins by successive reactions resulting in the transformation of phenylalanine into *p*-coumaroyl CoA, which is the common precursor of stilbenes, coumarins, phenylpropenes, 2-phenylchroman-containing flavonoids, and monolignols. The pathways leading to the production of hydroxycinnamic acids were obtained based on phenylpropanoid analysis in *Arabidopsis* mutants.

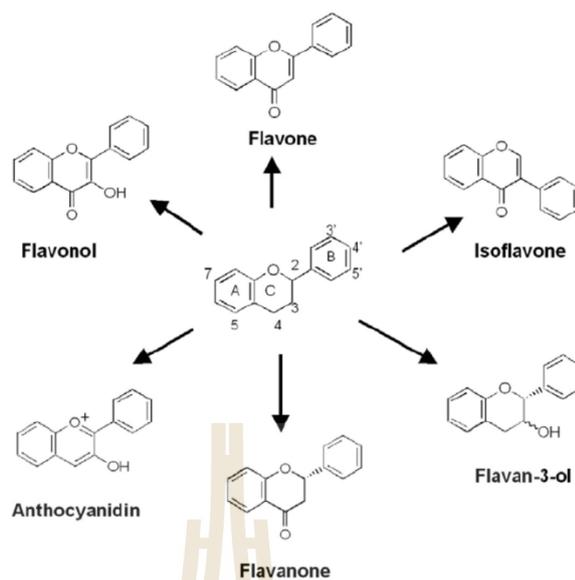


Figure 2.7 General structure of major flavonoids.

2.10 Detection of metabolites by liquid chromatography and mass spectrometry

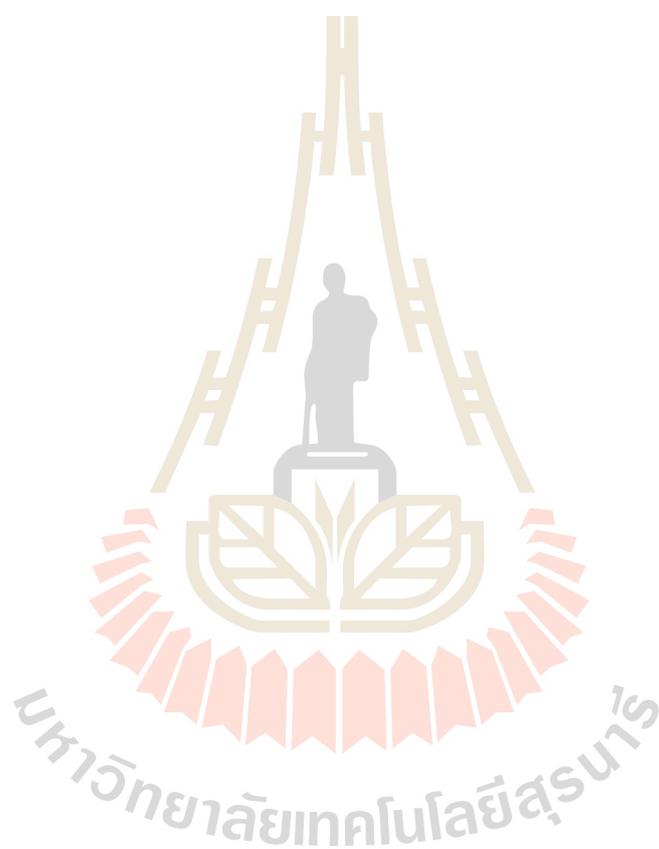
The measurement of lower abundance secondary metabolites has been achieved by chromatography and mass spectrometry. There are thousands of different chemicals produced in plants and the broad scale extraction and identification of plant secondary metabolites has led to the development of metabolomics (Weckwerth et al., 2007). Plant extracts contain a large number of analytes, which could be separated by chromatography followed analysis by high resolution mass spectrometry (Hill et al., 2009). In high performance liquid chromatography (HPLC) the most common stationary phases are reverse phase columns that bind to analytes by hydrophobic interaction and are eluted by increasingly less polar solvent (mobile phase) flow (Snyder et al., 1997). Scanning wavelength spectrometry with a detector diode array

can be applied to measure the elution of analytes from HPLC, for instance plant pigments such as anthocyanins (Matsuba et al., 2010). However, the most useful general detector for LC is the selective mass detector attached to mass spectrometry ion sources (electrospray and atmospheric pressure chemical ionization) (Dass et al., 2007).

A mass spectrometer, in simple terms, functions to detect the mass-to-charge ratio (m/z) and abundance of the various analytes generated during ionization of a sample in an extract or chromatographic fraction. Detection and quantification of secondary metabolites in LC-MS could be followed by use of selected ion monitoring (SIM), where the system is set up to collect only the specific mass (Hill et al., 2009). MS mass analyzers include quadrupole (Q-MS), time-of-flight (TOF-MS), and ion trap (IT-MS), which can select or scan ions in the sample. Quadrupole mass analyzers focus ions of a certain m/z ratio in a path to the detector, while ion traps retain ions of a certain m/z ratio before sending them to the detector, and time of flight mass analyzers simply measure the amount of time it takes for ion accelerated at a certain voltage to reach the detector, which is related to their m/z ratio. Tandem mass spectrometers combine mass analyzers in series to selectivity of scan parent ions and/or fragment ions (Dass et al., 2007).

LC-MS provides efficient ionization for different types of molecule, such as polar, high molecular mass drugs and metabolites (Kostiainen et al., 2003). Most quantitative work in metabolite analysis is carried out by triple quadrupole tandem mass spectrometry (QQQ), in which the parent ion and daughter ion can be selected at the same time. This makes it accurate, sensitive, and comprehensive for quantitative analysis (Clarke et al., 2001). However, unknown secondary metabolites are generally

identified by LC-Q-TOF-MS, which has higher detection sensitivity and mass resolution when scanning a range of masses (Allwood et al., 2010).



CHAPTER III

MATERIALS AND METHODS

3.1 General materials

3.1.1 Chemicals and reagents

Chemical reagents used in this work and their sources are shown in

Table 3.1

Table 3.1 Chemical reagents and sources.

Reagent	Source
• Brilliant blue R250	Acros Organic
• Dithiothreitol (DTT)	
• Ethidium bromide (EtBr)	
• 2-Mercaptoethanol	
• Triton X-100	
• Acetic acid	Carlo Erba
• Ammonium sulfate	
• Calcium chloride	
• Citric acid	
• Chloroform	
• Di-sodium hydrogen phosphate	
• Di-sodium hydrogen phosphate anhydrous	

Table 3.1 Chemical reagents and sources (Continued).

Reagent	Source
<ul style="list-style-type: none"> • Ethylenediaminetetraacetic acid disodium salt (EDTA) • Ethanol • Formic acid • Glass beads 500-750 um • Gsucose • Glycerol • Hydrochloric acid • Isopropanol • Methanol • Sodium acetate anhydrous • Sodium carbonate anhydrous • Sodium dodecyl sulfate (SDS) • Sodium chloride • Tris (hydroxymethyl)-aminomethane 	Carlo Erba
<ul style="list-style-type: none"> • Bacto-agar • Peptone • Yeast Extract • Skim Milk 	HiMedia
<ul style="list-style-type: none"> • HPLC grade Water • Acetonitrile 	Labscan
<ul style="list-style-type: none"> • <i>p</i>-Nitrophenol 	Merck

Table 3.1 Chemical reagents and sources (Continued).

Reagent	Source
• Sodium hydroxide	Merck
• Restriction enzymes: <i>HindIII</i> , <i>NcoI</i> , <i>SacI</i> , <i>SalI</i> , <i>PstI</i> , <i>XbaI</i> , <i>XhoI</i>	New England and Biolabs
• QIAQuick Gel purification kit	QIAGEN
• QIAQuick PCR purification kit	
• <i>Pfu</i> polymerase	Promega
• Deoxyribonucleotide triphosphate (dATP, dGTP, dCTP, and dTTP)	
• T4-DNA ligase	
• 4-Coumaric acid	Sigma
• 4- hydroxybenzoic acid	
• Abscisic acid	
• Apigenin	
• Caffeic acid	
• Chloramphenicol	
• Dimethyl sulfoxide (DMSO)	
• Ferulic acid	
• Isopropyl- β -D-thiogalactopyranosid (IPTG)	
• Indole-3-acetic acid (IAA)	
• Kaempferol	
• Phenylmethanesulphonylfluoride (PMSF)	

Table 3.1 Chemical reagents and sources (Continued).

Reagent	Source
<ul style="list-style-type: none"> • p-Nitrophenyl β-D-glucopyranoside (4NPGlc) • Syringic acid • Tetracycline • Trans-cinnamic acid 	Sigma
<ul style="list-style-type: none"> • Agarose • <i>Taq</i> DNA polymerase • Protein maker 	Vivantis

3.1.2 Plasmids and bacterial strains

A cDNA encoding Os5BGlu19 (Genbank accession number AAS79738.1) optimized for expression in *P. pastoris* was synthesized and inserted into the pUC57 vector by GenScript Corporation (Piscataway, NJ, USA). The expression vector pET32a(+) (Novagen, WI, USA) and pPICZ α BNH8/eGFP (derived from pPICZ α BNH8 (Toonkool et al., 2006) by insertion of an optimized eGFP coding segment by Manasachanok Kongdin) was used for expression of Os5BGlu19. *E. coli* stains DH5 α and XL1-Blue were used for cloning, while BL21(DE3) and Origami B(DE3) were used for protein expression. *P. pastoris* strain SMD1168H, which has the *pep4* mutant so it does not produce protease A and is thus used to minimize protease degradation of expressed proteins, was used for expression in yeast.

3.1.3 Oligonucleotide primers

Custom oligonucleotides were purchased from Pacific Science (Thailand) and are shown in Table 3.2 and 3.3.

Table 3.2 Oligonucleotide primers for Os5BGlu19 cloning.

Primer name	Sequence (5'-3')
5'M13_Fwd	GGTTTTCCCAGTCACGAC
BGlu19XbaRev	GGTGGTCTAGATACTGAGCGTGGAAGGCTGTG
BGlu19_truncFwd	CCATGGCTGCAGCACAGTTTACTAGAGACGACTTC
BGlu19_truncRev	CGAAGCTTGCTAGCGTTTTTCAAGAAGTCGGAGTAC

Table 3.3 Oligonucleotide primers for site-directed mutagenesis of Os9BGlu31.

Primer name	Sequence (5'-3')
W243G_fwd	GGGCTCACATTGCTCGGTGGGTGGTACGAGCCCGGGACG
W243G_rev	CGTCCCGGGCTCGTACCACCCACCGAGCAATGTGAGCCC
W243H_fwd	GGGCTCACATTGCTCGGTCATTGGTACGAGCCCGGGACG
W243H_rev	CGTCCCGGGCTCGTACCAATGACCGAGCAATGTGAGCCC
W243K_fwd	GGGCTCACATTGCTCGGTAAGTGGTACGAGCCCGGGACG
W243K_rev	CGTCCCGGGCTCGTACCAATGACCGAGCAATGTGAGCCC
W243Q_fwd	GGGCTCACATTGCTCGGTCAGTGGTACGAGCCCGGGACG
W243Q_rev	CGTCCCGGGCTCGTACCACTGACCGAGCAATGTGAGCCC
W243R_fwd	GGGCTCACATTGCTCGGTAGATGGTACGAGCCCGGGACG
W243R_rev	CGTCCCGGGCTCGTACCATCTACCGAGCAATGTGAGCCC
W243S_fwd	GGGCTCACATTGCTCGGTAGCTGGTACGAGCCCGGGACG
W243S_rev	CGTCCCGGGCTCGTACCAGCTACCGAGCAATGTGAGCCC
W243V_fwd	GGGCTCACATTGCTCGGTGTGTGGTACGAGCCCGGGACG
W243V_rev	CGTCCCGGGCTCGTACCACACACCGAGCAATGTGAGCCC

3.2 General methods

3.2.1 Preparation of *E. coli* strains DH5 α , XL1-Blue, BL21(DE3) and Origami B(DE3) competent cells

DH5 α , XL1-Blue and BL21(DE3) strains were streaked on Luria – Bertoli lysogeny broth agar (LB agar, 10 g/l peptone, 5 g/l yeast extract, 10 g/l NaCl, 15 g/l agar), while Origami B(DE3) was streaked on an LB agar plate containing 15 μ g/ml kanamycin and 12.5 μ g/ml tetracycline and incubated at 37 °C for 16-18 h. A single colony was picked and inoculated into 5 ml of LB, which was then incubated with shaking at 37 °C, 200 rpm, for 16 - 18 h as the starter culture. Then, 1 ml of starter culture was added to 100 ml of LB broth and shaken at 37 °C, 200 rpm until the optical density at 600 nm (OD₆₀₀) reached 0.4-0.6. The cell culture was chilled on ice for 10 min in sterile polypropylene tube and collected at 4,000 rpm at 4 °C for 10 min. The cell pellets were re-suspended in 10 ml ice-cold sterile 0.1 M CaCl₂ and centrifuged to collect the cell pellets again. Finally, the pellets were re-suspended with 1 ml of 0.1 M CaCl₂ containing 15% glycerol and 50 μ l aliquots were stored at -80 °C.

3.2.2 Preparation of *P. pastoris* SMD1168H competent cells

P. pastoris strain SMD1168H glycerol stock was streaked on a yeast extract peptone dextrose (YPD) plate without antibiotic, which was then incubated at 28 °C for 2-3 days. A single colony was inoculated into 5 ml YPD broth and grown at 28 °C with shaking at 220 rpm about 16 h. One milliliter of starter culture was transferred into 100 ml YPD broth and grown until the OD₆₀₀ reached 1.3-1.5. The cells were collected by centrifugation at 1,500 rpm for 5 min at 4 °C. The pellet was washed 2 times in 100 ml then 50 ml of ice-cold sterile water and collected by centrifugation at 1,500 rpm for 5 min at 4 °C each time. Next, the pellet was re-suspended with 10 ml of

ice-cold 1 M sorbitol and centrifuged at 1,500 rpm for 5 min at 4 °C. Finally, the pellet was re-suspended and kept in 0.5 ml of ice-cold 1 M sorbitol and 80 µl aliquots were used for transformation.

3.2.3 Transformation of plasmids into *E. coli* competent cells

DNA plasmids (50 – 100 mg) were added into 50 µl of competent *E. coli* cells then gently mixed. The competent cell suspensions containing plasmid were incubated on ice for 30 min before transformation by heat shocking the cells at 42 °C for 45 s and quickly chilling on ice for 5 min. Then, 200 µl LB was added to the transformed competent cells, which were then incubated at 37 °C for 1 h. The mixtures were spread on LB agar plates containing appropriate antibiotics for each *E. coli* strain and plasmid in (Table 3.4). The plates were then incubated at 37 °C overnight.

3.2.4 Transformation of plasmids into *P. pastoris* competent cells

The optimized pPICZαBNH8/Os5BGlu19 or pPICZαBNH8/Os5BGlu19/eGFP plasmid was linearized with *SacI* or *PmeI*, as recommended by the supplier (New England Biolabs, Beverly, MA, USA). The restriction enzyme was inactivated by heating at 65 °C for 10 min. Linearization of the plasmid was checked by electrophoresis on a 1% agarose gel. Then, linear DNA was precipitated by mixing with 0.1 volume of 3 M sodium acetate and 2-3 volumes of 100% ethanol. The DNA pellet was dissolved in 5-10 µl of sterile de-ionized water. The linearized recombinant pPICZαBNH8/Os5BGlu19/eGFP vector was transformed into SMD1168H competent cells by electroporation (Bio-Rad) with the parameters of 1.5 kV, 25 µF and 400 Ω (*Pichia* manual, Invitrogen, Agilent Corp, CA, USA). The transformed cells were selected on Yeast Extract Peptone Dextrose medium with Sorbitol (YPDS) plates containing 100 µg/ml zeocin. The YDPS plates were incubated

at 28 °C for 3-5 days. The transformed cell colonies were selected again on a YPDS plate containing 250 µg/ml zeocin incubated in the same manner.

3.2.5 Plasmid isolation by alkaline lysis method

A single colony on the plate containing plasmid was picked into 5 ml of LB with appropriate antibiotic and incubated at 37 °C with shaking at 200 rpm for 16-18 h. The cultured cells were collected by centrifugation at 12,000 rpm, 1 min. The supernatant was removed and the cells were resuspended in 100 µl of lysis buffer I (50 mM glucose, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0). Then, 200 µl of freshly prepared lysis buffer II (0.2 N NaOH, 1% (w/v) SDS) was added, inverted 4-6 times and chilled on ice for 3 min. After that, 150 µl of ice-cold lysis buffer III (3 M potassium acetate, pH 4.8) was added and the tube was mixed by inverting 4-6 times. The alkaline lysis reaction was incubated on ice for 5 min and the clear solution containing the plasmids was separated from the cell debris by centrifugation at 13,000 rpm, 10 min. The supernatant was transferred to a new tube and one volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added into the tube and vortexed or shaken by hand thoroughly for approximately 20 s. The mixture was centrifuged at room temperature for 5 min at 13,000 rpm. The upper aqueous phase was carefully removed into a fresh tube and precipitated with 2 volumes absolute ethanol for 10 min at 4 °C. The precipitated DNA was collected by centrifugation at 13,000 rpm for 10 min. The left over ethanol was removed by speed vacuum. Then, the DNA pellet was re-suspended in 100 µl TE buffer containing 2 µg RNase A and incubated at 37 °C for 10 min. The RNase A-treated plasmids were further purified by adding 70 µl of ice-cold precipitation solution (20% PEG 6000, 2.5 M NaCl) and chilling on ice for 1 h. The precipitated DNA was collected by centrifugation at 13,000 rpm for 10 min. The

supernatant was removed and the pellet was washed by adding 0.5 ml of 70% ethanol and inverting the tube twice, after which the ethanol solution was removed and the tube dried by speed vacuum. Finally, the DNA was re-dissolved with 30 μ l of TE buffer or sterile water.

Table 3.4 Appropriate antibiotics for each plasmid and *E. coli* strain.

Component	Antibiotic
pUC57, PET32a	50 mg/ml Ampicillin
pPICZ α BNH8, pPICZ α NH8/eGFP	25 mg/ml Zeocin
Origami B(DE3)	50 mg/ml Ampicillin
	15 mg/ml Kanamycin
	12.5 mg/ml Tetracycline

3.2.6 QIAGEN plasmid miniprep

The QIAprep® spin miniprep kit (QIAGEN, Hilden, Germany) was used to purify recombinant plasmid DNA according to the manufacturer's instructions. A single colony was picked and inoculated in 5 ml LB broth with appropriate antibiotics. The cultured cells were pelleted by centrifugation at 12,000 rpm for 1 min. The cell pellet was resuspended completely in 250 μ l P1 buffer (100 mg/ml RNaseA in 10 mM EDTA, 50 mM Tris-HCl, pH 8.0). Two hundred fifty microliters of P2 buffer (200 mM NaOH, 1% (v/v) SDS) was added to the re-suspended cells, and mixed by inverting the tube gently 4-6 times until the solution became viscous and slightly clear. After that, 350 μ l of P3 buffer (3 M potassium acetate, pH 5.5) was added and mixed immediately, to avoid localized precipitation, by inverting the tube gently 4-6 times.

The solution was centrifuged at 12,000 rpm for 10 min to compact the white pellet. The supernatant was applied to a QIA prep column by pipetting and centrifuging at 12,000 rpm for 1 min, and then the flow through was discarded. To protect against nuclease activity or carbohydrate content, 0.5 ml of PB buffer (1.0 M potassium acetate, pH 5.0) was added to the column and centrifuged at 12,000 rpm for 1 min. The column was washed 2 times by applying 0.75 ml PE buffer (1.0 M NaCl, 50 mM MOPS, pH 7.0, 15% (v/v) isopropanol) and centrifuging at 12,000 rpm for 1 min. The flow-through was discarded, and the column was centrifuged for an additional 1 min to remove residual wash buffer. Lastly, the column was placed in a new 1.5 ml micro-tube and 50 μ l distilled water was added to the center of column. The column was allowed to stand for 1 min, and then centrifuged at 12,000 rpm for 1 min to elute the plasmid DNA.

3.2.7 Agarose gel electrophoresis for DNA

The purified plasmids and PCR products were checked by agarose gel electrophoresis. One percent agarose gels were prepared in TAE 1X buffer (40 mM Tris HCl, pH 8.0, 40 mM acetic acid, 1 mM EDTA, pH 8.0) or in TBE buffer (90 mM Tris-HCl, pH 8.0, 89 mM boric acid, 2.5 mM EDTA, pH 8.0). The DNA samples were mixed 5:1 with 6X loading dye (0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol, 30% (v/v) sterilized glycerol). Agarose gel electrophoresis was performed in a Gel Electrophoresis Apparatus (Bio Rad) at a constant voltage of 120 V for 30 min. The DNA bands on the agarose gel were detected by staining with ethidium bromide (0.1 μ g/ml) 30 s and de-stained in distilled water for 5 min. The DNA bands were visualized by UV irradiation on a transilluminator (Bio-Rad). The sizes of the DNA bands were estimated by comparing their migration with those of 1 kb DNA ladder (Fermentas, Ontario, Canada).

3.2.8 Purification of DNA bands from gels

The correct size DNA bands that had been separated on agarose gel electrophoresis were purified with QIAquick® Gel Extraction Kit (QIAGEN). The agarose gel containing the target DNA band was sliced with a blade cutter and not more than 300 mg was transferred to a micro-tube. The agarose gel purification was done according to the manufacturer's instructions.

3.2.9 SDS-PAGE electrophoresis

The protein profile and the apparent molecular weights of proteins in various fractions were determined by sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE). The 12% SDS-PAGE separating gel consisted of 12% acrylamide diluted from a 30% acrylamide stock (29% (w/v) acrylamide and 1% (w/v) N, N' methylene bis acrylamide), 375 mM Tris HCl, pH 8.8, 0.1% SDS, 0.05% ammonium persulfate and 0.05% TEMED, while the 4% stacking gel consisted of 4% (w/v) acrylamide diluted from the 30% stock, 125 mM Tris-HCl, pH 6.8, 0.1% SDS, 0.05% ammonium persulfate and 0.05% TEMED. Protein samples were mixed 5:1 with 6X loading buffer (50 mM Tris-HCl, pH 6.8, 10% SDS, 0.2 mg/ml bromophenol blue, 50% glycerol, 20% β -mercaptoethanol) and boiled for 5 min to denature proteins. Twenty microliters of protein samples were loaded into sample wells, and electrophoresed through the polymerized gel at 170 V with Tris-glycine electrode buffer (50 mM Tris base, 125 mM glycine and 0.1% SDS, pH 8.3) until the bromophenol blue dye front reached the bottom of the gel plate. The gels were subsequently stained in staining solution containing 0.1% (w/v) Coomassie Brilliant Blue R250, 40% (v/v) methanol, and 10% (v/v) acetic acid in water for 30 min and de-stained with de-staining solution [40% (v/v) methanol and 10% (v/v) acetic acid in

water] for 1-2 h. The molecular masses of protein bands were determined by comparison to standard low molecular weight protein markers (GE Healthcare, Uppsala, Sweden), which consist of phosphorylase b (97.4 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), bovine carbonic anhydrase (31 kDa), trypsin inhibitor (21.5 kDa) and bovine α lactalbumin (14.0 kDa).

3.2.10 Determination of protein concentration

The protein concentration was determined by the Bio-Rad (Bradford) assay (Hercules, CA, USA). Bovine serum albumin (BSA) was used as a standard ranging from 0-5 μ g. Each assay tube contained 200 μ l of Bio-Rad protein assay solution and standard or sample and was made up to 1 ml with distilled water. The mixture was incubated at room temperature for 10 min. The absorbance was measured at a wavelength of 595 nm (A_{595}) with the protein Bradford program of a NanoDrop 2000 spectrophotometer (Thermo Scientific, MA, USA). For purified protein, the protein concentration was also determined by measuring the A_{280} spectrophotometrically and calculating with the following equation:

$$\text{Protein conc. (mg/ml)} = [\text{OD}_{280} / \text{extinction coefficient}] \times \text{dilution-fold} \times 1/\text{path length}$$

The protein concentration calculated from this absorbance with the extinction coefficient for each specific variant, which was calculated with the PROTEIN PARAMETERS program on the EXPASY website (www.expasy.org).

3.3 Cloning and expression of Os5BGlu19

3.3.1 Cloning of optimized Os5BGlu19 cDNA into pET32a vector

The optimized Os5BGlu19 cDNA was synthesized and inserted into the pUC57 vector by GenScript Corporation (Piscataway, NJ, USA). The pUC57/BGlu19 plasmid containing the optimized Os5BGlu19 cDNA was transformed into DH5 α competent cells as described in Section 3.2.3 and spread onto LB agar containing 50 μ g/ml ampicillin. A colony was picked, inoculated into LB media containing 50 μ g/ml ampicillin and incubated at 37 °C with shaking overnight before pUC57/BGlu19 plasmid was extracted by QIAprep Spin miniprep kit. The pUC57/Os5BGlu19 and pET32a plasmids were digested with *Nco*I and *Xho*I and the predicted size bands were purified by QIAquick® gel extraction kit. Then, Os5BGlu19 insert was ligated into pET32a digested with the same enzymes by T4 ligase reaction (Invitrogen). The mixture was incubated at 15 °C for 18 h and the ligation product was transformed into DH5 α competent cells and selected on a 50 μ g/ml ampicillin LB agar plate as described in Section 3.2.3. The colonies were cultured and plasmids extracted by alkaline lysis method, as described in section 3.2.5. The presence of the gene insert was checked by cutting the prepared plasmids with the *Nco*I and *Xho*I restriction endonucleases and evaluating the digested DNA bands by agarose gel electrophoresis. Finally, the plasmid insert sequence was verified by automated DNA sequencing at Macrogen Corp (Seoul, Korea).

3.3.2 Cloning of optimized Os5BGlu19 cDNA into pPICZ α BNH8/eGFP

An optimized Os5BGlu19 cDNA with an *Xba*I site was amplified from pUC57/OsBGlu19 as a template with 5'M13 forward and OptBG19_XbaRev primers (Table 3.2). The reactions were carried out by *Pfu* DNA polymerase with the

temperature cycling parameters shown in Table 3.5. The PCR product (1.5 kb) was checked by electrophoresis on a 1% agarose gel.

Table 3.5 Cycling parameters for amplification of Os5BGlu19 including *XbaI* site.

Segment	Cycles	Temperature (°C)	Time
1	1	95	2 min
		95	30 s
2	30	60	30 s
		72	3 min
3	1	72	7 min

The PCR product and pPICZ α NH8 or pPICZ α NH8/eGFP were digested with *PstI* and *XbaI*. The ligation of vector and insert followed the procedure described for ligation of pET32a/Os5BGlu19 (section 3.3.1). Two microliters of the reaction was transformed into DH5 α competent cells and selected on a 25 μ g/ml zeocin LB agar plate. Colonies from the plates were cultured for plasmid extraction and the plasmids containing Os5BGlu19 insert were identified by the *PstI* and *XbaI* restriction digestion. Finally, these clones were confirmed again by sequencing at MacroGen Corp.

3.3.3 Expression of Os5BGlu19 in *E. coli* cells

The recombinant pET32a/Os5BGlu19 plasmid was transformed into Origami B(DE3) and BL21(DE3) competent cells and spread onto LB-agar containing 50 μ g/ml ampicillin for BL21(DE3), and 15 μ g/ml kanamycin and 12.5 μ g/ml tetracycline in addition to ampicillin for Origami B(DE3). After incubating the plates at 37 °C overnight, the colonies were picked and incubated in 5 ml of LB media with

the same antibiotic at 37 °C, 16-18 h to make a starter. To express Os5BGlu19, 500 µl of starter was added into 50 ml of the same type of media and cultured at 37 °C with 200 rpm shaking until the optical density at 600 nm of the culture reached 0.4. The cultures were induced by adding IPTG and the temperature was shifted to the expression temperature and grown for 18 h. The optimum expression conditions were determined by varying the final concentration of IPTG at 0, 0.1, 0.2, 0.3, 0.4 and 0.5 mM, and the temperature at 20 °C, 25 °C, 30 °C and 37 °C. The cell pellets will be collected by centrifugation at 4000 rpm at 4 °C for 15 min. The cell pellets were kept at -80 °C to allow freeze-thaw breakage before extraction. The bacterial cell pellets were thawed on ice and then re-suspended in freshly prepared extraction buffer (20 mM Tris-HCl buffer, pH 8.0, 200 µg/ml lysozyme, 1% Triton-X 100, 1 mM PMSF, 25 µg/ml DNase I and 0.1 mg/ml soy bean trypsin inhibitor) in a ratio of 5 ml extraction buffer per gram fresh weight of cell pellets. The re-suspended cells were incubated at room temperature for 30 min. Then, the insoluble proteins were removed by centrifugation. Twenty microliters of the supernatant fractions were used to test hydrolysis activity with 2 mM 4NPGlc in 50 mM acetate buffer, pH 5.0, at 30 °C for 30 min. The insoluble and soluble protein fractions were analyzed by SDS-PAGE with a 12% separating gel.

3.3.4 Expression of pPICZαBNH8/Os5BGlu19/eGFP in *P. pastoris*

After transformed cells were selected again on a YPD plate containing 250 µg/ml zeocin. The colonies were screened for protein production in small scale cultures, as described in the *P. pastoris* manual (Invitrogen). Protein expression was induced by adding methanol to 1% final concentration every day for 4 days at 20 °C. The media of each clone was collected after 4 days to measure fluorescence absorbance

with excitation and emission wavelengths at 480 and 509 nm, respectively. For protein production, a single colony that has been selected on a 250 µg/ml zeocin YPD plate and was inoculated into 500 ml of BMGY medium containing 100 µg/ml zeocin and grown in a shaking incubator (220 rpm) at 28 °C until the cell culture OD₆₀₀ reached 2-3. Cells were harvested by centrifugation and resuspended in 1000 ml BMMY medium at the final OD₆₀₀ of 1. Protein expression was induced by adding methanol to 1% final concentration every 24 h for 3 days at 20 °C (Luang et al., 2010). The protein expression was assessed by measurement of eGFP by fluorescence in both media culture and cell pellets. The cell pellets were extracted in lysis buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% Triton-X 100, 1 mM phenylmethylsulfonylfluoride (PMSF), 25 µg/ml DNase I) containing glass beads. The cells were mixed gently by vortex and sonicated before centrifuging to collect the supernatant to run SDS-PAGE.

For localization of Os5BGlu19/eGFP in the cell pellet, pichia cells were prepared in 50% glycerol in TBS buffer for fluorescent microscopy. Other cells were stained with 300 nM DAPI stain solution. The *P. pastoris* cell suspensions were dropped in the center of the slide and the cover slip added before examination by microscopy and confocal microscopy. Fluorescence microscopy and confocal microscope were set up at 488 and 358 nm excitation wavelengths and 509 and 461 nm emission wavelengths to detect GFP and DAPI, respectively.

Media and crude cell extract were loaded onto immobilized metal-affinity chromatography (IMAC) on cobalt-equilibrated TALON resin. These columns were washed with 0, 5 and 10 mM imidazole and eluted with 250 mM of imidazole. All of fractions were analyzed by SDS-PAGE (on 12% acrylamide separating gels) and

western blot with specific anti-Os5BGlu19 anti-peptide antiserum and peroxidase conjugated goat anti-rabbit IgG secondary antibody.

3.4 Cloning, expression, purification of truncated Os5BGlu19

We designed primers (Table 3.2) to amplify truncated Os5BGlu19 (TrOs5BGlu19) without a short sequence located outside the region homologous to known GH1 structures, based on their protein sequence alignment and homology modeling. The PCR product was amplified with the BGlu19TrcFwd and BGlu19TrcRev primers. Both reactions were carried out with *Pfu* DNA polymerase with the same temperature cycling parameters and annealing temperature as described in section 3.3.2. The PCR product and pET32a vector were digested with *NcoI* and *HindIII* and purified by QIAquick® Gel Extraction Kit. The TrOs5BGlu19 insert and pET32a vector were ligated together by mixing 1:3 vector:DNA insert and heating at 45 °C for 5 min to melt any cohesive termini that had reannealed, then immediately chilled on ice. Then, 1X ligation buffer and T4 DNA ligase were added and the tube was mixed well but gently and incubated at 15 °C for 18 h. One microliter of the reaction was transformed into DH5 α competent cells, which were then selected on a 50 μ g/ml ampicillin LB agar plate. The colonies were cultured and plasmids prepared as described for the pET32/Os5BGlu19 plasmid (Section 3.2.1). The recombinant gene insert was checked by cutting with *NcoI* and *HindIII* and the sequence verified by automated DNA sequencing at MacroGen Corp.

The construction pET32a/TrOs5BGlu19 was transformed into BL21(DE3) and Origami B(DE3) competent cells and spread onto LB-agar containing 50 μ g/ml ampicillin, 15 μ g/ml kanamycin and 12.5 μ g/ml tetracycline for Origami B(DE3), and

50 µg/ml ampicillin alone for BL 21(DE3). The transformed cells were incubated at 37 °C overnight. For small scale expression, a single colony was picked and incubated at 37 °C, shaking at 200 rpm, overnight to make a starter. Then, 50 µl of starter was added into 50 ml medium containing the same antibiotics and incubated until the OD₆₀₀ reached 0.6. After that, the cell culture was induced with 0.1 – 0.5 mM IPTG with 200 rpm shaking and incubated 16-18 h at 20 °C. Then, the large scale 1 liter culture was grown similarly and induced with 0.3 mM of IPTG at 20 °C for 16 h. The cells were collected, kept and extracted as described in section 3.3.1.

The protein was purified from the soluble extract by IMAC on cobalt-equilibrated TALON resin. The resin was washed with the equilibration buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl) and eluted by a discontinuous concentration gradient of 10, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240 and 260 mM imidazole in the equilibration buffer. The presence and purity of protein of the appropriate size was evaluated by SDS-PAGE (12% separating gel and 4% stacking gel) with 10 µl of each fraction. The fractions that contained TrOsBGlu19 were pooled and the buffer exchanged to remove imidazole in 30 kDa MWCO Amicon® Ultra-15 centrifugal filters.

3.5 Immunoblotting

3.5.1 Antibody production

We designed an antigen CLTESTEDIAATERV from a sequence in the C-terminal part of Os5BGlu19 sequence based on its uniqueness in comparison to other GH1 sequences. Anti-peptide antibodies specific for Os5BGlu19 were raised by Genscript Corp. The peptide antigen was designed by optimization software and

chemically synthesized at Genscript with a cysteine added to the N-terminus of the peptide and used to conjugate the peptide to Keyhole limpet hemocyanin (Genscript, USA). The conjugates were mixed with Freund's complete adjuvant and injected into two New Zealand white rabbits, according to the supplier's protocol. The collected blood was be clotted and the serum tested for recognition of peptides and recombinant proteins by an indirect Enzyme-Linked Immuno Sorbent Assay (ELISA) assay in which the proteins are bound to the bottom of microtiter plate wells. Further antisera (anti-Os5BGlu19) dilutions were be tested once a response was detected.

3.5.2 Western blotting

Protein samples were mixed 5:1 with 6X loading buffer (50 mM Tris-HCl, pH 6.8, 10% SDS, 0.2 mg/ml bromophenol blue, 50% glycerol, 20% β -mercaptoethanol) and heated at 95 °C for 5 min, cooled, and centrifuged briefly before loading on 12% SDS-PAGE and blotted onto a nitrocellulose membrane (GE Healthcare Life Science, Little Chalfont, UK) by electro-blotting in a wet-blot apparatus. The membrane was then blocked with 5% skim milk in TBS (50 mM Tris-Cl, pH 7.6, 150 mM NaCl) and incubated with the rabbit anti-peptide anti-serum developed against Os5BGlu19 (Genscript, USA) at a dilution 1:500 in 5% skim milk in TBS at 4 °C for 1 h. After that, the membrane was washed 3 times with TBS and immediately incubated with a 1:3000 dilution of horse-radish-peroxidase-conjugated anti-rabbit IgG antibodies (Bio-Rad, CA, USA) in 5% skim milk at room temperature for 1 h. After washing three times with TBS, the membrane was development with the ECL western blotting analysis system (GE Healthcare Life Science) for 10 min before exposure to Hyperfilm ECL in a film cassette in the dark room, followed by film development.

3.6 Identification of Os5BGlu19 substrates from rice leaf extraction of KDML105 rice

Four and ten week old rice leaves and ten week old stem (*Oryza sativa* L. *ssp. Indica* cv. KDML105) were extracted for identification of donor substrates for Os5BGlu19. One hundred milligrams of rice 4 week and 10 week old leaves and 10 week old stems were ground separately in liquid nitrogen with a mortar and pestle. The ground samples were extracted in 500 μ l of 70% methanol in water by vortexing for 5 min and sonication for 15 min on ice. The extracts were centrifuged at 12,000 rpm for 10 min and the supernatants collected and evaporated by speed vacuum for 45 min to vaporize off the methanol (Komvongsa et al., 2015).

3.7 Assay of Os5BGlu19 activity

Initially, Os5BGlu19 activity was assayed with 5 mM 4-nitrophenyl β -D-glucopyranoside with and without 0.25 mM ferulic acid in 50 mM acetate buffer, pH 5, for 3 h. Since no activity was detected, rice extracts were tested as glycosyl donor. The TrOs5BGlu19 enzyme was assayed with rice extracts with ferulic acid as an acceptor in order to identify the activity of Os5BGlu19. The reactions with 100 μ g rice extract dissolved in water were set up with 0.25 mM ferulic acid and 5 μ g of TrOs5BGlu19 in 50 mM citrate buffer, pH 4.5. After incubating at 30 °C overnight, the reactions were stopped with 1% formic acid and centrifuged at 12,000 rpm for 10 min. The supernatants were then analyzed by ultraperformance liquid chromatography (UPLC) triple quadrupole tandem mass spectrometry analysis.

3.8 UPLC electrospray triple quadrupole tandem mass spectrometry (UPLC-QQQ-MS)

Ferulic acid (FA) and 1-O-feruloyl- β -D-glucose (FAG) were qualified on an Agilent 1290 UPLC system (Agilent, USA) in line with an Agilent 6490 triple quadrupole mass spectrometer. The reaction components were separated by a ZORBAX SB C18 (1.8 μ m, 2.1 \times 150 mm) column (Agilent, USA) eluted with mobile phases that contained 0.2% formic acid in water (solvent A) and acetonitrile (solvent B). In order to elute the compounds, a linear gradient was set up from 5% to 50% B (v/v) in 13 min, 50% to 70% B (v/v) in 1 min, and 70% to 5% B (v/v) in 2 min at a flow rate of 0.3 ml/min. The UV visible spectrum was scanned from 190-500 nm with a diode array detector (DAD). Electrospray ionization (ESI) was used in the negative ion mode with multiple reaction monitoring (MRM) to collect multiple product ions from precursor ions. For the negative ion mode, a capillary voltage of 3 kV, and a gas temperature of 300 $^{\circ}$ C were used. The gas flow was set as 16 liters/min, nebulizer at 45 psi with the sheath gas heater at 300 $^{\circ}$ C and sheath gas flow at 11 l/min. The major precursor ion m/z at 355 ([FAG-H⁺]⁻) with the selected product ions were at m/z 193 ([FA-H⁺]⁻) and 175 ([FA-H⁺-H₂O]⁻) for FAG and at m/z 193 with the product at m/z 134 for FA. FAG and FA had elution times of 7.2 and 9.2 min, respectively. Standard reactions containing 0.25 mM ferulic acid, 5 mM 4NPGlc with and without 5 μ g of Os9BGlu31 in 50 mM citrate buffer, pH 4.5, were used to set a standard curve for FAG with four fold serial dilutions to demonstrate the linearity of the ion abundance response.

3.9 Substrates purification by LH20 column

To purify the substrate, 20 g of ground 10 week old rice leave were exacted with 100% methanol (HPLC grade) and the mixture stirred 3 h at room temperature. The mixed extract was filtered through a 0.45 µm filter and methanol evaporated off by freeze drying for 5 h. Then, five milliliters of supernatant was loaded into a Sephadex LH-20 resin column (GE Healthcare, USA). The column was eluted with 5 CV of 100% methanol. Each fraction was evaluated via silica gel TLC developed with a chloroform/methanol/30% ammonia [7:2.8:0.2 (v/v/v)] solvent, stained with 10% sulfuric acid in ethanol, and charred at 110 °C. Five micrograms of Os5BGlu19 and control reaction (without Os5BGlu19) were assayed with eluted fractions and 0.25 mM ferulic acid in 50 mM citrate buffer, pH 4.5. The reactions were incubated at 30 °C, overnight, stopped by 1% formic acid and quantified by UPLC-QQQ-MS.

3.10 Characterization of relative activities of Os9BGlu31 transglucosidase wild type and its mutants

3.10.1 Construction of pET32a/DEST/TEV/Os9BGlu31

The wild type pET32a/DEST/TEV/Os9BGlu31 expression vector was constructed by Sukanya Luang and it is as yet unpublished. This construct, which I used to produce new mutations, was constructed as follows. The cDNA fragment encoding mature Os9BGlu31 was amplified with the AK121679F (5' CACCATGGCGGCGGGGATCACCAG 3') and AK121679R (5' CTCGAGAACCTTGATCACTGGGAGTAGGCTC 3') primers by *Pfu* DNA polymerase with the AK121679 clone plasmid as template. The Os9BGlu31 cDNA

fragment was amplified following the condition in Table 3.6. The PCR product (~1.5 kb) was purified from the agarose gel, and cloned into pENTR/TEV/D-TOP Gateway® entry vector (Invitrogen) by mixing 5 ng PCR product per 1 µl of 15-20 ng vector and incubating at 22-23 °C for 16-18 h. The reaction (2µl) was transformed into XL1-Blue competent cells and selected on a 15 µg/ml kanamycin LB agar plate. The entry clone size was cut checked with *SacI* restriction endonuclease enzyme. Then, the pENTR/TEV/D-TOPO/Os9BGlu31 was recombined into the pET32a/DEST destination vector with Gateway LR clonase using 75 ng entry clone per 150 ng destination vector, and the reaction was incubated at 25 °C for 18 h, as described in the LR Clonase instructions (Invitrogen). A 2.5 µl aliquot of the reaction was transformed into DH5α competent cells and plasmid-containing cells were selected on a 50 µg/ml ampicillin LB agar plate. The recombinant expression vector was cut checked within the insert gene and the plasmid with *EcoRI* restriction endonuclease and DNA sequencing was done at MacroGen Corp.

Table 3.6 Cycling parameters for Os9BGlu31 amplification.

Segment	Cycles	Temperature (°C)	Time
1	1	95	1 min
		95	30 s
2	30	63.5	1 min
		72	3 min
3	1	72	5 min

3.10.2 Site directed mutagenesis of Os9BGlu31

The tryptophan residue W243 in wild type Os9BGlu31 was changed to the amino acids glycine (G), histidine (H), lysine (K), glutamine (Q), serine (S) and valine (V) by site directed mutagenesis. The pET32a/TEV/Os9BGlu31 plasmid was used as a template for amplification of full-length mutant plasmid with the specific primers for the desired point mutations (Table 3.7). Following the Quikchange® site-directed mutagenesis method (Agilent), the reactions thermocycling conditions were set up as shown in Table 3.4 with reaction containing 5-50 ng template, 100 ng/ μ l forward and reverse primers, 0.2 μ M dNTP mix, 1X *Pfu* Ultra HF reaction buffer (20 mM Tris-HCl (pH 8.8), 10 mM (NH₄)₂SO₄, 10 mM KCl, 0.1% (v/v) Triton X-100, 0.1 mg/ml BSA and 2 mM MgSO₄), and 2.5 U of *Pfu* polymerase (which has proofreading activity) to synthesize the mutated plasmid DNA during the temperature cycling. After amplification, 10 U of *DpnI* was added to the PCR products and the reaction was incubated overnight in order to digest the methylated and hemi-methylated DNA template. Then, the products were transformed into XL1-Blue competent cells and selected on 50 mg/ml of ampicillin LB agar plates. All mutations were confirmed by DNA sequencing after the plasmids were extracted by alkaline lysis.

Table 3.7 Cycling parameters for mutation of pET32a/Os9BGlu31 by the QuikChange® Site-Directed Mutagenesis method.

Segment	Cycles	Temperature (°C)	Time
1	1	95	2 min
		95	50 s
2	18	60	50 s
		68	15 min
3	1	68	7 min

3.10.3 Expression and purification of Os9BGlu31 and its mutants

The pET32a/DEST/TEV/Os9BGlu31 wild type plasmid and its mutants W243 G, H, K, N, Q, R, S and V were transformed into Origami B(DE3) and selected on 50 mg/ml ampicillin 15 mg/ml kanamycin and 12.5 mg/ml tetracycline in LB agar plates. The cells were grown in LB media containing the same antibiotics, at 37 °C with 200 rpm shaking until the OD₆₀₀ reached 0.6. The protein expression was induced by 0.4 mM IPTG, at 20 °C for 18 h. The cells were collected, kept and extracted as described in section 3.3.1. Os9BGlu31 wild type protein and its mutants were purified by IMAC on cobalt-equilibrated TALON resin, as described in section 3.5. All of the purified fractions were assayed in 150 µl total volume with 2 mM 4NPGlc in 50 mM acetate buffer, pH 4.5, at 30 °C for 30 min. The reactions were stopped by addition of 75 µl of 2M Na₂CO₃ and the absorbance at 405 nm measured by spectrometer. The fractions containing activity with 4NPGlc were pooled and imidazole removed by buffer exchange in 30 kDa MWCO Amicon® Ultra-15 centrifugal filters. The N-terminal fusion tag was removed by cleavage with 1 mg TEV protease per 50 mg of

fusion protein at 4 °C for 16 h. Then, the digested proteins were loaded onto the second IMAC column, flow-through fractions containing the 4NPGlc cleavage activity, assayed as described above, were pooled and concentrated with a 30 kDa MWCO Centricon filter.

3.10.4 Relative activities of Os9BGlu31 wild type and mutants

In order to compare the activities of Os9BGlu31 and its mutants on a range of acceptor substrates, reactions were set up with 0.25 mM of acceptor and 5 mM of 4NPGlc as donor and 1 µg of Os9BGlu31 wild type and W243H, G, Q, S and V mutants, 0.2 µg of Os9BGlu31 W243N mutant or 3 µg of Os9BGlu31 W243K and R mutants, in 50 mM citrate buffer, pH 4.5. The reactions were incubated at 30 °C for 15 min, then stopped by addition of 1% formic acid. The reactions were centrifuged at 12,000 rpm, 10 min to collect supernatants. Then, two microliters of reactions were injected into the ZORBAX SB-C18 (1.8 µm, 2.1 x 150 mm) column on an Agilent 1290 UPLC. The mobile phase and linear gradient were as described in Section 3.7.3. The absorbance was measured at wavelengths between 190 and 500 nm, monitored by DAD, to detect the glucoconjugated compounds and the relative activities were evaluated from the amount of released 4-nitrophenol (4NP) that eluted at 10.5 min and was quantified from its absorbance at 360 nm.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Cloning and expression of Os5BGlu19

4.1.1 Cloning and expression of pET32a/Os5BGlu19 in *E. coli*

Os5BGlu19 was originally predicted from the rice genome sequences, based on homology with GH1 proteins and was predicted to results from a 1590 nucleotide open reading frame, encoding a 530 amino acid precursor protein, including a signal sequence of 31 amino acids (Opassiri et al., 2006). Currently, NCBI annotation of the Rice Genome Locus Os050365600, from which Os5BGlu19 was predicted, as the originally predicted cDNA encoding Os5BGlu19 (AAS79738.1), 3 predicted isoforms, X1-X3, and one experimental full-length cDNA, AK105546. However, the AK105546 sequence has a frame-shift, suggesting it results from missplicing or a sequencing error. After removal of a predicted 31 amino acid residue signal sequence, the originally predicted Os5BGlu19 sequence would produce a protein of 499 amino acids with an expected molecular weight (MW) of 57 kDa and pI of 5.05.

Initial attempts to clone the rice Os5BGlu19 gene from the rice seedling RNA were unsuccessful, so a gene optimized for expression in yeast was obtained from gene synthesis at Genscript Corporation. When the optimized gene encoding the mature Os5BGlu19 in pUC57 vector was excised with *NcoI* and *XhoI* and ligated into the pET32a vector, a clone with a correct insert of size about 1.5 kb was obtained, as indicated by digestion with the same restriction enzymes (Figure 4.1).

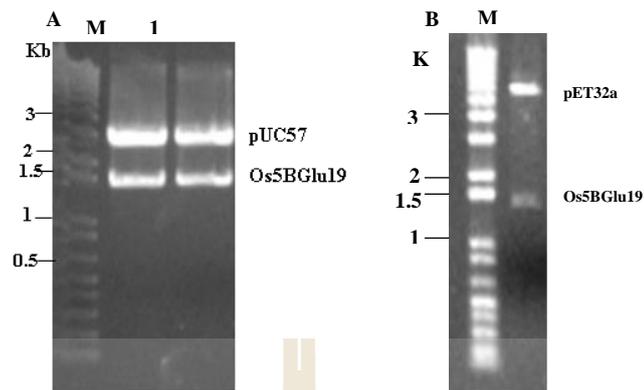


Figure 4.1 Agarose gel electrophoresis analysis of cloning optimized Os5BGlu19 into pET32a. Lane M, GeneRuler 1kb DNA ladder; (A) lanes 1-2, the digest of the optimized Os5BGlu19 in pUC57 with *NcoI* and *XhoI* and (B) lane 1, the digestion of pET32a/Os5BGlu19 with the same restriction enzymes.

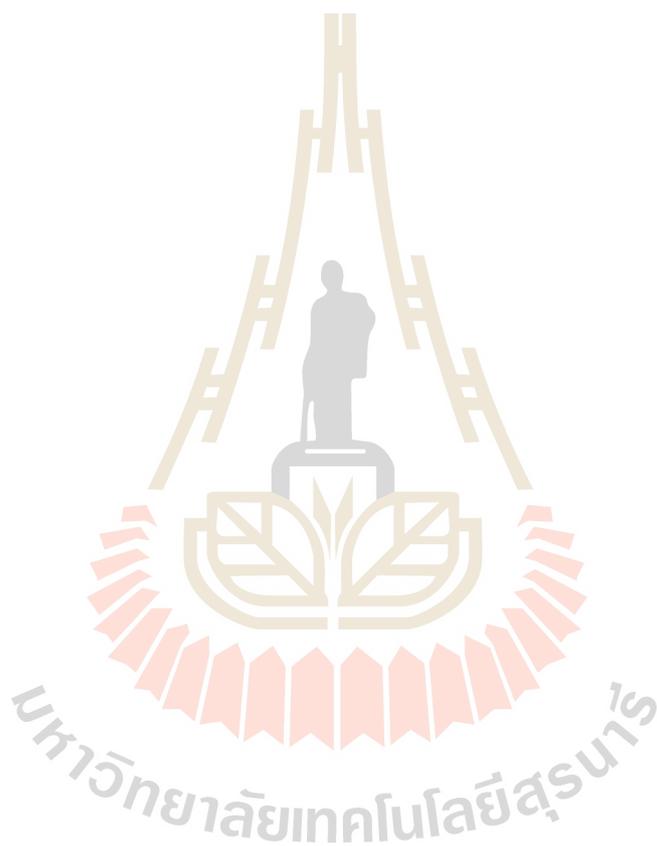
The amino acid alignment of Os5BGlu19 was 53% identical to rice Os9BGlu31 transglucosidase, 52% to anthocyanin 7-O-glucosyltransferase (AA7GT) from delphinium (*Delphinium grandiflorum*), 47% to anthocyanin 5-O-glucosyltransferase (AA5GT) from carnation (*Dianthus caryophyllus*) and 42% to rice Os3BGlu6, respectively, as shown in Figure 4.2 (Seshadri et al., 2009; Luang et al., 2013; Matsuba et al., 2010). In the GH1 family 1, the highly conserved peptide motifs TFNEP and TENG containing glutamic acids that act as the catalytic acid/base and nucleophile, respectively, play critical roles in the glycosidase activity (Sanz-Aparicio et al., 1998). The acyl glucoside dependent anthocyanidin glucosyl transferases and Os9BGlu31 have the same sequence, HENG, around the catalytic nucleophile, but these enzymes function to transfer glucose with different substrate specificities. In contrast,

the most similar enzyme with a solved crystal structure, Os3BGlu6 β -glucosidase, contains the SENG sequence at the catalytic nucleophile and was found to show no significant transglucosylation activity (Seshadri et al., 2009). However, its X-ray crystal structure with 2-deoxyl-2-fluoroglucoside was used in prediction of the homology model structures of Os5BGlu19 and Os9BGlu31.

To characterize the function of Os5BGlu19, we expressed the full-length mature Os5BGlu19 in both *E. coli* and *P. pastoris* systems. pET32a/Os5BGlu19 was transformed into BL 21(DE3) and Origami B(DE3) *E. coli* strains and Os5BGlu19 expression conditions were optimized, including testing a variety of temperatures, concentrations of IPTG and induction times. SDS-PAGE analysis of fractions from expression in strain Origami B(DE3) which was induced at different IPTG concentrations at 20 °C for 18 h is shown in Figure 4.3. There were similar patterns of protein in SDS-PAGE of extracts of cells induced at different temperatures and in the two types of host cells. Os5BGlu19 formed inclusion bodies during expression in *E. coli*, while no soluble protein band was detected by western blot with specific anti-Os5BGlu19 anti-peptide antiserum (Figure 4.3).

Moreover, no activity could be detected in enzyme assays of the soluble cell lysate with 2 mM of 4NPGlc in 50 mM sodium acetate, pH 5.0, suggesting that full-length Os5BGlu19 β -glucosidase was not expressed in soluble, active form. 4NPGlc is a convenient synthetic chromogenic substrate to determine the presence of most β -glucosidases during expression and purification steps (Ketudat Cairns et al., 2015). However, some isoenzymes β -glucosidases are not able to hydrolyze 4NPGlc, for example, Arabidopsis scopolin β -glucosidases AtBGLU21, AtBGLU22, and AtBGLU23 could not hydrolyze either 4NPGlc or 2NPGlc; or Sorghum dhurrinase-1

catalyzes hydrolysis of dhurrin, but not 4NPGlc. (Ahn et al., 2010; Hosel et al., 1987). Matsuba et al., 2010 also indicated that 4NPGlc cleavage was not detectable for AAGTs, glucotransferases that could catalyze transglucosyation but not hydrolysis, even though they belong to the glycoside hydrolase family GH1.



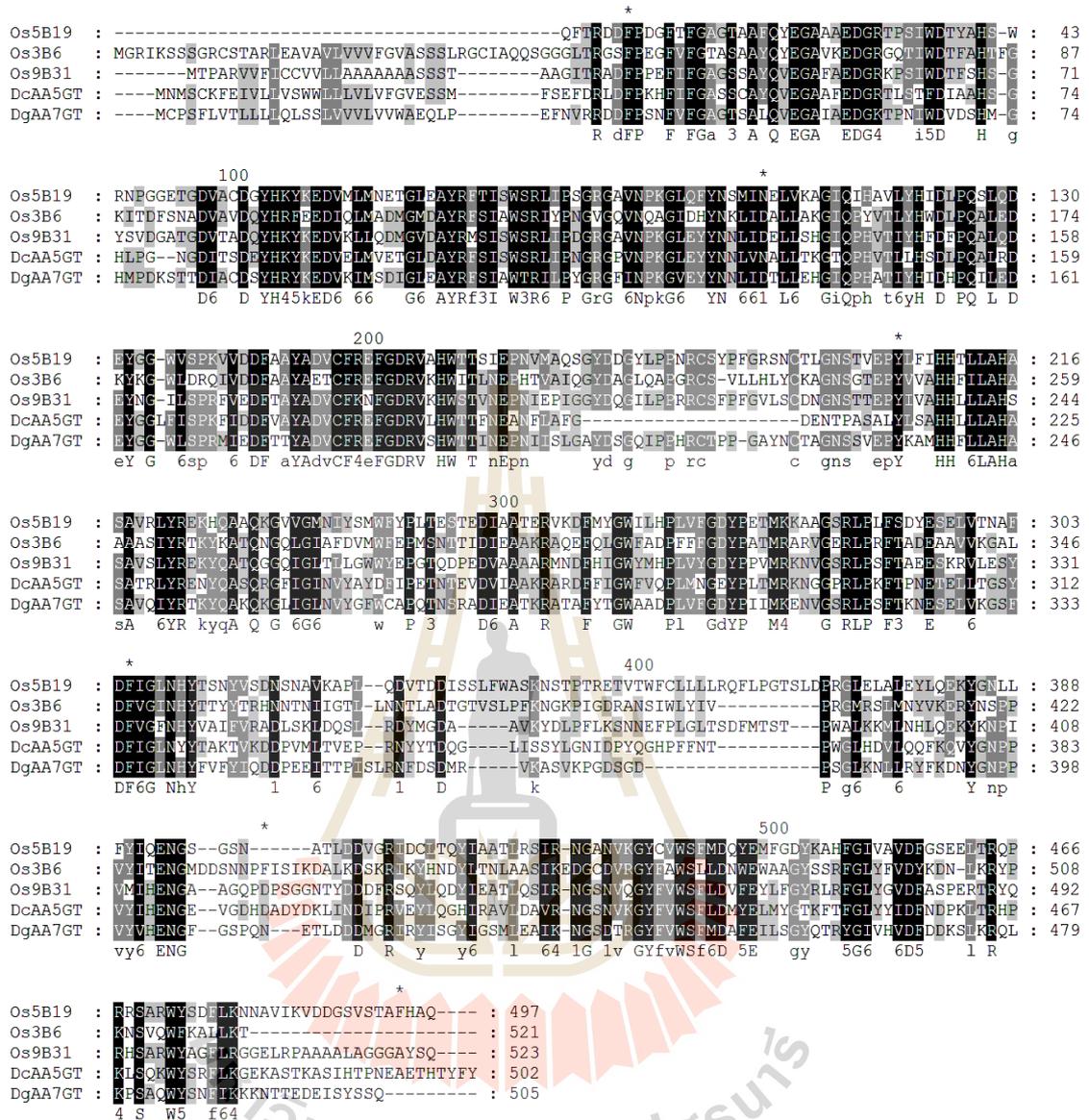


Figure 4.2 Amino acid sequence alignment of predicted rice β -glucosidase Os5BGlu19, rice Os3BGlu6, rice Os9BGlu31 transglucosidase and acyl-glucose-dependent glucosyltransferases from the petals of carnation and delphinium (DcAA5GT and DgAA7GT).

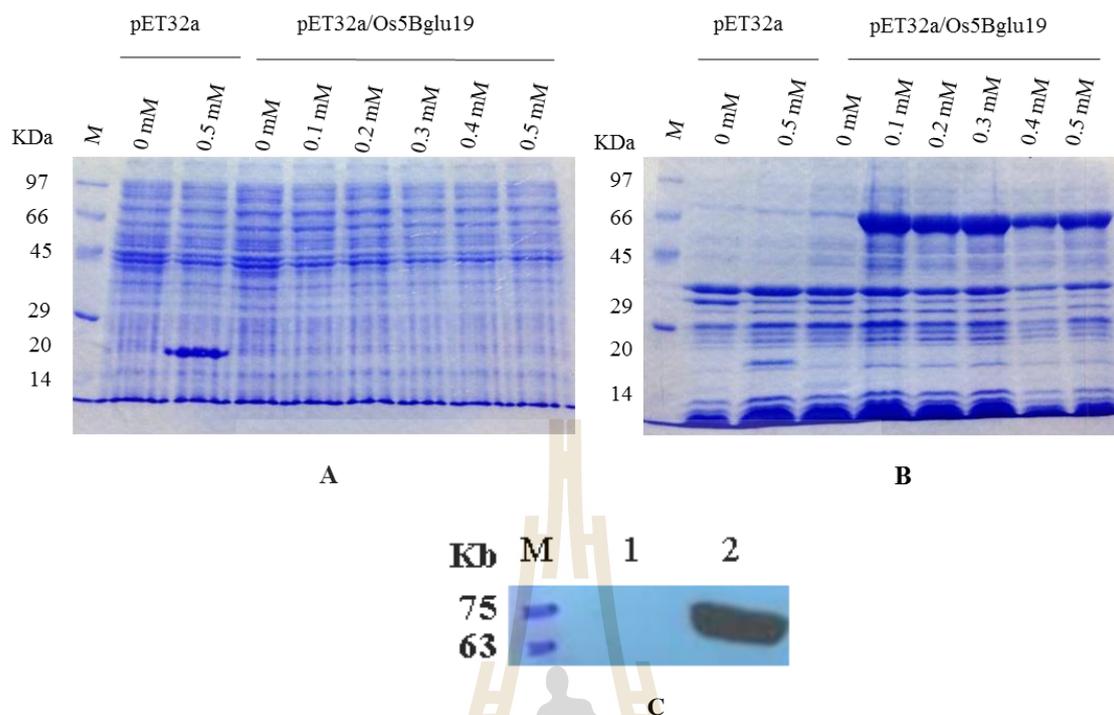


Figure 4.3 SDS-PAGE analysis and western blot of Os5BGlu19 expression. SDS-PAGE analysis of soluble (A) and insoluble (B) fractions of crude extracts of pET32a/Os9BGlu19 containing *E. coli* strain Origami B(DE3) cells that were induced with 0 - 0.5 mM IPTG. pET32a containing cells were also induced with 0 and 0.5 mM IPTG and analyzed in the same manner for comparison. (C) Os5BGlu19 in soluble and insoluble fractions of crude extracts of Origami B(DE3) *E. coli* containing pET32a/Os5BGlu19 induced with 0.4 mM IPTG were detected by western blot analysis with anti-Os5BGlu19 anti-peptide antiserum and peroxidase conjugated goat anti-mouse IgG with ECL chemiluminescent substrate. Lane M is pre-stained protein marker; lane 1, soluble fraction; and lane 2, insoluble fraction.

4.1.2 Cloning and expression of pPICZ α BNH8/Os5BGlu19/eGFP in *P. pastoris*

In order to construct a *P. pastoris* expression vector with Os5BGlu19 in pPICZ α BNH8/eGFP, Os5BGlu19 was amplified with *Pfu* polymerase from the 5' AOXFwd and BGlu19XbaRev primers (Table 3.1) to incorporate *Pst*I and *Xba*I sites. After the PCR product was inserted into pPICZ α BNH8/eGFP, agarose gel analysis showed the correct size for Os5BGlu19 (~1.5 kb) and Os5BGlu19/eGFP (2.3 kb) bands obtained in the digestion of pPICZ α BNH8/Os5BGlu19/eGFP with *Pst*I and *Xba*I or *Pst*I and *Sal*I, respectively (Figure 4.4). Os5BGlu19 was expressed in *P. pastoris* to overcome the problem of only insoluble protein being produced in the *E. coli* system. *P. pastoris* was developed as a heterologous protein expression system using the strong and tightly regulated AOX1 promoter (Cregg et al., 1985). Unfortunately, we found that *P. pastoris* could secrete other enzymes, which could release 4-nitrophenol from 4NGlc in the media in 5-7 days induction and the enzyme with only a His-tag could not be clearly identified when expressed in this system (data not shown). Therefore, the Os5BGlu19 insert was cloned in-frame with a C-terminal eGFP fusion to facilitate expression screening. The presence of eGFP could also show the protein localization in the media of cell culture or inside the pichia cells. The pPICZ α BNH8/Os5BGlu19/eGFP plasmid was then transformed into *P. pastoris* SMD1168H and the initial screen was conducted on YPD plates containing the antibiotic zeocin.

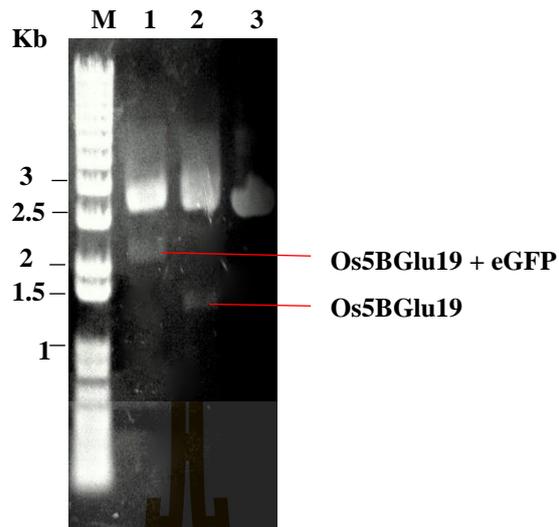


Figure 4.4 Agarose gel electrophoresis analysis of pPICZ α BNH8/Os5BGlu19/eGFP double restriction enzyme digests. Lane M is DNA ladder, pPICZ α BNH8/Os5BGlu19/eGFP digested with *Pst* I and *Sal* I (lane 1) and *Pst*I, *Xba*I (lane 2) and uncut pPICZ α BNH8/Os5BGlu19/eGFP (lane 3).

Twenty colonies were grown in BMMY media and protein expression induced with 1% methanol for 4 days at 20 °C. Twenty clones were examined for fluorescence in the media for the first and fourth day of expression, as shown in Figure 4.5. The Os5BGlu19-eGFP in the *P. pastoris* cells was visualized by fluorescence microscopy and confocal microscope using the wavelengths at 488 nm for excitation and 509 nm for emission to detect eGFP. Green (508) fluorescence intensity was clearly observed inside the cells, proving that the measured fluorescence of the cell pellets from intracellular eGFP from the Os5BGlu19-eGFP fusion protein. Moreover, nuclear staining with DAPI was mixed with the pichia cells to compare the intracellular accumulation to the location of the DNA in the nucleus. No band of secreted

Os5BGlu19 protein was found in western blot analysis of the media. Only crude cell extract contained the diffuse bands expected for heterogeneously glycosylated glycoprotein binding with the specific anti-peptide anti-Os5BGlu19 antiserum, as shown in Figure 4.6. When protein was expressed with a larger scale to purify Os5BGlu19 from *P. pastoris* cells, the protein failed to bind in IMAC, since the western blot could detect protein in crude cell extracts and IMAC flow through fractions, but could not detect any protein band in the eluted fractions (Figure 4.7).

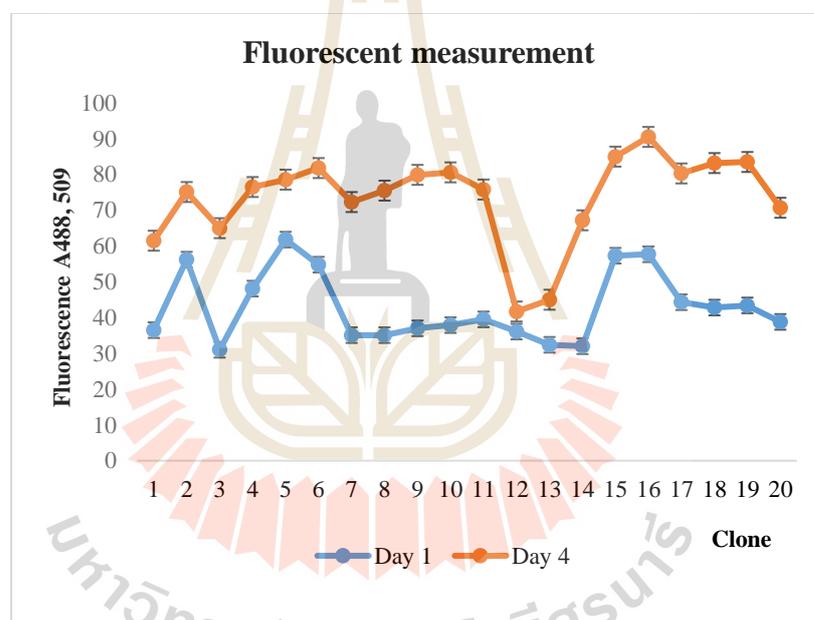


Figure 4.5 Fluorescence measurement of clones from twenty colonies that were expressed in BMMY with 1% methanol induction for 4 days. The fluorescence was measured with 488 nm excitation and 509 nm emission in a fluorescence spectrometer.

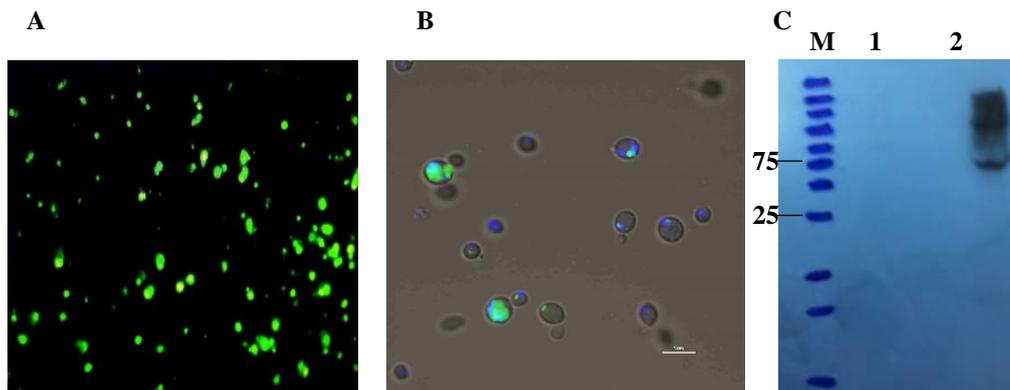


Figure 4.6 Detection of Os5BGlu19-eGFP in *P. pastoris* cells by fluorescence and confocal microscopy and by western blot analysis of Os5BGlu19 after pPICZ α B/Os5BGlu19/eGFP was expressed in *P. pastoris*. **A**, *P. pastoris* cells under fluorescence microscope. **B**, *P. pastoris* cells expressing Os5BGlu19-eGFP viewed in a confocal microscope after DAPI staining (wavelength 358-461 nm). **C**, Western blot analysis of Os5BGlu19 expression in *P. pastoris*. Lane M, pre-stained protein marker; lane 1, media from the culture; and lane 2 crude extract from pichia cells.



Figure 4.7 Western blot analysis of Os5BGlu19 fractions from attempted purification by IMAC. W0, W5, W10 and E1-E8 represent wash fractions with 0, 5, 10 mM of imidazole and elute fractions with 250 mM of imidazole, respectively.

4.2 Cloning, expression and purification of truncated Os5BGlu19

Initial attempts to express of Os5BGlu19 in *E. coli* and *Pichia pastoris* resulted in low or no soluble protein in *E. coli* and failure to bind to the IMAC column in *P. pastoris*. To overcome the problem of a probable protease cleavage site between the His-tags and protein and possible vacuole-targeting sequences in the *Pichia* system, we endeavored to remove unnecessary sequence from the N- and C- termini of Os5BGlu19. We predicted a homology model structure of Os5BGlu19 (Figure 4.8) and truncated both termini to remove any extra sequence that did not match the Os3BGlu6 β -glucosidase as a template for this structure. TrOs5BGlu19 (~1.3 kb) was amplified and inserted into the expression cassette of pET32a/TrOs5BGlu19, as shown in Figure 4.9.

Since truncation might also improve solubility of protein expressed in *E. coli*, we first tried to express the truncated protein construct in *E. coli*. The pET32a/TrOs5BGlu19 expression vector was transformed in BL21(DE3) and Origami B(DE3) and expression optimized by inducing with different of IPTG concentrations from 0.1 to 0.5 mM at 20 °C for 16 h. The soluble protein (~66 kDa) was observed by SDS-PAGE and western blotting using anti-peptide anti-Os5BGlu19 antiserum (Figure 4.10). The gels showed no clear relationship between IPTG concentration and soluble and insoluble protein. However, based on the western blot of the soluble protein, 0.3 mM IPTG induction of Origami B(DE3) host cells gave the best yield of soluble protein. The recombinant TrOs5BGlu19 was partially purified by IMAC and eluted at the concentrations of imidazole from 40–260 mM. All of purified protein fractions were check by SDS-PAGE and confirmed by western blotting (Figure 4.11).

TrOs5BGlu19 fusion protein with N-terminal thioredoxin, His-tag and S-tags, which was predicted to have a molecular weight of 72 kDa and pI of 5.05, was optimally

expressed in *E. coli* strain Origami B(DE3) induced with 0.3 mM IPTG at 20 °C for 16 h, but there was no 4NPGlc hydrolysis activity detected in the crude cell extract and all of purified fractions. Protein purification by IMAC produced a protein is approximately 70 kDa that was pooled from 40-260 mM imidazole fractions and used for further experiments.

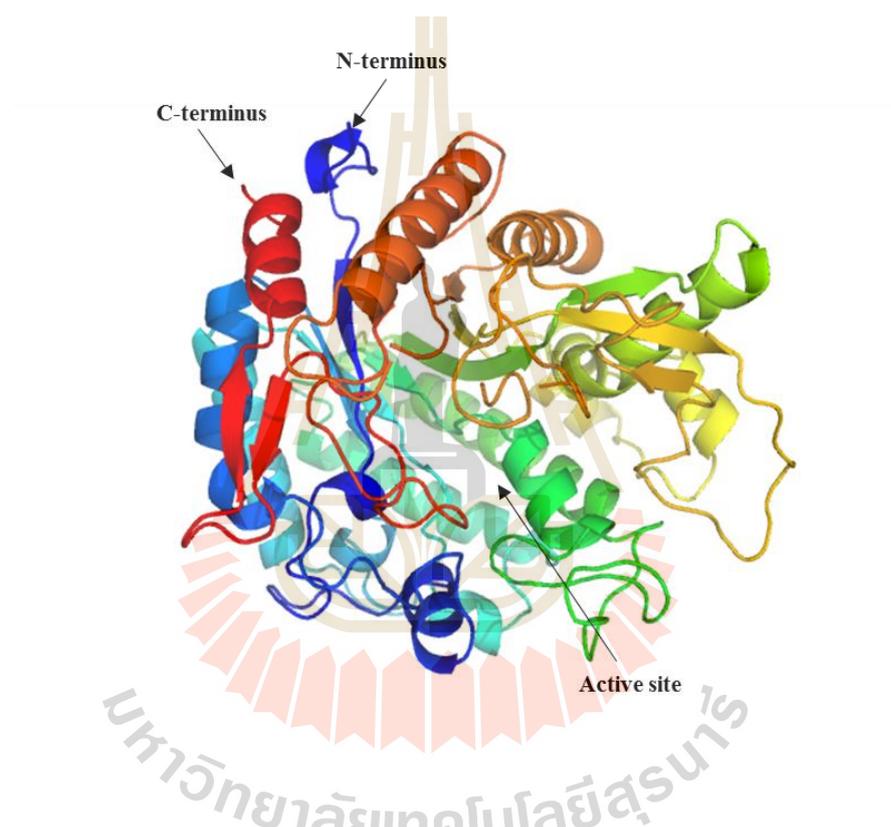


Figure 4.8 A homology model of Os5BGlu19 based on the X-ray crystal structure of the Os3BGlu6 β -glucosidase covalent intermediate with 2-fluoro- α -D-glucoside (PDB code 3GNR).

```

                                *
TrBGlu19 : -----QFTRDDFPDGFTFGAGTAAFQYEGAAAEDGRTPSIWDTYAHSWRNPGETGDVACDGY : 58
2.       : LEVLFQGP|QFTRDDFPDGFTFGAGTAAFQYEGAAAEDGRTPSIWDTYAHSWRNPGETGDVACDGY : 67
                                QFTRDDFPDGFTFGAGTAAFQYEGAAAEDGRTPSIWDTYAHSWRNPGETGDVACDGY

                                80                                *
TrBGlu19 : HKYKEDVMLMNETGLEAYRFTISWSRLIPSGRGAVNPKGLQFYNSMINELVKAGIQIHAVLYHIDL : 125
2.       : HKYKEDVMLMNETGLEAYRFTISWSRLIPSGRGAVNPKGLQFYNSMINELVKAGIQIHAVLYHIDL : 134
                                HKYKEDVMLMNETGLEAYRFTISWSRLIPSGRGAVNPKGLQFYNSMINELVKAGIQIHAVLYHIDL

                                160                                *
TrBGlu19 : QSLQDEYGGWVSPKVDDFAAYADVCFREFGDRVAHWTTISIEPNVMAQSGYDDGYLPPNRCSPYFGR : 192
2.       : QSLQDEYGGWVSPKVDDFAAYADVCFREFGDRVAHWTTISIEPNVMAQSGYDDGYLPPNRCSPYFGR : 201
                                QSLQDEYGGWVSPKVDDFAAYADVCFREFGDRVAHWTTISIEPNVMAQSGYDDGYLPPNRCSPYFGR

                                240
TrBGlu19 : SNCTLGNSTVEPYLFIHHTLLAHASAVRLYREKHQAAQKGVVGMNIYSMWFYPLTESTEDIAATERV : 259
2.       : SNCTLGNSTVEPYLFIHHTLLAHASAVRLYREKHQAAQKGVVGMNIYSMWFYPLTESTEDIAATERV : 268
                                SNCTLGNSTVEPYLFIHHTLLAHASAVRLYREKHQAAQKGVVGMNIYSMWFYPLTESTEDIAATERV

                                *                                320
TrBGlu19 : KDFMYGWIHLHPLVFGDYPETMKKAAGSRLPLFSDYSESELVTNADFDFIQLNHYTSNYVSDNSNAVKAP : 326
2.       : KDFMYGWIHLHPLVFGDYPETMKKAAGSRLPLFSDYSESELVTNADFDFIQLNHYTSNYVSDNSNAVKAP : 335
                                KDFMYGWIHLHPLVFGDYPETMKKAAGSRLPLFSDYSESELVTNADFDFIQLNHYTSNYVSDNSNAVKAP

                                *                                400
TrBGlu19 : LQDVTDDISSLFWASKNSTPTRETVTWFCLLLLRQFLPGTSLDPRGLELALAYLQEKYGNLLFYIQE : 393
2.       : LQDVTDDISSLFWASKNSTPTRETVTWFCLLLLRQFLPGTSLDPRGLELALAYLQEKYGNLLFYIQE : 402
                                LQDVTDDISSLFWASKNSTPTRETVTWFCLLLLRQFLPGTSLDPRGLELALAYLQEKYGNLLFYIQE

                                *
TrBGlu19 : NGSNSATLDDVGRIDCLTQYIAATLRSIRNGANVKGVCVWSFMDQYEMFGDYKAHFGIVAVDFGSE : 460
2.       : NGSNSATLDDVGRIDCLTQYIAATLRSIRNGANVKGVCVWSFMDQYEMFGDYKAHFGIVAVDFGSE : 469
                                NGSNSATLDDVGRIDCLTQYIAATLRSIRNGANVKGVCVWSFMDQYEMFGDYKAHFGIVAVDFGSE

                                480
TrBGlu19 : ELTRQPRRSARWYSDFLKN----- : 479
2.       : ELTRQPRRSARWYSDFLKNNAVIKVDGGSVSTAFHQAQ : 506
                                ELTRQPRRSARWYSDFLKN

```

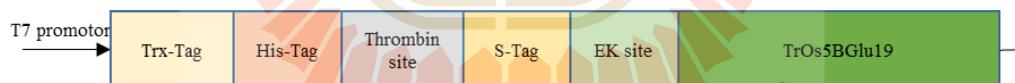


Figure 4.9 Amino acid sequence alignment and construction of TrOs5BGlu19. **A.** Amino acid sequence alignment of mature Os5BGlu19 and the truncated version, TrOs5BGlu19. An artificial protease site was also removed from the N-terminus of the original expression construct. **B.** Schematic diagram of the final expression cassette for Os5BGlu19, showing it is fused to N-terminal thioredoxin, His₆- and S- tags with thrombin and enterokinase (EK) protease sites for removal of the tags.

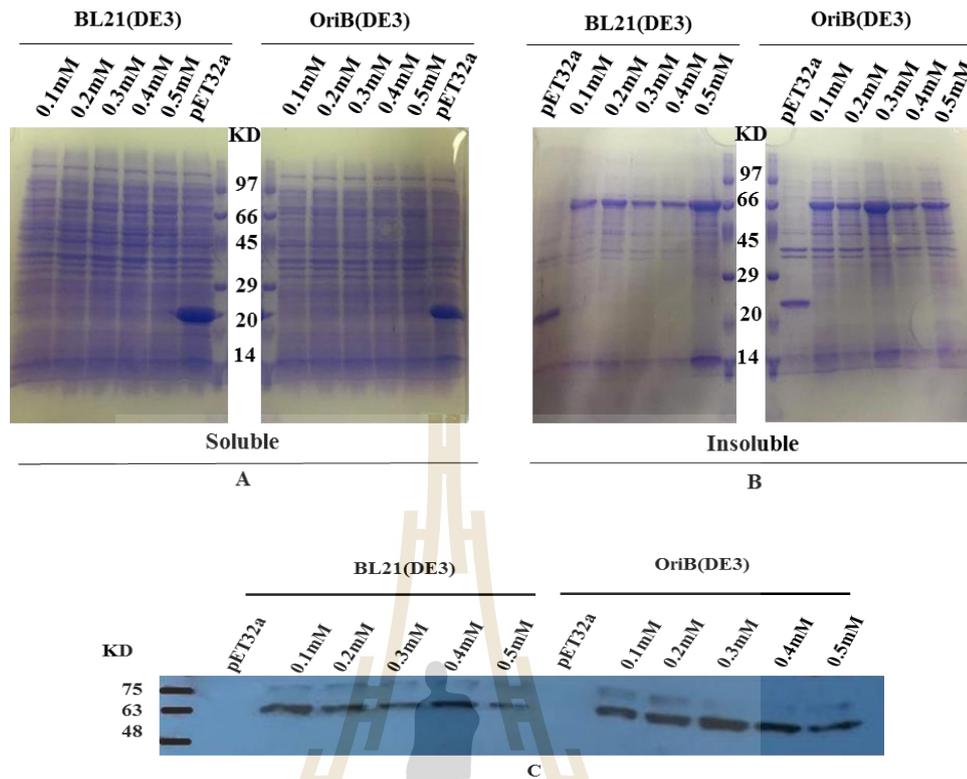


Figure 4.10 SDS-PAGE and western blot analysis of expression of TrOs5BGlu19 in *E. coli* induced with different IPTG concentrations. **A.** SDS-PAGE analysis of soluble protein in extracts of *E. coli* cells containing pET32a/TrOs5BGlu19 were induced with 0.1 – 0.5 mM IPTG and cells containing empty pET32a vector were induced with 0.5 mM IPTG. **B.** SDS-PAGE analysis of insoluble proteins in BL21(DE3) and Origami B(DE3) cells containing pET32a/TrOs5BGlu19. Soluble cell lysates of cells containing control pET32a plasmid induced with 0.5 mM IPTG are included in the first lane of each gel. The gels were stained with Coomassie brilliant blue. **C.** Western blot analysis of soluble TrOs5BGlu19 in the cells described in **A** with specific anti-peptide anti-Os5BGlu19 antiserum.

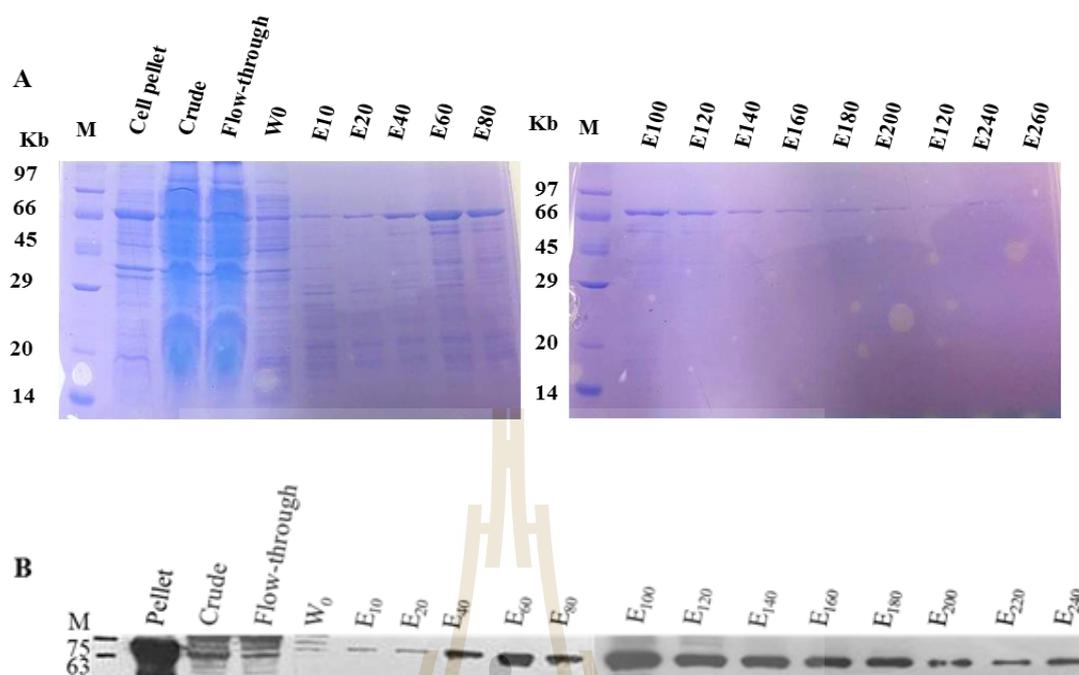


Figure 4.11 TrOs5BGlu19 purification fraction analysis. A. SDS-PAGE of TrOs5BGlu19 IMAC purification fractions. W0 and E10 to E260 represent wash and elution fractions at different imidazole concentrations (with the number representing the imidazole concentration in units of mM). B. Western blot of TrOs5BGlu19 IMAC purification fractions with anti-peptide anti-Os5BGlu19 antiserum.

4.3 Identification of Os5BGlu19 substrates from rice leaf extract

Based on the origin tissues of expressed sequence tag (EST) cDNA clones of *japonica* rice in the database, we could predict the expression of β -glucosidase in particular tissues and different growth stages of rice. There are sources libraries of rice GH1 gene ESTs includes callus, seedling (stem, root and leaf), mature plants (leaf), panicle at flowering and ripening stage, and immature seeds (Opassiri et al., 2006). The Rice Expression Profile Database (RiceXpro) provided the expression profile of

Os5BGlu19, as shown in Figure 4.12. Os5BGlu19 is mostly expressed in young leaves, and is also found in the stem of reproductive and ripening stages and some parts of flowers.

Another GH1 At/Os cluster 6 enzyme, Os9BGlu31 showed high transglucosylation activity with ferulic acid to produce 1-O-feruloyl- β -D-glucose (FAG), which was qualified by UPLC and determined the mass and fragmentations by QToF-MS (Komvongsa et al., 2015). Therefore, the TrOs5BGlu19 enzyme was assayed with rice extracts and ferulic acid as an acceptor in order to identify the activity of Os5BGlu19. Standard reactions containing ferulic acid and 4NPGlc with and without Os9BGlu31 were used to set a standard curve for FAG with four fold serial dilutions that showed a linearity of the ion abundance measured versus concentration (Figure 4.13). The amount of FAG increased in the reactions of Os5BGlu19 with 0.25 mM FA and rice extracts (4 week old, 10 week old leaves and 10 week old stem) (Figure 4.14). Quantification of the peaks in the reactions containing rice extracts showed that the relative abundance of FAG increased over 30-fold in 4-week-old rice leaf extracts when TrOs5BGlu19 was added to the reaction (Table 4.1). The highest levels of FAG were achieved with 10-week-old rice leaf extracts (20.6 mM), although this was only 2.7 times the concentration in extract reactions without enzyme, due to the high basal level of FAG. Nonetheless, the increase in FAG demonstrated the activity of Os5BGlu19 to transfer glucose to FA from glucosyl donors in the rice extract.

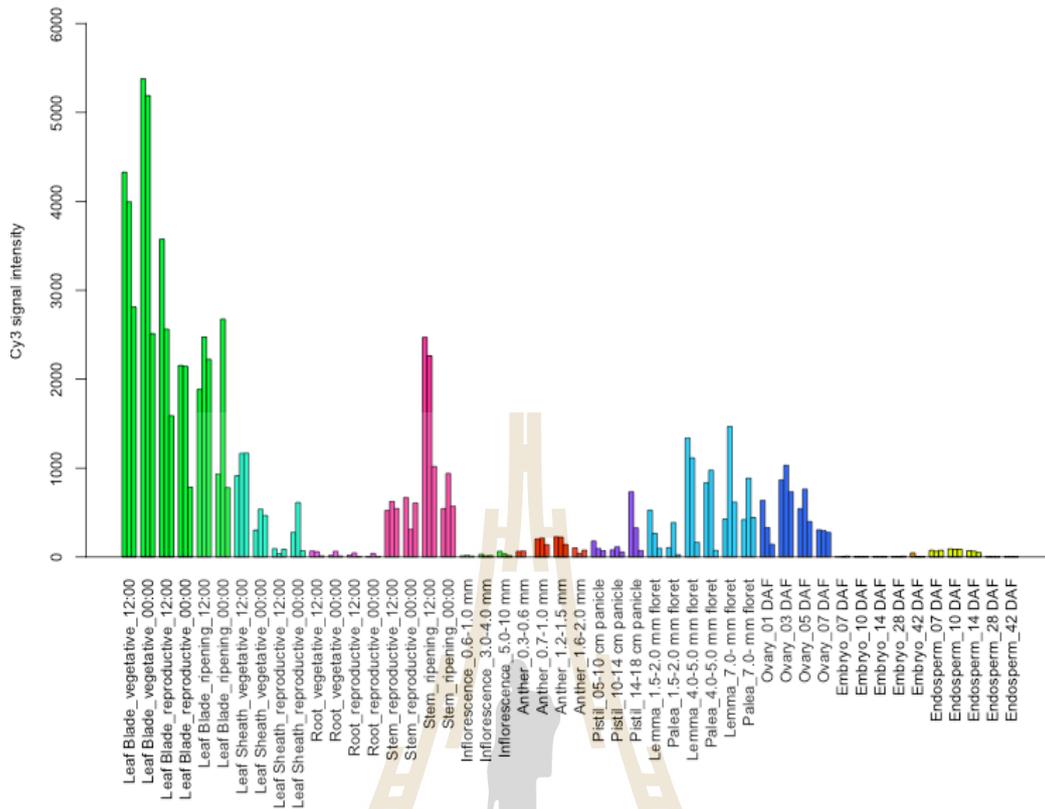


Figure 4.12 Expression profile of Os5BGlu19 in different organs and tissues. The data was collected from the Rice Expression Profile Database.

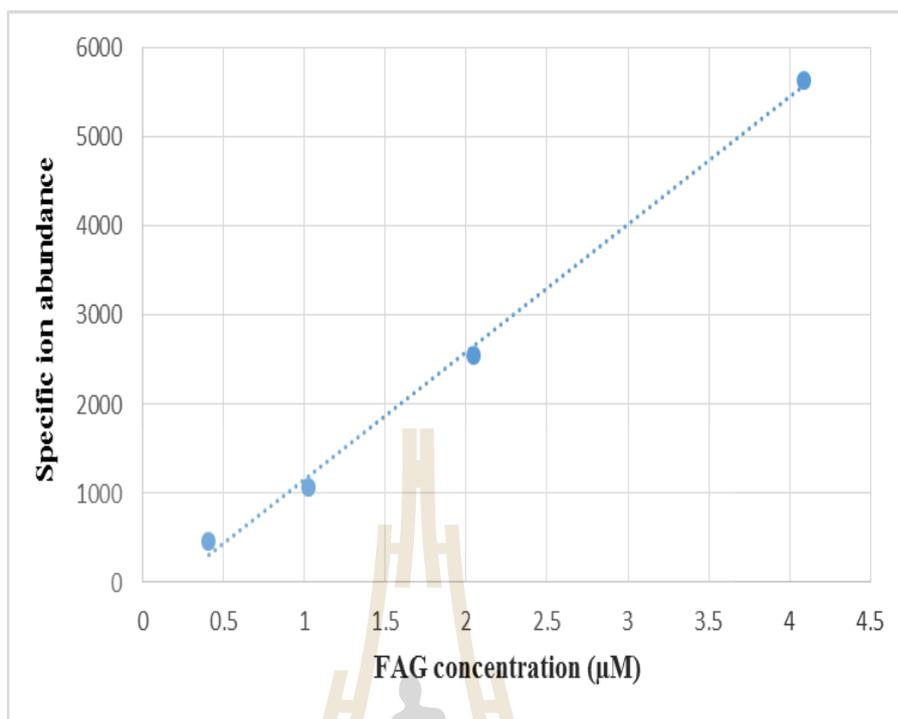


Figure 4.13 The standard curve of specific fragment ion response for FAG concentration at m/z 175 from the parent ion at 355 m/z .

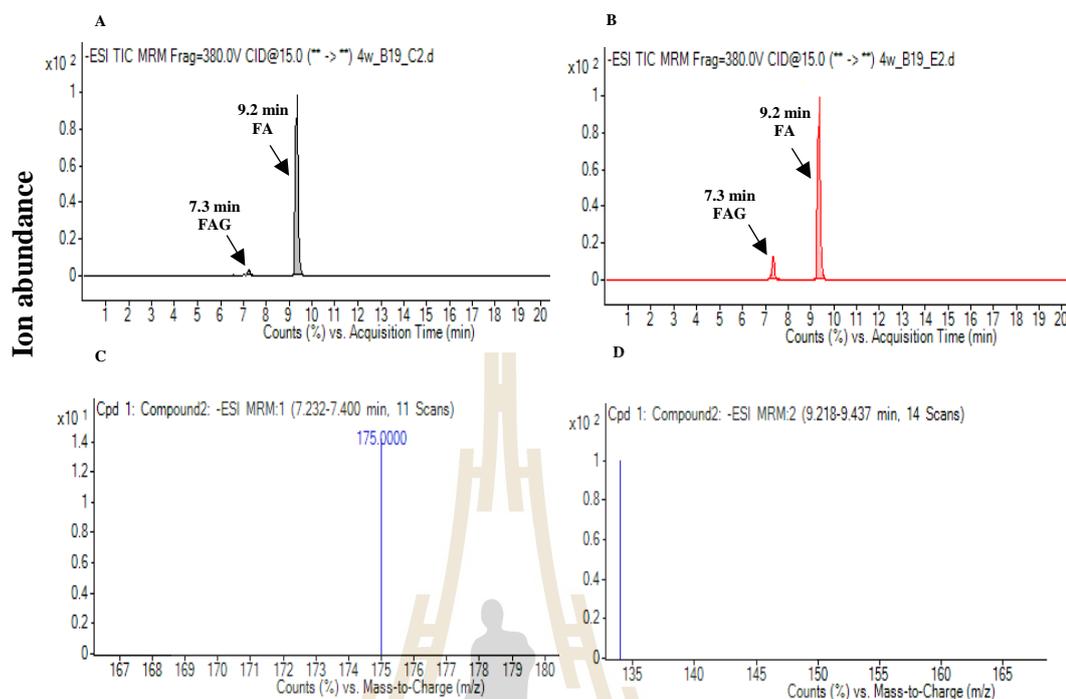
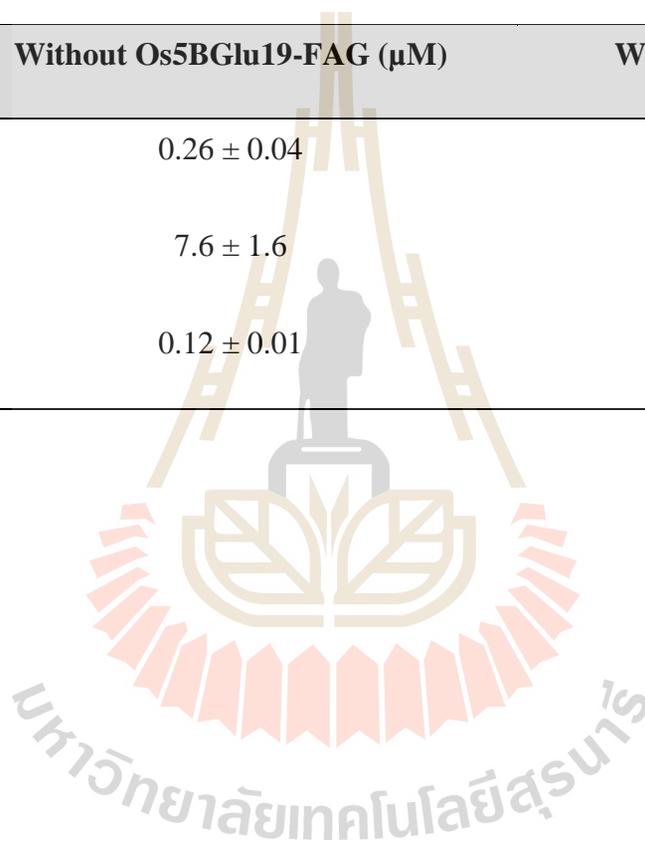


Figure 4.14 Chromatograms and specific fragment ion spectra from UPLC triple quadrupole tandem mass spectrometry in MRM mode chromatograms. The reaction mixtures of rice leaf extract and ferulic acid acceptor without Os5BGlu19 (A) and with Os5BGlu19 (B), The 1-O-feruloyl- β -D-glucose peak elutes at 7.3 min with a fragment ion at 175 m/z generated from the 355 m/z parent ion (C), while the ferulic acid elutes at 9.3 min and was measured from the ion at 134 m/z fragmented from the 193 m/z precursor ion (D).

Table 4.1 Concentrations of 1-O-feruloyl- β -D-glucose ester (FAG) in reactions of Os5BGlu19 with rice extracts and ferulic acid (FA) determined by UPLC-QQQ-MSMS. Reactions were run in triplicate and means and standard deviations are shown.

Sample	Without Os5BGlu19-FAG (μ M)	With Os5BGlu19-FAG (μ M)
4 week old leaves	0.26 ± 0.04	8.9 ± 1.6
10 week old leaves	7.6 ± 1.6	20.6 ± 2.4
10 week old stem	0.12 ± 0.01	0.78 ± 0.08



4.4 Donor substrate purification and identification by LH20 column

In the initial evaluation, Os5BGlu19 is a ferulic acid β -transglucosidase based on reaction with rice extract. To identify any compound in rice extract that could be a donor for Os5BGlu19, the 10 week old rice extract was fractionated by LH20 column and the fractions with compounds detected by TLC were assayed with and without Os5BGlu19 in 50 mM citrate buffer, pH 4.5. The reactions were analyzed by UPLC-QQQ-MS to check the increase of FAG level. There was no significant difference from fractions 10 to 16 in the ion abundance of reactions with and without enzyme. However, the FAG amount was increased dramatically in fraction 18 (12 folds) between Os5BGlu19 and control reactions (Figure 4.15). To further purify the donor substrate, we pooled fractions that resulted in increased FAG levels in the assay in two groups, which were 17, 18 and 19 in the first and 20 and 21 in the second. These pools were separated again by TLC and the individual fractions were eluted with alcohol and analyzed with the same condition as LH20 fractions determination. There were fluctuations in the peak area of FAG on UPLC-QQQ-MSMS in both pooled fraction kinds after separating by TLC, as shown in Figure 4.16. Fraction 7 and fraction 9 resulted in slight differences between control and BGlu19 reactions, but they were not statistically significant. In contrast, most of the fractions had a low FAG ion abundance, which were comparable to those without BGlu19. The ferulic acid β -transglucosidase activity was reduced, suggesting that the glycosyl donor disappeared during the purification steps. The donor substrate from rice extract could be degraded or reduced to too low amount to determine by UPLC-QQQ-MS.

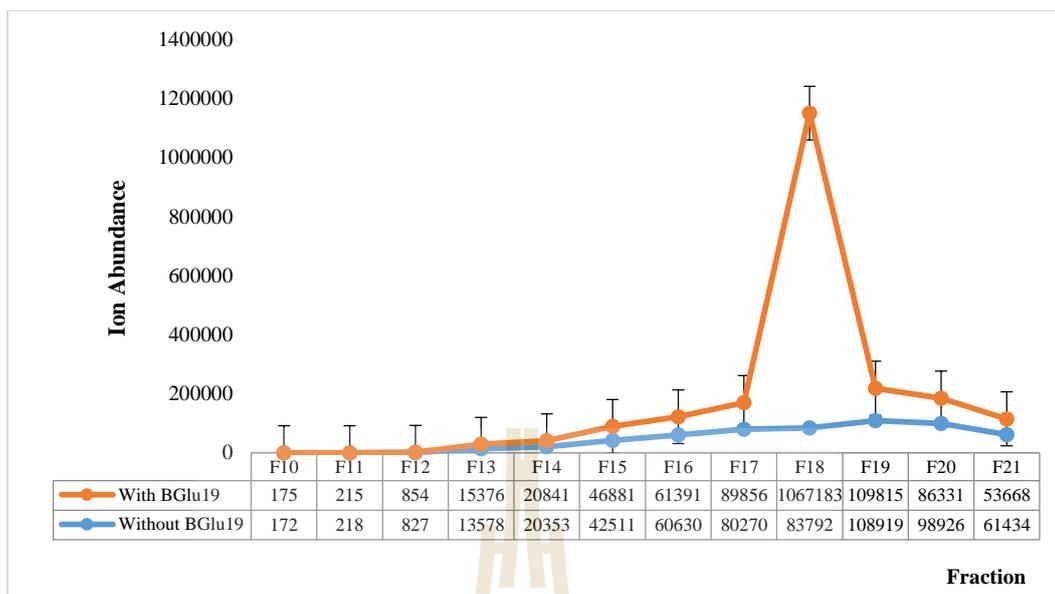


Figure 4.15 FAG ion abundances of reactions containing one of 11 elution fractions from LH20 column chromatography determined by UPLC-QQQ-MS. The fractions were assayed with 0.25 mM ferulic acid with 5 μ g Os5BGlu19 and without enzyme as a control reaction. The FAG ion abundances were collected as those of the 175 m/z ion fragmented from the 355 m/z precursor ion in the negative ion mode.

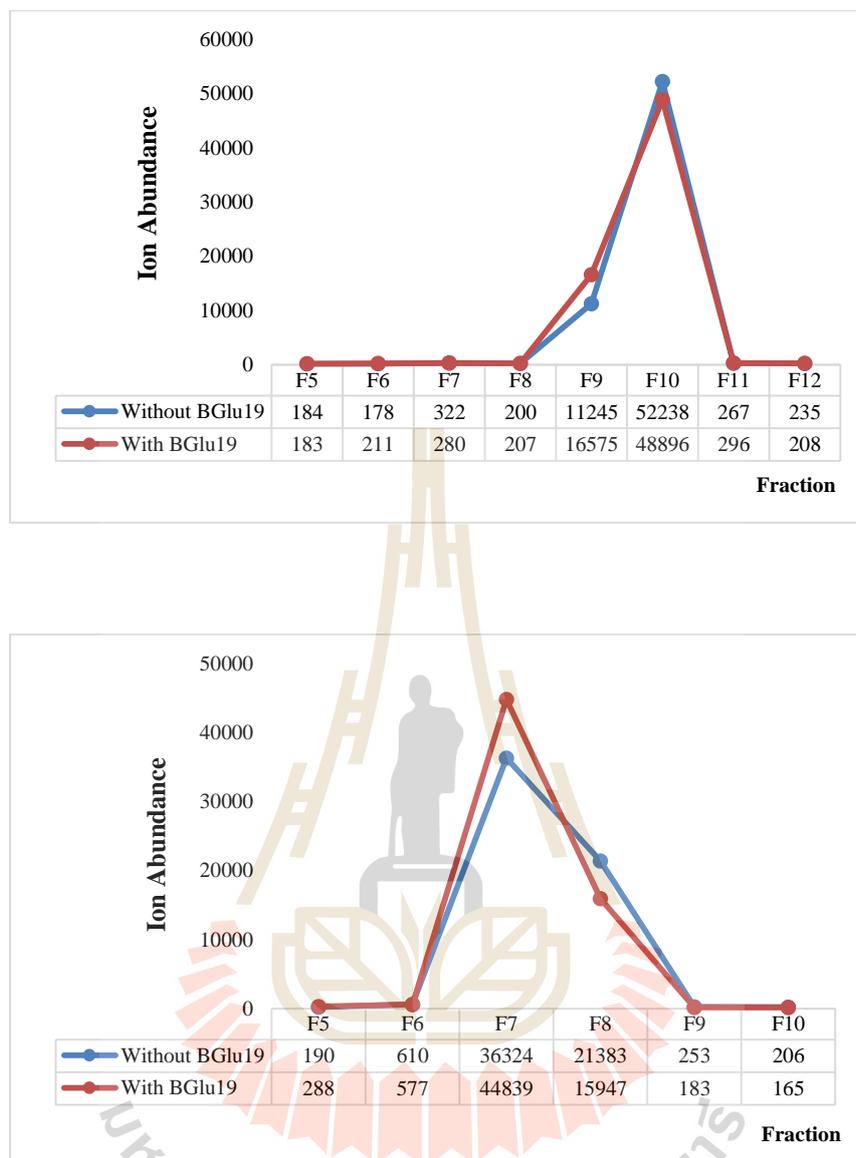


Figure 4.16 FAG ester peak ion abundances in assays of fractions extracted from TLC determined by UPLC triple quadrupole-MS/MS. (A) F5 to F10 were separated from pooling of F17, F18 and 19; and (B) F5 to F12 were separated from the pool F20 and F21 from LH20 chromatography, respectively. These fractions were assayed with 0.25 mM ferulic acid in control and Os5BGlu19 reactions. The ion abundances for FAG were collected at 7.3 min and 175 m/z as product ion of FAG.

There may be several substrates all in low amounts, making it difficult to determine the biological substrates of Os5BGlu19. The donor substrates could work with other cofactors or complex reactions, which may be catalyzed by transglucosidase enzymes. Furthermore, a large amount of material may be required for substrate purification. Brassinolide is plant growth promoting englongation of bean internodes. The extraction of 40 kg of *Brassica napus* L. (rape) pollen produced 4 mg brassinolide, which was used in order to determine biological activity and structure of this hormone (Mitchell et al., 1970; Grove et al., 1979). So, the twenty grams or rice leaves that was extracted may not be enough to continue with several purification steps.

Although the truncated Os5BGlu19 could be expressed as a soluble protein form as a thioredoxin-His tagged fusion protein in *E. coli*, no activity could be detected in standard β -glucosidase assay with 4NPGlc, or transglucosidase assay using 4NPGlc as a glucosyl donor. However an increase in FAG in rice extracts upon addition of TrOs5BGlu19 confirmed that it could act as a transglucosidase to glycosylate ferulic acid to FAG using glycosyl donors. Development of more efficient activity assays may be required to understand the biological and biochemical functions of Os5BGlu19.

4.5 Characterization of relative activities of Os9BGlu31 transglucosidase wild type and its mutants

4.5.1 Expression and purification of Os9BGlu31 wild type and its mutants

The expression vector for wild type Os9BGlu31 was formerly constructed from a rice cDNA in pET32a/DEST with an enterokinase site to cleave the fusion tag from Os9BGlu31 protein. In this experiment, we constructed an expression

vector including a tobacco etch virus (TEV) site, since TEV protease demonstrates specific proteolytic activity under a wide range of parameters, including salt concentrations, temperatures and pH values (Polayes et al., 1998) and can be produced in the laboratory. The pET32a/DEST/TEV/Os9BGlu31 wild type was confirmed by DNA sequencing and used as a template to change amino acid residues by site directed mutagenesis. In previous a study, we proved active site cleft residues surrounding the aglycon and acceptor binding site play an important role in the transglycosidase specificity of Os9BGlu31 (Komvongsa et al., 2015). Particularly, the mutations of tryptophan at the position 243 (W243) to several amino acids suggested that increasing the polarity and decreasing the size of amino acid side chains around the active site could increase the transglycosylation of phenolic and flavonoid compounds. In order to further clarify this case, we mutated this W243 site to additional amino acids, including basic (H, K, R), non-polar (G, V) and polar (Q, S) residues to evaluate the Os9BGlu31 transglucosidase activity.

In the experiment, we expressed wild type, the previously generated mutant W234N and six new mutants in Origami B(DE3). The proteins were first purified by TALON® IMAC resin and the fusion tag was cleaved from the protein by TEV and Os9BGlu31 without tag (~50 KD) was collected by a second IMAC purification. Proteins could be obtained with more than 90% purity, as indicated by SDS-PAGE (Figure 4.17).

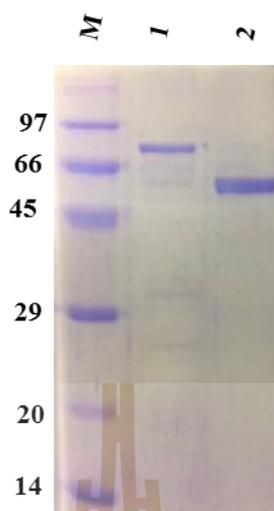


Figure 4.17 Coomassie blue SDS PAGE gel of Os9BGlu31 W243V throughout purification. Lane M, protein marker; lane 1, the N-terminal thioredoxin/His6-tagged Os9BGlu31 W243V fusion protein after the first IMAC step; lane 2, the purified Os9BGlu31 W243V mutant protein after cleavage of the fusion protein with TEV protease and removal of the tag with a 2nd IMAC step. 1 μ g protein was loaded in each of lane.

4.5.2 Relative activities of Os9BGlu31 wild type and its mutants with various acceptors

Os9BGlu31 wild type and W243N mutant enzymes were characterized previously for transglucosylation to transfer the glucosyl moiety from a donor substrate, such as 4NPGlc, to an acceptor (Luang et al., 2013; Komvongsa et al., 2015). Various aromatic glycosides and glucose esters were produced in the reaction of Os9BGlu31 transglycosidase in the presence of acceptors. The relative activities of Os9BGlu31 wild type and its mutants were measured by the quantification of 4NP released from 4NPGlc

in the reaction of different acceptors by UPLC. W243 mutations had differential effects on glycosylation of twenty-two acceptors, including water (for hydrolysis) determined by the relative rate of 4NP (Figure 4.18). Most of the Os9BGlu31 variants gave higher rates of 4NP release with most of the phenolic compound than by hydrolysis with water. The wild type Os9BGlu31 and most of mutants prefer ferulic acid among the acceptor substrates tested, while the W243K mutant has the highest activity with chrysin and W243Q accepted indole-3-acetic acid as the best acceptor. W243N also has highest activity with 4-coumaric acid and 6-hydroxyflavone, and similar activity with other compounds compared to ferulic acid.

In the relative activity analysis, 4NP release was less than water for certain variants with certain acceptor substrates, but glucoconjugates were still formed by transglucosylation, such as in flavanones and phytohormes. For example, there was a low amount of 4NP released by the reaction between Os9BGlu31 and gibberellin A₄, however the UPLC chromatogram analysis showed the GA₄-glucose ester product obtained by transglycosylation (Figure 4.19). Gibberellin A₄ glucose ester (GA₄GE) was eluted at 11.15 min and absorbed at 210 nm measured by the DAD detector. The product was compared with the standard GA₄GE which was synthesized by Hua et al, 2013.

Furthermore, the Os9BGlu31 transglucosidase wild type and mutants were able to transfer glucose to chloramphenicol, and potentially improve its solubility and pharmaceutical properties, as shown in Figure 4.20. Chloramphenicol glucoside has retention time at 9.2 min and was detected at 280 nm. Of the Os9BGlu31 variants tested, the highest product yield was obtained in the reaction with W243N, and lowest amount obtained with wild type, based on the peak areas.

Interestingly, several acceptors that have more than one hydroxyl group accepted glucose from the W243 mutants at multiple positions to produce multiple glucoconjugates compounds, such as 4-hydroxybenzoic acid (Figure 4.21), vanillic acid (Figure 4.22), luteolin (Figure 4.23) and keampferol. There are no glucoside and glucose ester standards to determine the kinds of extra products in this experiment, so we the elution positions of products were used to tentatively identify them. It was assumed that most glycoside compounds elute before the glucose ester product, due to their higher polarity caused by the unconjugated carboxyl group which is blocked by glucose in the esters. Further clues to the identity can be obtained by LCMSMS on an electrospray ionization (ESI)-QToF MSMS (Komvongsa et al., 2015), but we were unable to do the UPLC and QToF MS on the same column and condition. Although, the results are similar with Komvongsa et al., 2105, Os9BGlu31 wild type and mutants were tested for transglucosylation of a greater number of acceptors, especially secondary metabolite compounds from plants.

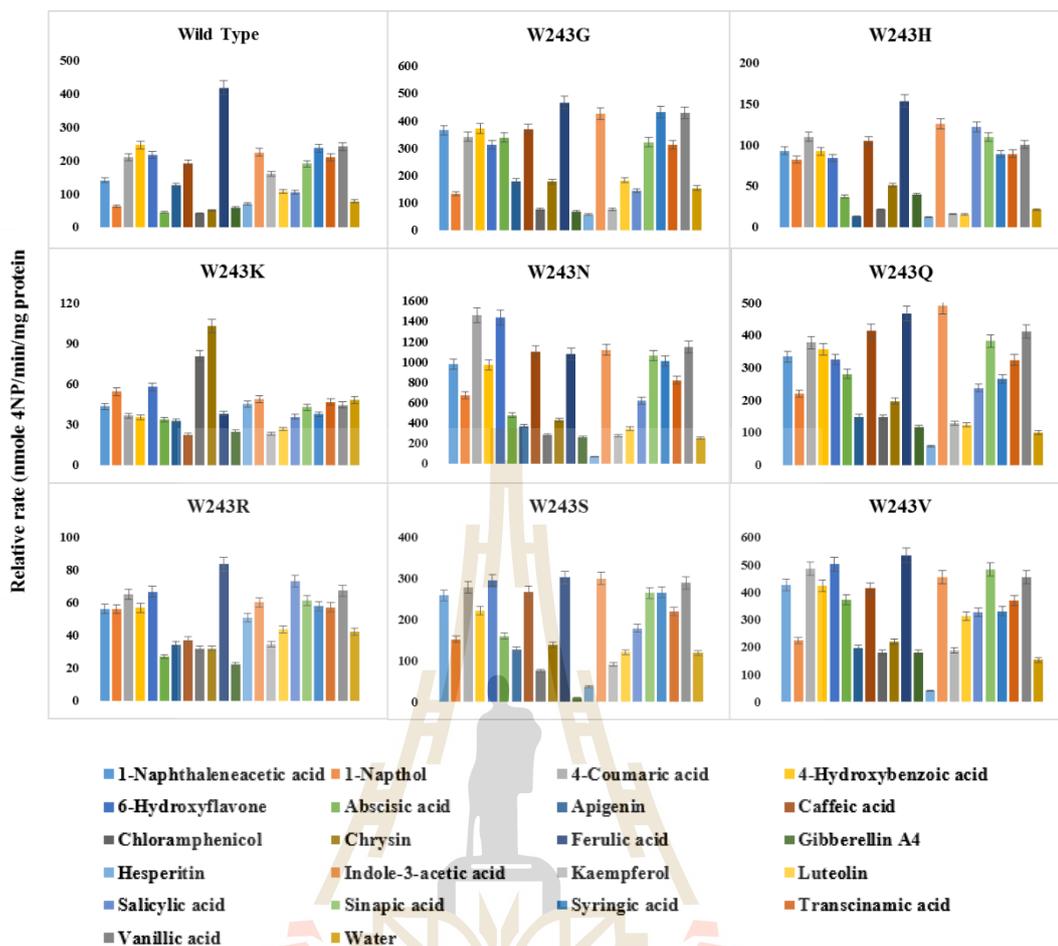


Figure 4.18 Relative rates of Os9BGlu31 enzyme and W243 mutants with various acceptors. Activities were assayed with 5 mM 4NPGlc as donor and 0.25 mM acceptor substrates. The relative activities with twenty-two different acceptors, including water for hydrolysis reactions, are shown in for wild type and variants with different W243 mutations.

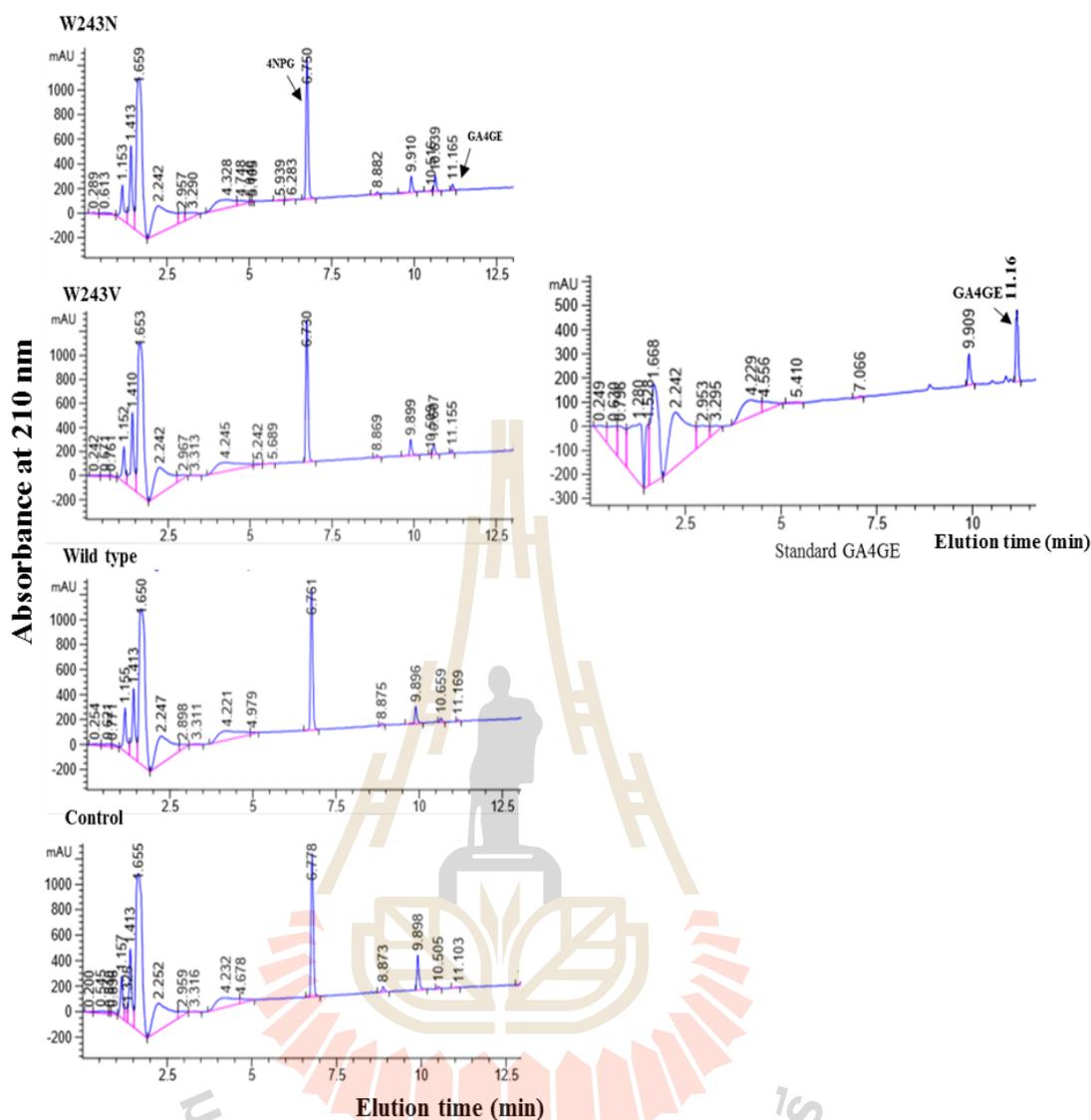


Figure 4.19 Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants for transfer of glucose to gibberellin GA₄ as phytohormone acceptor. The product gibberellin GA₄ glucose ester eluted at 11.16 min and its absorbance at 210 nm was measured by the DAD detector from the UPLC. The standard synthesized GA₄GE (1 mM) was run the same condition in UPLC to confirm that it gave the same elution time in the comparison to transglycosylation reactions.

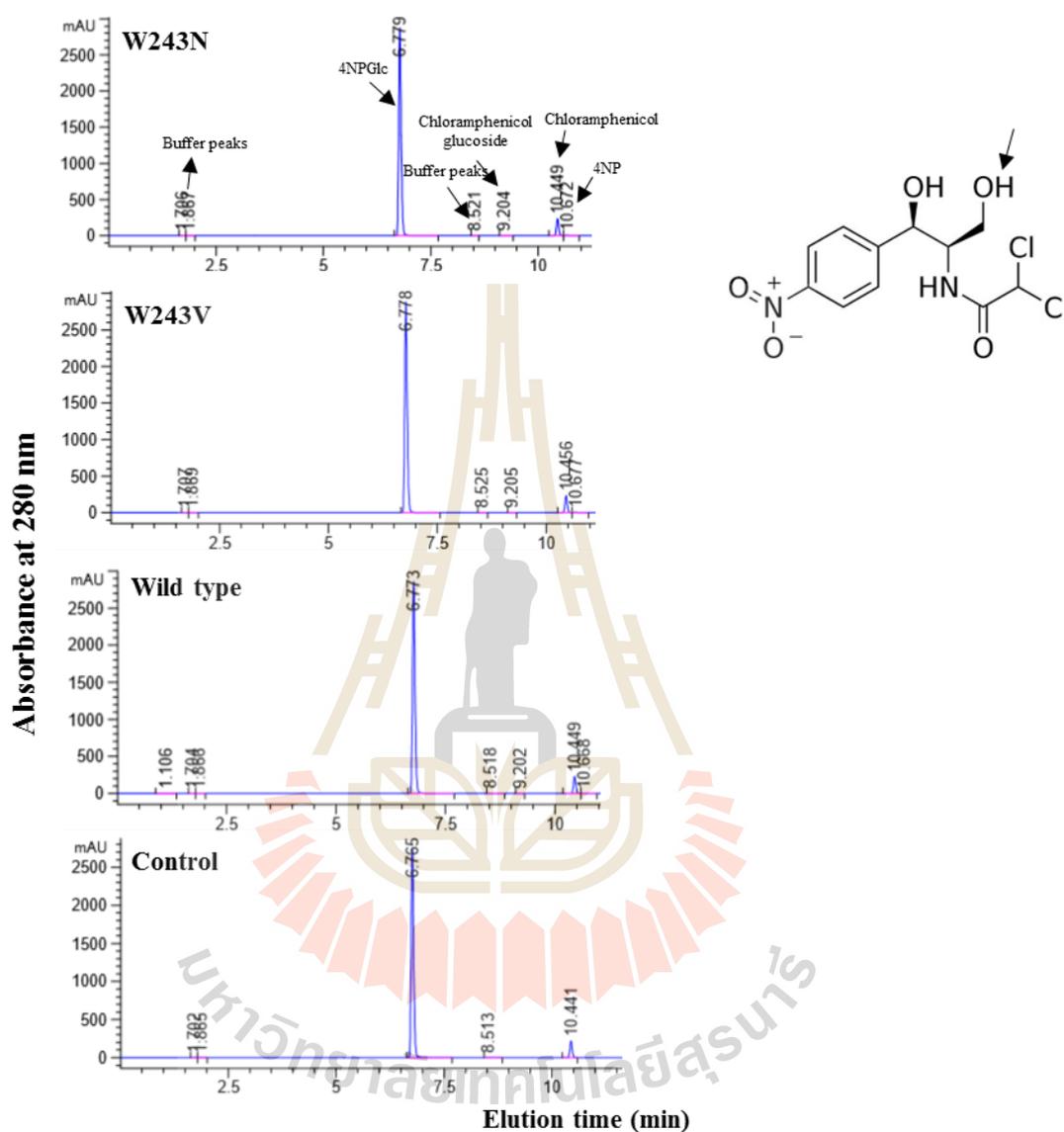


Figure 4.20 Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants with 4NPGlc and chloramphenicol, an antibiotic acceptor. Os9BGlu31 wild type and the mutants could transfer to a hydroxyl group to make chloramphenicol glucoside.

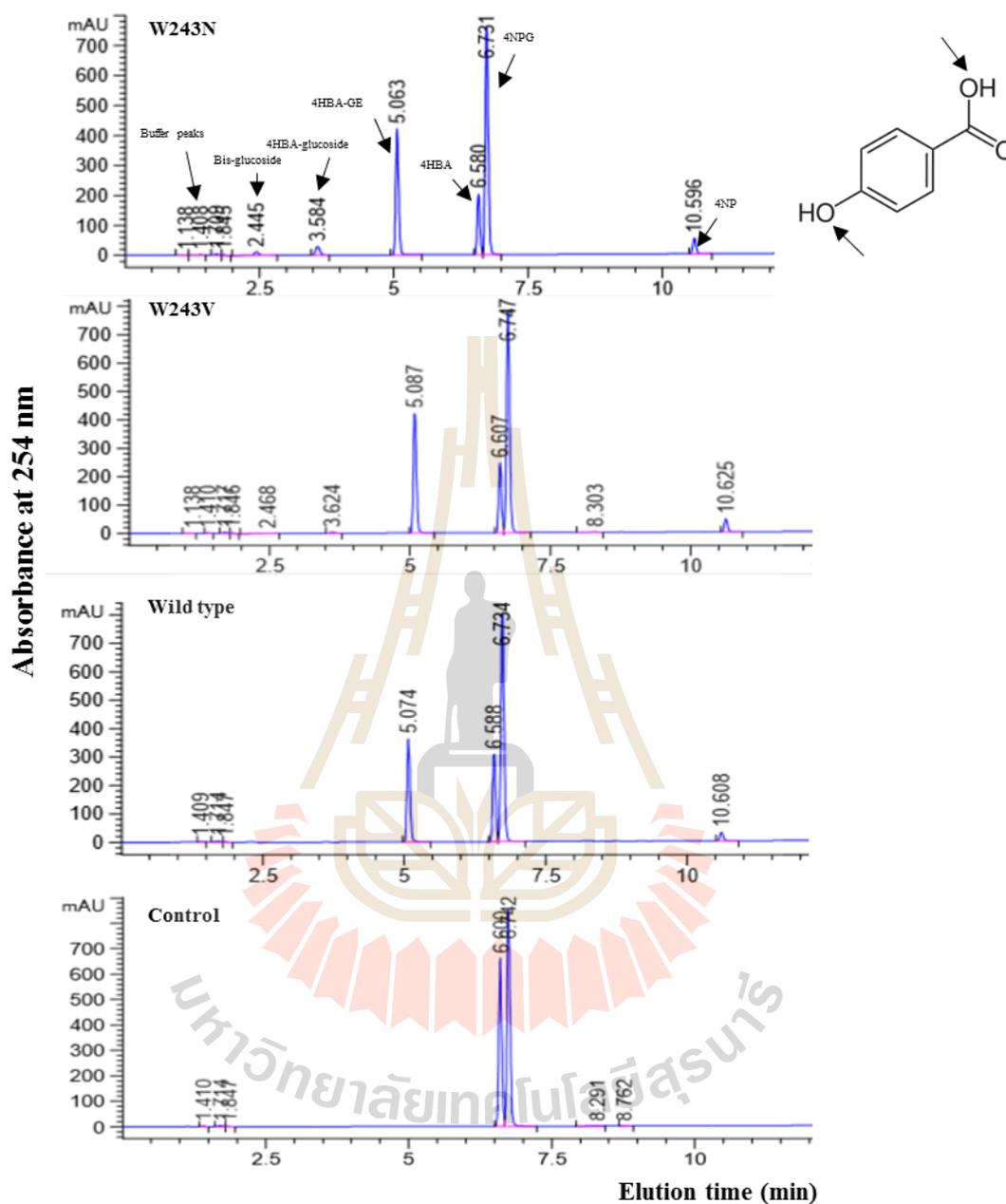


Figure 4.21 Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants (shown as examples) transfer of glucose to 4-hydroxybenzoic acid. Os9BGlu31 mutants can transfer to both the hydroxyl and carboxylic acid groups.

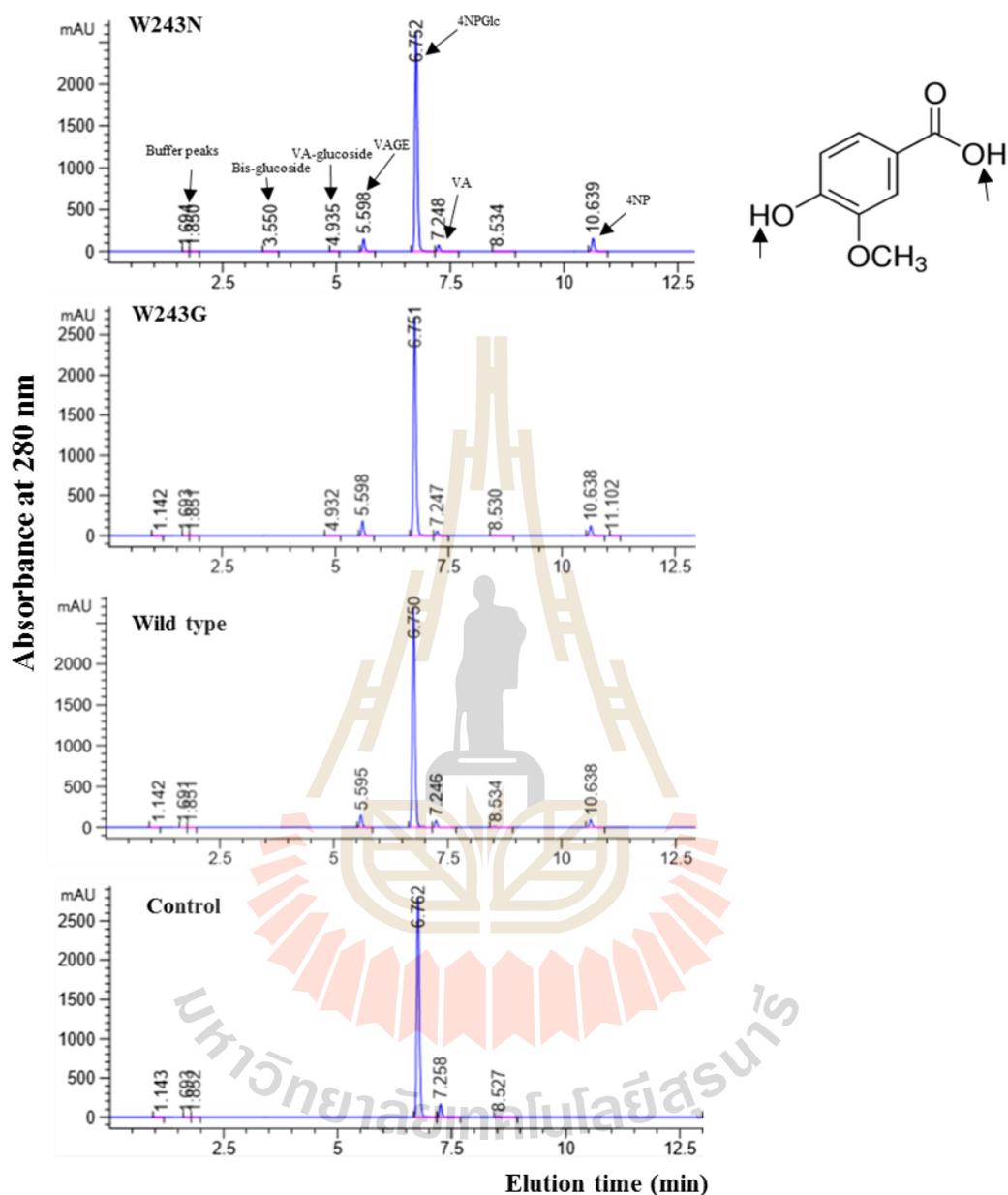


Figure 4.22 Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243G mutants (shown as examples of variants with high activity to this substrate) transfer of glucose to vanillic acid. Os9BGlu31 mutants can transfer to both the hydroxyl and carboxylic acid groups.

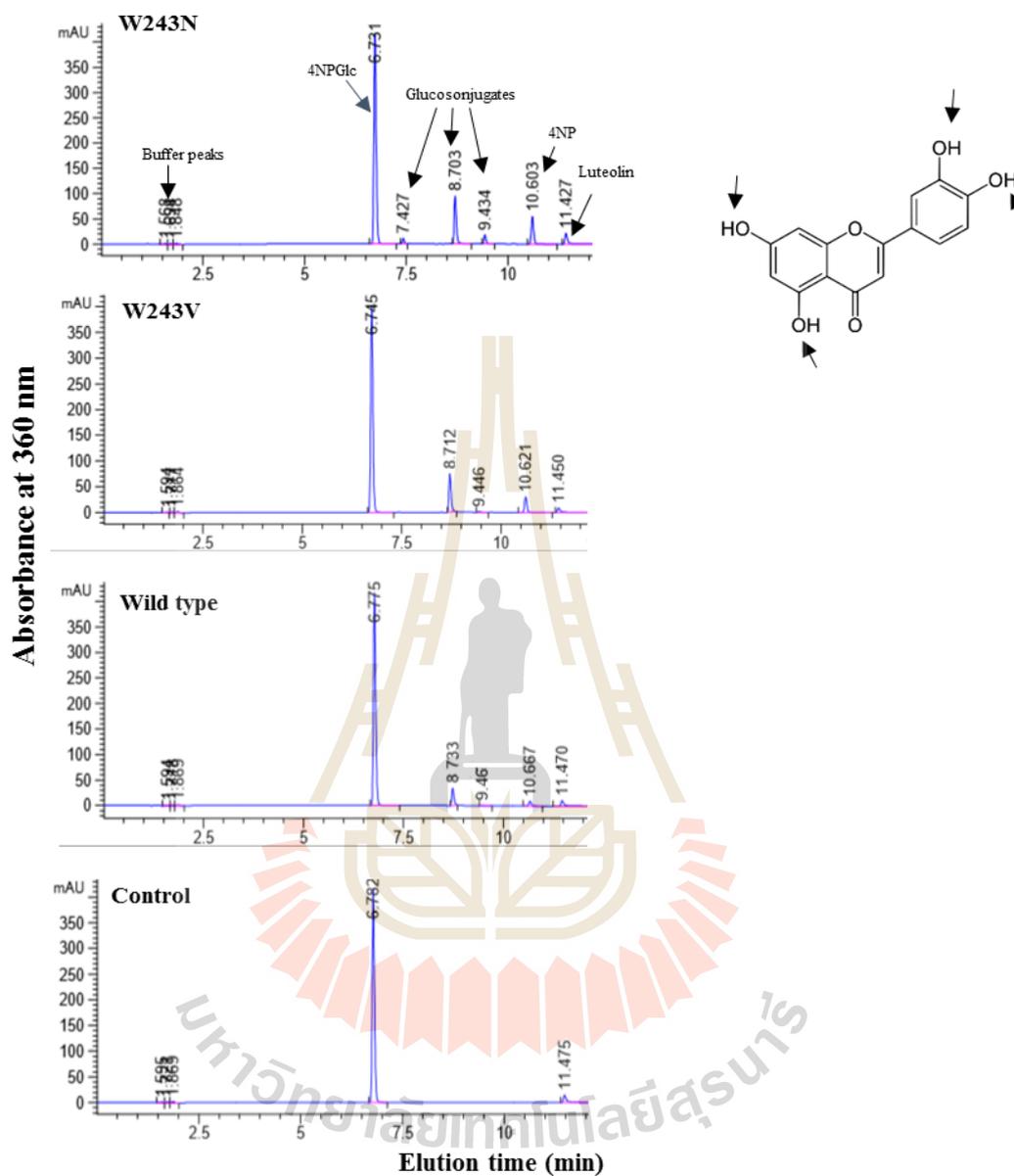


Figure 4.23 Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants (given as examples of variants with high activity toward this acceptor) transfer of glucose to luteolin. Os9BGlu31 mutants can transfer to multiple positions to provide different glucosides. The possible glycosylation positions are marked on the structure.

CHAPTER V

CONCLUSION

Rice Os5BGlu19 is classified in the phylogenetic cluster At/Os 6 of GH1 and, as such, is closely related to rice Os9BGlu31 transglucosidase, which transglycosylates secondary metabolites, flavonoids and phytohormones. A cDNA sequence was optimized to encode the 506 amino acids of the predicted mature Os5BGlu19 protein (1518 ORF nucleotides) for expression in both *P. pastoris* and *E. coli*. The Os5BGlu19 cDNA was inserted into the pET32a expression vector to express an Os5BGlu19 fusion protein with N-terminal thioredoxin and His₆ tags. Nevertheless, Os5BGlu19 protein could not be expressed in *E. coli* as an active β -glucosidase and the insoluble Os5BGlu19 was identified by western blot using anti-peptide anti-Os5BGlu19 serum. As an alternative, Os5BGlu19 was also cloned into the pICZ α BNH8 and pICZ α BNH8/eGFP plasmids, which were used to express Os5BGlu19 as a secreted protein in *P. pastoris* strain SMD1168H at 20 °C for 4 days of induction with 1% methanol. Os5BGlu19 apparently could not be efficiently secreted out, and the protein was located inside the cells, which was seen by fluorescence of the C-terminal eGFP tag. The activity of Os5BGlu19 expressed in pichia was unclear since some enzymes could hydrolyzed 4NPGlc in the media and the enzyme was unsuccessfully purified with IMAC from pichia cell extraction.

Based on the homology model structure of Os5BGlu19, we attempted to truncate unnecessary sequence that did not match the template structure of Os3BGlu6

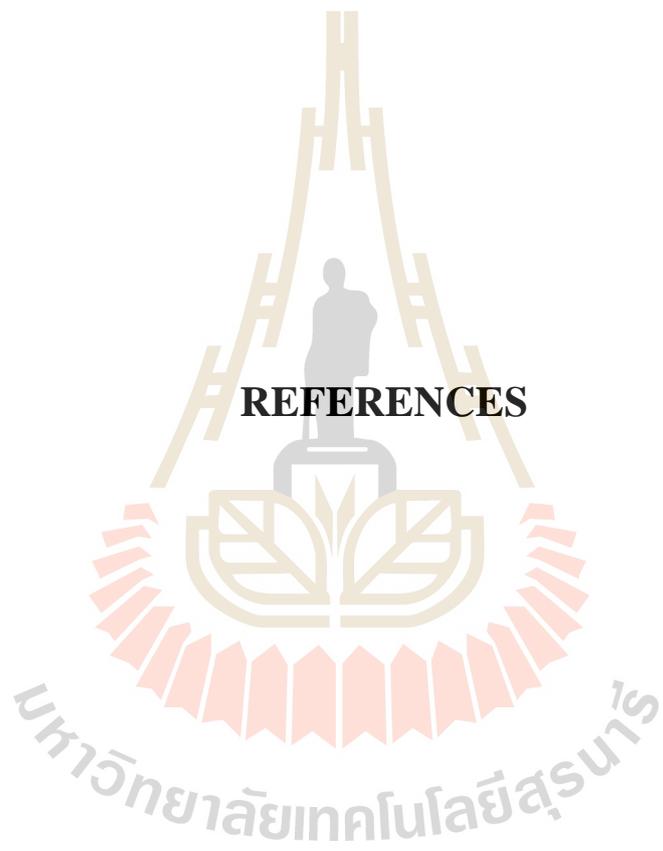
from the N- and C- terminal parts of the full-length Os5BGlu19. The truncated Os5BGlu19 included 1.3 kb and was inserted into the pET32a plasmid. pET32a/TrOs5BGL19 was expressed in Origami B(DE3) at 20 °C for 16 hours of 0.3 mM IPTG induction. The protein was partially purified by IMAC eluted with an imidazole concentration gradient. No activity could be detected in the standard β -glucosidase assay with 4NPGlc.

Substrate identification of Os5BGlu19 was based on its glycosylation of ferulic acid (FA) to produce 1-O-feruloyl glucose- β -D-ester (FAG) with the method developed to measure FAG produced by Os9BGlu31 in our previous work. Typically, Os5BGlu19 transferred glucose to ferulic acid as an acceptor from glucosyl donors in the rice leaf extracts to produce FAG. The product was measured with UPLC-QQQ-MS and the ion abundance interpreted from the standard curve of ion abundance, which was made from reactions of Os9BGlu31 with ferulic acid and 4NPGlc. There was a significant increase in reactions containing extracts of rice leaf (4 week old and 10 week old), upon the addition of TrOs5BGlu19. This result suggested that Os5BGlu19 could act as a ferulic acid β -transglucosidase to glycosylate ferulic acid to FAG using glucosyl donors in the rice extract. Fractionation of rice extraction was carried out over an LH20 column to identify the glucosyl donors. Several fractions could produce a high of FAG amount in the presence of TrOs5BGlu19. However, the purification of donor substrates is still unsuccessful, since further separation resulted in a decrease in FAG in the reaction in comparison with the control reaction, suggesting that a loss of active substrates occurred during the purification steps.

The rice (*Oryza sativa*) Os9BGlu31 transglucosidase transfers glucosyl moieties between a broad range of phenolic acids, flavonols, and their 1-O acyl- β -D-

glucose esters (Luang et al., 2013). The residue W243 of the active site plays an important role in the specificity of the transglucosidase activity of Os9BGlu31 (Komvongsa et al., 2015). Seven mutations were produced in the same position, W243, by site directed mutagenesis to investigate the difference in transglycosylation activity in these variants of Os9BGlu31 transglucosidase. The relative activities of Os9BGlu31 wild type and its mutant were measured with twenty-two acceptors, by release of 4NP from 4NPGlc as donor glucoside substrate in UPLC analysis. The W243N mutant produced by Komvongsa et al. (2015) still gave the highest transglucosidase and hydrolase activities in comparison to wild type and the other mutants. However, new glucoconjugate products were found in the transglycosylation of Os9BGlu31 W243 mutants with the phytohormone gibberilin A4 and the antibiotic chloramphenicol, as well as various plant phenolic compounds. Moreover, Os9BGlu31 mutants could produce multiple glucoconjugates with flavonoids, such as luteolin, or phenolic compounds such as 4-hydroxybenzoic acid and vanillic acid. Further work may continue characterization of Os9BGlu31 specificity determinants and application of glucoconjugate production by changing the W243 position to more kinds of amino acids or combining the W243N or other W243 mutations with other mutations in the active site.

REFERENCES



REFERENCES

- Ahn, Y.O., Shimizu, B., Sakata, K., Gantulga, D., Zhou, C., Bevan, D.R., and Esen, A. (2010). Scopolin-hydrolyzing β -glucosidases in roots of Arabidopsis. **Plant Cell Physiol.** 51(1): 132-43.
- Ahn, Y.O., Mizutani, M., Saino, H., and Sakata, K. (2004). Furcatin hydrolase from *Viburnum furcatum* blume is a novel disaccharide-specific acuminosidase in glycosyl hydrolase family 1. **J. Biol. Chem.** 279: 23405-23414.
- Allen, H.J., and Kisailus, E.C. (1992). Glycoconjugates "Composition, structure and function". 313-314 pp.
- Baiya, S., Hua, Y., Ekkhara, W., and Ketudat Cairns, JR. (2014). Expression and enzymatic properties of rice (*Oryza sativa* L.) monolignol β -glucosidases. **Plant Sci.** 227: 101-109.
- Allwooda, J.W., and Goodacrea, R. (2010). An introduction to liquid chromatography mass spectrometry instrumentation applied in plant metabolomic analyses. **Phytochem Anal.** 21(1): 33-47.
- Bojarova, P., and Kren, V. (2009). Glycosidases: A key to tailored carbohydrates. **Trends Biotechnol.** 27(4): 199-209.
- Bowles, D. (2005) Glycosyltransferases: managers of small molecules. **Curr Opin Plan Bio.** 8(3): 254-263.
- Braman, J., Papworth, C., and Greener, A. (1996). Site-directed mutagenesis using double-stranded plasmid DNA templates. **Methods. Mol. Biol.** 57: 31-44.

- Burmeister, W.P., Cottaz, S., Rollin, P., Vasella, A., and Henrissat, B. (2000). High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base. **J. Biol. Chem.** 275(50): 39385-39393.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Chen, M., Xia, L., and Xue, P. (2007). Enzymatic hydrolysis of corncob and ethanol production from cellulosic hydrolysate. **Int Biodeterior Biodegradation.** 59: 85-89.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. **Nucleic Acids Res.** D233-8.
- Clarke, N.J., Rindgen, D., Korfmacher, W.A., and Cox, K.A. (2001) Systematic LC/MS metabolite identification in drug discovery. **Anal. Chem.** 73: 430A-439A.
- Collendavello, J., Legrand M., Geoffroy, P., Barthelemy, J., and Fritig, B. (1981). Purification and properties of the three-diphenol O-methyltransferases of tobacco leaves. **Phytochemistry.** 20: 611-616.
- Coutinho, P.M. (2003). An evolving hierarchical family classification for glycosyltransferases. **J. Mol. Biol.** 328: 307-317.
- Cregg, J.M., Cereghino, J.L., Shi, J., and Higgins, D.R. (2000). Recombinant protein expression in *Pichia pastoris*. **Mol. Biotechnol.** 16(1): 23-52.
- Crozier, A., Jaganath, I.B., and Clifford, M.N. (2007). Phenols, polyphenols and tannins: An overview. In: **Plant Secondary Metabolites: Occurrence, Structure and Role in the Human Diet.** Ch 1: 1-24 pp.

- Crout, D.H., and Vic, G. (1998). Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis. **Curr. Opin. Chem. Biol.** 2(1): 98-111.
- Czjzek, M., Cicek, M., Zamboni, V., Burmeister, W.P., Bevan, D.R., Henrissat, B., and Esen A. (2000). The mechanism of substrate (aglycone) specificity in β -glucosidases is revealed by crystal structures of mutant maize β -glucosidase-DIMBOA, DIMBOAGlc, and -dhurrin complexes. **Proc. Natl. Acad. Sci. USA.** 97: 13555-13560.
- Dass, C. (2007). Fundamentals of contemporary mass spectrometry. **Hoboken, N.J: Wiley-Interscience.** 151-187 pp.
- Fraenkel, G.S. (1959). The raison d'être of secondary plant substances. **Science.** 129: 1466-1470.
- Gantt, R.W, Peltier-Pain, P., Cournoyer, W.J., and Thorson, J.S. (2011). Using simple donors to drive the equilibria of glycosyltransferase-catalyzed reactions. **Nat. Chem. Biol.** 7: 685-691.
- Grove, M.D., Spencer, G.F., Rohwedder, W.K., Mandava, N., Worley, J.F., Jr. J.D.W., Steffens, G.L., Flippen-Anderson, J.L., and Carter Cook, J. (1979). Brassinolide, a plant growth-promoting steroid isolated from *Brassica napus* pollen. **Nature.** 281: 216-217.
- Hammer, M.B., Eleuch-Fayache, G., Schottlaender, L.V., Nehdi, H., Gibbs, J.R., Arepalli, S.K., Chong, S.B., Hernandez, D.G., Sailer, A., Liu, G., Mistry, P. K., Cai, H., Shrader, G., Sassi, C., Bouhlal, Y., Houlden, H., Hentati, F., Amouri, R., and Singleton, A.B. (2013). Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity. **Am. J. Hum. Genet.** 92: 245-251.

- Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochem. J.** 280(Pt2): 309-316.
- Henrissat, B. (2000). Glycoside hydrolases and glycosyltransferases families, modules, and implications for genomics. **Plant Physiol.** 124: 1515-1519.
- Hoffman, K., Bucher, P., Falquet, L., and Bairoch, A. (1999). The PROSITE database. **Nucl. Acids. Res.** 27: 215-219.
- Hill, L., and Wang, T.L. (2009). Approaches to the analysis of plant-derived natural products. In: Osbourn A.E., Lanzotti V. **Plant-Derived Natural Products.** 97-126 pp.
- Hrmova, M., Varghese, J.N., Høj, P.B., and Fincher, G.B. (1998). Crystallization and preliminary X-ray analysis of β -D-glucan exohydrolase isoenzyme ExoI from barley (*Hordeum vulgare*). **Acta. Crystallogr.** D54: 687-698.
- Hruska, K.S., LaMarca, M.E., Scott, C.R., and Sidransky, E. (2008). Mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). **Hum Mutat.** 29: 567-583.
- Hösel, W., Tober, I., Eklund, S.H., and Conn, E.E. (1987). Characterization of beta-glucosidases with high specificity for the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) moench seedlings. **Arch. Biochem. Biophys.** 252(1): 152-162.
- Hua, Y., Sansenya, S., Saetang, C., Wakuta, S., and Ketudat Cairns, J.R. (2013). Enzymatic and structural characterization of hydrolysis of gibberellin A4 glucosyl ester by a rice β -d-glucosidase. **Arch. Biochem. Biophys.** 537: 39-48.
- Hua, Y., Ekkhara, W., Sansenya, S., Srisomsap, C., Roytrakul, S., Saburi, W., Takeda, R., Matsuura, H., Mori, H., and Ketudat Cairns, R.J. (2015). Identification of

- rice Os4BGlu13 as a β -glucosidase which hydrolyzes gibberellin A4 1-O- β -D-glucosyl ester, in addition to tuberonic acid glucoside and salicylic acid derivative glucosides. **Arch. Biochem. Biophys.** 583: 36-46.
- Izumi, T., Piskula, M.K., Osawa, S., Obata, A., Tobe, K., Saito, M., Kataoka, S., Kubota, Y., and Kikuchi, M. (2000). Soy isoflavone aglycones are absorbed faster and in higher amounts than their glucosides in humans. **J. Nutr.** 130(7): 1695-1699
- Ishii, T.M., Zerr, P., Xia, X.M., Bond, C.T., Maylie, J., and Adelman, J.P. (1998). Site-directed mutagenesis. **Methods Enzymol.** 293: 53-71.
- Jahn, M., Chen, H.M., Müllegger, J., Marles, J., Warren, R.A.J., and Withers, S.G. (2004). Thioglycosynthases: Double mutant glycosidases that serve as scaffolds for thioglycoside synthesis. **Chem. Commun.** 7(3): 274-275
- Jenkins, J., Leggio, L.L., Harris, G., and Pickersgill, R. (1995). β -glucosidase, β -galactosidase, family A cellulases, family F xylanases and two barley glucanases from a superfamily of enzymes with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β strands four and seven. **FEBS. Lett.** 362: 281-285.
- Juge, N., Andersen, J.S., Tull, D., Roepstorff, P., and Svensson, B. (1996). Overexpression, purification and characterization of recombinant barley α -amylases 1 and 2 secreted by the ethylophilic yeast *Pichia pastoris*. **Protein Expre. Purif.** 8: 204-214.
- Ketudat Cairns, J.R., Champattanachai, V., Srisomsap, C., Wittman-Liebold, B., Thiede, B., and Svasti, J. (2000). Sequence and expression of Thai Rosewood

beta-glucosidase/beta-fucosidase, a family 1 glycosyl hydrolase glycoprotein.

J. Biochem. 128: 999-1008.

Ketudat Cairns, J.R., and Esen, A. (2010) β -Glucosidases. **Cell Mol. Life Sci.** 67: 3389-3405.

Ketudat Cairns, J.R., Pengthaisong, S., Luang, S., Sansenya, S., Tankrathok, A., and Svasti J. (2012). Protein-carbohydrate interactions leading to hydrolysis and transglycosylation in plant glycoside hydrolase family 1 enzymes. **J. Appl. Glycosci.** 59: 51-62.

Khaw, O., and Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. **Asian Pac. J. Trop. Biomed.** 2(2): 159-162.

Komvongsa, J., Luang, S., Marques, J.V., Phasai, K., Davin, L.B., Lewis, N.G., and Ketudat Cairns, J.R. (2015). Active site cleft mutants of Os9BGlu31 transglucosidase modify acceptor substrate specificity and allow production of multiple kaempferol glycosides. **Biochim. Biophys. Acta.** 1850(7): 1405-1414.

Komvongsa, J., Mahong, B., Phasai, K., Hua, Y., Jeon, J.S., and Ketudat Cairns, J.R. (2015) Identification of fatty acid glucose esters as Os9BGlu31 transglucosidase substrates in rice flag leaves. **J. Agric. Food Chem.** 63(44): 9764-9769.

Koshland, D.E.J. (1995). Stereochemistry and the mechanism of enzymatic reactions. **Biol. Rev.** 28(4): 416-436.

Kostiainen, R., Kotiaho, T., Kuuranne, T., and Auriola, S. (2003) Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies. **J. Mass. Spectrom.** 38: 357-372.

- Kuchner, O., and Arnold, F.H. (1997). Directed evolution of enzyme catalysts. **Trends Biotechnol.** (12): 523-530.
- Kuntothom, T., Luang, S., Harvey, A.J., Fincher, G.B., Opassiri, R., Hrmova, M., and Ketudat Cairns, J.R. (2009). Rice family GH1 glycoside hydrolases with β -D-glucosidase and β -D-mannosidase activities. **Arch. Biochem. Biophys.** 491: 85-95.
- Lairson, L.L. and Withers, S.G. (2004). Mechanistic analogies amongst carbohydrate modifying enzymes. **Chem. Commun.** 20: 2243-2248.
- Lehfeldt, C., Shirley, A.M., Meyer, K., Ruegger, M.O., Cusumano, J.C., Viitanen, P.V., Strack, D., and Chapple, C. (2000). Cloning of the SNG1 gene of Arabidopsis reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. **Plant Cell.** 12: 1295-1306.
- Le Roy, J., Huss, B., Creach, A., Hawkins, S., and Neutelings, G. (2016). Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants. **Plant. Sci.** 7: 735.
- Lieberman, R.L., Wustman, B.A., Huertas, P., Powe Jr, A.C., Pine, C.W., Khanna, R., Schlossmacher, M.G., Ringe, D., and Petsko G.A. (2007). Structure of acid beta-glucosidase with pharmacological chaperone provides insight into Gaucher disease. **Nat. Chem. Biol.** 3: 101-107.
- Lodish, H., Berk, A., and Zipursky, S.L. (2000). Molecular Cell Biology. 4th edition. New York: **Chapter 17, Protein Sorting: Organelle Biogenesis and Protein Secretion.** 719-727 pp.
- Luang, S., Cho, J., Mahong, B., Opassiri, R., Akiyama, T., Phasai, K., Komvongsa, J., Sasaki, N., Hua, J., Matsuba, J., Ozeki, Y., Jeon, J.S., and Ketudat Cairns, J.R.

- (2013). Rice Os9BGlu31 Is a transglucosidase with the capacity to equilibrate phenylpropanoid, flavonoid, and phytohormone glycoconjugates. **J. Biol. Chem.** 288: 10111-10123.
- Mackenzie, L.F., Wang, Q, Warren, R.A.J., and Withers, S.G. (1998). Glycosynthases: Mutant glycosidases for oligosaccharide synthesis. **J. Am. Chem. Soc.** 120(22): 5583-5584.
- Marana, SR., Jacobs-Lorena, M., Terra, W.R., and Ferrieira, C. (2001). Amino acid residues involved in substrate binding and catalysis in an insect digestive β -glycosidase. **Biochim. Biophys. Acta.** 1545: 41-52.
- Martin, E., Schule, R., Smets, K., Rastetter, A., Boukhris, A., Loureiro, J. L., Gonzalez, M. A., Mundwiller E., Deconinck T., Wessner M., Jornea L., and Oteyza, A.C. (2013). Loss of function of glucocerebrosidase. GBA2 is responsible for motor neuron defects in hereditary spastic paraplegia. **Am. J. Hum. Genet.** 92: 238-244.
- Matern, H., Boermans, H., Lottspeich, F., and Matern, S. (2001). Molecular cloning and expression of human bile acid beta-glucosidase. **J. Biol. Chem.** 276(41): 37929-37933.
- Matsuba, Y., Sasaki, N., Tera, M., Okamura, M., Abe, Y., Okamoto, E, Nakamura, H., Funabashi, H., Takatsu, M., Saito, M., Matsuoka, H., Nagasawa, K., and Ozek, Y. (2010). A novel glucosylation reaction on anthocyanins catalyzed by acyl-glucose-dependent glucosyltransferase in the petals of carnation and delphinium. **Plant Cell.** 22: 3374-3389.
- Matthews, C.E., Van Holde, K.E., and Ahern, K.G. (1999). **Biochemistry.** 3rd edition

- Mitchell, J.W., Mandava, N., Worley, J.F., Plimmer, J.R., and Smith, M.V. (1970). Brassins--a new family of plant hormones from rape pollen. **Nature**. 225(5237): 1.
- Meyer, Y., Grosset, J., Chartier, Y., and Cleyet-Marel, J.C. (1988). Preparation by two-dimensional electrophoresis of proteins for antibody production: antibodies against proteins whose synthesis is reduced by auxin in tobacco mesophyll protoplasts. **Electrophoresis**. 9: 704-712.
- Miyahara, T., Sakiyama, R., Ozeki, Y., and Sasaki, N. (2013) Acyl-glucose-dependent glucosyltransferase catalyzes the final step of anthocyanin formation in Arabidopsis. **J. Plant Physiol**. 170: 619-624.
- Mizutani, F.M., Nakanishi, H., Ema, J., Ma, S.J., Noguchi, N., Ochiai, M.I., Mizutani, M.F., Nakao, M., and Sakata, K. (2002). Cloning of β -primeverosidase from tea leaves, a key enzyme in tea aroma. **Plant Physiol**. 130: 2164-2176.
- Moellering, E.R., Muthan, B., and Benning, C. (2010). Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. **Science**. 330: 226-228.
- Mugford, S.T., Oi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R.K., Papadopoulou, K., Melton, R., Philo, M., Sainsbury, F., Lomonossoff, G.P., Roy, A.D., Goss, R.J.M., and Osbourn, A. (2009). A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. **Plant Cell**. 21: 2473-2484.
- Müllegger, J., Jahn, J., Chen, H-M., Warren, R.A.J., and Withers, S.G. (2005). Engineering of a thioglycoligase: randomized mutagenesis of the acid/base

residue leads to the identification of improved catalysts. **Protein Eng. Des. Sel.** 18: 33-40.

Nguyen, T.T., Mathiesen, G., Fredriksen, L., Kittl, R., Nguyen, T.H., Eijsink, V.G.H., Haltrich, D., and Peterbauer, C.K. (2011). A food-grade system for inducible gene expression in *Lactobacillus plantarum* using an alanine racemase-encoding selection marker. **J. Agric. Food. Chem.** 59: 5617-5624.

Nishizaki, Y., Yasunaga, Y., Okamoto, E., Okamoto, M., Hirose, Y., Yamaguchi, M., Ozeki, Y., and Sasak, N. (2014). *p*-Hydroxybenzoyl-glucose is a zwitter donor for the biosynthesis of 7-polyacylated anthocyanin in Delphinium. **Plant Cell.** 25: 4150-4165.

Opassiri, R., Ketudat Cairns, J.R., Akiyama, T., Wara-Aswapati, O., Svasti, J., and Esen, A. (2003). Characterization of a rice β -glucosidase highly expressed in flower and germinating shoot. **Plant. Sci.** 165: 627-638.

Opassiri, R., Hua, Y., Wara-Aswapati, O., Akiyama, T., Svasti, J., Esen, A., and Ketudat Cairns, J.R. (2004). Beta-glucosidase, exo-beta-glucanase and pyridoxine transglucosylase activities of rice BGlu1. **Biochem. J.** 379(Pt 1): 125-131.

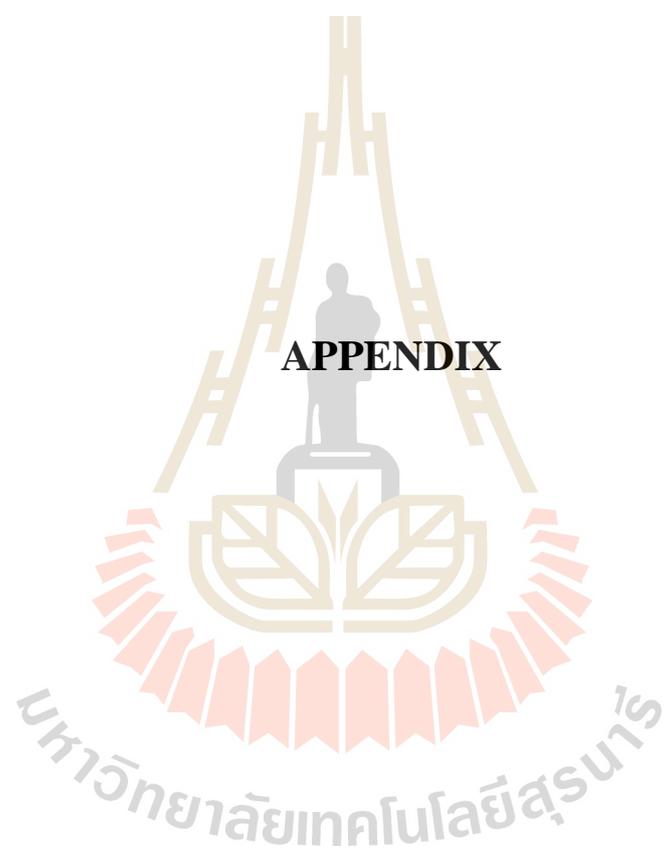
Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A., and Ketudat Cairn, J.R. (2006). Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β -glucosidase. **BMC. Plant. Biol.** 10.1186/1471-2229-6-33.

Opassiri, R., Maneesan, J., Akiyama, T., Pomthong, B., and Jin, S. (2010). Rice Os4BGlu12 is a wound-induced β -glucosidase that hydrolyzes cell wall- β -glucan-derived oligosaccharides and glycosides. **Plant. Sci** 179: 273-280.

- Palackal, N., Brennan, Y., Callen, W.N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G.P., Healey, S., Kang, S.E., Kretz, K.A., Lee, E., Tan, X., Tomlinson, G.L., Verruto, J., Wong, V.W.K., Mathur, E.J., Short, J.M., Robertson, D.E., and Stee, B.A. (2007). An evolutionary route to xylanase process fitness. **Protein Sci.** 13(2): 494-503.
- Peracchi, A. (2001) Enzyme catalysis: removing chemically 'essential' residues by site-directed mutagenesis. **Trends Biochem. Sci.** 26(8): 497-503.
- Pereira, D.M, Valentão, P., Pereira, J.A., and Andrade, P.B. (2009). Phenolics: From chemistry to biology. **Molecule.** 14: 2202-2211.
- Pigman, W., and Anet, E.F.L.J. (1972). "Chapter 4: Mutarotations and actions of acids and bases". **Chemistry and Biochemistry.** Vol 1A. 165-194 pp.
- Polayes, D.A. (1998) **Methods in Molecular Medicine.** Vol.13: 169-183 pp.
- Raychaudhuri, A., and Tipton, P.A. (2002). Cloning and expression of the gene for soybean hydroxyisourate hydrolase. Localization and implications for function and mechanism. **Plant Physiol.** 130: 2061-2068.
- Rich, J.R., and Withers, S.G. (2009). Emerging methods for the production of homogeneous human glycoproteins. **Nat. Chem. Biol.** 5(4): 206-215.
- Robyt, J.F. (1998). Essentials of Carbohydrate Chemistry. Chapter 1 Beginnings. 1-18 pp.
- Rouyi, C., Baiya, B., Lee, S.K., Mahong, B., Jeon, J.S., Ketudat-Cairns, J.R., and Ketudat-Cairns, M. (2014). Recombinant expression and characterization of the cytoplasmic rice β -glucosidase Os1BGlu4. **PLoS One.** 9(5): e96712.
- Sanz-Aparicio, J., Hermoso, J.A., Martínez-Ripoll, M., Lequerica, J.L., and Polaina, J. (1998). Crystal structure of beta-glucosidase A from *Bacillus polymyxa*:

- insights into the catalytic activity in family 1 glycosyl hydrolases. **J. Mol. Biol.** 275: 491-502.
- Schmidt, H., Gleave, E.S., and Carter, A.P. (2012). Insights into dynein motor domain function from a 3.3-Å crystal structure. **Nat. Struct. Mol. Biol.** 19: 492-497.
- Seshadri, S., Akiyama, T., Opassiri, R., Kuaprsert, B., and Ketudat Cairns, J. R. (2009). Structural and enzymatic characterization of Os3BGlu6, a rice β -glucosidase hydrolyzing hydrophobic glycosides and (1 \rightarrow 3) and (1 \rightarrow 2)-linked disaccharides. **Plant Physiol.** 151: 47-58.
- Shewale, J.G. (1982). Beta-glucosidase: its role in cellulase synthesis and hydrolysis of cellulose. **Int. J. Biochem.** 14(6): 435-443.
- Shewry, P. R., and Gutierrez, S. (1992). Plant protein engineering. **Cambridge University Press 3-4.** 3-16 pp.
- Sinnott, M.L. (1990). Catalytic mechanisms of enzymic glycosyl transfer. **Chem. Rev.** 90: 1171-1202.
- Snyder, L.R, Kirkland J.J., and Glajch, J.L. (1997). **Practical HPLC Method Development, 2nd ed. Wiley, New York.** 1-21 pp.
- Strack, D., and Wray, V. (1992). Anthocyanins. In J.B. Harborne (ed.), **The Flavonoids: Advances in Research Since 1986.** Chapman & Hall, London. 1-22 pp.
- Strack, D. (2001). Enzymes involved in hydroxycinnamate metabolism. **Methods Enzymol.** 335: 70-81.
- Sumner, L.W., Huhman, D.V., Wochniak, E.V., and Lei, Z. (2007). Methods, applications and concepts of metabolite profiling: Secondary metabolism. **Plant Sys. Biol.** 97: 195-212.

- Teze, D., Hendrickx, J., Czjzek, M., Ropartz, D., Sanejouand, Y.H., and Tran, V., Teller, C., Dion, M. (2013). Semi rational approach for converting a GH1 β -glycosidase into a β -transglucosidase. **Protein Eng. Des. Sel.** 27(1): 13-19.
- Tobimatsu, Y., Elumalai, S., Grabber, J.H., Davidson, C.L., Pan, X., and Ralph, J. (2012). Hydroxycinnamate conjugates as potential monolignol replacements: in vitro lignification and cell wall studies with rosmarinic acid. **Chem. Sus. Chem.** 5(4): 676-686.
- Vaistij, F.E., Lim, E.K., Edwards, R., and Bowles, D.J. (2009). Glycosylation of secondary metabolites and xenobiotics. In: Osbourn A.E., Lanzotti V. **Plant-Derived Natural Products.** 209-228 pp.
- Zagrobelny, M., Bak, S., and Møller, B.L. (2008). Cyanogenesis in plants and arthropods. **Phytochemistry.** 69: 1457-1468.
- Wakuta, S., Hamada, S., Ito, H., Matsuura, H., and Nabeta, K (2010). Identification of a β -glucosidase hydrolyzing tuberonic acid glucoside in rice (*Oryza sativa* L.). **Phytochemistry.** 71(11): 1280-1288.
- Weckwerth, W. (2007). Metabolomics: methods and protocols. **Humana Press Totowa,NJ.** 39-53 pp.



APPENDIX

QUANTIFICATION OF 4-NP BY UPLC METHOD

1. Methodology

An Agilent 1294 UPLC system with an Agilent SB-C18 RRHD 1.8 μm , 2.1x150 mm column (Agilent Technologies, CA, USA) was used to detect 4-NP released from the reaction of Os9BGlu31 wild type and its mutants with 4NPGlc as donor and acceptors. Four wavelengths (210 nm, 254 nm, 280 nm and 360 nm) were set up in the UV detector.

2. Standard 4-Nitrophenol

4-NP was eluted at 10.5 min and 4-NP standard curve was shown as Table A1, Figure A1 and A2.

Table A1 4-Nitrophenol standard concentration and area peak at 360 nm.

4-NP concentration (mM)	Area [mAU*s]
0.005	4.41
0.01	8.77
0.05	43.54
0.1	83.33
0.2	175.52

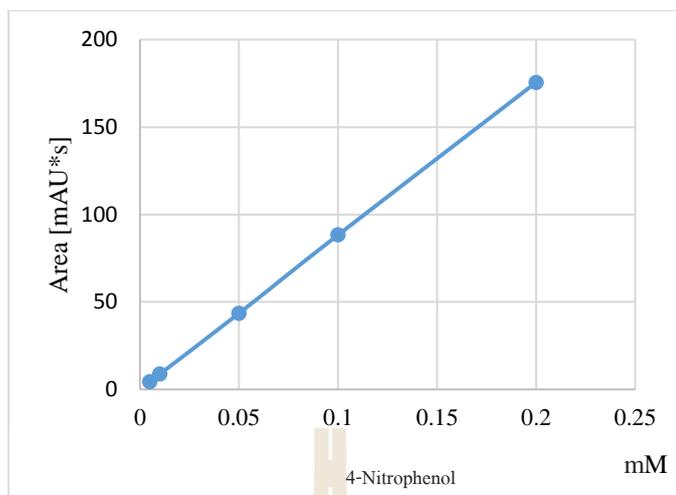


Figure A1 4-Nitrophenol standard curve from UPLC.

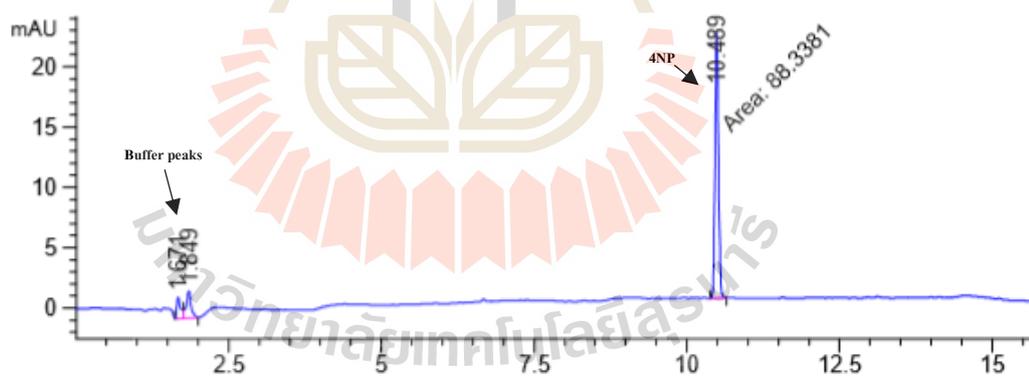


Figure A2 Chromatogram of 4NP obtained by UPLC with detection of the absorbance at 360 nm.

CURRICULUM VITAE

NAME Ms. Linh Thuy Tran

DATE OF BIRTH December 05, 1990

PLACE OF BIRTH Ninh Binh, Viet Nam

EDUCATION

2014-Present Master of Science in School of Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand.

2008-2012 Bachelor in Crop Science, Faculty of Agronomy, Viet Nam National University of Agriculture, Hanoi, Viet Nam

มหาวิทยาลัยเทคโนโลยีสุรนารี