INVESTIGATION OF RICE SERINE CARBOXYPEPTIDASE-LIKE (SCPL) PROTEINS

ISTIFTAKHUN NIKMAH

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Biochemistry
Suranaree University of Technology
Academic Year 2021

# การตรวจสอบโปรตีนซีรีนคาร์บอกซีเพปทิเดส-ไลค์ (SCPL) ในข้าว

นางสาวอีสตีฟตาคุล นิคมาห์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาชีวเคมี
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2564

INVESTIGATION OF RICE SERINE CARBOXYPEPTIDASE-LIKE (SCPL) PROTEINS

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

Thesis Examining Committee
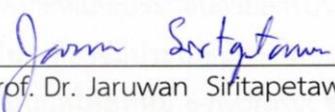
_____
(Assoc. Prof. Dr. Jaruwan Siritapetawee)
Chairperson

_____
(Prof. Dr. James R. Ketudat-Cairns)
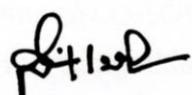Member (Thesis Advisor)

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

_____
(Dr. Rung-Yi Lai)
Member

_____
(Assoc. Prof. Dr. Chatchai Jothiyangkoon)
Vice Rector for Academic Affairs
and Quality Assurance

_____
(Prof. Dr. Santi Maensiri)
Dean of Institute of Science

อีสตีฟตาคุล นิคมาห์ : การตรวจสอบโปรตีนซีรีนคาร์บอกซีเพปทิเดส-ไลค์ (SCPL) ในข้าว (INVESTIGATION OF RICE SERINE CARBOXYPEPTIDASE-LIKE (SCPL) PROTEINS) อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.เจมส์ เกตุทัต-คาร์นส์, 88 หน้า.

คำสำคัญ: สารประกอบฟีนอล/ โปรตีนกลุ่มคาร์บอกซีเพปทิเดส/ แอซิติลทรานสเฟอเรส/ ไฮโดรไลสิส

ข้าว (*Oryza sativa* L.) โดยเฉพาะข้าวสีสามารถสังเคราะห์และประกอบด้วยสารประกอบ ฟีนอลิกเป็นจำนวนมาก เช่น แอนโธไซยานินและฟลาโวนอยด์อื่นๆ ที่อาจมีประโยชน์ทางโภชนาการ ฟีนอลิกไกลโคไซด์อย่างง่ายบางครั้งถูกอะซิเลต ตัวอย่างเช่น บนกลูโคสที่หมู่ 6-hydroxyl ของไซ ยานินบางชนิดเพื่อสร้างสารประกอบที่ซับซ้อนขึ้น เช่น เม็ดสี ในพืช ยีนของตระกูลหนึ่งซึ่งถอดรหัส เป็นเอ็นไซม์ที่เกี่ยวกับ acylation คือ ซีรีนคาร์บอกซีเพปทิเดส-ไลค์ อะซิลทรานสเฟอเรส (serine carboxypeptidase-like acyltransferase, SCPL-AT) SCPL-AT สามารถถ่ายโอน acyl moieties จาก 1-*O*-β-D-glucose esters ไปยังสารตั้งต้นตัวรับ ได้แก่ สารประกอบฟีนอลิก ในช่วงสองทศวรรษ ที่ผ่านมา SCPL-AT บางตัวได้มีการศึกษาลักษณะเฉพาะใน *Arabidopsis thaliana, Avena strigose (A. strigose), Brassica napus, Delphinium grandiflorum (D. grandiflorum)* และอื่น ๆ อีกมากมาย แต่ยีน SCPL-AT ในข้าวยังไม่เคยมีการอธิบายอย่างชัดเจน ในงานวิจัยนี้ ผู้วิจัยพยายาม ตอบคำถามว่ายีนใดบ้างที่ถอดรหัสเป็น SCPL-AT ของข้าว และเพื่อทดสอบว่าเอ็นไซม์สามารถ ดัดแปลงสารประกอบฟีนอลิกในหลอดทดลองได้หรือไม่

ผู้วิจัยทำการวิเคราะห์ทาง *in silico* เพื่อเลือกยีนที่อาจจะถอดรหัสเป็น SCPL-AT ของข้าว จากการวิเคราะห์สายวิวัฒนาการพบ SCPL ของข้าวที่น่าสนใจสองชนิดคือ OsSCPL2a และ OsSCPL7 มีความสัมพันธ์อย่างใกล้ชิดกับ SCPL-AT ที่รู้จักจาก *A. strigosa* (SCPL1) และ *D. grandiflorum* (DgSCPL2) ซึ่งอยู่ใน clade 1A ของตระกูลซีรีนคาร์บอกซีเพปทิเดส (SCP) รวมทั้ง SCPL-AT ที่มีการศึกษาลักษณะเฉพาะก่อนหน้านี้ ผู้วิจัยได้ทำการแสดงออกของโปรตีน OsSCPL ด้วย ระบบการแสดงออกในยีสต์และแบคทีเรีย โดยใช้ *Pichia pastoris* (*P. pastoris*) และ *Escherichia coli* (*E. coli*) ตามลำดับ โปรตีนถูกแยกให้บริสุทธิ์โดยวิธี immobilized affinity chromatography (IMAC) ที่จับกับ Ni^{2+} ตรวจสอบความบริสุทธิ์ด้วยเจล SDS-PAGE และตรวจสอบลักษณะเฉพาะโดย โครมาโตกราฟีแบบของเหลวและ tandem mass spectrometry (LC/MS/MS) โปรตีนขนาด 75 kDa จากการเพาะเลี้ยง *P. pastoris* เพื่อเหนี่ยวนำให้แสดงออกของ OsSCPL7 ถูกระบุว่าเป็น แอลกอฮอล์ออกซิเดส 1 (AOX1) ซึ่งเป็นโปรตีน peroxisomal ที่มักจะไม่ถูกปล่อยสู่อาหารเลี้ยงเชื้อ เอ็นไซม์ AOX1 สามารถจับกับเรซิน IMAC ได้ ซึ่งนำไปสู่การทำให้บริสุทธิ์แทน OsSCPL7 พบว่า AOX1 สามารถเปลี่ยน cyanidin-3-O-glucoside (Cy3G) ไปเป็นสารประกอบที่ไม่สามารถดูดกลืน แสงในช่วง 520 นาโนเมตร แต่ดูดกลืนแสงได้ค่อนข้างดีที่ 360 นาโนเมตร โดยวิธีโครมาโตกราฟี ของเหลวประสิทธิภาพสูงพิเศษ (UHPLC)

OsSCPL7 พบว่า AOX1 สามารถเปลี่ยน cyanidin-3-O-glucoside (Cy3G) ไปเป็นสารประกอบที่ไม่สามารถดูดกลืนแสงในช่วง 520 นาโนเมตร แต่ดูดกลืนแสงได้ค่อนข้างดีที่ 360 นาโนเมตร โดยวิธีโครมาโตกราฟีของเหลวประสิทธิภาพสูงพิเศษ (UHPLC)

ในทางกลับกัน การแสดงออกและการทำให้โปรตีน SCPL ใน *E. coli* บริสุทธิ์ค่อนข้างยากเนื่องจากโปรตีนที่ผลิตในระบบแบคทีเรียไม่สามารถละลายน้ำได้ การทำให้บริสุทธิ์โดยวิธี IMAC บนเรซินที่จับกับ Co^{2+} ให้แถบโปรตีนที่มีนัยสำคัญของ OsSCPL2a เท่านั้น แต่ไม่มีแถบโปรตีน OsSCPL7 ผู้วิจัยได้เพียงแถบขนาดเล็กของโปรตีนเป้าหมายเมื่อเทียบกับแถบโปรตีนของ *E. coli* และพบว่าเป็น OsSCPL2a จากการวิเคราะห์ด้วย LC/MS/MS ของทริปติกเปปไทด์ แม้ว่าจะมีสถานการณ์เหล่านั้น ผู้วิจัยได้ทำการทดสอบการทำงานของเอ็นไซม์โดยการเพิ่ม glucose esters ซึ่งเป็นผู้ให้อะซิลกับฟลาโวนอยด์หลายชนิด หลังจากทดสอบด้วย UHPLC พบว่า OsSCPL2a ดูเหมือนจะทำหน้าที่เป็นเอ็นไซม์ไฮโดรเลสและมีความจำเพาะต่อการปล่อยกลูโคสที่เชื่อมกับ 7-O ออกจากฟลาโวนอยด์แทนที่จะเป็น acyltransferase ที่ acylates ฟลาโวนอยด์กลูโคไซด์เหล่านั้น ผลลัพธ์เหล่านี้อาจเปิดโอกาสในการค้นหายีนบางตัวที่ถอดรหัสเป็น SCPL-glycosidases ทำให้เข้าใจว่า SCPL ของข้าวบางตัวอาจมีหน้าที่ต่างกัน อย่างไรก็ตาม เพื่อให้ได้โปรตีนที่บริสุทธิ์กว่านี้ซึ่งจะทำให้ได้ข้อสรุปที่ชัดเจนขึ้น ควรมีความพยายามในการหาวิธีที่จะผลิต OsSCPL ที่ละลายน้ำได้มากกว่าในระบบการแสดงออกแบบรีคอมบิแนนท์

สาขาวิชาเคมี

ปีการศึกษา 2564

ลายมือชื่อนักศึกษา _____

ลายมือชื่ออาจารย์ที่ปรึกษา _____
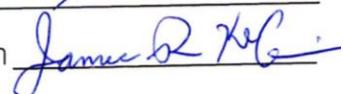
ISTIFTAKHUN NIKMAH : INVESTIGATION OF RICE SERINE CARBOXYPEPTIDASE-LIKE (SCPL) PROTEINS. THESIS ADVISOR : PROF. JAMES R. KETUDAT-CAIRNS, Ph.D. 88 PP.

Keyword: Phenolic Compounds/ Serine Carboxypeptidase-Like Protein/ Acyltransferase/ Hydrolysis

Rice (*Oryza sativa L.*), especially colored rice, synthesizes and contains a large group of phenolic compounds, such as anthocyanins and other flavonoids, that have shown nutritional benefits. The simple phenolic glycosides are sometimes acylated, to form more complex compounds, such as pigments. In plants, one family of genes encoding enzymes responsible for such acylation are known as serine carboxypeptidase-like acyltransferase (SCPL-AT), which are able to transfer acyl moieties from a 1-*O*-β-D-glucose esters to acceptor substrates, including phenolic compounds. Over the last two decades, some SCPL-ATs have been characterized in *Arabidopsis thaliana, Avena strigosa, Brassica napus, Delphinium grandiflorum,* and many others. As yet, the SCPL-AT genes in rice have not been clearly described. Here, we sought to test genes encoding rice SCPL-AT to see whether the encoded enzymes modify the phenolic compounds *in vitro*.
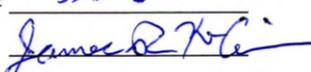
In silico analysis revealed that two putative rice SCPLs, OsSCPL2a and OsSCPL7, are closely related to the known SCPL-AT from *A. strigosa* (SCPL1) and *D. grandiflorum* (DgSCPL2). First, OsSCPL proteins were expressed in *Pichia pastoris*. The proteins were purified by immobilized metal affinity chromatography (IMAC) resin bound to Ni^{2+}, excised from SDS-PAGE gels, and identified by liquid chromatography and tandem mass spectrometry (LC/MS/MS). A 75 kDa protein from culture of *P. pastoris* induced to express OsSCPL7 was identified as alcohol oxidase 1 (AOX1), a peroxisomal protein that is not usually released to the medium. The AOX1 protein was found to convert cyanidin-3-O-glucoside (Cy3G) to compounds that no longer exhibit absorbance in the 520 nm range, but absorb relatively strongly at 360 nm, as detected in ultra-high-performance liquid chromatography (UHPLC).

On the other hand, the expression and purification of SCPL protein in *E. coli* yielding a small amount of OsSCPL2a soluble protein after purification by Co^+-IMAC resin and detected on LC/MS/MS analysis. After testing the semi-purified OsSCPL2a protein in transglycosylation in reactions monitored by UHPLC, we found that OsSCPL2a appeared to act as a hydrolase enzyme and specifically release 7-O-linked glucose from flavonoids, instead of as an acyltransferase that acylates those flavonoid

glucosides. These results may indicate a new divergent role for SCPL proteins as glycosidases, thereby improving our understanding of this gene family. However, to obtain purer protein from which to make stronger conclusions, efforts should be made to find a way to produce more soluble OsSCPL in the recombinant expression system.
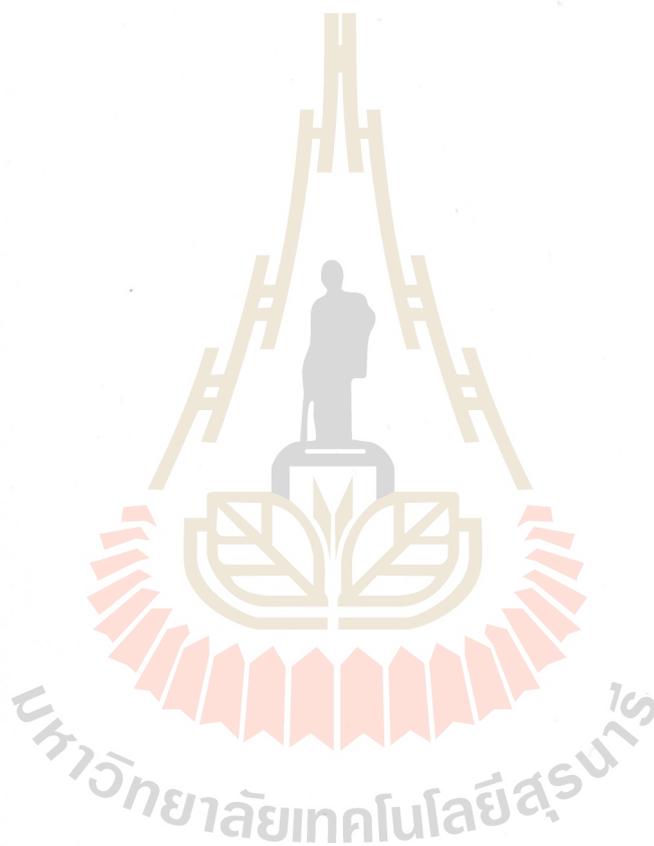
School of Chemistry

Academic Year 2021

Student's signature _____

Advisor's signature _____

ACKNOWLEDGEMENTS

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF ABBREVATIONS

(μ/m)L	(Nano, Micro) Liter
(μ/m)M	(Micro, Milli) Molar
(k)bp	Kilo Base pair DNA
(n/μ/m)g	(Nano, Micro, Milli) Gram
°C	Degree (s) Celsius
A7G	Apigenin-7-*O*-glucoside
APS	Ammonium persulfate
AT	Acyltransferase
Bp	Base pair(s)
C7G	Cyanidin-3-*O*-glucoside
cDNA	Complementary deoxynucleic acid
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease Electrophoresis
ER	Endoplasmic reticulum
ESI	Electrospray ionization
FA	Ferulic acid
FAG	Feruloyl-glucose
GH	Glucose hydrolase
GT	Glucosyltransferase
h	Hour (s)
K7G	Kaempferol-7-*O*-glucoside
kDa	Kilo Dalton
L7G	Luteolin-7-*O*-glucoside
LB	Luria-Bertani lysogeny broth
LC	Liquid chromatography
min	Minute
MS	Mass spectrometry
MW	Molecular weight
nm	Nano meter
OD	Optical density
pCA	*p*-coumaric acid

LIST OF ABBREVATIONS (Continued)

pCAG	*p*-coumaroyl-glucose
PEG	Polyethylene glycol
pHBA	*p*-hydroxybenzoic acid
pHBAG	*p*-hydroxybenzoyl-glucose
PMSF	Phenylmethylsulfonyl fluoride
Q7G	Quercetin-7-*O*-glucoside
RNA	Ribonucleic acid
RNAse	Ribonuclease
rpm	Revolutions per minute
s	second (s)
SCPL	Serine carboxypeptidase-like
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel
TEMED	Tetramethylenediamine
TLC	Thin layer chromatography
Tris	Tris-(hydroxymethyl)-aminoethane
UHPLC	Ultra-high-performance liquid chromatography
v/v	volume per volume
VA	Vanillic acid
VAG	Vanillyl-glucose
w/v	weight per volume

CHAPTER I

INTRODUCTION

1.1 General introduction

Rice is one of the most important commodities and is consumed as a staple food for the majority of the world's population. According to the international organization of Food and Agriculture (FAO) in 2018, the largest rice producers in the Asian countries are led by China, India, Bangladesh, Indonesia, Thailand, and Vietnam.By the pigment contained in the protective pericarp or seed, rice can be distinguished in several types, including white rice, red rice, and black rice (Ciulu *et al.*, 2018). Recently, the consumption of colored rice is increasing, due to public awareness in maintaining health, along with the results of research on the nutritional content in pigmented rice. Nutrients contained in white rice, such as sterols, γ-oryzanol, tocopherols, tocotrienols, and phenolic compounds, can be found in pigmented rice, often in higher amounts than in white rice (Iqbal *et al.*, 2005; Samad, 2015).

In addition, pigmented rice has proanthocyanidins and anthocyanins in red rice and black rice, respectively. Among the types of rice classified by color, black rice varieties have the highest antioxidant activities due to anthocyanins, which are mainly found in the bran (Laokuldilok *et al.*, 2011; Goufo and Trindade, 2014). Furthermore, studies recently revealed that anthocyanins isolated from black rice are able to prevent diabetes (Yawadio *et al.*, 2007), decrease cholesterol, LDL, and tri-acylglycerol in rat plasma (Zawistowski *et al.*, 2009)reduce body fat and act as anti-inflammatories (Bowen-Forbes *et al.*, 2010) and cardio protective agents, based on *in vivo* testing (Rechner and Kroner, 2005; He *et al.*, 2011).

Some enzymes involved in the formation and modification of anthocyanin compounds in black rice may include transglucosidases, and acyl-transferases, as well as glycosyltransferases (Matsuba *et al.*, 2010). The transglucosidases that have been studied are related to β-glucosidases, and particularly those that are members of glycoside hydrolase family GH1 (Matsuba *et al.*, 2010; Miyahara *et al.*, 2013; Nishizaki *et al.*, 2013). So far, much research has been reported about the content of the GH1 β-glucosidases in rice (Cairns *et al.*, 2015), with a few studies on Os9BGlu31 transglucosidase (Luang *et al.*, 2013; Komvongsa *et al.*, 2015), but none of these were reported to act on anthocyanin pigments.

Indeed, those studies were completed on white rice, which lacks anthocyanins. However, anthocyanin transglucosidases have been reported for carnation and delphinium, as well as Arabidopsis (Matsuba *et al.*, 2010; Miyahara *et al.*, 2013; Nishizaki *et al.*, 2013). The large and complex structure of delphinium pigments has been attributed to the successive action of acyl glucose-dependent glucosyl transferases and acyl transferases, the latter of which are serine carboxypeptidase-like (SCPL) proteins.

To date, some SCPL acyltransferases have been fully characterized. In *Lycopersicum pennellii,* a glucose acyltransferase (GAC) catalyzed the production of , a glucose isobutyryl polyesters, 1,2-di-*O*-acyl-β-glucose, which protects the plants from insect (Li and Steffens, 2000). In *Arabidopsis thaliana*, four SCPL acyltransferases synthesized sinapate esters using 1-*O*-sinapoyl-β-D-glucose to form sinapoyl malate (Lehfeldt *et al.*, 2000). Moreover, 1-*O*-sinapoyl:choline sinapoyltransferase (SCT) catalyzes choline acylation from 1-*O*-sinapoyl-β-D-glucose to form sinapoylcholine (sinapine) in seeds (Shirley *et al.*, 2001). Two SCT orthologs (BnSCT1 and BnSCT2) were identified in *Brassica napus* (Milkowski *et al.*, 2004; Weier *et al.*, 2008). 1-*O*-sinapoyl-β-D-glucose is also used as a donor molecule by sinapoylglucose: sinapoylglucose sinapoyltransferase (SST) to synthesize 1-*O*-sinapoyl-β-D-glucose, and by sinapoylglucose: anthocyanin sinapoyltransferase (SAT) to synthesize sinapoylated anthocyanins (Fraser *et al.*, 2007). In *Clitoria ternatea,* CtAT1 coumaroylates an anthocyanin called ternatin C5 (delphinidin-3-*O*-(6"-*O*-malonyl)-β-glucosidase-3',5'-di-*O*-β-glucosidase to produce ternatin C3 (Kogawa *et al.*, 2007) (Patent WO/2007/046148). In addition, Mugford *et al*. (2009) identified SCPL1, an acyltransferase that is able to synthesize a triterpene glycoside (avenacin) in *Avena strigose*. AsSCPL1 has the ability to transfer acyl group to des-acyl avenacin by using *N*-methyl anthraniloyl-*O*-glucose as the substrate.

As yet, the content of acyl-glucose-dependent acyl-transferase and transglucosidases in black rice have not been clearly demonstrated, although both of these enzymes play a role for anthocyanin synthesis in other plants, as noted above. Additionally, the expression and purification of SCPL protein is quite challenging due to insolubility of the protein produced in bacterial systems. Ciarkowska *et al* (2018) reported that the recombinant of IAInos synthase, a member of SCPL-acyltransferase family, was expressed using yeast and bacterial system with an additional step of refolding from inclusion bodies. Thus, the aim of this research study to express genes encoding rice SCPL protein and to characterize the function of rice SCPL protein. By

conducting this set of experiments, we expect to clarify the roles of SCPL proteins in rice, including possible acylation of anthocyanins.

1.2   Research objectives

The purpose of this project is to determine whether SCPL proteins are involved in acylation of anthocyanins or other compounds in rice.   The specific objectives include:

1.1   To identify the genes encoding SCPL proteins in rice, including those expressed in tissues that are pigmented in black rice, and test expression of two of these in yeast and bacterial systems.

1.2   To characterized the function of SCPL protein toward the formation of phenolic compounds in rice.
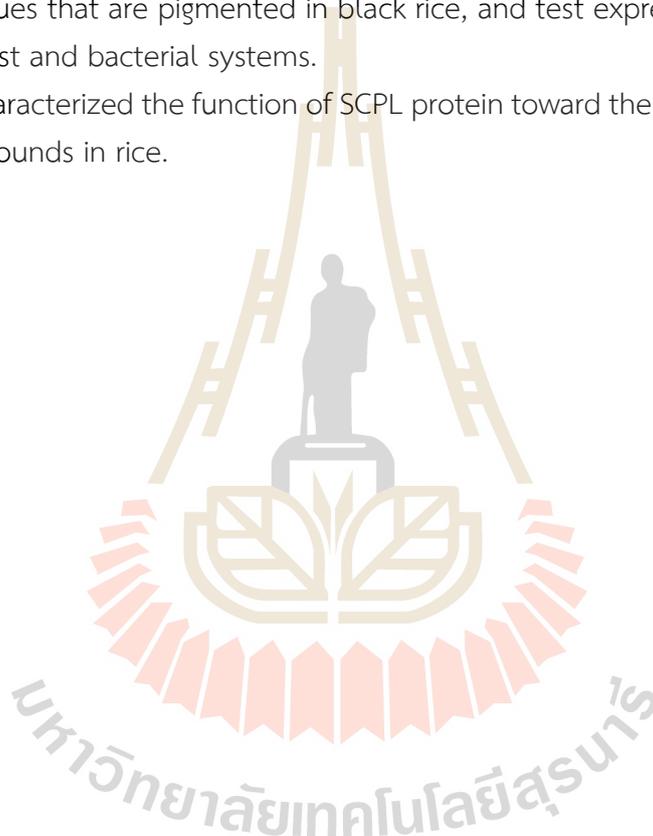
CHAPTER II
LITERATURE REVIEW

2.1 Rice (*Oryza sativa*)

Rice (*Oryza sativa*) is the monocot plant, which is the second most produced and most consumed cereal in the world (Goufo and Trindade, 2014). Rice is important in the human diet and health. The consumption of rice has increased significantly in the past two decades and is projected increase by 64 Mt to 562 Mt, with 84% of this increase occurring in Asian countries, led by China (146.73 Mt), India (118.87 Mt), Bangladesh (35.85 Mt), Indonesia (34.7 Mt), Vietnam (27.1 Mt) and Thailand (17.66 Mt), respectively (Statistica, 2021) **(Figure 2.1)**. Hence, it is also essential for economic development, in addition to food security.

Based on the various pigmentations of the outer layer, rice can be distinguished as white, red, black, etc. (Ciulu *et al.*, 2018). Purple rice is generally derived from black rice, but has less pigmentation. Several compounds with antioxidant activity have been identified in rice by high performance liquid chromatography (HPLC) analysis, including phenolic compounds, tocopherols, tocotrienols and γ-oryzanol (Iqbal *et al.*, 2005). Among the types of rice by color, black rice bran has the highest antioxidant activities, due to its anthocyanin content, which is only found in black rice (Laokuldilok *et al.*, 2011; Goufo and Trindade, 2014). The phenolic compounds are mainly associated with the pericarp in rice, and darker pericarp color contain higher amounts of polyphenols (Tian *et al.*, 2004; Zhou *et al.*, 2004). In black rice, pigments are primarily located in the aleurone layer, which is characterized as dark purple to black in color (Hu *et al.*, 2003).
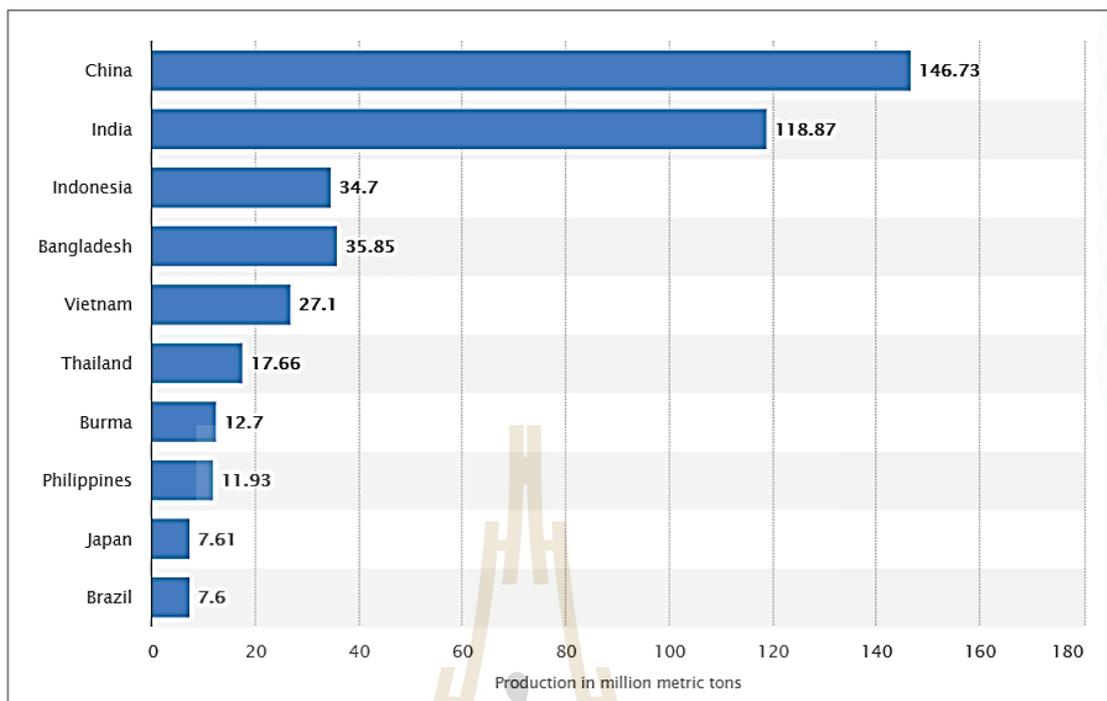
Figure 2.1 Top 10 countries that contribute to produce rice in world (Statista, 2021).

2.2   Phenolic compound in rice

Phenolic compounds are secondary metabolites that are generally distributed in the plant kingdom. For plants, phenolic compounds have many roles, such as protection against pathogens, predators, and ultraviolet radiation, as well as attraction of pollinating animals and mechanical support (Parr and Bolwell, 2000). These compounds also potentially have antioxidant properties and free radical scavenging capabilities. Moreover, previous studies revealed that phenolic compounds are able to exert various physiological effects in humans, such as preventing oxidative damage of lipids and low-density lipoproteins, as well as reducing the risk of coronary heart disease and cancer (Morton *et al.*, 2000). Fruits and vegetables are known to be major dietary sources of phenolic compounds. However, cereal consumption is also an excellent way to increase phenolic compound intake (Scalbert and Williamson, 2000).

The major phenolic compounds found in rice are divided into the simple molecules, such as phenolic acids (Al-Farsi and Lee, 2008) and their aldehydes (Qiu *et al.*, 2010), to highly polymerized compounds, such as lignin and tannins. However, the main phenolics in the diet are the phenolic acids, flavonoids and tannins. The phenolic acids in rice are derived from hydroxycinnamic acid and hydroxybenzoic acid. Hydroxycinnamic acid derivatives include ferulic, *p*-coumaric, chlorogenic, caffeic and sinapic acids (Li *et al.*, 2010; Lin and Lai, 2011). Ferulic and *p*-coumaric acids exist in the free form, the soluble conjugate form, or the insoluble bound form, which is found in dietary fiber (Krygier *et al.*, 1982). Protocatechuic, *p*-hydroxybenzoic, vanillic, and syringic acids are other compounds that are derived from hydroxybenzoic acid that have a C6-C1 structure **(Figure 2.2)** (Goffman and Bergman, 2004; Tian *et al.*, 2004; Vichapong *et al.*, 2010; Walter and Marchesan, 2011). Additionally, aldehydes, such as vanillin (Qiu *et al.*, 2010), protocatechuic aldehyde and *p*-hydroxybenzaldehyde, are also found in rice grains (Setyaningsih *et al.*, 2016).

Caffeic acid: R_1=OH, R_2=OH, R_3=OH, R_4=H
p-Coumaric acid: R_1=H, R_2=OH, R_3=H, R_4=H
Ferulic acid: R_1= OCH$_3$, R_2=OH, R_3= H, R_4=H
Sinapic acid: R_1=OCH$_3$, R_2=OH, R_3=OCH$_3$, R_4=H
Chlorogenic acid: R_1= OH, R_2=OH, R_3=H, R_4=quinate

Protocatechuic acid: R_1=H, R_2=OH, R_3=OH
Hydroxybenzoic acid: R_1=H, R_2=OH, R_3=H
Vanillic acid: R_1=H, R_2=OH, R_3=OCH$_3$
Syringic acid: R_1=OCH$_3$, R_2=OH, R_3=OCH$_3$
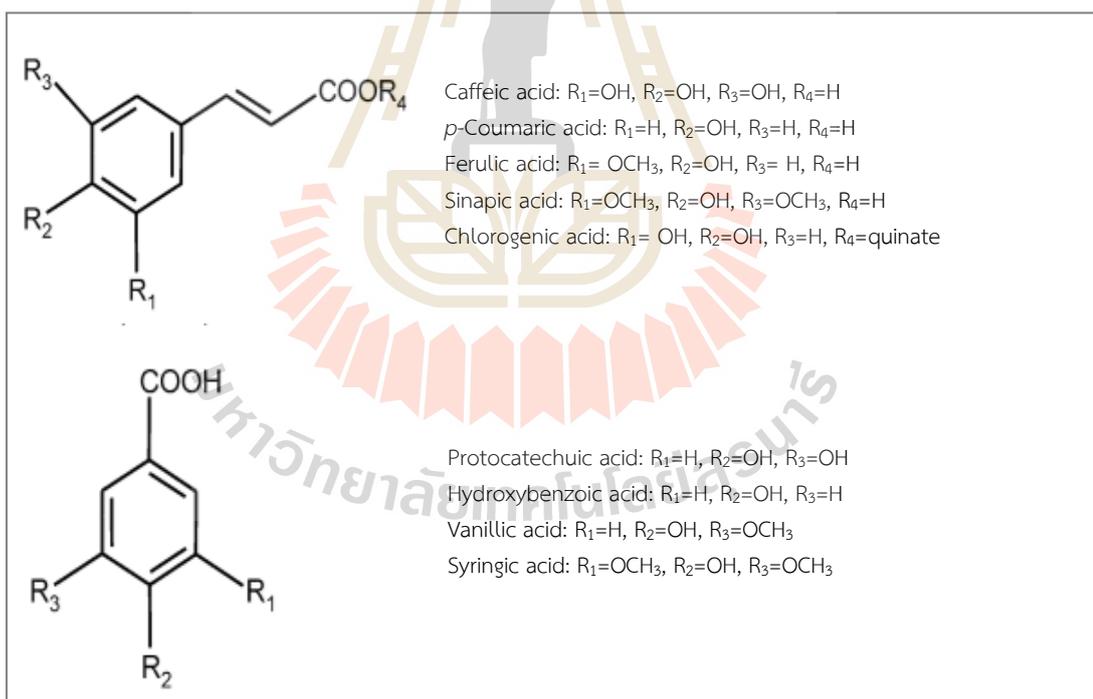
Figure 2.2 Chemical structures of simple phenolic compounds found in rice (Tian *et al.*, 2004).

There are some factors that affect the phenolic compounds in rice. Yawadio *et al*. (2007) reported that the pericarp color in the rice grain reflects the type and concentration of phenolic compounds. Pigmented rice varieties have higher concentrations of phenolic compounds compared to white rice. There is also variation in the content of total phenolics among rice with the same pericarp color (Goffman and Bergman, 2004), for instance, japonica rice varieties were found to be richer compared with Indica rice varieties (Goufo and Trindade, 2014).

Tian *et al*. (2014) revealed that polishing, germination, and storage processes influence the concentration of phenolic compounds. Polishing significantly reduces the concentration of phenolics, since these compounds are predominantly in the external layers of the grain, while germination results in a loss of approximately 70% of feruloyl-sucrose and sinapoyl-sucrose due to their hydrolysis during germination. Furthermore, Zhou *et al*. (2004) concluded that the reduction in the content of bound phenolic acids in brown and white rice occurs during storage, and the reduction was higher at 37 °C than at 4 °C.

2.2.1 Flavonoids in rice

Flavonoids are formed by 15 carbons, organized in two aromatic rings linked by a three-carbon chain (structure C6-C3-C6) (Ross and Kasum, 2002). Flavonoids can be divided into 13 classes, the most common being the flavonols, flavanols, flavones, isoflavones, anthocyanidins and flavanones (Scalbert and Williamson, 2000). In rice, there are some flavonoids commonly found, such as catechin, epigallocatechin, quercetin, kaempferol, isorhamnetin, apigenin, tricin and genistein **(Figure 2.3)** (Ciulu *et al.*, 2018). They occur as sulfated and methylated derivatives, conjugated with monosaccharides and disaccharides and form complexes with oligosaccharides, lipids, amines, carboxylic acids and organic acids (Duthie *et al.*, 2003). Furthermore, complex flavonoids in rice consist of quercetin-3-*O*-glucoside, isorhamnetin-3-*O*-glucoside, dihydroquercetin-3-*O*-glucoside, dihydroisorhamnetin-3-*O*-glucoside, rutin, isorhamnetin-3-*O*-rutinoside, luteolin-7-*O*-glucoside as well as apigenin-7-*O*-glucoside **(Figure 2.4)** (Ciulu *et al.*, 2018). Flavonoids are essential for human diet and admitted as the most potent antioxidants from plants (Jovanovic *et al.*, 1994).
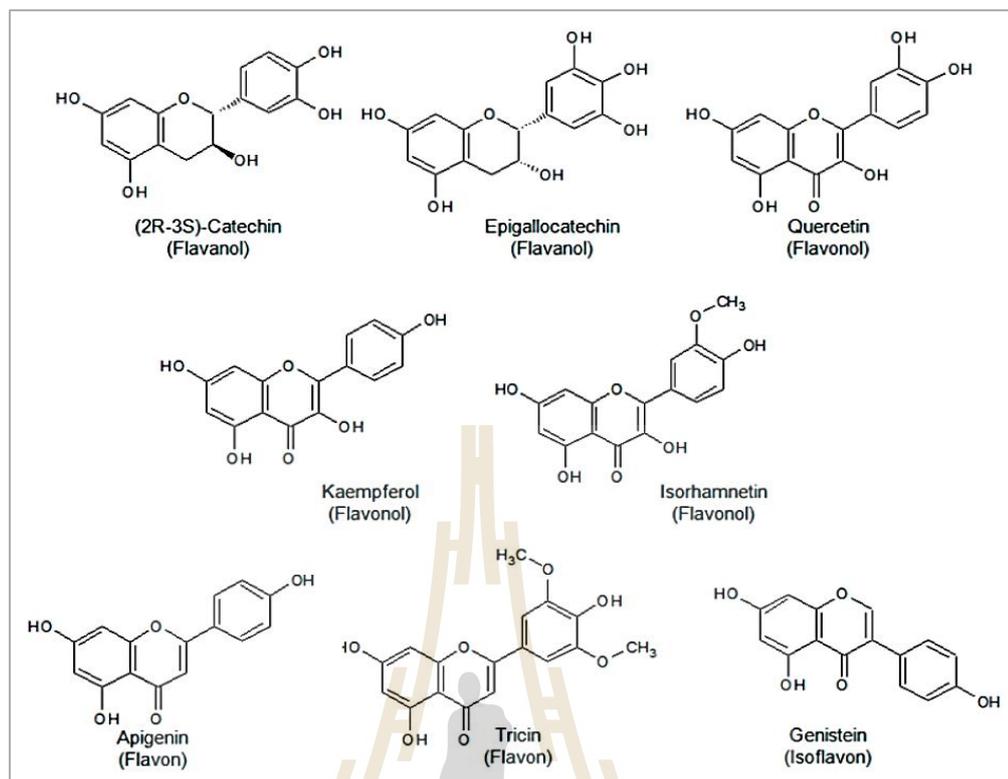
Figure 2.3 Some of the simple flavonoid commonly found in rice (Ciulu *et al*., 2018).
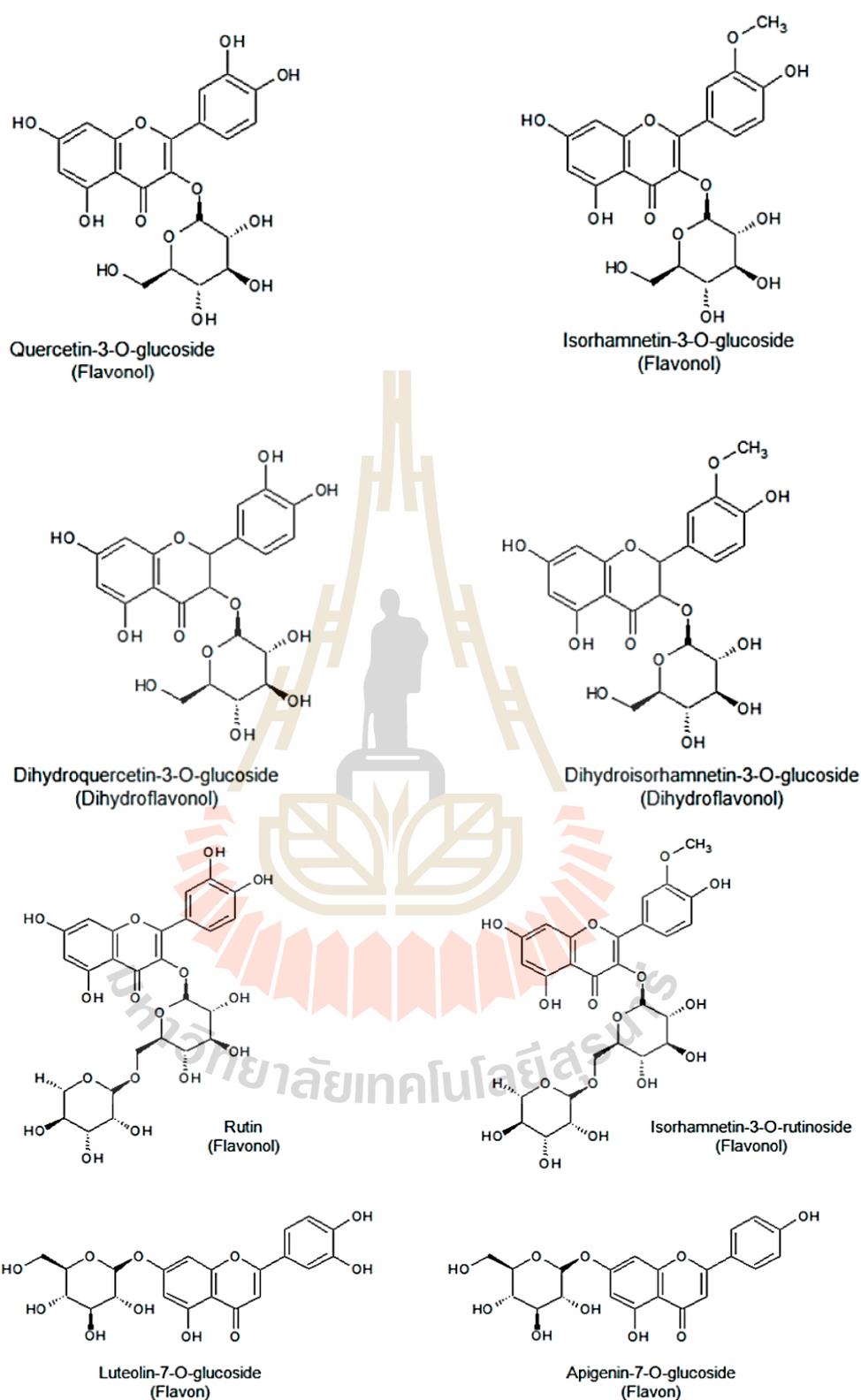
Figure 2.4 Some of the glycosylated flavonoids found in rice (Ciulu *et al*., 2018).

2.2.2 Anthocyanin in rice

Anthocyanins are colored water-soluble pigments and belong to the flavonoid family. These compounds are responsible for red, purple, and blue pigments found in plants, especially flowers, fruits, and tubers. Berries, currants, grapes, and some tropical fruits have high anthocyanin contents, while red to purplish blue-colored leafy vegetables, grains, roots, and tubers are the edible vegetables that contain a high level of anthocyanins (Khoo *et al.,* 2017).
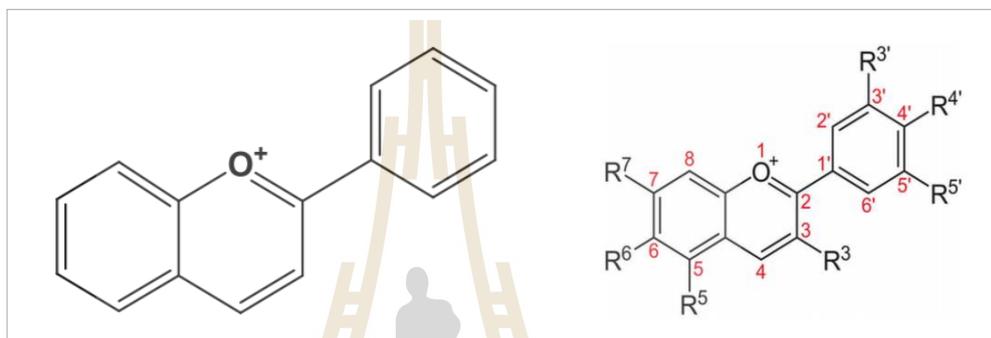
Figure 2.5 Structure of flavylium ion (left) and basic anthocyanin structure (right) (Khoo *et al.,* 2017).

Anthocyanins are the glycosylated forms of anthocyanidins and have a positive charge at the oxygen atom of the C-ring of the basic flavonoid structure, namely a flavylium (2-phenyl chromenylium) ion **(Figure 2.5)**. The empirical formula for the flavylium ion of anthocyanidin is $C_{15}H_{11}O^+$, which has a molecular weight of 207.24 g/mol. Anthocyanins are in the forms of anthocyanidin glycosides and acylated anthocyanins, while anthocyanidins, known as the aglycone, are divided into 3-hydroxy anthocyanidins, 3-deoxy anthocyanidins, and O-methylated anthocyanidins (Sasaki *et al.,* 2014; Khoo *et al.,* 2017). The anthocyanidins found in plants include cyanidin, delphinidin, malvidin, pelargonidin, peonidin, and petunidin. Their molecular structures are shown in **Figure 2.6**.
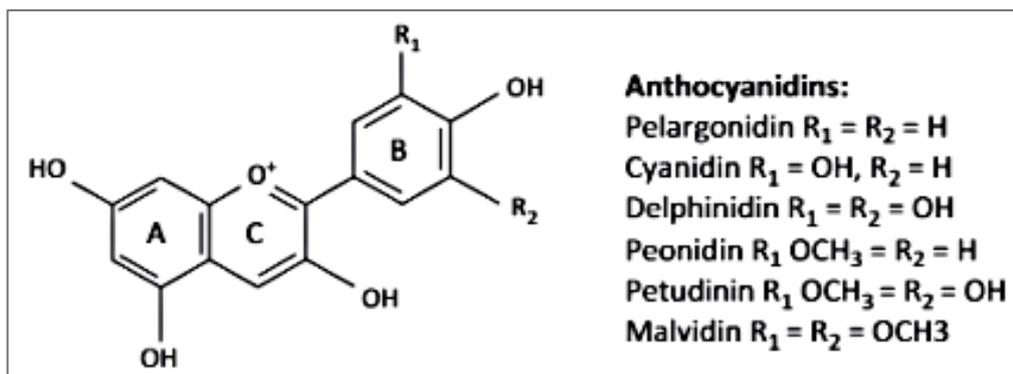
Figure 2.6 Molecular structure of major anthocyanin in plants (Jaakola, 2013).

The flavonoid biosynthetic pathway leading to anthocyanins via the phenylpropanoid pathway is well described **(Figure 2.7)**. The specific flavonoid pathway begins with the condensation of one molecule of 4-coumaroyl-coenzyme A (CoA) and three molecules of malonyl-CoA, which results in naringenin chalcone. This reaction is catalyzed by chalcone synthase (CHS). Flavanone-3-hydroxylase (F3H), flavonoid $3'$-hydroxylase (F3$'$H), and flavonoid $3',5'$-hydroxylase (F3`5`H) then direct the route towards cyanidin and delphinidin anthocyanidins (Koes *et al.*, 2005; Hichri *et al.*, 2011).

Anthocyanidins are converted from leucoanthocyanins by leucoanthocyanidin dioxgenase/anthocyanidin synthase (LDOX/ANS) and further glycosylated by uridine diphosphate (UDP)-glucose: flavonoid-*O*-glycosyltransferase (UFGT). *O*-Methyltransferases (OMTs) catalyze the formation of *O*-methylated anthocyanins, such as malvidin, peonidin, and petunidin (Koes *et al.*, 2005; Hichri *et al.*, 2011). The enzymes involved in the flavonoid biosynthesis pathway are localized in the cytosol. After biosynthesis, flavonoids are transported to vacuoles or cell walls (Koes *et al.*, 2005). In addition, further glycosylation and acylation increase the color stability of anthocyanins (Patras *et al.*, 2010).

Furthermore, the modification that occurs on anthocyanins, such as the position within the molecule of the modifying moiety, the type of attached sugar or organic acid, or a combination of these variable factors can lead to structural and functional diversity. These modifications include glycosylation via glycosyltransferases (GTs), and acylation catalyzed by acyltransferases (ATs). These modifications play important roles in the chemical stabilization of anthocyanins in the vacuolar sap and determining the color of the flowers in

response to variations in vacuolar sap pH or to interaction with metal ions (Yoshida *et al.*, 2009).
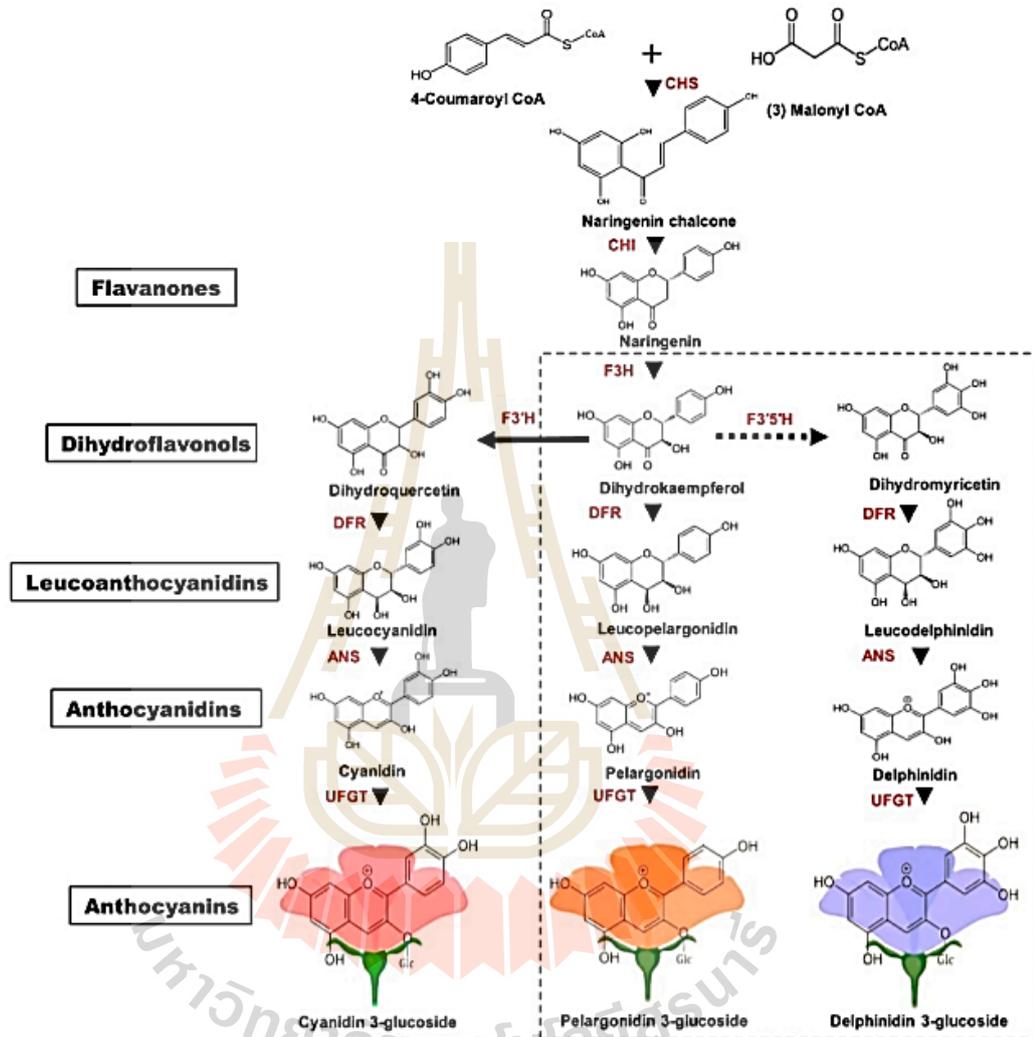
Figure 2.7 Schematic of anthocyanin biosynthesis pathway via general phenylpropanoid pathway (Mekapogu *et al.*, 2020).

There are two types of these enzymes, namely cytoplasmic AT and GT, and vacuolar AT and GT, respectively. Cytoplasmic AT and GT utilize acyl-CoA and UDP-sugar as donor molecules. Glycosylation of anthocyanin is catalyzed by family 1 glycosyltransferases (UGTs) (Yonekura-Sakakibara and Hanada, 2011) while acylation by AT that belong to the BAHD family protein (D'Auria, 2006). On the other hand, both vacuolar AT and GT use acyl-glucoses as donor molecules

(Nishizaki *et al.*, 2013). Acyl-glucose-dependent ATs involved in diacyl-glucose biosynthesis in the wild tomato (*Lycopersicon pennellii*) and sinapoyl-malate biosynthesis in *Arabidopsis thaliana* were analyzed and their amino acid sequences determined.

These acyl-glucose dependent acyltransferase sequences belong to the serine carboxypeptidase-like (SCPL) protein family (Lehfeldt *et al.*, 2000; Li and Steffens, 2000). The vacuolar GT are really transglycosidases (TG) that belong to glycoside hydrolase family 1 (GH1), members of which characteristically hydrolyze glycosides. These enzymes transfer a glucosyl group to an anthocyanin using an acyl-glucose as the glucosyl donor. Examples of this process were described in delphinium and Arabidopsis, where *p*-hydroxybenzoylglucose and sinapoyl-glucose are used as zwitter donors (Miyahara *et al.*, 2013; Nishizaki *et al.*, 2013).

2.3   Serine carboxypeptidase-like (SCPL) acyltransferases

Serine carboxy peptidase-like acyltransferases (SCPL-ATs) are enzymes that using phenolic glycoconjugates to conduct acylation, such as in the formation and modification of anthocyanins (Bontpart *et al.*, 2015). SCPL-ATs belong to the α/β hydrolases superfamily. These enzymes use 1-*O*-β-glucose esters as energy rich donor molecules. Previously, SCPL protein was annotated as peptidases based on sequence similarity. However, recent studies have reported that some of them function not as peptidases, but as acyltransferases and lyases as well (Lehfeldt *et al.*, 2000; Shirley *et al.*, 2001).

Several SCPL proteins have been shown to catalyze the production of plant secondary metabolites involved in herbivory defense and UV protection. For example, Arabidopsis catalyzes the formation of sinapoyl-malate, a UV-protectant in seedlings by sinapoyl-glucose: malate sinapoyltransferase (SMT) and sinapoyl-glucose: choline sinapoyltransferase (SCT) is required for the formation of sinapoyl-choline in Arabidopsis seeds (Lehfeldt *et al.*, 2000; Shirley *et al.*, 2001). An SCPL-AT in delphinium has been shown to have a function in the formation of complex anthocyanin (Nishizaki *et al.*, 2013).

Blue coloration in delphinium sepals is derived from the 7-polyacylated anthocyanins violdelphin and cyanodelphin **(Figure 2.8)** (Hashimoto *et al.*, 2002). Nishizaki *et al*. (2013) revealed the polyacyl moieties in delphinium are composed of glucose and *p*-hydroxybenzoic acid (pHBA) molecules in which each pHBA in linked to

glucose moieties through an ester bond to form *p*-hydroxybenzoyl-glucose (pHBAG) as the donor. In delphinium, AA7G-AT and AA7GBG-AT, which is an SCPL protein, transfers *p*-hydrobenzoic acid to the anthocyanin in the glucosyl position of anthocyanin. It was also previously report also describe that acyl-glucosyl-dependent acyltransferase catalyzes and transfer *p*-hydroxybenzoic acid to anthocyanin in butterfly pea (Noda, 2006).
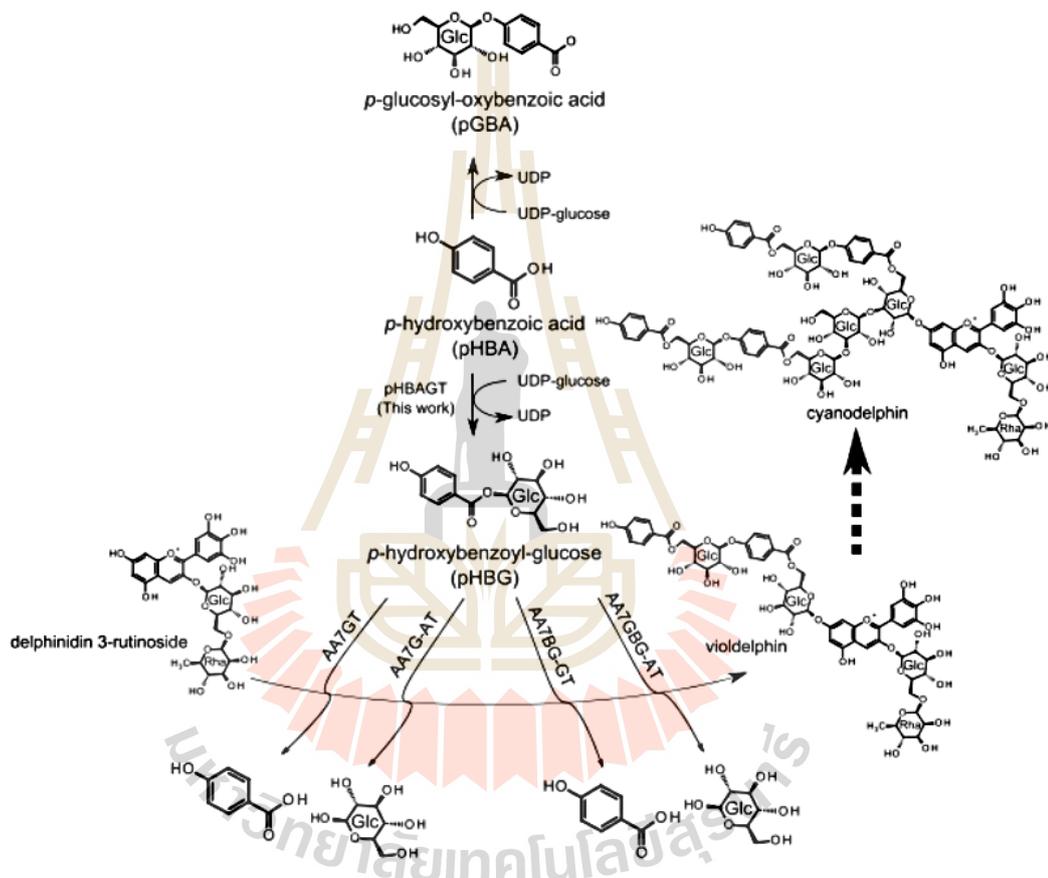
Figure 2.8 Biosynthesis of complex anthocyanin in delphinium (cyanodelphin) (Nishizaki *et al*., 2013).

Bontpart *et al*. (2015) summarized the localization and function of SCPL and BAHD acylation in the cell. SCPL acyltransferases are predicted to contain a signal peptide that allows the protein to follow the secretory pathway before finally being targeted to the vacuole (Lehfeldt *et al.*, 2000; Mugford *et al.*, 2009). The signal peptide is recognized and cleaved at the endoplasmic reticulum (ER). The protein is then transported to the Golgi apparatus (GA), where it probably undergoes

glycosylation. Finally, the protein is targeted to the vacuole to catalyze acylation with glucose esters acting as the acyl donor **(see Figure 2.9)**.
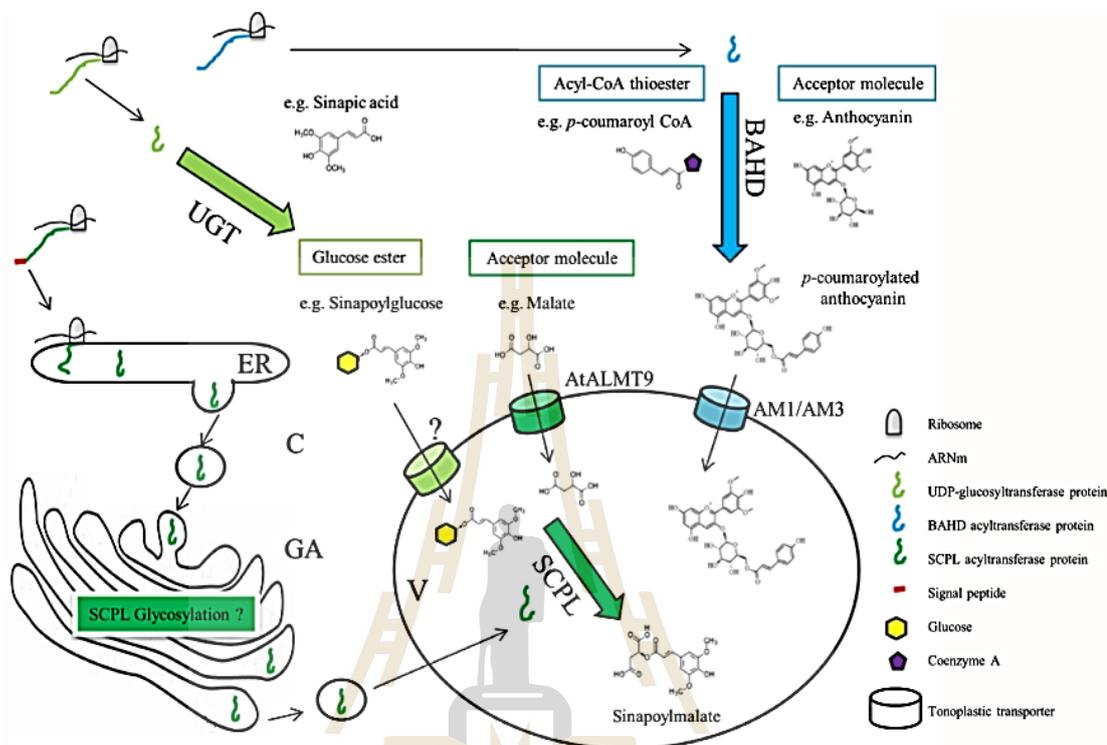
Figure 2.9 The function of SCPL and BAHD acylation in plant cells (Bonpart *et al.*, 2015).

Meanwhile, in the cytosolic compartment, UDP glucosyltransferase (UGT) uses UDP-glucose to glycosylate secondary metabolites and form glucose esters, for example sinapoyl glucose. Glucose esters are then imported into the vacuole via tonoplastic transporters. In the vacuole, glucose esters are donor molecules used by SCPL acyltransferases to acylate acceptor molecules, such as malate. Malate can be imported into the vacuole by the aluminum-activated malate transporter in *Arabidopsis thaliana* (AtALMT9) **(Figure 2.9)** (Bontpart *et al.*, 2015).

Due to the post-translational modifications, which are only suitably accomplished in eukaryotic hosts that carry suitable cellular machinery for such modifications, the characterization of SCPL-ATs must often use eukaryotic hosts. Heterologous expression of SCPL-ATs in yeast has proven to be a good strategy for obtaining functional SCPL-ATs (Stehle *et al.*, 2008). As of 2015, only nine SCPL-ATs had been fully characterized (**Table 2.1**; Bontpart *et al.*, 2015) with just SCPL1 identified in

a monocot plant (oat, *Avena strigose* AsSCPL1), where it catalyzes production of avenacin (Mugford *et al*., 2009).

Table 2.1 Characterize and putative functional SCPL in plants (Bontpart *et al*., 2018).

Name	Species	NCBI number	Function or putative function
LpGAC	*Lycopersicon pennellii*	AAF64227	Isobutyryl acylation
CtAT1	*Clitoria ternatea*	BAF99695	Anthocyanin acylation
AtSMT	*Arabidopsis thaliana*	AAF78760	Malate acylation
AtSCT	*Arabidopsis thaliana*	AAK52316	Choline acylation
BnSCT1	*Brassica napus*	AAQ91191	Choline acylation
BnSCT2	*Brassica napus*	CAM91991	Choline acylation
AtSAT	*Arabidopsis thaliana*	AEC07395	Anthocyanin acylation
AtSST	*Arabidopsis thaliana*	AEC07397	Sinapoyl glucose acylation
AsSCPL1	*Avena strigosa*	ACT21078	Avenacin acylation
SCPL17*	*Arabidopsis thaliana*	AAS99709	Glucosinolate acylation
DkSCPL1*	*Diospyros kaki*	BAF56655	Flavan-3-ol galloylation
DkSCPL2*	*Diospyros kaki*	BAH89272	Flavan-3-ol galloylation
DgSCPL1*	*Delphinium grandiflorum*	BAO04182	Anthocyanin acylation
DgSCPL2*	*Delphinium grandiflorum*	BAO04183	Anthocyanin acylation
DgSCPL3*	*Delphinium grandiflorum*	BAO04184	?
CsSCPL*	*Camellia sinensis*	AIW39897	Associated with abiotic stress
HvSCPI	*Hordeum vulgare*	P07519	Protein degradation
HvCPDWII	*Hordeum vulgare*	P55748	Storage protein mobilization
HvSCPIII	*Hordeum vulgare*	P21529	Protease
TaCPD-WII	*Triticum aestivum*	P08819	Protease
ScCPY	*Saccharomyces cerevisae*	NP_014026	Protease
SbHNL1	*Sorghum bicolor*	CAD12888	Hydroxynitrile lyase
BRS1	*Arabidopsis thaliana*	Q9M099	Protease, brassinosteroid signaling
LeSCP	*Lycopersicon esculentum*	Q9M513	SCP Type I, wound inducible
MtSCP1	*Medicago truncatula*	AAQ63884	Mycorrhiza specific

Based on phylogenetic analysis, there are many SCPL acyltransferases in both dicot and monocot species **(Figure 2.10)**. Members of Clade 1A of SCPL proteins are established as acyltransferases with roles in plant secondary metabolism, such as

producing extended anthocyanin. The expression pattern of the gene encoding the closest relative to AsSCPL1 in rice (Os10g01134) was analyzed using publicly available transcriptome data sets to investigate the possible functions of other monocot members. The available *in silico* data indicated that this gene is expressed principally in the leaves, developing inflorescence and seed, while the oat AsSCPL1 gene is expressed exclusively in the epidermal cells of oat root tips (Mugford *et al*., 2010).
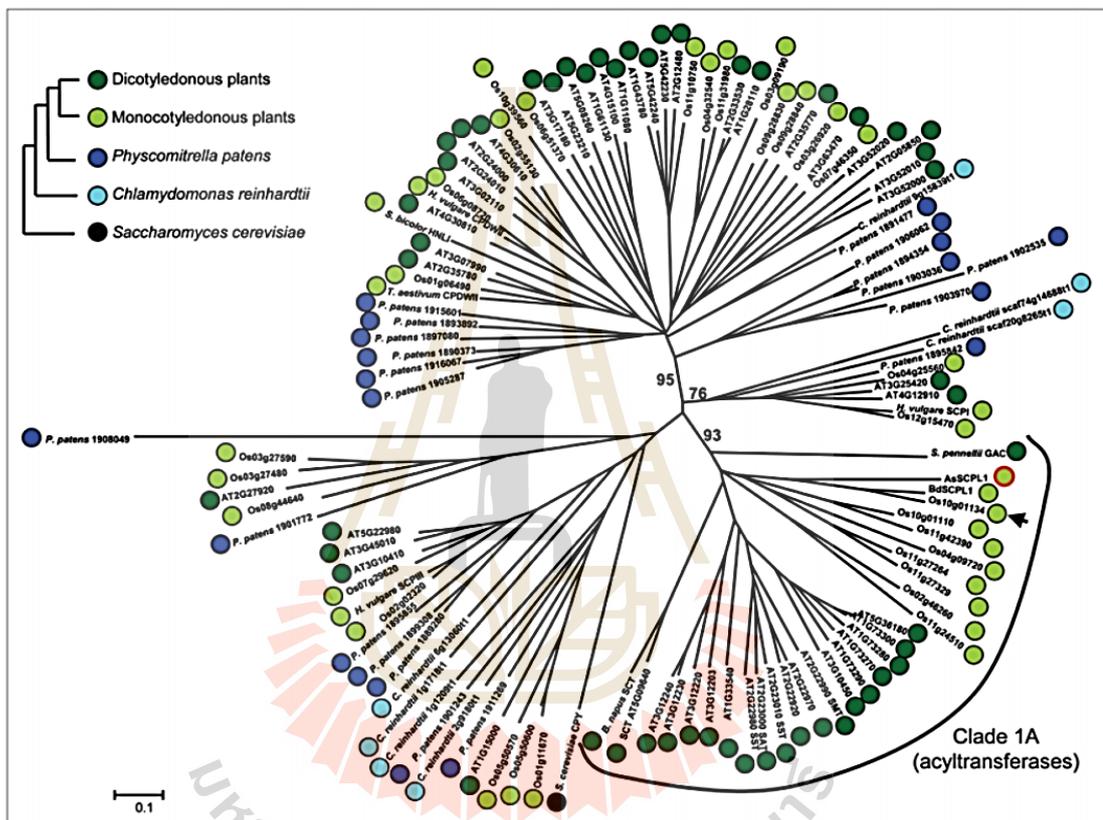
Figure 2.10 Phylogenetic tree of plant SCPL protein family (Mugford *et al*., 2010).

CHAPTER III
MATERIALS AND METHODS

3.1 General materials

3.1.1 Chemicals and reagents

Chemical and reagents used in this work, along with their sources, are shown in Table 3.1.

Table 3.1 Chemicals, reagents and sources.

Chemicals and Reagents	Sources
Acetic acid	Carlo Erba
Acetic acid glacial	
1-Butanol	
Calcium chloride	
Chloroform	
Citric acid	
Cobalt (II) chloride	
Ethanol	
Ethylenediaminetetraacetic acid disodium salt	
Formic acid	
Glycerol	
Hydrochloric acid	
Isopropanol	
Methanol	
Sodium chloride	
Sodium dodecyl sulphate (SDS)	
Tris (hydroxymethyl)-aminomethan	
2-mercaptoethanol	Acros Organics
Aminocaproic acid 99%	
Benzamidine hychloride 99%	
Coomassie brilliant blue R250	
Dithiothreitol (DTT)	
Iodoacetamide	
Triton X-100	

Table 3.1 Chemicals, reagents and sources (Continued).

Chemicals and Reagents	Sources
Acrylamide	Pacific Science
Ammonium persulfate (APS)	Amresco
Agarose	Vivantis
Protein marker	
Vivantis plasmid miniprep	
GeneJET Gel Extraction Kit	ThermoFisher
T4-DNA ligase	Promega
Sodium hydroxide	Merck
Trifluoroacetic acid	
Acetonitrile	Labscan
HPLC Grade Water	Bio Basic
Isopropyl-β-D-thiogalactopyranoside (IPTG)	
Kanamycin sulfate	
Tetracycline HCl	
Tetramethyl ethylenediamine (TEMED)	
Lysozyme	
p-Coumaric acid	
Phenylmethylsulfonyl fluoride (PMSF)	
Agar bacteriological	
Glycine	
Peptone	
Yeast extract	
Ampicillin, sodium salt	
Bis-acrylamide	
Imidazole	Affymetrix
DNA 1 kb marker	Gene Ruler
Trypsin	USB
Vanillic acid	Fluka
Apigenin-7-*O*-glucoside	Extrasynthase
Kaempferol-3-*O*-glucoside	
Kaempferol-7-*O*-glucoside	
Luteolin-7-*O*-glucoside	
Quercetin-3-*O*-glucoside	
Quercetin-7-*O*-glucoside	

Table 3.1 Chemicals, reagents and sources (Continued).

Chemicals and Reagents	Sources
Quercetin	TCI
Os9BGlu31 protein	Lab's stock
Zeocin	Invitrogen
IMAC sepharose 6 fast flow resin	GE Healthcare
Buffer 3.1	New England Biolabs
Enterokinase	
Restriction enzymes: *Nco*I and *Sal*I	
4-hydroxybenzoic acid 99%	Sigma-Aldrich
4-nitrophenyl-β-D-glucopyranoside	
Ammonium bicarbonate	
Apigenin	
Dimethyl sulfoxide (DMSO)	
Ferulic acid	
Kaempferol	
Luteolin	

3.1.2 Instruments and equipment

Instruments and equipment that were used for this research include of: laminar air flow cabinet, incubator, shaking incubator, autoclave, centrifuge, weighing balance, freezers (-80, -40, -20, and 4 °C), micropipettes, vacuum condenser, hot air oven, column chromatography system, UV light, sonicator, nanodrop, spectrometer, G4220 ultra-high-performance chromatography (UHPLC) with tandem quadrapole-time of flight mass spectrometer (LC-MSMS) (Agilent, USA), personal computer and printer.

3.1.3 Plasmid, bacterial strain and vectors

Two optimized *Oryza sativa* SCPL2a (OsSCPL2a) and SCPL7 (OsSCPL7) cDNA was synthesized and inserted into pPICZαBNH8 vector at the *Pst*I and *Xba*I sites by GenScript Corporation (Piscataway, NJ USA). The cDNAs were inserted into pET32a using restriction sites of *Nco*I and *Sal*I. The corresponding recombinant plasmid was transformed into *Escherichia coli* (*E. coli*) strain XL1-Blue for cloning, increasing the amount of recombinant plasmid and preparing for sequencing purpose. Competent cells of Origami(DE3) were used to express in bacterial system and *Pichia pastoris (P. pastoris)* strain SMD1168H for yeast expression system.
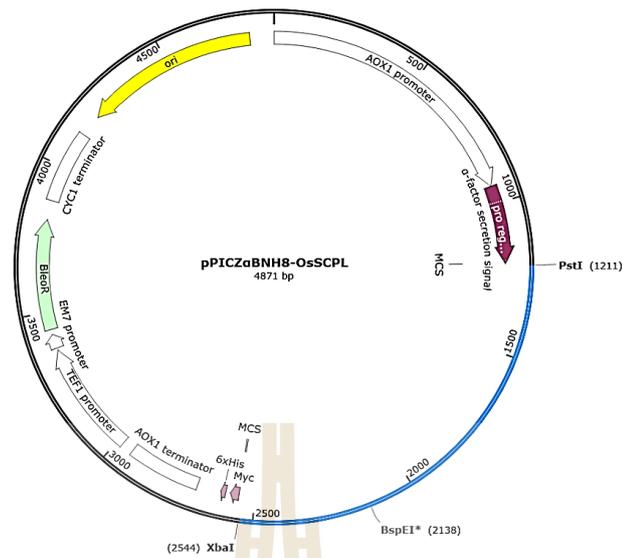
Figure 3.1 Constructs of recombinant plasmids pPICZαBNH8 with optimized gene OsSCPLs (blue line) for yeast expression protein system. The corresponding cDNAs were ligated at PstI and XbaI position.
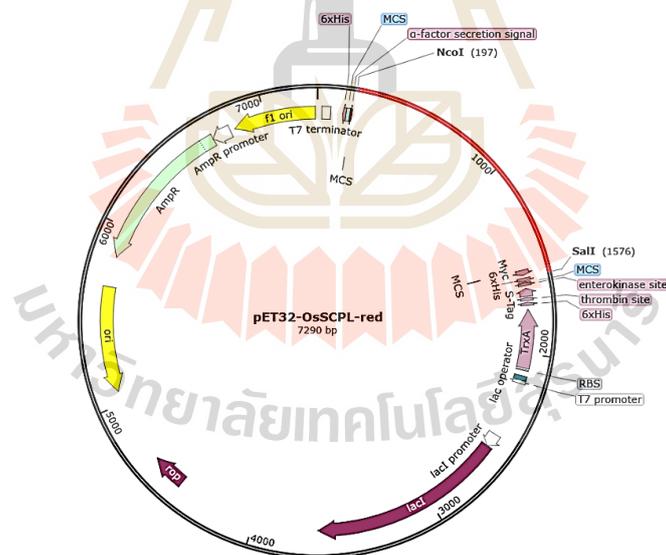
Figure 3.2 Constructs of recombinant plasmids pET32a with optimized gene OsSCPLs (red line) for yeast expression protein system. The corresponding cDNAs were ligated at NcoI and SalI position.

3.2 General methods

3.2.1 Preparation of *E. coli* strain XL1-Blue competent cells

The XL1-Blue competent cells were streaked from a frozen glycerol stock onto the Lennox broth (LB) agar (1% peptone, 0.5% sodium chloride, 0.5% yeast extract and 1.5% agar, w/v) without antibiotics and incubated at 37 °C overnight. A single colony was picked and inoculated into a 5 mL of LB broth (1% peptone, 1% sodium chloride and 0.5% yeast extract, w/v) and this culture incubated at 37 °C, with 200 rpm shaking overnight. To make a starter culture, 1 mL of overnight culture was added to 99 mL fresh LB media (1:100 dilution, without antibiotics) and incubated at 37 °C, with 200 rpm shaking for 3-4 h until the optical density (OD) reached 0.4-0.6 at 600 nm.

The culture was transferred to a 50 mL pre-cooled sterilized polypropylene centrifuge tube and chilled on ice for 5-10 min. The cell pellet was harvested by centrifuging at 4,000 rpm at 4 °C for 10 min and resuspended with 10 mL of prechilled 0.1 M $CaCl_2$, then placed on ice for 20 min. The mixture was centrifuged to collect the pellet and that pellet was resuspended with 0.5 mL of prechilled 0.1 M $CaCl_2$, then placed on ice for 1 h. The mixture was mixed with 15% volume glycerol and the cell suspension was aliquoted to 50 µL in each tube and kept at -80 °C.

3.2.2 Preparation of *P. pastoris* strain SMD1168H competent cells

Pichia pastoris SMD1168H competent cell was streaked in YPG agar media (1% yeast extract, 2% peptone, 2% glucose and 2% agar, w/v) without antibiotics and incubated at 30 °C for 2-3 days. A single colony was inoculated to 50 mL of YPG broth (1% yeast extract, 2% peptone and 2% glucose, w/v) and incubated at 30 °C, 200 rpm overnight. The cells were collected by centrifugation at 3,500 rpm, 4 °C for 10 min and resuspended with 25 mL of prechilled sterilized HPLC H_2O, followed by incubating on ice for 5 min.

The cells were harvested by centrifugation, resuspended with 20 mL of prechilled 1 M sorbitol and incubated on ice for 5 min. The cells were collected again by centrifugation, resuspended with 10 mL of ice-cold 1 M sorbitol and incubated on ice for 5 min. At the final step, the harvested cells were resuspended with 0.5 mL of prechilled 1 M sorbitol and 50 µL was aliquoted into each tube after addition glycerol to a final concentration 15% (v/v). The aliquoted competent cells were kept at -80 °C before using for plasmid transformation.

3.2.3 Transformation of pPICZ**α**BNH8/OsSCPLs and pET32a/OsSCPLs plasmid into XL1-Blue and Origami(DE3)

E. coli strain XL1-Blue competent cell aliquot was taken out from the -80 °C freezer and thawed on ice for approximately 20-30 min. The LB media agar was prepared with 25 µg/mL zeocin for pPICZ**α**BNH8/OsSCPLs transformed to XL-1 Blue; 50 µg/mL ampicillin for pET32a/OsSCPLs transformed to XL1-Blue, and 12.5 µg/mL tetracycline, 15 µg/mL kanamycin, 50 µg/mL ampicillin for pET32a/OsSCPLs transformed to Origami(DE3). One microliter of plasmid was added into the competent cells and gently mixed by flicking at the bottom of the tube. The mixture was incubated on ice for 30 min and transformed by placing the tube into the 42 °C water bath for 45 s. The tube was put back on ice for 2 min and then 300 µL LB broth media was added to the tube. The mixture was grown at 37 °C for 45 min. All of the transformed cells were spread onto the LB agar plate and incubated at 37 °C overnight.

3.2.4 pPICZ**α**BNH8/OsSCPL plasmid extraction by alkaline lysis

For isolation of pPICZ**α**BNH8/OsSCPL plasmid, alkaline lysis was performed (Zhou *et al.*, 1997). A freshly grown colony of *E. coli* XL1-Blue harboring the pPICZ**α**BNH8/OsSCPL plasmid was inoculated into 5 mL LB broth media containing appropriate antibiotics and grown at 37 °C with 200 rpm shaking overnight. After overnight culture, the cells were harvested by centrifugation at 2,500 rpm at 4 °C for 30 min and resuspended in 150 µL of the resuspension buffer (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0) completely by pipetting on ice.

The resuspended pellet was transferred into a new tube and 2x volumes of freshly prepared lysis buffer (0.2 M NaOH, 1% sodium dodecyl sulfate) was added quickly. The mixture was inverted gently and briefly before incubated on ice for 5 min. The cellular debris was precipitated with 2x volume of prechilled 3 M CH_3COOK, pH 4.8, and inverted briefly before incubating on ice for 5 min.

Then, the mixture was centrifuged at top speed for 10 min. The supernatant was transferred into new tubes, followed by adding 1x volume of phenol: chloroform (1:1) and briefly inverting, before centrifugation at maximum speed and room temperature in a microcentrifuge for 5 min. The aqueous upper layer was transferred into the new tubes and 2 volumes of absolute ethanol was added, briefly inverted and incubated at room temperature for 20 min. Subsequently, the mixture was kept on ice for 10 min,

followed by centrifugation at 12,000 rpm at 4 ℃ for 10 min. The supernatant was removed. The pellet was washed with 500 µL of 70% ethanol, mixed and centrifuged at 12,000 rpm for 5 min.

All residual supernatant was aspirated for 5-10 min followed by adding 100 µL TE buffer, pH 8.0, + 2 µL of 1 mg/mL RNase and incubated at 37 ℃ for 10 min. Then, 70 µL of prechilled 20% PEG 6000 + 2.5 M NaCl was added and incubated on ice for 1 h. The mixture was centrifuged at 12,000 rpm at 4 ℃ for 10 min. After removing the supernatant, the pellet was rinsed with 500 µL of 70% ethanol and centrifuged at maximum speed, 4 ℃ for 10 min before discarding the supernatant.

The pellet was dried for 10 min and dissolved with 20 µL of autoclaved HPLC-grade H_2O or TE buffer. Furthermore, the concentration of DNA plasmid absorbance at 260 nm and 280 nm by nanodrop 2000 spectrophotometer (Thermo Scientific-Massachusetts, USA). The ratio of absorbance at 260 and 280 nm is used to assess DNA purity. The ratio of ±1.8 is generally accepted as "pure DNA". If the ratio is appreciably lower (≤1.6), it may indicate the presence of protein, phenol, or other contaminants that absorb strongly at or near 280 nm. In contrary, since absorbance readings cannot discriminate between DNA and RNA, the presence of RNA can lead to the ratio increasing and this possibility must be considered to avoid DNA over quantification. The plasmid DNA was kept at -20 ℃ for further analysis.

3.2.5 Plasmid linearization and transformation into SMD1168H competent cells

The recombinant plasmids were linearized by digestion with *SacI* for integration into the alcohol oxidase (AOX) gene promotor. A linear DNA plasmid was analyzed by 1% agarose gel electrophoresis. Approximately 5-10 ng of DNA linearized plasmids were transformed into 50 µL of Pichia pastoris SMD1168H competent cells by electroporation. The contents of the tube were mixed and transferred into prelabelled and prechilled electroporation cuvettes. The cuvettes were wiped with tissue paper to remove any moisture, water droplets, or any other contaminants. Electric pulse was given by Gene Pulser II (Bio-Rad, California-USA) using the following setting (voltage: 1500 VH, capacitor: 25 µF using BTX, ECM 20 and time: 5 s) (Kumar, 2019).

Immediately, 1 mL of ice-cold 1 M sorbitol (as recovery medium) was added into the cuvettes and mixed. The transformed *Pichia* was transferred into the new tube and incubated at 30 ℃ for 1 h. The mixture was divided into

three plates which received the following volumes of cell suspensions (200 µL, 300 µL and 500 µL), spread on YPGS agar (1% yeast extract, 2% peptone, 2% filtered glucose, 2% agar and 18.2% sorbitol, w/v) containing 100 µg/mL zeocin and incubated at 30 °C for 2-3 days. The colonies were selected and screened in YPG agar containing 250 µg/mL zeocin.

3.2.6 Agarose gel electrophoresis

The quality of DNA plasmid was observed by electrophoresis in a 1% agarose gel (1 g agarose gel in 100 mL of 1x TAE buffer) (Kotchoni *et al.*, 2003). The agarose was microwaved for 1-3 min until completely dissolved and cooled down to about 50 °C at room temperature (about 5-10 min). After that, the agarose gel was poured slowly to avoid bubbles into a gel tray with the well comb in place. Any bubbles could be pushed away from the well comb or towards the edges of the gel with a pipette tip. A newly poured gel was placed at room temperature for 30 min until it had completely solidified. The solid gel was placed onto the gel box and filled with 1x TAE buffer until the gel is covered. A molecular weight ladder (10 µL of 1 kb DNA marker, 10 µL loading dye and 20 µL HPLC H_2O) was carefully loaded into the first lane of the gel. The DNA sample was prepared by diluting with a 6x loading buffer to two-fold concentration of DNA and also carefully loaded into the additional wells of the gel.

The gel was run at 60 V for 1 h or until the dye line is approximately 75-80% of the way down the gel. Subsequently, the gel was placed into a container filled with 100 mL of 1x TAE running buffer and 5 µL ethidium bromide for 30 s and destained with water three times for 5 min each. The DNA fragment was visualized on a UV transilluminator.

3.3 Protein expression in the yeast system and purification

Protein induction and expression was conducting by inoculating a single colony of *Pichia pastoris* SMD1168H selected to contain genome-integrated pPICαBNH8/OsSCPL2, pPICαBNH8/OsSCPL7, and empty vector pPICαBNH8 into BMGY media (YPG broth + 100 mM potassium phosphate buffer, pH 6.0, 1.34% (w/v) yeast nitrogen base (YNB) w/o amino acids, 1% (v/v) glycerol and 4×10^{-5}% biotin) (Inan *et al.*, 2005). The culture was grown in the shaking incubator for 24 h at 30 °C at 250 rpm and the cells were harvested by centrifugation at 3500 rpm, 4 °C for 20 min.

The BMGY media was replaced with BMMY media (YPG broth + 100 mM potassium phosphate buffer pH 6.0, 1.34% (w/v) YNB w/o amino acids, 1% (v/v)

methanol and 4×10^{-5}% biotin). The culture was incubated at 20 °C with constant shaking at 250 rpm. Thereafter, the protein production was induced by adding methanol every 24 h to a final concentration of 1% for a total period of 5 days. The supernatant was collected by centrifugation at 3500 rpm, 4 °C for 30 min and added phenylmethylsulfonyl fluoride (PMSF) to 1 mM. The pH of the supernatant was adjusted to 7.5 by adding 2 M dibasic potassium phosphate and filtered through a 0.45 µm membrane (Millipore, Massachusetts-USA).

For protein purification, immobilized metal affinity chromatography (IMAC) was employed (Hays *et al*., 2010). The IMAC resin sepharose-Ni^{2+} was equilibrated with equilibrium buffer (Eq buffer, 20 mM Tris-HCl, pH 7.5, 150 mM NaCl). The supernatant was incubated with IMAC resin at 4 °C for 1 h with moderate shaking and transferred to a column. Equilibrium buffer was used to wash the column twice, followed by W_1 (5 mM imidazole in Eq buffer) and W_2 (10 mM imidazole in Eq buffer), respectively. The protein was eluted from the column by an elution solution (100 mM and 250 mM imidazole in Eq buffer) and the fractions were collected and analyzed by 12% SDS-PAGE gel with Coomassie brilliant blue staining. To assess glycoprotein, the protein was digested with endoglycosidase-H at 4 °C overnight.

3.4 Protein expression and purification by bacterial system

3.4.1 Plasmid digestion with restriction enzyme

pPICZ𝛂BNH8/OsSCPL7 and pPICZ𝛂BNH8/OsSCPL2a plasmids were double digested to the 10 µL mixture containing approximately 100 ng plasmid DNA, 1x buffer 3.1 (NEB), 0.5 unit each of two restriction enzymes *Nco*I and *Sal*I. The mixture was incubated at 37 °C overnight and the digested plasmid was observed by running agarose gel electrophoresis. Once the digested plasmid was confirmed, agarose gel electrophoresis was performed again with plenty of starting material since some of the DNA sample will be lost during the gel purification step.

3.4.2 Gel purification using GeneJET purification kit

The gel purification in this work was performed based on the GeneJET kit (Thermo Fisher Scientific, Massachusetts-USA). The gel slice containing the digested DNA plasmid of pPICZ𝛂BNH8/OsSCPL was excised using a clean razor blade. The gel was cut as close to the DNA target as possible to minimize the gel volume, then placed into a pre-weighed 1.5 mL and weighed. The weight of the gel slice was recorded. A binding buffer was added 1:1 (v/w) to the gel slice. The gel mixture was incubated at 50-60 °C for 10 min or until the gel slice

was completely dissolved and mixed by inverting the tube every minute to facilitate the melting process. The gel mixture was vortexed briefly before loading on the column. The color of the solution must be checked. A yellow color indicated an optimal pH for DNA binding.

Approximately, 800 μL of the solubilized gel solution was transferred to the GeneJET purification column and centrifuged at 12,000 rpm for 1 min. The flow through was discarded and the column was placed back into the same collection tube. After each application, the column was centrifuged for 30-60 s and the flow-through solution was discarded after each spin. The step was repeated until the entire volume had been applied to the column membrane. Then, 700 μL of washing buffer that was diluted with ethanol was added to the purification column and it was centrifuged for 1 min. The flow-through solution was discarded and the column was placed back into the same collection tube. The empty GeneJET purification column was centrifuged for an additional 1 min to completely remove the residual wash buffer. This step was essential to avoid residual ethanol in the purified DNA solution.

The GeneJET purification column was transferred into a clean 1.5 mL microcentrifuge tube and 50 μL of elution buffer was added to the center of the purification column membrane. Then, the column was centrifuged for 1 min and incubated at room temperature for 5 min. The amount of the elution buffer can be reduced to increase DNA concentration. The GeneJET purification column was discarded and the purified DNA was stored at -20 °C.

3.4.3 Ligation of OsSCPL cDNA to pET32a vector

The concentration of DNA from gel purification was measured by nanodrop. The volume of pET32a (vector DNA, 37 ng/μL) and OsSCPL (insert DNA, 60 ng/μL) used in the ligation depended on the size of each and their concentration. However, for most standard cloning (where the insert is smaller than the vector), a 1 vector: 3 insert molar ration. Ten microliter of reaction mixture consist of 3 μL of DNA vector, 1.5 μL DNA insert, 1 μL buffer, 1 μL T4 ligase, and 3.5 μL water was warmed at 45 °C for 5 min to avoid reannealing and chilled at 4 °C. A 10 μL ligation reaction was set containing 3 μL of vector DNA, 1.5 μL of insert DNA, 1 μL of ligation buffer (1 μL/10μL reaction of 10x buffer), 1 μL of T_4 DNA ligase and 3.5 μL sterile HPLC H_2O. The ligation reaction was mixed well gently by flicking the tube and incubated at 15 °C overnight, followed by bacterial transformation (Section 3.2.3) into *E. coli* XL1-Blue and selection on 50 μg/mL ampicillin LB plates.

3.4.4 Extraction of pET32a/OsSCPL plasmid by Vivantis plasmid miniprep and sequencing

The extraction of pET32a/OsSCPL plasmid was conducted using Vivantis miniprep kit. A single colony of XL1-Blue competent cell containing two recombinant plasmids pET32a/OsSCPL2a and pET32a/OsSCPL7 was inoculated in 5 mL LB medium with appropriate antibiotics (50 µg/mL ampicillin) overnight (12-16 h) at 37 °C with agitation. In the following day, the culture was harvested by centrifugation at 3,000 rpm for 10 min. The cell pellet was resuspended and moved to the small tube, then centrifuged at 12,000 rpm for 1 min. After decanting the supernatant, 250 µL of S1 solution was added to the pellet and the cells were resuspended completely by vortexing or pipetting. The suspension pellet was transferred to a clean 1.5 mL microcentrifuge tube and added with 250 µL of S2 solution. The mixture was gently mixed by inverting the tube 4-6 times to obtain a clear lysate and incubated on ice for not longer than 5 min.

To neutralize the lysate, 400 µL of buffer NB was added to the mixture and gently mixed by inverting the tube 6-10 times until a white precipitate formed. After that, the mixture was centrifuged at 12,000 rpm for 10 min, followed by transferring 650 µL of supernatant into a column assembled in a clean collection tube. The column was centrifuged at 12,000 rpm for 1 min, and the flow-through solution was discarded. This step was repeated for the remaining sample.

The column was washed with 650 µL wash buffer and centrifuged at 12,000 rpm for 1 min. The flow-through solution was discarded, and the column was centrifuged again to remove any residual ethanol. The column was placed into a clean microcentrifuge tube and 50 µL elution buffer, sterilized HPLC H_2O was added directly into column membrane and incubated for 1-5 min at room temperature. Then, the column was centrifuged at 12,000 rpm for 1 min to elute DNA. The concentration of DNA was checked by measuring the 260 nm and 280 nm absorbance on a nanodrop spectrophotometer and stored at -20 °C. The quality of the DNA was confirmed by 1% agarose gel electrophoresis. Subsequently, the plasmid DNA was transformed into competent cells Origami(DE3), protocol described at section 3.2.3.

3.4.5 Protein expression by bacterial system and purification

Protein induction and expression was conducting by inoculating a single colony of pET32a/OsSCPL into 50 mL LB media with 50 µg/mL ampicillin, 15

µg/mL kanamycin and 12.5 µg/mL tetracycline and growing it in the shaking incubator at 37 ℃ at 250 rpm shaking overnight. To make a starter culture, 2% of culture volume was transferred into 2 L new LB media with appropriate antibiotics and shaking incubated for 2-4 h at 37 ℃ until OD at 600 nm reached 0.4-0.6.

The culture was induced with vary concentrations of isopropyl β-D-1-thiogalactopyranoside (IPTG) with range 0.0; 0.2; 0.4; 0.6; and 0.8 mM. Then, the culture was incubated at 20 ℃ for 18 h with shaking. The cells were harvested by centrifugation at 4,000 rpm, 4 ℃ for 20 min and kept at -80 ℃ at least overnight to break the cell. In the meantime, the 12% polyacrylamide gel was prepared and kept in the cold room until use. Furthermore, the frozen pellet was thawed at room temperature for 30 min.

Fresh extraction buffer (0.4 mg/mL lysozyme, 1 mM aminocaproic acid, 1 mM benzamidine-HCl, 1 mM PMSF, 1% Triton X-100, 5 µg/mL DNase I and 20 mM Tris-HCl, pH 7.5, + 150 mM NaCl) was prepared and mixed well. The pellet was resuspended with extraction buffer and incubated for 30 min, followed by sonication on ice thrice of 30 s followed by intervals of 1 min for cooling. The mixture was centrifuged at 12,000 rpm, 4 ℃ for 30 min to separate the pellet and supernatant.

The supernatant was transferred into the new tube and kept on ice. The pellet was resuspended with an Eq buffer and kept on ice for SDS-PAGE analysis. For protein purification, immobilized metal affinity chromatography (IMAC) was employed. The IMAC sepharose resin equilibrated with Co^{2+} was prepared by loading resin and 5x volume of 20% ethanol to the sterilized column and washed with 5x volume of water.

The column was washed with 5x volume of 0.5 M EDTA to remove some contaminants and water. The new IMAC resin was added into the column, followed by washing with 5x volume of 0.1 M $CoCl_2$ and water. Then, the column was washed with 5x volume of 0.3 M NaCl and water before adding with 5x volume of Eq buffer. After the resin in the column was ready, the cell extract supernatant was loaded into the column.

Equilibrium buffer was used to wash the column twice, followed by W_1 (5 mM imidazole in Eq buffer) and W_2 (10 mM imidazole in Eq buffer), respectively. The protein was eluted from the column by an elution solution (100 mM and 250 mM imidazole in Eq buffer) and the fractions were collected and analyzed by 12% SDS-PAGE gel with Coomassie brilliant blue staining.

The fractions that have the targeted protein were collected and concentrated using 30 kDa Amicon ultra centrifugal filters (Merck, Germany), that was rinsed with Eq buffer. To remove imidazole, the protein was added with Eq buffer and centrifuged at 3,500 rpm, 4 °C for 30 min to discard the flow-through filtrate, then Eq buffer was added again to 15 mL of volume. This step was conducted 3-4 times. After the last centrifugation, the concentrated protein was transferred to the new tubes and the 280 nm absorbance was measured by nanodrop.

The protein was digested with 1 µL enterokinase enzyme and incubated in the cold room overnight. The digested protein was diluted five times with Eq buffer before loading to the new IMAC resin for a second purification. For the second purification, the IMAC resin was washed with an Eq buffer and low concentrations of imidazole. All the fractions were collected to check by SDS-PAGE with a 12% polyacrylamide gel. After concentrating, the protein was measured with nanodrop and kept at -40 °C for enzymatic assay purpose.

3.4.6 Sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE)

The quality of protein was observed by polyacrylamide gel electrophoresis on a 12% separating gel (3.2 mL of distilled water, 4 mL of 30% solution acrylamide/bis-acrylamide (37.5:1, w/w), 2.6 mL of 1.5 M Tris-HCl, pH 8.8, 100 µL of 10% SDS, 100 µL of 10% APS and 10 µL of TEMED added as the last ingredient) for final volume of 10 mL. The separating gel was poured in the glass plates assembled with spacers of the gel casting unit. Distilled water was overlaid on the gel's surface to maintain an even and horizontal surface of the gel and limit oxygen access.

The gel was allowed to set for 30 min at room temperature. A 5% stacking gel (2.975 mL of distilled water, 0.67 mL of 30% acrylamide, 1.25 of 0.5 M Tris-HCl pH 6.8, 50 µL of 10% SDS, 50 µL of 10% APS and 5 µL of TEMED added as the last ingredient) for final volume of 5 mL was prepared. The overlaid water was discarded from the separating gel before adding stacking gel until it overflowed.

Then, the comb was inserted immediately and it was ensured that no air bubbles are trapped in the gel or near the walls. The gel was allowed to set for 30 min at room temperature. The samples were prepared by adding of protein from purification and one-fourth volume of 5x loading buffer (50 mM Tris-HCl pH 6.8, 10% SDS, 0.2 mg/mL bromophenol blue, 50% glycerol, 20% β-

mercaptoethanol). The above samples were boiled at 95 °C for 5 min to denature the protein and centrifuged before being loaded to the SDS-PAGE gel, which was fastened in an electrophoresis apparatus filled well with 1x running buffer. 10 μL protein samples or 3 μL of pre-stained protein ladder were loaded in each well and electrophoresed at a constant voltage of 150 V for 60 min or until the tracking dye reached the bottom of the gel plate.

After that, the gel was removed from the glass plates and moved to a tank containing a staining solution (0.1% (w/v) Coomassie Brilliant Blue R-250, 40% (v/v) methanol, and 10% (v/v) acetic acid). The staining step was allowed to set for at least 1 h or overnight if necessary and followed by adding destaining solution (40% ethanol and 10% acetic acid, v/v) for 1 h or more to remove staining solution.

3.4.7   Tryptic Digestion and LC/MS/MS

The major purified protein bands from yeast system expression and purification were excised from SDS-PAGE gel and washed with purified water three times. In-gel digestion with trypsin was employed to expose the protein to LC/MS/MS analysis, following the protocol as described (Sevchenko *et al*., 1996 and Wall *et al*., 2011). The gel was cut into small pieces and mixed with 100 μL destaining solution (100 mM NH_4HCO_3, 50% (v/v) acetonitrile), followed by briefly vortexing and incubating at 30 °C for 20 min. The destaining treatment was repeated three times or more to get rid of stain solution in the gel. The destaining solution was discarded and the gel was dried in a speed vacuum.

A 100 μL solution of 100 mM NH_4HCO_3, 10 mM DTT and 1 mM EDTA was added. The mixture was incubated at 60 °C for 45 min and centrifuged at 10,000 rpm for 5 min. After discarding the supernatant, 1x volume of alkylation solution (100 mM iodoacetamide, 100 mM NH_4HCO_3) was added to the gel and it was incubated in the dark room for 30 min. Then, the gel mixture was centrifuged at 10,000 rpm for 5 min and the supernatant was discarded before rinsing the gel for 3-4 times with wash solution (50 mM Tris-HCl pH 8.5, 50% (v/v) acetonitrile).

While drying the gel, the digestion buffer, consisting of 1:9 (v/v) trypsin solution (0.1 mg/mL trypsin, 1% (v/v) acetic acid) and buffer solution (50 mM Tris-HCl pH 8.5, 10% (v/v) acetonitrile, 1 mM $CaCl_2$) was prepared. The dried gel was suspended with 2 volumes of digestion solution and incubated at 37 °C overnight, followed by centrifugation at 1000 rpm for 5 min. The supernatant was transferred into the new tube (master tube). Subsequently, 2 volumes of

2% (v/v) trifluoroacetic acid was added into the old tube and it was incubated at 60 °C for 30 min to stop enzymatic reaction.

Centrifugation at 10,000 rpm for 5 min was performed to remove the gel from the supernatant, which was transferred into the master tube. The gel pellet was resuspended with 100 µL of a solution containing 50 mM Tris-HCl, pH 8.5, 1 mM $CaCl_2$ and incubated at 30 °C for 10 min. The mixture was sonicated briefly before adding acetonitrile, followed by second incubation and sonication. Another step of centrifugation was conducted to isolate the supernatant and transfer into the master tube. A hundred microliter of 5% (v/v) formic acid and acetonitrile was added to gel in the old tube. After that, the mixture was centrifuged to remove the supernatant and transfer to the master tube. Finally, a total solution in the master tube was dried and ready for further analysis, which was conducted at the National Center for Genetic Engineering and Biotechnology, Pathumthani, Thailand.

On the other hand, in-gel digestion with trypsin was similarly employed to identify the protein that was expressed and purified by the bacterial system, but it was sent to Laboratory Biochemistry, Chulabhorn Research Institute, Bangkok, Thailand for LC/MS/MS analysis. Then, the peptide mass data was analyzed by MASCOT MS/MS ions search (http://matrixscience.com) using raw MS/MS data, trypsin as the digesting enzyme and carbamidomethyl (C) was selected as a fixed modification. The detail parameter for analysis MS/MS data was explained in **Table 3.2**.

Table 3.2 Parameter input at mascot online using MS/MS ion search.

Parameter	Input
Databases	Contaminant (AA)
	SwissProt (AA)
	UP59680_O_sativa
Enzyme	Trypsin
Cleavages	Allow up to 1
Fixed modifications	Carbamidomethyl (C)
Peptide tol. ±	1.2 Da
MS/MS tol. ±	0.6 Da
Peptide charge	2+
Monoisotopic	Yes
Data format	Mascot generic
Instrument	ESI-TRAP

3.4.8   Enzymatic Assay

The enzyme activity was determined by setting up 100 µL reaction consist of 1 µg of OsSCPL protein, 100 mM citrate buffer, pH 5.0, and 300 µM substrate containing cyanidin-3-*O*-glucoside, apigenin, apigenin-3-*O*-glucoside, apigenin-7-*O*-glucoside, luteolin, luteolin-3-*O*-glucoside, luteolin-7-*O*-glucoside, kaempferol, kaempferol-3-*O*-glucoside, kaempferol-7-*O*-glucoside, quercetin, quercetin-3-*O*-glucoside or quercetin-7-*O*-glucoside. For initial assays, a glucosyl donor was added at 500 µM. The mixture was incubated at 30 °C overnight and the reaction was terminated with 1x volume of 0.2% formic acid in methanol. The mixture was kept at 4 °C before analyzing with thin layer chromatography (TLC) and ultra-high-performance liquid chromatography (UHPLC).

3.4.9   Thin layer chromatography (TLC)

The enzymatic reaction was spotted on silica gel 60 F254 TLC plate (Merck, Kenilworth, NJ, USA) and resolved with a mobile phase system of *n*-butanol: acetic acid: water (3:1:1 v/v/v) (Liaqat *et al.*, 2018 with modification). The developed plates were observed under visible and UV light, stained with 10% sulfuric acid in ethanol and charred at 120 °C to observe carbohydrate.

3.4.10 Ultra-high-performance liquid chromatography (UHPLC)

In this work, the product from the enzyme reaction and the standard were filtered through a 0.2 μm filter and detected by G4220 ultra-high-performance liquid chromatography (UHPLC) (Agilent Technologies-California, USA) with 2.1x150 mm SB-C18 column (Agilent, USA). The column temperature was maintained at 40°C. The flow rate was 0.2 mL min^{-1} with injection volume of 10-15 µL. The signal was captured by a diode array detector (DAD) with wavelengths of 260, 280, 325, 360, 520, and 530 nm. For the gradient elution, 0.2% formic acid in HPLC water was used as solvent A and 100% acetonitrile as solvent B. The gradient elution program was as follows: 0-2 min, 95% A; 2-13 min, 95% to 50% A; 13-14 min 50% to 30% A, 14-16 min, 30% to 0% A; 16-20 min, 0% to 95% A; 20-25 min, 95% A.

CHAPTER IV
RESULT AND DISCUSSION

4.1 Phylogenetic and expression pattern analysis of rice serine carboxypeptidase-like (OsSCPL) proteins

Phylogenetic analysis was performed with rice OsSCPL protein sequences retrieved from the National Center for Biotechnological Information (NCBI) Genbank protein database. The proteins were identified by a BLASTp search using the acyl-glucose-dependent anthocyanin acyltransferase of *Delphinium grandiflorum* (DgSCPL2, ID: BAO04183.1) as a query to find putative anthocyanin acyltransferases. The phylogenetic tree in **Figure 4.1** shows that *Oryza sativa* serine carboxypeptidase-like, abbreviated as OsSCPLs, are equally diverged from DgSCPL2.

Interestingly, only OsSCPL2a (ID: XP_015614311.1) was closely related to SCPL from *Avena strigosa.* **(Figure 4.1)**. Nishizaki *et al*. (2013) reported that DgSCPL2 is a candidate for an acyl glucose-dependent anthocyanin 7-O-glucoside acyl transferase (AA7G-AT) and/or an acyl glucose-dependent anthocyanin 7-O-glucosyl-oxybenzyl glucoside acyl transferase (AA7GBG-AT) protein. Those two enzyme activities are distinguished based by their acceptor substrate specificity, and are responsible for modification and formation of complex anthocyanins in delphinium. They both use *p*-hydroxybenzoyl-glucose (pHBG) as an acyl donor.

As we mentioned above, rice serine carboxypeptidase-like 2a (OsSCPL2a, ID: XP_015614311.1) is likely similar to serine carboxypeptidase-like acyltransferase 1 from *Avena strigosa* (AsSCPL1) **(Figure 4.1)**. AsSCPL1 protein, which is encoded by the Sad7 gene, is required for synthesis of saponins as defense compounds in oat by forming an acylated avenacin, avenacin A-1, from des-acyl avenacin A as acceptor and *N*-methyl anthraniloyl-glucose as acyl donor substrate (Mugford *et al*., 2009).

Both of AsSCPL1 and DgSCPL2 have acyltransferase activity that is responsible to modify and form complex compounds for second metabolism and defense system in plants. Although all the isoenzymes of rice SCPL proteins shown in Figure 4.1 appear equally diverged from the AA7G-AT, we chose rice OsSCPL2a and OsSCPL7 as our proteins of interest based on their expression patterns (see below). The amino acid sequences of these two OsSCPLs were aligned with known SCPL acyltransferase by

phytohormones, and specific cell types or tissues isolated by laser microdissection (LMD), based on full-length white rice cDNA.

```
DgSCPL2    SKQGYIFIRKWLVDHQKFLSNPLYISGDSYSGITIPPIGQYIFDGIESGDELLINFKGYL    213
OsSCPL2a   GQQLVVFLTKWLGDHPEFASNPLYIGGDSYSGYTVPVTALQIA--NDDDARARLNLKGYL    226
OsSCPL7    TLQLTKFVNKWFSQHREFLSNPLYVGGESYAGKLVPFLLQKISEDVEAGVKPVLNLKGYL    225
AsSCPL1    GRHVRVFLQKWMAQHPELASNPLYIGGDSFSGYTVTVSALEVA-NHPAASSELNLKGYM    210
SpGAC      VTLTYDFLRKWLMDHPEYLNNPLYVGGDSYSGIFVALLTRKIYDGIEVGDRPKVNIKGYI    205
CtAT1      VHDAHEFLRKWLIDHPKFLSNEVYIGGDSYSGITVPAIVQEISQGNEKGIQPSINLQGYI    217
AtSCT      MHQIDQFLRSWFVKHPEFISNPFYVGGDSYSGKIVPGAVQQILLGNEKGLTPLINIQGYV    209
BnSCT1     IHQIDQFLRSWFVDHPEFISNSFYVGGDSYSGKIVPGVVQQISLGNEKGLTPLINIKGYV    206
BnSCT2     IHQIDQFLRSWFVDHPEFISNSFYVGGDSYSGKIVPGVVQQISLGNEKGLTPLINIQGYV    206
AtSMT      VKRTHEFLQKWLSRHPQYFSNPLYVVGDSYSGMIVPALVQEISQGNYICCEPPINLQGYM    204
AtSAT      VKRIHEFLQKWLSKHPQFFSNPFYVTGDSYSGMIVPALVQEISKGNYICCKHLINLQGYV    206
AtSST      VKKIHEFLQKWLIKHPQFLSNPFYVVGDSYSGMIVPALVHEISKGNYICCNPPINLQGYV    206
              *:  .*:    *  :   .*  .*:  *:*::*    :             :*::**:

DgSCPL2    GLNYLYDNISSIGYHLYLSRKG-CRSLIYSGDHDFLVPFLSTETWITSLNYSITDDWRPW    420
OsSCPL2a   LF--RHDVDSAIPYHAELTQRG-YRALVYNGDHDLIMTFVGTQQWIRTLGYNVVTAWRPW    421
OsSCPL7    GLPYSQDIESSIKYHQNLTSRG-YRVLVYSGDHDSVVPFLGTQAWVRSLNYPIVDDWRAW    429
AsSCPL1    KLHLTNDVPTTVPYHRRLTQRG-YRALVYNGDHDLLMTHIGTHAWIRSLGYPVVAPWRAW    442
SpGAC      TESYVYDVPSVIDDHQHLTSKS-CRALIYSGDHDMVVPHLSTEEWIETLKLPIADDWEPW    411
CtAT1      NLPYTEDIPSSFEYHVNLSGKGYYRSLVYSGDHDLMVPFLGTQAWIRSLNYSIVDDWRQW    419
AtSCT      NIPYTFEIFNAVPYHVNNSLKG-FRSLIYSGDHDSMVPFSSTQAWIRALNYSIVDDWRPW    414
BnSCT1     NIPYTYDIHNAIPYHVNNSRKG-FRALIYSGDHDMMIPFSSTEAWIKSLNYSIVDDWRPW    416
BnSCT2     NIPYTYDIHNAIPYHVNNSRKG-FRALIYSGDHDMMIPFSSTEAWIKSLNYSIVDDWRPW    416
AtSMT      TIPYNHDIVSSIPYHMNNSISG-YRSLIYSGDHDIAVPFLATQAWIRSLNYSPIHNWRPW    384
AtSAT      TIPYDNNIISSVPYHMDNSING-YRSLIYSGDHDITMPFQATQAWIKSLNYSIVDDWRPW    388
AtSST      GIPYKSDIRSSIPYHMNNSING-YRSLIFSGDHDITMPFQATQAWIKSLNYSIIDDWRPW    388
             :  .  .  *  :   .  *  *:::.****: .  .*. *: :*    *: *

DgSCPL2    FVG-GQVAGYTRSYAN---NMTFATVKGAGHTAPEYKPYECFSMMKKWLSYEPL-       470
OsSCPL2a   YSN-RQVAGFTTVYDH---NLTFATVKGGGHTAPEYRPKECLIMLDRWTSPAGEL       472
OsSCPL7    HID-GQSAGFTITYAN---DLTFATVKGGGHTAPEYQPERCLAMFGRWISEKSL-       479
AsSCPL1    YSN-NEVAGFTVEYSN---NLTFATVKGAGHMAPESRPKQCLIDMVRRWISPAGKL      493
SpGAC      FVD-DQVAGYKVKYLQNDYEMTYATVKGAGHTAPEYKPEQCLPMVDRWFSGDPL-       464
CtAT1      NTN-GQVAGYTRTYSN---RMTFATVKGGGHTAPEFKPEECFAMYSRWISKRPL-       469
AtSCT      MMSSNQVAGYTRTYAN---KMTFATIKGGGHTA-EYTPDQCSLMFRRWIDGEPL-       464
BnSCT1     MMNSNQVAGYTRTYAN---KMTFATIKGGGHTA-EYNPDQCSLMFRRWIDGESL-       466
BnSCT2     MMTSNQVAGYTRTYAN---KMTFATIKGGGHTA-EYNPDQCSLMFKRWIDGESL-       466
AtSMT      MIN-NQIAGYTRAYSN---KMTFATIKGGGHTA-EYRPNETFIMFQRWISGQPL-       433
AtSAT      MIN-DQIAGYTRTYSN---KMTFATVKGGGHTA-EYLPNESSIMFQRWISGQPL-       437
AtSST      MIK-GQIAGYTRTYSN---KMTFATVKGGGHTA-EYLPEESSIMFQRWISGQPL-       437
             **:.  *  :  :*:**:**.** * * :.  *    * .*:.*
```

Figure 4.2 Partial multiple sequence alignment of OsSCPL amino acid sequences with those of characterized SCPL acyltransferases from plants. The red highlight indicates the catalytic triad of amino acid residues.

The expression profile result showed that both OsSCPL2 and OsSCPL7 are well expressed in ovary, as well as lemma and palea, even though they also showed expression in other stages. OsSCPL2a is likely expressed at initial growth in leaf blade, while OsSCPL7 is significantly expressed in the final growth step, specifically in embryo and endosperm stage **(Figure 4.3)**. Indeed, the highest expression in OsSCPL2a occurred in ovary stages and it may correlate to the seed coloring in pericarp. The expression of both enzymes in white rice and in tissues other than those that most

often contain anthocyanins suggests that they may act on other compounds in these tissues.
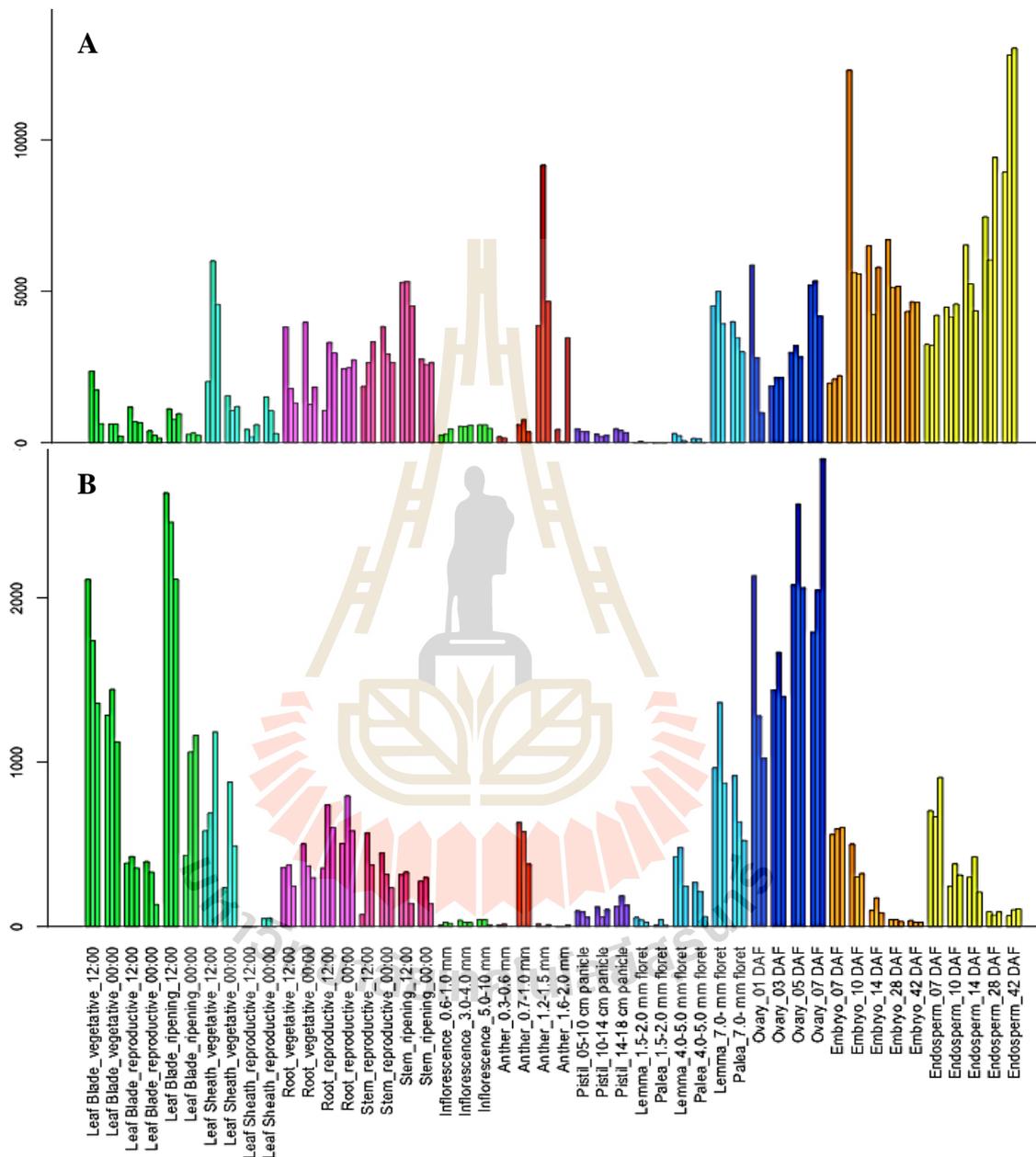
Figure 4.3 Expression of rice OsSCPL7 (A) and OsSCPL2a (B) from leaf blade of the vegetative stage to the endosperm of the reproductive tissue. Both proteins were well expressed in lemma, palea and ovary. However, OsSCPL7 is significantly expressed in embryo and endosperm.

The expression pattern of genes involved in anthocyanin synthesis, such as chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3'-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), flavonoid 3'5'-hydroxylase (F3'5'H), dihydroflavonol reductase (DFR), and anthocyanidin synthase (ANS), as well as glucosyltransferase (GT) and acyltransferase (AT) were found in some parts of plant which produce the anthocyanin pigments (Lu *et al*., 2021). In another plant, delphinium, for instance, DgSCPL2 showed high expression in certain developmental stages of sepal, similar with the expression pattern of DgCHS, DgF3'5'H, and DgANS. In addition, the expression occurred in blue cultivars, such as Triton Light Blue and Blue Magic, but not in Ariel White and White Candle cultivars, which lack of anthocyanin (Nishizaki *et al*., 2013).

4.2   Rice SCPL (OsSCPL) profile, gene optimization, and construct of recombinant protein

Both OsSCPL2 (ID: XP_015614311.1) and OsSCPL7 (ID: XP_015617342.1) have putative signal peptides which were predicted online by SignalP version 5.0 (source: http://www.cbs.dtu.dk/services/SignalP/). This server also provides the probable location of the cleavage site in proteins from archaea, bacteria or eukaryotes (Armenteros *et al*., 2019). Based on the signal result, the putative signal peptides from both proteins were present in the N-terminal amino acid sequence. In OsSCPL2a, the cleavage site was detected after the first 34 amino acid residues of signal sequence (MHLLASSSSIPMATTRATSTLLLLLLLLVSATWAA) **(Figure 4.4A)**. The cleavage site by the signal peptide in OsSCPL7 was detected after the first 22 amino acid residues comprising the signal sequence (MSLRLRLAFRLLFCYCLLLAAA) **(Figure 4.4B)**.

Both of putative signal peptides were deleted to produce optimized genes encoding the mature protein for expression. The protein's profile including predicted mass, isoelectric point (pI), and number of amino acid residues was calculated using ProtParam tool-Expasy (https://web.expasy.org/protparam/). OsSCPL2a has 446 amino acid residues, with an approximately 49 kDa of molecular weight and a theoretical pI of 5.24. Meanwhile, OsSCPL7 has 465 amino acid residues, for 52 kDa of mass and theoretical pI of 5.39. Sequence alignments of the amino acid sequences encoded by the expression gene and the natural one are shown to emphasize the mature protein in **Figures 4.5** and **4.6**.

Figure 4.4 Signal peptides prediction of OsSCPL2a (XP_015614311.1) (a) and OsSCPL7 (XP_015614311.1) (b).

```
XP_015614311.1   MHLLASSSSIPMATTRATSTLLLLLLLLVSATWAASAPTTSRARNVITHVKGFQGRLPFHL   60
Opt_OsSCPL2a     ---------------------------AAMASAPTTSRARNVITHVKGFQGRLPFHL       30
                                            :  ****************************

XP_015614311.1   ETGYVEVDNTNTVELFYYFIQSERSPADDPLILWITGGPGCSALSGLLFEIGPLKFDVAG   120
Opt_OsSCPL2a     ETGYVEVDNTNTVELFYYFIQSERSPADDPLILWITGGPGCSALSGLLFEIGPLKFDVAG    90
                 ************************************************************

XP_015614311.1   YTEGFPQLFYFQDSWTKVSNVIFLDAPVGTGFSYAREEQGYNVTLTQTGQQLVVFLTKWL   180
Opt_OsSCPL2a     YTEGFPQLFYFQDSWTKVSNVIFLDAPVGTGFSYAREEQGYNVTLTQTGQQLVVFLTKWL   150
                 ************************************************************

XP_015614311.1   GDHPEFASNPLYIGGDSYSGYTVPVTALQIANDDDARARLNLKGYLVGNAATDVKYDSGG   240
Opt_OsSCPL2a     GDHPEFASNPLYIGGDSYSGYTVPVTALQIANDDDARARLNLKGYLVGNAATDVKYDSGG   210
                 ************************************************************

XP_015614311.1   KVPFMHGMGLISDEMYEAARSSCRGDYVSTPTNADCANALQAISMATFAINPVHILEPIC   300
Opt_OsSCPL2a     KVPFMHGMGLISDEMYEAARSSCRGDYVSTPTNADCANALQAISMATFAINPVHILEPIC   270
                 ************************************************************

XP_015614311.1   GFALRGRAMPETTMDQRLRLGLPVECRDNGYRLSYLWADDPEVRATLGIHEGSIASWSRC   360
Opt_OsSCPL2a     GFALRGRAMPETTMDQRLRLGLPVECRDNGYRLSYLWADDPEVRATLGIHEGSIASWSRC   330
                 ************************************************************

XP_015614311.1   TALPLFRHDVDSAIPYHAELTQRGYRALVYNGDHDLDMTFVGTQQWIRTLGYNVVTAWRP   420
Opt_OsSCPL2a     TALPLFRHDVDSAIPYHAELTQRGYRALVYNGDHDLDMTFVGTQQWIRTLGYNVVTAWRP   390
                 ************************************************************

XP_015614311.1   WYSNRQVAGFTTVYDHNLTFATVKGGGHTAPEYRPKECLDMLDRWTSPAGEL           472
Opt_OsSCPL2a     WYSNRQVAGFTTVYDHNLTFATVKGGGHTAPEYRPKECLDMLDRWTSPAGEL           446
                 ***************************************************
```

Figure 4.5 Alignment of the OsSCPL2a amino acids sequence from the NCBI database (ID: XP_015614311.1) and the optimization protein, which would be inserted to the yeast vector pPICZαBNH8 for yeast expression system. The first 34 amino acid residues corresponding to the predicted signal sequence were deleted in the protein from the optimized gene. Red boxes represent potential N-glycosylation sites.

```
XP_015617342.1    MSLRLRLAFRLLFCYCLLLAAAHPSSRRLLPLLVVSAAAERTNVASLPGLDGALPSRFETG    60
Opt_OsSCPL7       ----------------AAMAHPSSRRLLPLLVVSAAAERTNVASLPGLDGALPSRFETG      42
                                  *  ****************************************

XP_015617342.1    YVTVDEENGGELFYYFIESEGDPGADPVLLWINGGNRCSVLSALFFEIGPVKLAIEPYDG   120
Opt_OsSCPL7       YVTVDEENGGELFYYFIESEGDPGADPVLLWINGGNRCSVLSALFFEIGPVKLAIEPYDG   102
                  ************************************************************

XP_015617342.1    GVPRLRYNPYTWTKVASVLFVDSPVGAGFSFSRDPRGYDVGDVSSTLQLTKFVNKWFSQH   180
Opt_OsSCPL7       GVPRLRYNPYTWTKVASVLFVDSPVGAGFSFSRDPRGYDVGDVSSTLQLTKFVNKWFSQH   162
                  ************************************************************

XP_015617342.1    REFLSNPLYVGGESYAGKLVPFLLQKISEDVEAGVKPVLNLKGYLVGNPGTGESIDYESK   240
Opt_OsSCPL7       REFLSNPLYVGGESYAGKLVPFLLQKISEDVEAGVKPVLNLKGYLVGNPGTGESIDYESK   222
                  ************************************************************

XP_015617342.1    VPYAHGVGIISDQLYETILEHCGREDYANPKNATCAQALNRFSELMGEVSEAHILYKKCI   300
Opt_OsSCPL7       VPYAHGVGIISDQLYETILEHCGREDYANPKNATCAQALNRFSELMGEVSEAHILYKKCI   282
                  ************************************************************

XP_015617342.1    YVSPKPDDGTIGRKILEEIVVGNHRPPRPPMDCSTYPNYLSYFWANSNNTRENLGIKKGT   360
Opt_OsSCPL7       YVSPKPDDGTIGRKILEEIVVGNHRPPRPPMDCSTYPNYLSYFWANSNNTRENLGIKKGT   342
                  ************************************************************

XP_015617342.1    VDEWVRCHDDGLPYSQDIESSIKYHQNLTSRGYRVLVYSGDHDSVVPFLGTQAWVRSLNY   420
Opt_OsSCPL7       VDEWVRCHDDGLPYSQDIESSIKYHQNLTSRGYRVLVYSGDHDSVVPFLGTQAWVRSLNY   402
                  ************************************************************

XP_015617342.1    PIVDDWRAWHIDGQSAGFTITYANDLTFATVKGGGHTAPEYQPERCLAMFGRWISEKSL    479
Opt_OsSCPL7       PIVDDWRAWHIDGQSAGFTITYANDLTFATVKGGGHTAPEYQPERCLAMFGRWISEKSL    461
                  *********************************************************
```

Figure 4.6 Alignment of the OsSCPL7 amino acid sequence from the NCBI database (XP_015617342.1) and the optimization protein, which would be translated from the gene inserted to the yeast vector pPICZαBNH8 for yeast expression system. The first 22 amino acid residues corresponding to the predicted signal sequence were deleted in the protein from the optimized gene. Red boxes represent potential N-glycosylation sites.

Most of plant SCPL protein, including rice, are predicted to undergo posttranslational modification by N-glycosylation, disulfide bond formation and removal of the N-terminal signal peptide. Thus, prokaryotic expression may not be optimal for production of active acyltransferase protein (Shirley *et al*, 2001; Mugford and Milkowski, 2012, Ciarkowska *et al*., 2018). Potential N-glycosylation sites in OsSCPL2 and OsSCPL7 were predicted with NetNGlyc-1.0 server (source: http://www.cbs.dtu.dk/services/NetNGlyc/) (Grupta and Brunak, 2002).

As shown in **Figure 4.7**, two potentials N-glycosylation sites (N-X-T/S) can be identified in the OsSCPL2a amino acid sequence: Asn162 and Asn437, although it has strong potential for glycosylation at Asn162 only **(Figure 4.5,** marked in red box). Meanwhile, the OsSCPL7 amino acid sequence has three potential N-glycosylation sites: Asn272, Asn348, and Asn387 **(Figure 4.7** and **Figure 4.6**, marked in red box).

Finally, as we planned to express OsSCPL proteins in a yeast expression system, the optimized genes were inserted to the vector pPICZαBNH8 at the site of PstI and XbaI position, forming two constructs of pPICZαBNH8/OsSCPL2 and pPICZαBNH8/OsSCPL.
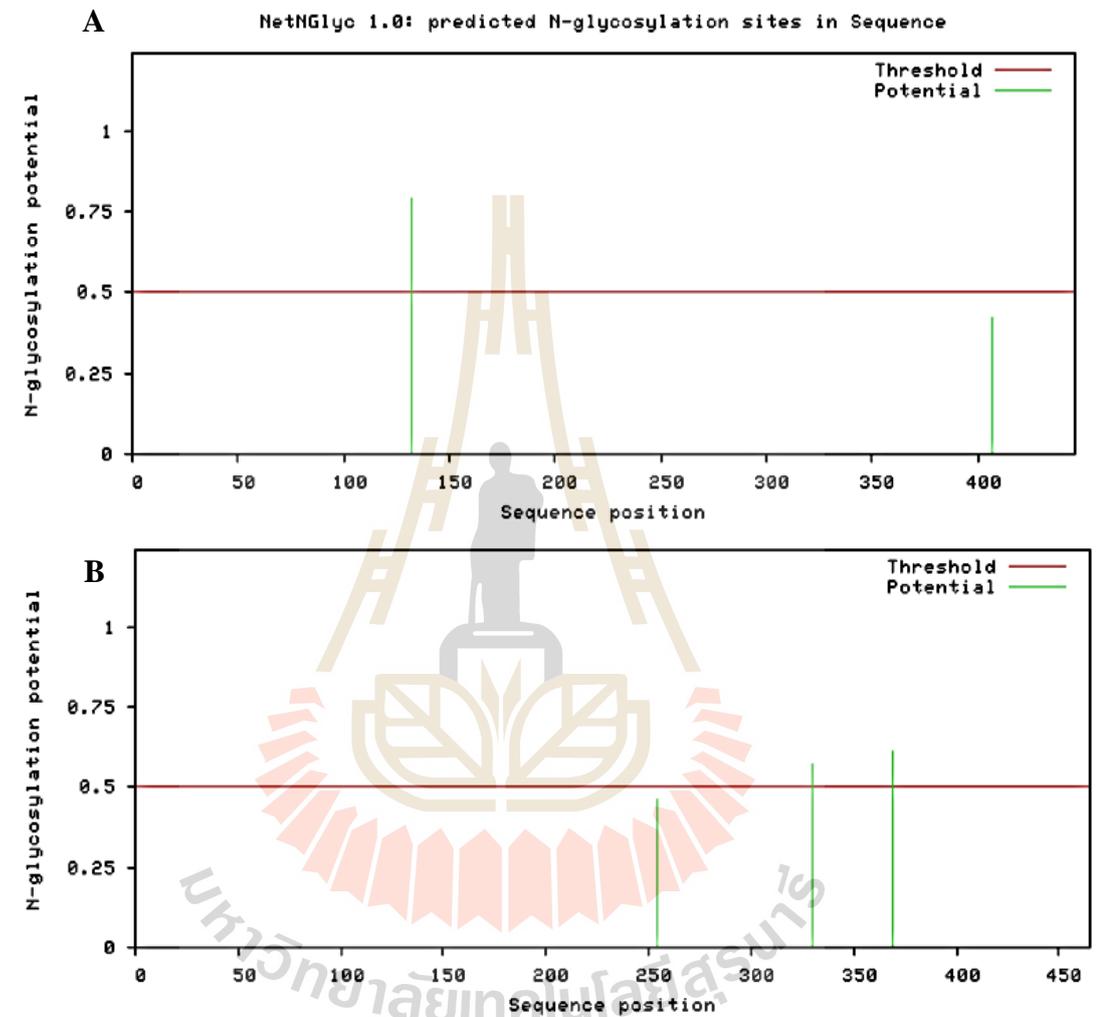

Figure 4.7 Probability of N-glycosylation sites predicted by NetNGlyc 1.0. (A) Prediction for OsSCPL2a identified possible sites at residues 162 and 437. (B) Three potentials N-glycosylation sites were identified in OsSCPL7, located at 272, 348 and 387 of the amino acid sequence.

4.3 Glucose ester as predicted acyl donor substrate for OsSCPL acyltransferase activity

To test the activity of the OsSCPL protein, the reaction mixture consisted of a glucose ester as acyl donor substrate, an acyl acceptor substrate, appropriate buffer and protein from OsSCPL2a or OsSCPL7 expression. I produced four kinds of glucose ester: *p*-hydroxybenzoyl-glucose (pHBG), feruloyl glucose (FG), *p*-coumaroyl-glucose (pCG), and vanillyl-glucose (VG). The purpose was to use various of glucose ester is to identify the preference of acyl donor for acyltransferase activity. All of the glucose esters were synthesized by Os9BGlu31 transglycosidase, which transfers the glucosyl group from *p*-nitrophenyl β-D-glucopyranoside (pNPGlc) to another compound, in these simple phenolic acids, to generate the phenolic 1-*O*-β-D-glucose ester (Luang *et al.*, 2013) **(Figure 4.8).**

The acyl donor substrates of SCPL acyltransferases range from the relatively simple *O*-isobutyrylglucose (Li and Steffens, 2000) to glucose esters of more complex aromatic compounds such as sinapate or *N*-methyl anthranilate (Milkowski *et al.*, 2004; Mugford *et al.*, 2009). Meanwhile, the range of acyl acceptor substrates covers a similarly wide spectrum from low molecular weight compounds, such as choline and L-malate (Lehfeldt *et al.*, 2000; Shirley *et al.*, 2001), through to large complex molecules, such as anthocyanin (Fraser *et al.*, 2007, Nishizaki *et al.*, 2013, Lu *et al.*, 2021) and triterpenoid glycosides (Mugford *et al.*, 2009). Various acyl acceptor substrates were tested in this experiment, including anthocyanin, flavonoid aglycone, and flavonoid glycoside. The expected mechanism in this work is forming acylated conjugation at a hydroxyl position of the substrate acceptor by OsSCPL acyltransferase activity.

The production of glucose ester required pNPGlc as substrate donor, phenolic compounds as substrate acceptor, 100 mM citrate buffer, pH 4.5, and 1 μg/μL of Os9BGlu31. The glucose esters were visualized on TLC plate, forming an extra spot below to the standard compound, indicating that glucosyl group successfully attached with phenolic compounds. (**Figure 4.9**). Subsequently, purification of glucose esters was needed to isolate the product. After purification, a single spot representing glucose ester was recorded on TLC plate (**Figure 4.10**).
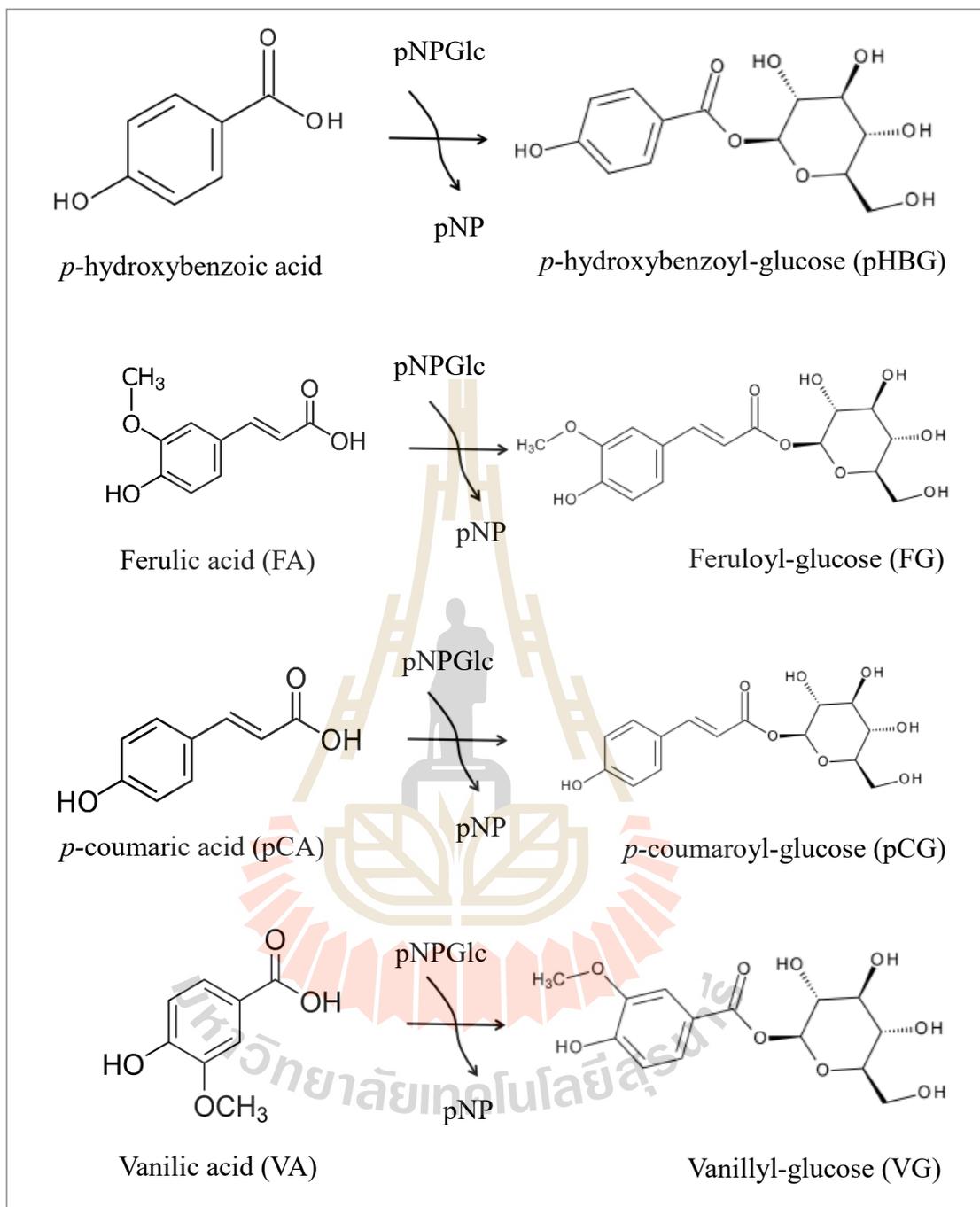
Figure 4.8 Schematic reaction of 1-O¬-glucose esters production by Os9BGlu31 transglucosidase, which transfers glucose from pNPGlc to phenolic acids.
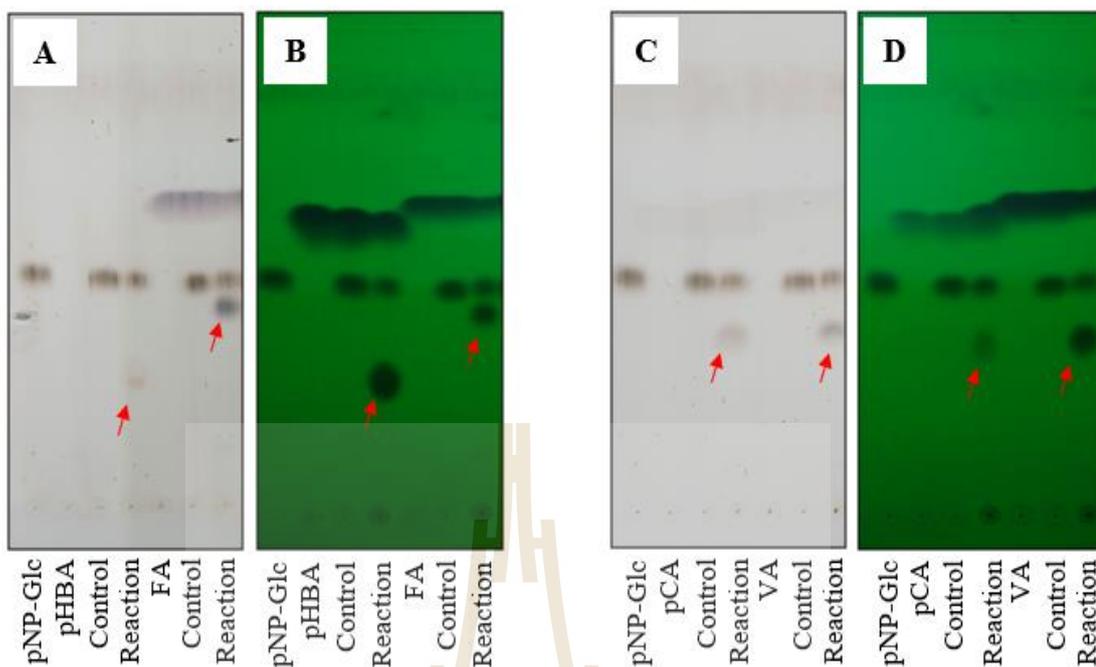
Figure 4.9 TLC analysis of glucose ester production from p-hydroxybenzoic acid (pHBA), ferulic acid (FA) (A and B), p-coumaric acid (pCA) and vanillic acid (VA) (C and D). The Os9BGlu31 enzyme was employed to transfer glucose from pNPGlc to the phenolic acids. Red arrows indicate the glucose ester products. A and C were detected by 10% H_2SO_4, while B and D were detected by fluorescence under UV light.
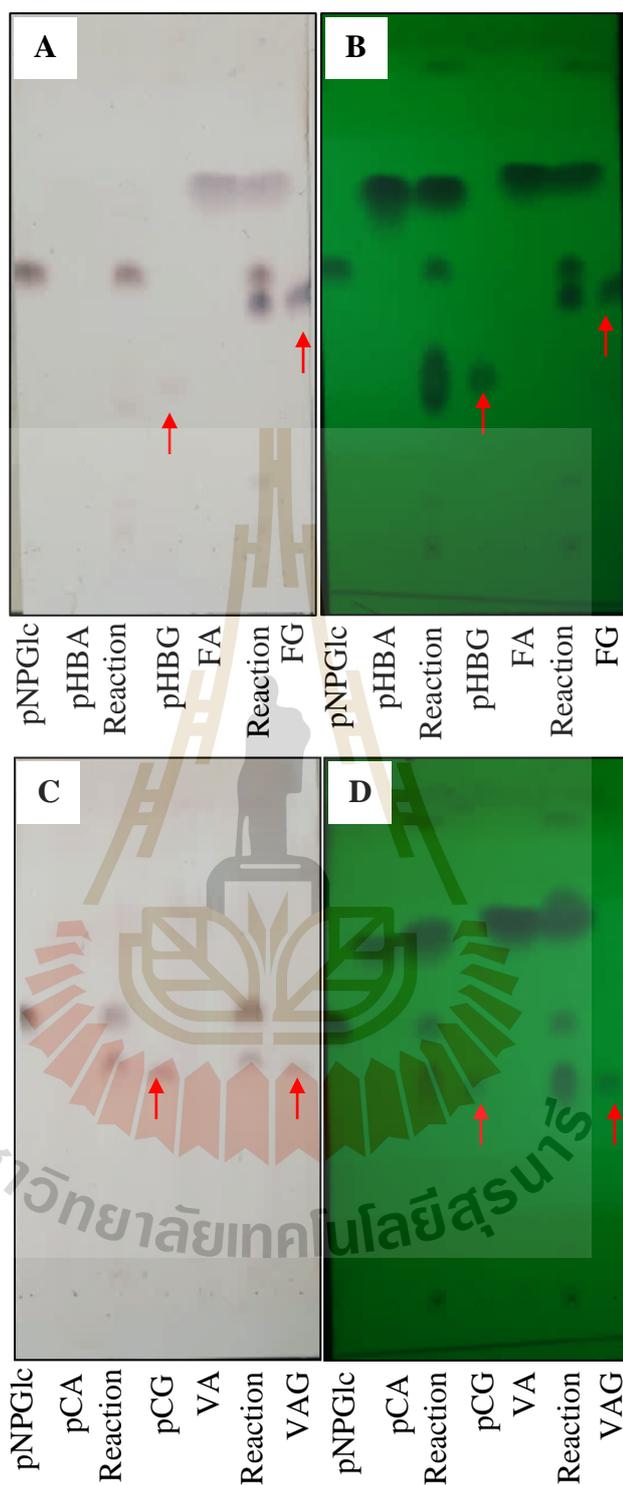
Figure 4.10 Purification of glucose ester products visualized under UV light (B and D) and stained with 10% sulfuric acid in ethanol (A and C). The transglucosidase products, 1-O-beta-glucose esters, including p-hydroxybenzoyl-glucose (pHBG), feruloyl glucose (FG), p-coumaroyl-glucose (pCG) and vanillyl-glucose (VG), are indicated by red arrows below the spots.

4.4   Plasmid linearization and yeast colony screening

Before introducing the recombinant expression vectors pPICZαBNH8/OsSCPL2a and pPICZαBNH8/OsSCPL7 into the host *P. pastoris* strain SMD1168H, linearization of the recombinant vector DNA was performed to ensure that the cloned target gene was stably integrated in the host genome. Integration of the gene into the yeast genome increases stability of the recombinant protein expression in the host cells (Cereghino and Cregg, 2000). Instability in the segregation step might result in partial and or complete loss of a recombinant vector from the host cells during the process of recombinant protein expression, particularly with high-density of host cells (Choi *et al*., 2006).

In addition, the integration process also minimizes the variability of copy number, which can occur in plasmid-based systems (Chan *et al*., 2018). When the recombinant plasmids pPICZαBNH8/OsSCPL2a and pPICZαBNH8/OsSCPL7 were linearized with *Sac*I. the undigested plasmids showed smear bands on agarose gel electrophoresis, while digested plasmids appeared as single bands with the size of approximately 5000 bp. This size is a combination of size of pPICZαBNH8 (3597 bp) and OsSCPL gene (1400 bp) **(Figure 4.11A)**.

After transformation, the colonies were picked and spotted in triplicate grew on YPD agar at the higher concentration of zeocin. The results showed that 96 clones of OsSCPL2a and 106 clones of OsSCPL7 were able to grow on 250 µg/mL of zeocin **(Figure 4.12)**. The purpose of screening colony with a higher concentration of antibiotic resistant is to isolate genomic integrants that are highly resistant to zeocin and could have multiple inserts (Chan *et al*., 2018).

Subsequently, 20 out of the colonies from each recombinant plasmid were randomly chosen to precipitate with 60% ammonium sulfate. After washing three times with equilibrium buffer, the protein was tested by SDS-PAGE electrophoresis to identify the presence of expected protein that be carried in every colonies. For OsSCPL2a, four colonies (3, 6, 7, 10) were chosen for further analysis. In addition, three colonies of OsSCPL7 (92, 94, 96) also treated as same experiment **(Figure 4.11B)**.
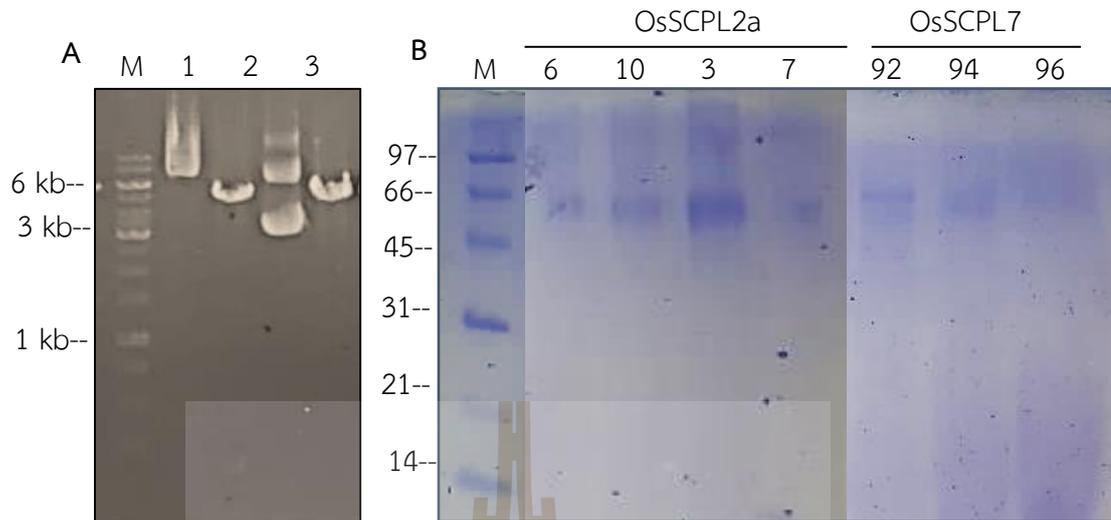
Figure 4.11 Plasmid preparation and expression in *Pichia pastoris*. (A) Plasmid linearization for transformation of Pichia pastoris. Lanes 1 & 2: undigested and digested pPICZαBNH8/OsSCPL2a. Lanes 3 & 4: undigested and digested pPICZαBNH8/OsSCPL7. (B) Several colonies from OsSCPL2a and OsSCPL7 were randomly picked and precipitated using 60% amonium sulfate and visualized by SDS-PAGE. Colonies 3, 6, 7, and 10 of OsSCPL2a were observed to have a band around 66 kDa. Colonies 92, 94, and 96 of OsSCPL7 had same visualization although the band not as obvious as OsSCPL2a.
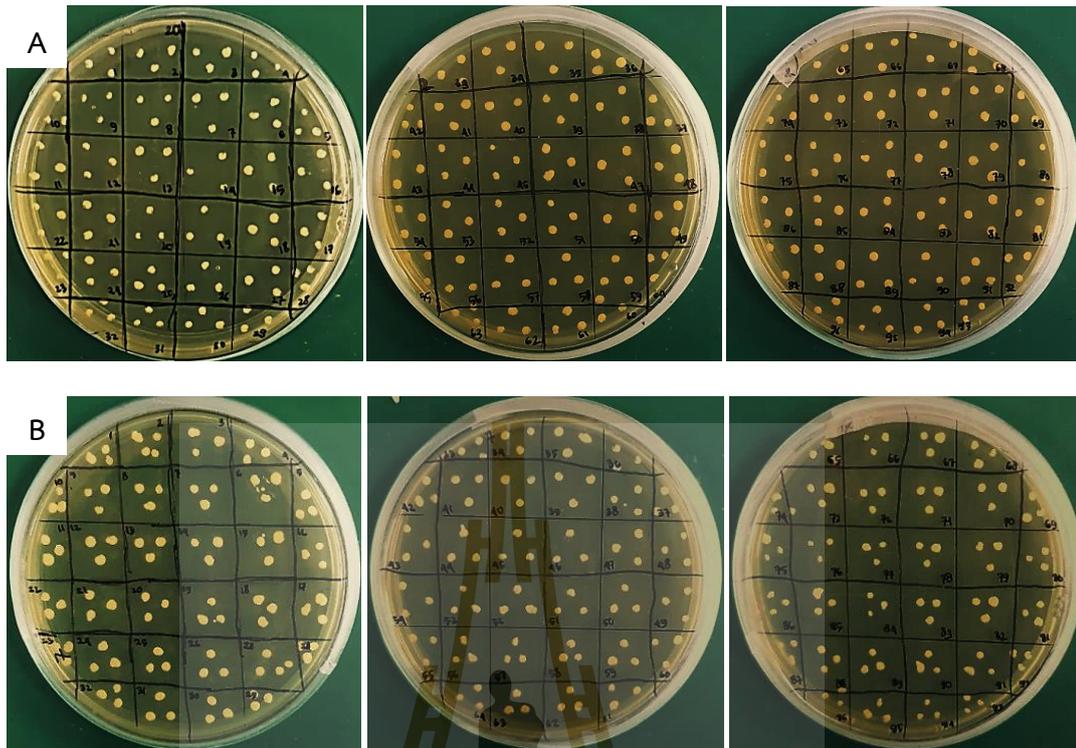
Figure 4.12 The yeast colonies were screened with 250 μg/mL zeocin and grown on YPD agar with triplicate spots for OsSCPL2a (a) and OsSCPL7 (b).

4.5 Isolation and purification of SCPL protein from the yeast expression system

Protein purification was performed with sepharose-Ni^{2+} IMAC resin. After collecting all fractions, the protein samples were observed by SDS-PAGE, which resulted in discrete bands of 75 and 45 kDa in crude and flow-through fractions from IMAC **(Figure 4.13)**. However, after washing with Eq buffer and low concentrations of imidazole (10 and 20 mM, respectively), the band of approximately 75 kDa was obtained during expression of OsSCPL7, but not in OsSCPL2a. After elution with 100 mM and 250 mM imidazole in Eq buffer, a smaller band around 50 kDa appeared, which is the expected protein size. The protein band was clearly observed in OsSCPL7.

OsSCPL7 fractions were divided into two parts, the washed fraction and elution fraction, before concentrating and digesting with endoglycosidase-H (EC 3.2.1.96) overnight to remove high mannose N-linked carbohydrate and see whether it was a glycoprotein. Glycosylation is a common post-translational modification of secreted proteins in eukaryotic cells and closely related to the folding, stability and activity of the protein. Therefore, abnormal glycosylation may affect protein function (Wang *et*

al., 2015). *P. pastoris*, which is a commonly used eukaryotic host for heterologous expression, has a complex glycosylation system that adds high mannose glycan (Cereghino and Cregg, 2000).
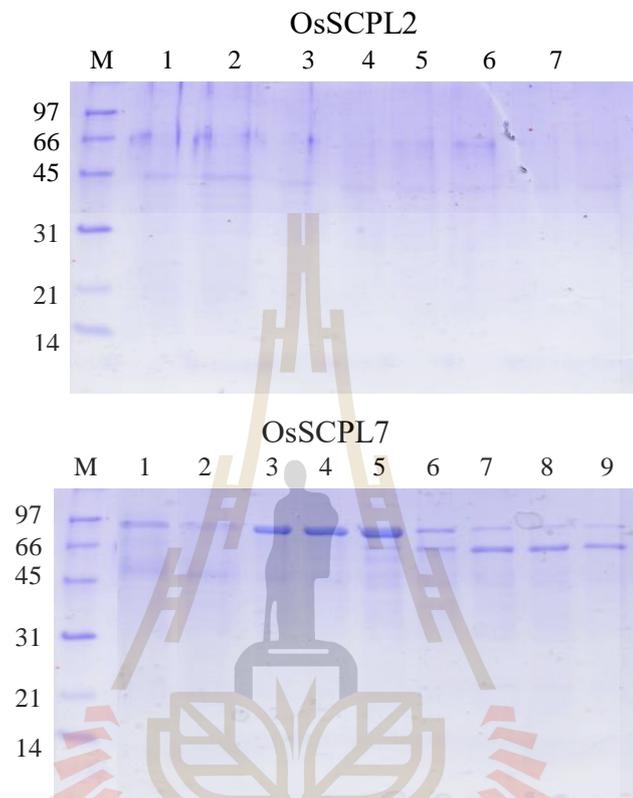
Figure 4.13 SDS-PAGE results from protein purification. Protein bands are seen in OsSCPL7 and OsSCPL2a expression. Only in OsSCPL7 expression, the 75 kDa band consistently appeared in elution fractions (E1 & E2 = elution with 100 mM imidazole in Eq buffer; E2 & E3: elution with 250 mM imidazole in Eq buffer).

A previous report indicated that hyper-glycosylation may enhance the thermodynamic and kinetic stability of recombinant protein (Shental-Bechor and Levy, 2009). However, some studies have reported that hyper-glycosylation that occurs in *P. pastoris* may compromise function (Guo *et al.*, 2002; Partridge *et al.*, 2004). Therefore, digestion of glycoproteins by endoglycosidase-H is an important test to check the effect of hyper-glycosylation on the function and stability of the protein, as well as to check the size of the protein without extensive glycosylation. As explained above, the expected protein OsSCPL size was around 50 kDa **(Figure 4.14).**
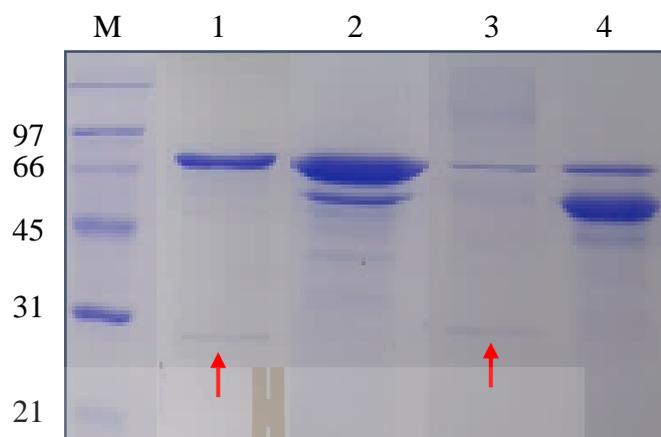
Figure 4.14 The SDS-PAGE gel of OsSCPL7 from IMAC purification, digested with endoglycosidase-H. M= marker; 1 and 2, digested and undigested washed fraction; 3 and 4, digested and undigested elution fraction. The endoglycosidase-H band is seen at 29 kDa (red arrows).

4.6   Identification of expressed protein by LC/MS/MS

After expressing the proteins in *Pichia pastoris* and purification by IMAC, the protein was loaded to SDS-PAGE gel and the protein band used for identification based on the masses of its tryptic peptides. In-gel digestion by trypsin enzyme was performed for OsSCPL7 that had been expressed in *Pichia* and partially purified. The final dried product of protein sample from gel purification was sent to BIOTEC, Thailand to be analyzed by LC/MS/MS.

Based on the tryptic oeptide MS/MS ion search, the main protein found in the medium when OsSCPL7 was expressed in *Pichia* system was identified as alcohol oxidase I (AOX1) protein from *Komagataella phaffii* (*K. phaffii*), with a score of 457 as a MASCOT protein hit **(Figure 4.15)**. Kurtzman (2009) reported that the strain of *P. pastoris* commonly used in gene expression studies is actually *K. phaffii*. Fourteen masses and nine sequences matched with this protein. Meanwhile, another protein also was detected as trypsin from *Sus scrofa,* with a hit score of 87, which matches the enzyme that was used to digest the protein and was left over as a contaminant (**Table 4.1**).
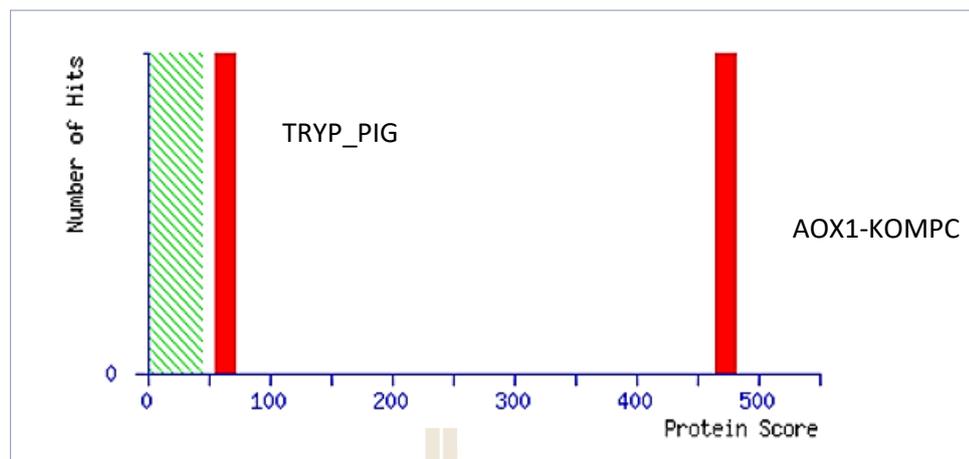
Figure 4.15 Protein score distribution of OsSCPL7. Protein score hit at 457 identified as alcohol oxidase 1, AOX1, from *Komagataella phaffii* (strain ATCC 76273/ CBS 7435/ CECT 11047/ NRRL Y-114, accession ID: XP_002494271.1).

Table 4.1 LC/MS/MS result of OsSCPL7 purification showed 14 matched peptides to AOX1 K. phaffii. Individual ion score >45 indicate identity or extensive homology.

Observed	Mr (expt)	Mr (calc)	Score	Peptide
900.2	2697.7	2697.4	27	KVGLIEAGENNLNNPWVYLPGIYPRN
583.6	1747.8	1747.8	46	KTASFYTSNPSPHLNGRR
875.0	1748.0	1747.8	87	KTASFYTSNPSPHLNGRR
838.4	1674.8	1674.7	73	RGSASDYDDFQAEGWKT
838.5	1674.9	1674.7	78	RGSASDYDDFQAEGWKT
834.9	1667.8	1667.8	73	RACNNPDIHGFEGPIKV
545.3	1088.6	1088.6	40	RSGFGDPIKLRA
724.4	1446.9	1446.9	54	RAAGVKPLVNLPGVGRN
483.3	1446.9	1446.9	42	RAAGVKPLVNLPGVGRN
594.3	1779.8	1779.7	50	RNFQDHYCFFSPYRI
890.9	1779.8	1779.7	56	RNFQDHYCFFSPYRI
803.5	2407.6	2407.2	72	RVFDQWYANGTGPLATNGIEAGVKI
1204.7	2407.4	2407.2	60	RVFDQWYANGTGPLATNGIEAGVKI
899.5	2695.3	2695.3	55	KVGDLSVCPDNVGCNTYTTALLIGEKT

The heterogenous expression in *Pichia* was driven by a promoter derived from the *alcohol oxidase 1* (*AOX1*) gene, which is suited to control the expression of foreign genes due to its induction by methanol (Cregg *et al.*, 2000 and Koch *et al.*, 2016). The *AOX1* gene encodes the AOX enzyme (EC 1.1.3.13), which belongs to the family of glucose-methanol-choline (GMC) oxidoreductases (Kiess *et al.*, 1998), and allows *P. pastoris* to live on methanol as a sole carbon source. Release of AOX1 protein into the medium during the expression of OsSCPL7 protein is contrary with the theoretical benefit of the extracellular expression system with the α-factor signal sequence, which is expected diminish the protein mixture complexity, due to the small number of endogenous proteins excreted in the extracellular medium by *Pichia* (Celik and Calik, 2012; Chen *et al.*, 2018).

P. pastoris AOX1 is normally found in the peroxisome, where it helps to metabolize methanol and other short chain alcohols (Ozimek *et al.*, 2005). The appearance of AOX1 in the medium is likely an indication of stress. Although the two OsSCPLs were expressed under identical conditions, AOX1 was only found in the medium from clones expressing OsSCPL7. Interestingly, AOX1 protein in the medium bound to the IMAC column and was found in every elution fraction, possibly reflecting the fact that it is a metalloenzyme, since the IMAC resin chelates divalent metal ions (Barnett *et al.*, 2012).

In the previous reports, a number of SCPL family enzymes, such as isobutyryltransferase from *Lycopersicon pennellii* (Li and Stefens, 2000), sinapoyl acyltransferase in *Arabidopsis thaliana:* AtSMT (Stehle *et al.*, 2006, 2008), AtSCT (Shirley and Chapple, 2003), and sinapoylglucose: sinapoylglucose sinapoylacyltransferase (Stehle *et al.*, 2013) were expressed using another species of yeast, *Saccharomyces cerevisiae,* as the heterologous host with different efficiencies. Although we employed different yeast strain, *P. pastoris*, we expected that the system and methods are not very different. Interestingly, instead of our target protein, another protein from the host background was released.

Although AOX1 in *P. pastoris* is known as an SCAO involved in methanol metabolism*,* and is naturally an intracellular enzyme (Koch *et al.*, 2016), aromatic alcohol oxidase (AAO) and secondary alcohol oxidase (SAO) are similar enzymes that are also mostly secreted to the medium (Goswami *et al.*, 2013). AAO (EC 1.1.3.7) has molecular weights around 63.5 to 78 kDa, suggesting that it could be included in the band from SDS-PAGE gel (around 75 kDa), although the tryptic peptide MS analysis only identified AOX1. In fact, we could find no study that has yet reported that AAO

can be found in *P. pastoris*, nor could we find a study reporting the trouble shooting to avoid the releasing of AOX1 from the host cells.

4.7   The AOX activity towards cyanidin-3-O-glucoside (Cy3G)

Reactions of the purified alcohol oxidase 1 (AOX1) with cyanidin-3-O-glucoside (Cy3G) were conducted before the protein was identified as AOX1. The full reaction consists of substrate Cy3G, an acyl glucose donor, 50 mM citrate buffer pH 4.5, and AOX1. Subsequently, reactions without the acyl glucose donor were found to give similar results. The reactions were incubated at 30 °C overnight without terminating the reactions. The reaction mixtures were subjected to UHPLC analysis. The UHPLC spectra shows that Cy3G has absorbance at 280 and 520 nm.

Anthocyanin, including C3G, has UV-visible absorption spectra with λ_{max}= 280 nm in the UV range and $\lambda_{max=}$ 520 nm in the visible range (Qin *et al*., 2010). The standard Cy3G was detected at the retention time of 7.6 in absorbance 520 nm, while two peaks were observed at 7.6 and 8.0 min by detection at 360 nm. Additional peaks were present in the UHPLC chromatogram in the reaction containing the enzyme, at the retention times of 9.5 and 10.5 min in absorbance 360 nm **(Figure 4.16a)**.

This result indicates that there are products from AOX activity, which presumably oxidized C3G. The additional peaks were not recorded in 520 nm absorbance, as would be expected for acylated anthocyanin, although the peak area for Cy3G was significantly decreased. However, the additional peaks were observed in absorbance at 360 nm, suggesting that AOX protein may catalyze the oxidation reaction that changes the Cy3G structure so that it is no longer absorbs at 520 nm.
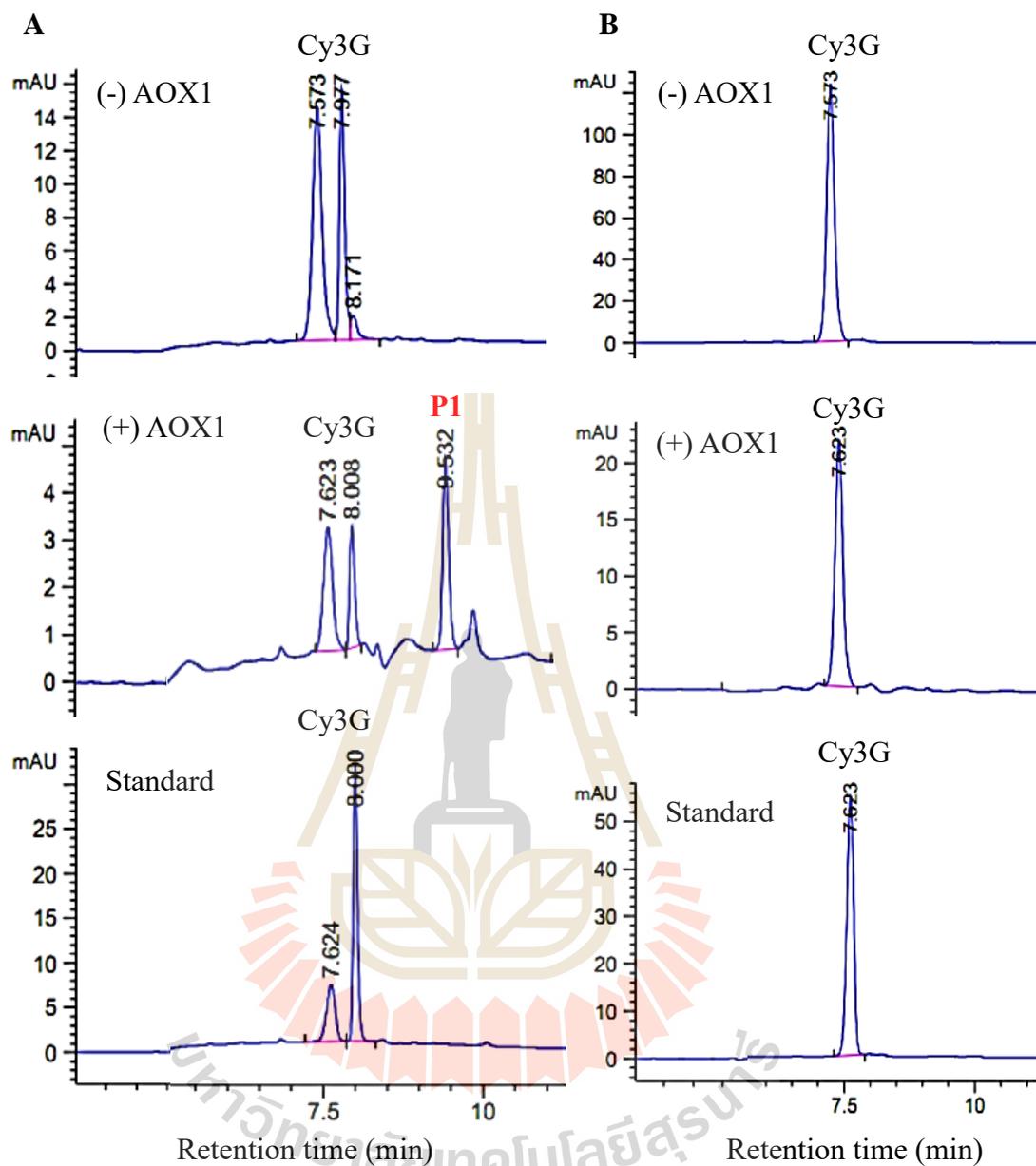
Figure 4.16 Effect of purified AOX1 activity upon anthocyanin. UHPLC chromatograms with detection by absorbance at 360 nm (A) and 520 nm (B), respectively.

4.8   Construct of recombinant plasmid for bacterial expression protein system

Since expression of the rice OsSCPL proteins in *Pichia* was unsuccessful, we attempted to use a bacterial expression protein system. To prepare the construct of recombinant protein in bacterial vector, cDNA from recombinant plasmid pPICZαBNH8/OsSCPL2a and pPICZαBNH8/OsSCP7 were cloned into the *Nco*I and *Sal*I

restriction sites in pET32a, creating new recombinant pET32a/OsSCPL2a and pET32a/OsSCPL7 plasmids, respectively.

A   **B**

pET32a/OsSCPL2a   pET32a/OsSCPL7
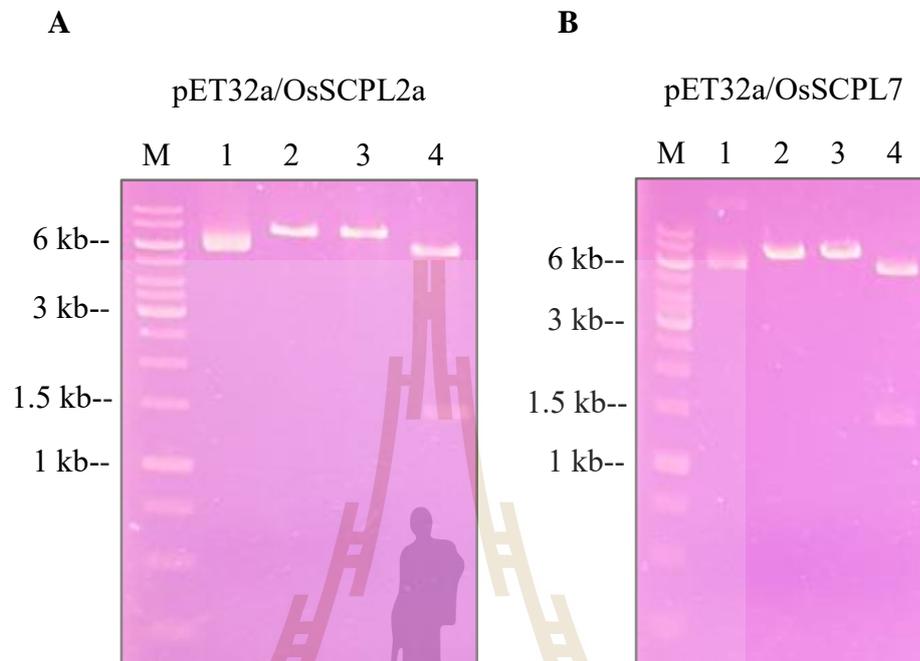
Figure 4.17 Confirmation of OsSCPL genes cloned into pET32a vector (5900 bp). (A) Electrophoresis on 1% gel agarose of pET32a/OsSCPL2a and (B) pET32a/OsSCPL7. M, marker; lane 1, undigested recombinant plasmid; lane 2, recombinant plasmid digested with NcoI; lane 3, plasmid digested with SalI; lane 4, double digestion of recombinant plasmid with NcoI and SalI to release the OsSCPL genes (approximately 1400 bp).

To confirm the presence of the expected insert sizes of 1343 bp for OsSCPL2a and 1400 bp for OsSCPL7, the new recombinant plasmids were checked by single and double digestion with *Nco*I and *Sal*I restriction endonucleases. Each recombinant plasmid was observed to have the expected size of around 7300 bp upon single digestion, and double digestion with *Nco*I and *Sal*I showed the expected size of around 1400 bp for the cDNA inserts **(Figure 4.17)**. The sequences were further verified by sequencing of the full inserts by automated DNA sequencing, which showed that they had the desired sequences and were in-frame with the N-terminal fusion tag sequences.

4.9 Rice OsSCPL protein expression in *E. coli* expression system

In this work, the pET system was employed for the expression of heterologous model protein, yielding pET32a/OsSCPL2 and pET32/OsSCPL7, which were transformed into *E. coli* strain Origami(DE3) and grown with appropriate antibiotics. This family of vectors contains a pMB1 origin of replication (medium-copy number replicon) and uses the T7 promoter for gene transcription. When the gene for T7 polymerase is under the control of the lac derivative operon, so isopropyl β-D-1-thiogalactopyranoside (IPTG) is usually added to induce protein expression (Rosano and Ceccarelli, 2014). This system is one of the most commonly used expression systems in *E. coli*, mainly due to its very high expression levels as the target protein can reach up to 50% of the total cell protein (Rosano and Ceccarelli, 2014; Gomes and Mergulhão, 2020). Unlike lactose and other galactosides, IPTG is a metabolic-free, or gratuitous inducer, because it is not metabolized by the cell (Silaban *et al.*, 2018).

The level of gene transcription can be regulated by using an appropriate concentration of IPTG. Thus, the protein induction was screened with various concentrations of IPTG: 0.0, 0.2, 0.4, 0.6, and 0.8 mM. The protein was tested by SDS-PAGE and compared with empty vector pET32a without induction with IPTG **(Figure 4.18)**. The result show that there was no soluble protein detected on SDS-PAGE (supernatant fraction) even after induction with various concentration of IPTG. However, the insoluble fractions are obviously presence in the pellet, especially in pET32a/OsSCPL2a induced with 0.6 mM IPTG. Only insoluble protein was observed with pET32a/OsSCPL7 as well, in which induction with IPTG resulted in less insoluble protein. Since there was no result which clearly indicated soluble protein in supernatant, 0.4 mM IPTG was chosen as inducer for further experiments.

Some experiments were conducted to express the soluble protein from OsSCPL2a and OsSCPL7, such as: changing the competent cell to Rosetagami(DE3) and BL21(DE3), optimizing the time and temperature of incubation, adjusting the extraction or lysis buffer composition, and growing cells in auto induction media (AIM). In some cases, auto-induction allows efficient screening of many clones in parallel for expression and solubility, as cultures have only to be inoculated and grown to saturation, and yields of target protein are typically several-fold higher than obtained by conventional IPTG induction (Studier, 2005). However, in this work, none of approaches gave a significant increase in the amount of soluble protein.
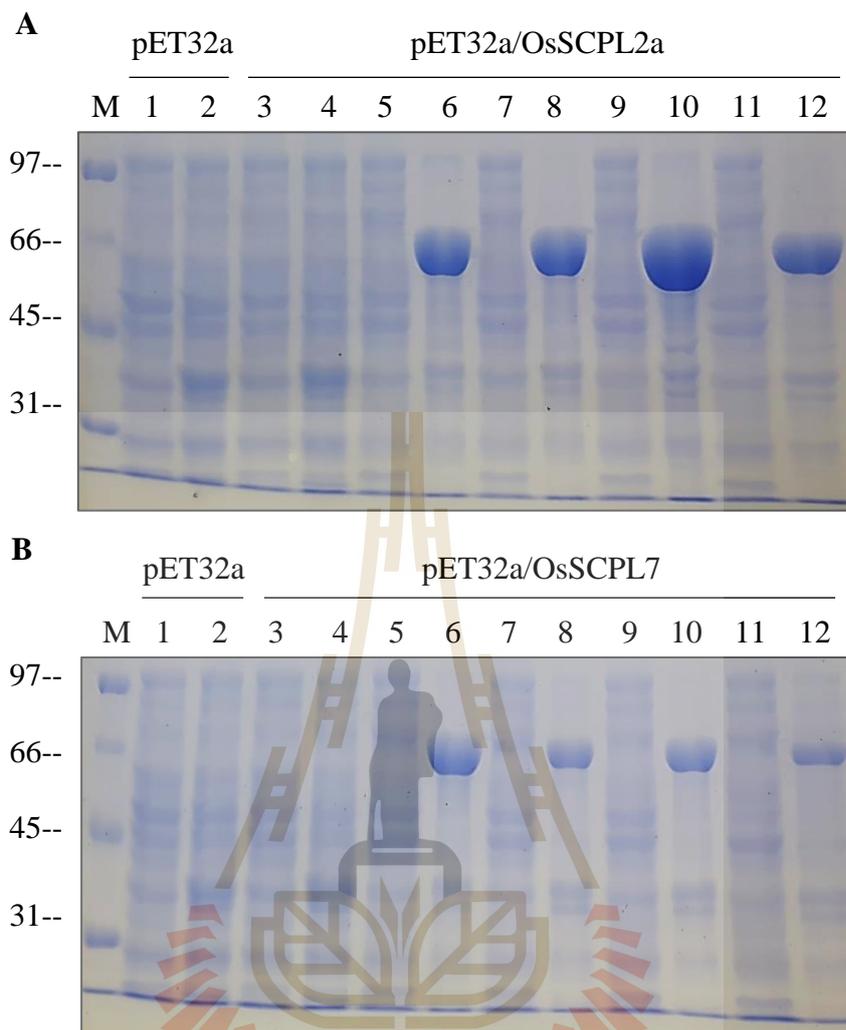
Figure 4.18 SDS-PAGE analysis of extracts of E. coli harboring pET32a and OsSCPLs expression vector pET32a/OsSCPL2a (A) and pET32a/OsSCPL7 (B). The odd numbers represent supernatant and the even numbers represent pellets. Lanes 1 and 2 are extracts of cells containi ng pET32a with OsSCPL insert. Lanes 3 and 4 are cells grown in the absence of IPTG; lanes 5-6, induced with 0.2 mM IPTG; lanes 7 -8, induced with 0.4 mM IPTG; lanes 9-10, induced with 0.6 mM IPTG; and lanes 11-12, induced with 0.8 mM IPTG.

Instead, the purification of protein was conducted to confirm the possibility of soluble protein expression. Two heterologous expressed protein pET32a/OsSCPL2a and pET32a/OsSCPL7 were purified by Co^{2+} IMAC resin. The same procedure was conducted for empty vector as negative control. After the first purification by Co^{2+} IMAC resin, analysis by SDS-PAGE of the soluble protein fraction revealed that OsSCPL2 were seen to be more convincing for further analysis compared to OsSCPL7. A big band

was found in the pellet fraction, indicating the presence of insoluble protein. In SDS-PAGE gels, a protein band corresponding to approximately 60 kDa consistently showed up from early washing fraction with Eq buffer, leading to elution with small concentration of imidazole. A 60 kDa of protein is appropriate with the calculated molecular mass of recombinant SCPL protein plus thioredoxin and 6x His tags produced from pET32a plasmid **(Figure 4.19A)**.

The second purification was subsequently performed with the same column resin and conducted after digestion of the OsSCPL2a protein with enterokinase **(Figure 4.19B)**. The concentrated protein from the pool of fractions eluted with 100- and 200-mM imidazole of first purification, appeared highly impure, as there were many bands in the SDS-PAGE of the recombinant protein. Thus, a second purification step was needed. Upon visualization by SDS-PAGE, increased protein purity was observed in some fractions, including the flow-through solution, wash with Eq buffer, and those eluted with low concentrations of imidazole. Interestingly, a 60 kDa protein band consistently appeared even after the digestion with enterokinase. Our target protein has molecular weight of 49 kDa, once the tag is removed. The corresponding band seems to be less intensive yet clearly observed in the final concentrated protein **(Figure 4.19B)**.

Despite its many advantages, there are also disadvantages for using *E. coli* as an expression host. In contrast to eukaryotic systems, transcription and translation are fast and tightly coupled. Since many eukaryotic proteins require longer times and/or the assistance of folding chaperones to fold into their native state, this rate enhancement often leads to a pool of partially folded, unfolded, or misfolded, insoluble proteins (Oberg *et al.*, 1994). Thus, some targets, especially larger multidomain and membrane proteins, either fail to express in *E. coli* or express insolubly as inclusion bodies (Francis and Page, 2010).

Although the majority of the recombinant protein was present in the insoluble fraction and the protein was undetectable in the supernatant fraction after SDS-PAGE. Lehfeldt *et al.* (2000) reported that the expression of SMT from Arabidopsis in *E. coli* had acyltransferase activity. SMT activity was readily measured in the soluble protein fraction. Moreover, there are some recombinant SCPL acyltransferases that successfully refolded from the inclusion body. Application of refolding procedures resulted in catalytically active SCT from *A. thaliana* and *Brassica napus* (Shirley *et al.*, 2001; Milkowski *et al.*, 2004).
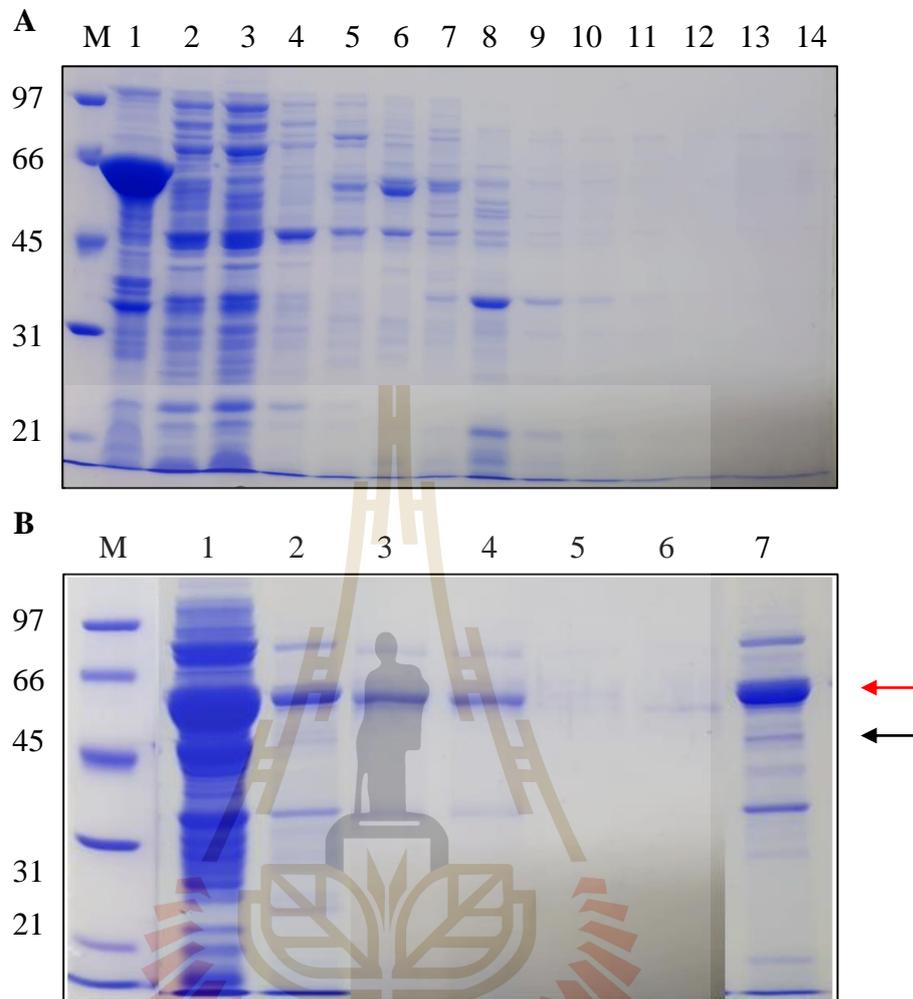
Figure 4.19 SDS-PAGE result of IMAC purification of OsSCPL2a protein from the bacterial expression system. (A) First purification by IMAC showed a large insoluble protein band in the pellet fraction. The protein was eluted with Eq buffer, 10, 20, 100- and 200-mM imidazole in Eq buffer. Lane 1, pellet; 2, supernatant; 3, flow through; 4, washed with Eq buffer; 5-6, washed with 10 mM imidazole; 7-8; washed with 20 mM imidazole; 9-11, eluted with 100 mM imidazole; 12-14 eluted with 200 mM imidazole. (B) Second purification that was performed after digestion of the pool of eluted fraction with 100- and 200-mM imidazole with enterokinase. The initial fractions were collected and concentrated. Lanes 1 and 2, undigested and digested with enterokinase, 3, flow through; 4, washed with Eq buffer, 5 and 6, washed with 10- and 20-mM imidazole; 7, final concentrated protein (red arrow, 60 kDa; black arrow, 52 kDa).

Ciarkowska *et al*. (2018) reported the expression of recombinant rice 1-*O*-indole-3-acetyl-β-D-glucoside (1-*O*-I AGlc): *myoinositol* acyltransferase (IAInose) synthase protein was synthesized in *E. coli* cells with high efficiency. However, the enzyme was in the form of insoluble inclusion bodies, even though induction of protein synthesis was at lower temperature (18 °C) overnight, a culture condition that is usually suitable for increasing production of recombinant proteins in soluble form (Gasser *et al*., 2008; Demain and Vaishnav, 2009). No IAInose synthase activity was detected either in the supernatant fraction or in inclusion bodies fraction. As mentioned above, we tried to increase the expression of recombinant protein in the soluble fraction by optimizing the conditions of expression. However, these attempts did not significantly improve the result. Thus, we tried to identify the protein from SDS-PAGE band by tryptic digestion and subsequently tested with mass spectrometry.

4.10 Identification of Rice OsSCPL2a protein expressed in *E. coli* by tryptic digest and MS/MS of the SDS-PAGE band

The major SDS-PAGE band and minor bands with masses similar to the expected mass of OsSCPL2a protein expressed in *E. coli* **(Figure 4.19B)** were submitted to tryptic digest and mass spectrometry to identify the proteins. The most intense band was identified as 60 kDa chaperone protein from *E. coli* (Accession number: WP_071686894.1) along with other proteins **(Table 4.2)**.

Table 4.2 The matched proteins based on MS/MS ions from the 60 kDa SDS-PAGE band from bacterial expression Of OsSCPL2a.

Organism Source	Score	Mass	Matches	Description
Bacteria				
- CH60_ECO24	17031	57464	579	60 kDa chaperonin *E. coli*
- CH60_COLMA	5636	57711	169	60 kDa chaperonin *C. maris*
- CH60_ERWAP	4513	56786	216	60 kDa chaperonin *E. aphidicola*
- CH60_PHOLL	2864	57541	151	60 kDa chaperonin *P. laumondii*
- CH60_CHRVO	1782	57496	72	60 kDa chaperonin *C. violaceum*
- CH60_VIBPA	113	56367	7	60 kDa chaperonin *V. parahaemolyticus*
- KPYK_ECOLI	1538	51553	77	Pyruvate kinase II *E. coli.*

Table 4.2 The matched proteins based on MS/MS ions from the 60 kDa SDS-PAGE band from bacterial expression Of OsSCPL2a (Continued).

Organism Source	Score	Mass	Matches	Description
Homo sapiens				
- K2C1-HUMAN	488	66170	21	Keratin type II
- K1C10_HUMAN	451	59020	16	Keratin type II
Other				
00021304.1	302	66110	13	Keratin, type II cytoskeletal epidermal

Table 4.3 The matched proteins based on MS/MS ions from the 49 kDa SDS-PAGE band from bacterial expression of OsSCPL2a.

Organism source	Score	Mass	Matches	Description
Escherichia coli				
- RHO_ECOLI	4754	47032	252	Transcription termination factor
- CH60_ECO24	2542	57464	102	Rho 60 kDa chaperonin
- KATG_ECODH	2329	80031	99	Catalase peroxidase
- IMDH-ECOLI	1178	52275	35	Inosine-5'-monophosphate dehydrogenase
- DHE4_ECOLI	1094	48778	32	NADP-specific glutamate dehydrogenase
- ENO_ECOBW	448	45683	19	Enolase
Oryza sativa				
- Q10A76	374	52498	12	Os10g0101200 protein
Homo sapiens				
- K2C1_HUMAN	282	66170	14	Keratin, type II cytoskeletal

The second protein from the band at approximately 49 kDa also gave several *E. coli* proteins, but limiting the search to rice resulted in a clear best match to Os10g0101200 of *Oryza sativa* subspecies japonica, also recognized as a serine carboxypeptidase-like 2a (SCPL2a) protein (ID: XP_015614311.1), designated here as OsSCPL2a. The protein has a protein match score of 374, twelve peptide masses from six peptide sequences matched OsSCPL2a **(Table 4.3)**. Indeed, most of the protein background from the host cell and contaminants were also detected as the top 7 hits when the search was not limited to rice proteins **(Table 4.4)**. The impurity of OsSCPL2a

protein obtained even after the second purification with subtractive IMAC indicated that OsSCPL2a is a low-level component in the soluble protein, with the SDS-PAGE of the pellet suggesting it is mostly in the insoluble fraction. However, I continued to test the activity of the protein in the soluble fraction.

Table 4.4 LC/MS/MS result of OsSCPL2 purification showed six matched peptides to Os10g0101200 protein of Oryza sativa. Individual ion score >45 indicate identity or extensive homology.

Observed	Mr (Exp)	Mr (Calc)	Score	Peptide
472.32	942.63	942.50	49	R.LGLPVECR.D
489.02	976.03	976.03	51	R.CTAPLFR.H
534.32	1066.63	1066.44	54	K.ECLDMLDR.W + Oxidation (M)
604.38	1206.74	1206.62	81	K.GYLVGNAATDVK.Y
732.41	1462.81	1462.71	93	R.LSYLMADDPEVR.A
1007.01	2012.00	2012.04	116	K.VSNIFLDAPVGTGFSYAR.E

4.11 Rice serine carboxypeptidase-like 2a protein (OsSCPL2a) reactions produced hydrolysis products

Acyltransferases (EC 2.3.1.x) catalyze the transfer of an acyl group from a donor molecule to the hydroxyl, amino, or thiol group of an acceptor molecule to form an acyl conjugate. In plant secondary metabolism, acylation is a widespread reaction. Together with glycosylation, hydroxylation, methylation, decarboxylation, and oxidation/reduction, acyltransferase reactions account for the remarkable diversity of plant natural products generated from a limited number of different central metabolic pathways (Osbourn and Lanzotti, 2009; Vogt, 2010).

We tested the activity of OsSCPL2a acyltransferase protein using the crude (pellet and supernatant extract), fractions collected from purification, as well as concentrated protein. As a preliminary test, acyl donor *p*-hydroxybenzoyl-glucose (pHBAG) and luteolin-7-*O*-glucoside (L7G) were included to the mixture along with citrate buffer pH 4.5. We also conducted the same treatment for the crude extract of induced cells carrying empty pET32a vector.

The enzymatic reaction was incubated overnight at 30 °C and analyzed by TLC, which did not detect activity in the reaction with crude protein, both from extracts of cells carrying empty and recombinant vector. Furthermore, the L7G's spots were slightly different to each other in the reaction with purification fractions, likely due to

buffer effects. Interestingly, the reaction with concentrated protein of OsSCPL2a after the second purification yielded an additional spot, which was observed above the L7G substrate spot **(Figure 4.20)**. This result suggested that, although we obtained impure protein even after the second purification, the protein has activity toward the substrates.
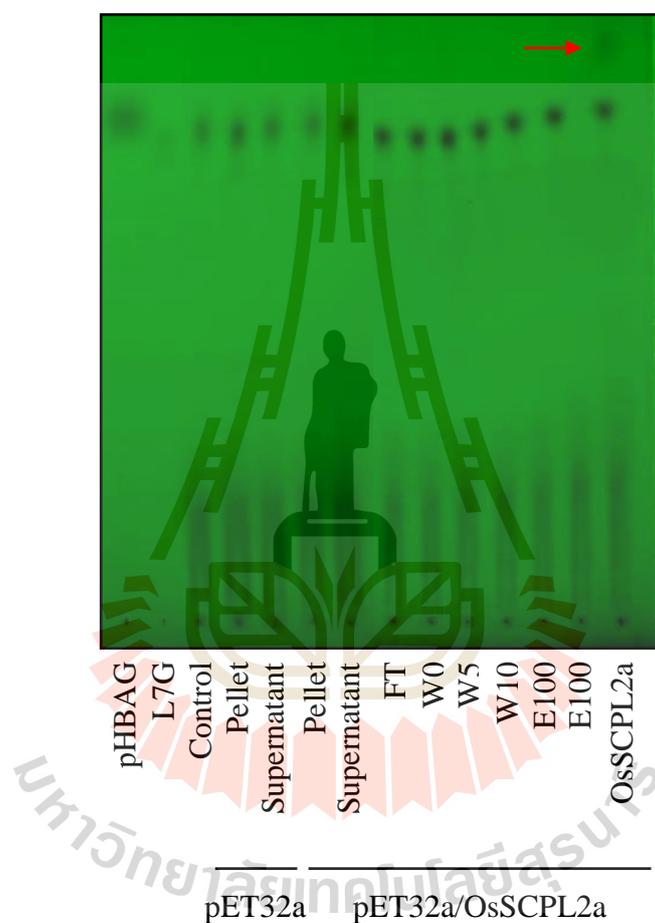
Figure 4.20 TLC analysis of OsSCPL2a activity compared with empty vector pET32a, visualized under UV light. The concentrated protein of OsSCPL2a after second purification resulting a new spot above the luteolin-7-O¬-glucoside (L7G) substrate spot (red arrow). Phbag p-hydroxybenzoyl-glucose (acyl donor); control, reaction without enzyme; FT, flow through; W0, washed with Eq buffer; W5-10, washed with 5- and 10-mM imidazole; E100-200, eluted with 100 and 200 mM imidazole, respectively.

More sensitive analysis to test for acyl transfer was conducted by UHPLC. The reaction mixtures consisted of acyl donor, acyl acceptor, 50 mM citrate buffer pH 4.5, and 1 µg OsSCPL2a protein from *E. coli* expression. The product mixture after the reaction was subjected to an UHPLC analysis. The result showed that an additional peak on the UHPLC chromatogram was poorly recorded in the reaction containing cyanidin-3-*O*-glucoside (Cy3G) in the presence of OsSCPL2a when it measured at 360 nm, followed by decreasing of the Cy3G peak area **(Figure 4.21)**.

Interestingly, Cy3G is an anthocyanin compounds which has absorbance at 280 and 520 nm. An acylated anthocyanin was expected to be detectable in the range of 520-530 nm. In fact, no additional peaks can be observed at that range, although the peak area represents Cy3G was significantly decreased. Additionally, another anthocyanin compounds such as peonidin-3-*O*-glucoside (P3G) and cyanidin-3,5-di-*O*-glucoside (Cy35G) were tested with the same condition. However, neither of those compounds resulting the extra peak in absorbance 360, 520, and 530 nm.

Furthermore, we tested the specificity of the enzyme with 3-*O*-glycosylated and 7-*O*-glycosylated flavonoid: apigenin-7-*O*-glucoside (A7G), luteolin-7-*O*-glucoside (L7G), kaempferol-3-*O*-glucoside (K3G), kaempferol-7-*O*-glucoside (K7G), quercetin-7-*O*-glucoside (Q3G), and quercetin-7-*O*-glucoside (Q7G). However, based on TLC and UHPLC analysis, the extra product was not indicated in those reactions with 3-*O*-glycosylated flavonoid as acceptor substrate **(Figure 4.22)**.

On the other hand, a UHPLC analysis of the products of the enzymatic reactions using recombinant protein showed that extra peaks were produced in reactions with 7-*O*-glycosylated flavonoids. The photodiode array spectrum of the chromatogram obtained in the reaction mixture had peaks in absorbance at 360 nm, that of acyl donors showed peaks at 260 nm (for pHBG and VG) and 280 nm (for pCG and FG). According to UHPLC chromatogram, the extra peaks had the same retention times as the flavonoid aglycones **(Figures 4.23 and 4.24)**. This suggests that OsSCPL2a may have hydrolysis activity instead of acyltransferase activity.
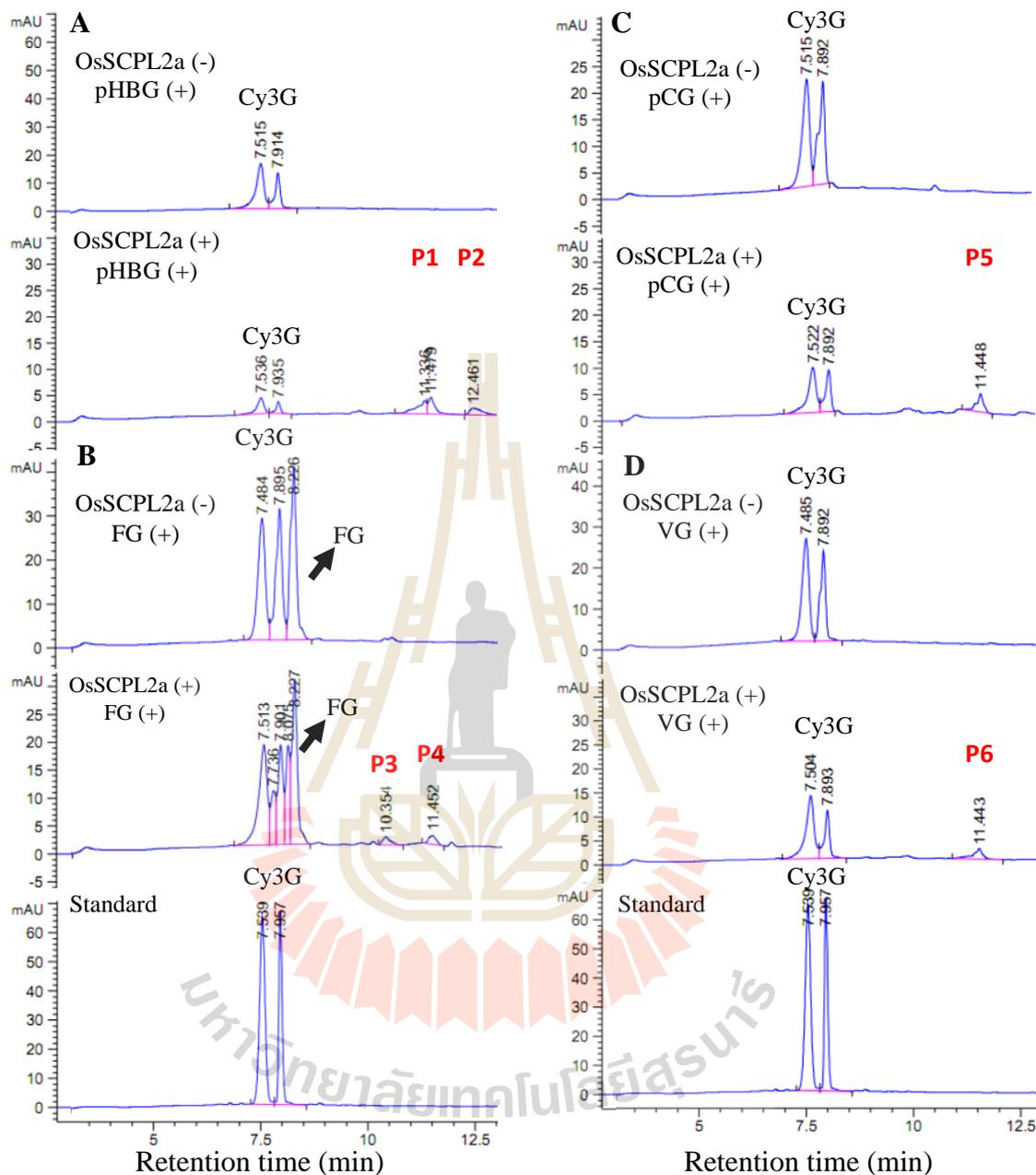
Figure 4.21 UHPLC chromatograms of the reaction products of recombinant OsSCPL2a protein in cyanidin-3-O-glucoside (Cy3G) with various glucose esters as acyl donors, such as: A) pHBG, B) FG, C) pCG, and D) VG. The small extra peaks were detected in the mixture treated with OsSCPL2a. The chromatogram was detected by 360 nm absorbance. mAU, miliabsorbance unit.
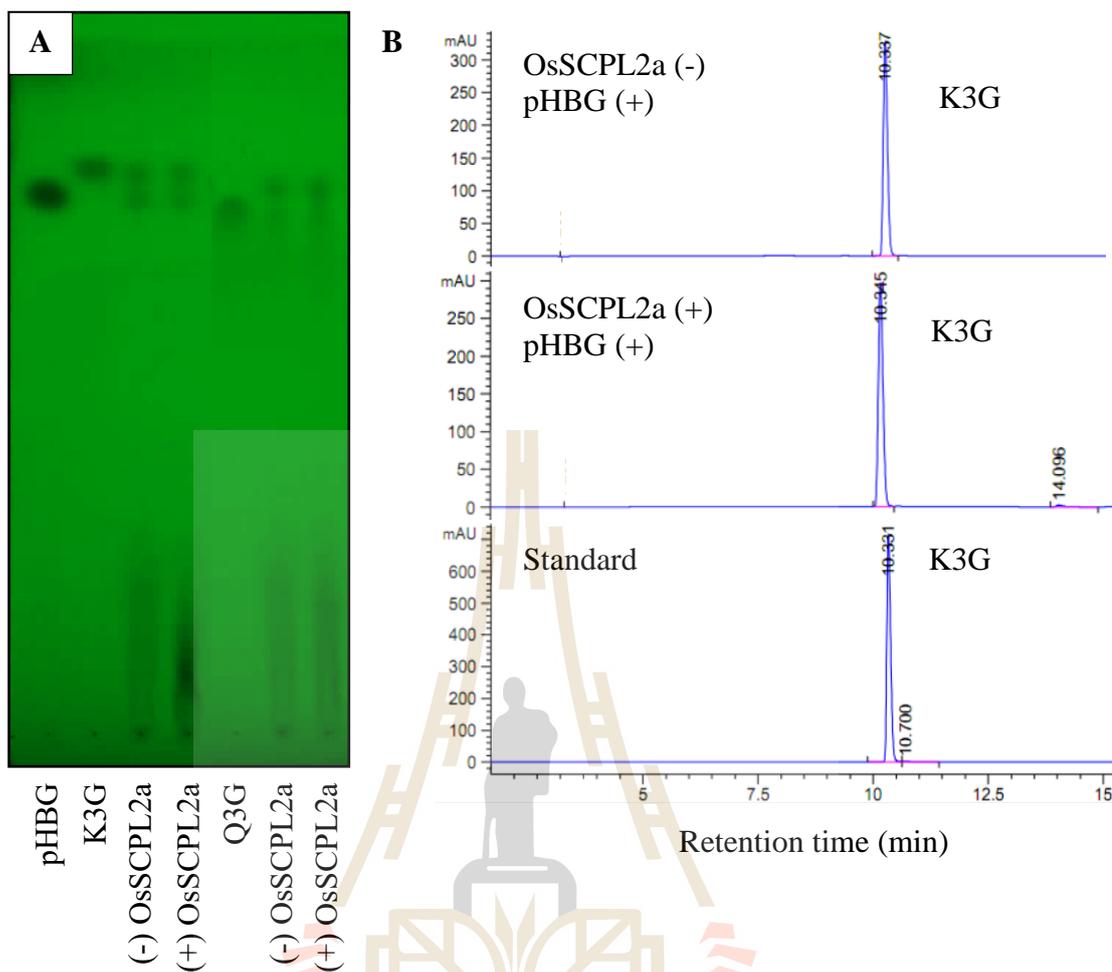
Figure 4.22 Figure 4.22 OsSCPL2a activity toward kaempferol-3-O-glucoside (K3G) and quercetin-3-O-glucoside (Q3G) in the presence of acyl donor pHBG, detected by A) TLC analysis using mobile phase B:A:W(3:1:1, v/v/v) and B) UHPLC analysis at 360 nm absorbance.

68
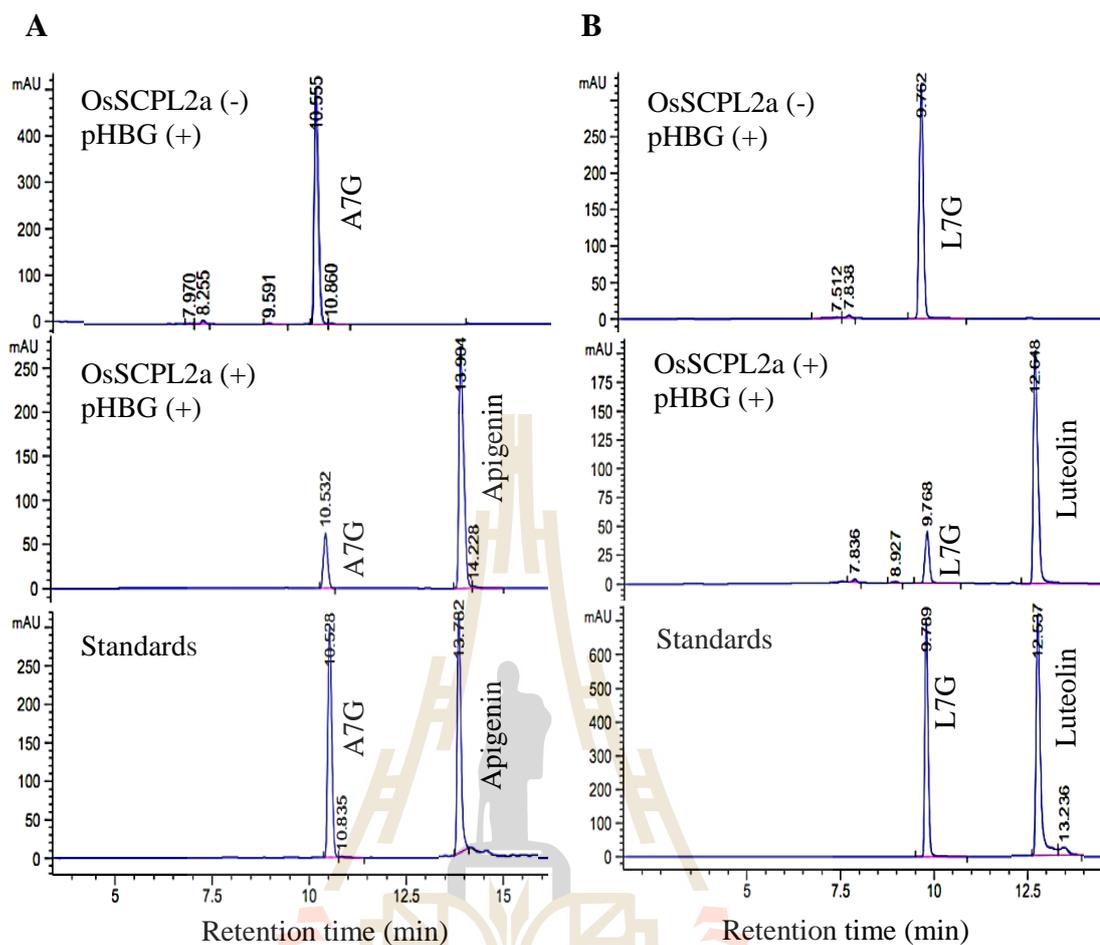
Figure 4.23 Detection of OsSCPL2a activity on the flavonoid glucosides apigenin-7-O-glucoside (A7G) (A) and luteolin-7-O-glucoside (L7G) (B). UHPLC analysis of the product of the reaction was performed after reaction with the enzyme (+) and in its absence (-). Absorbance was measured at 360 nm.
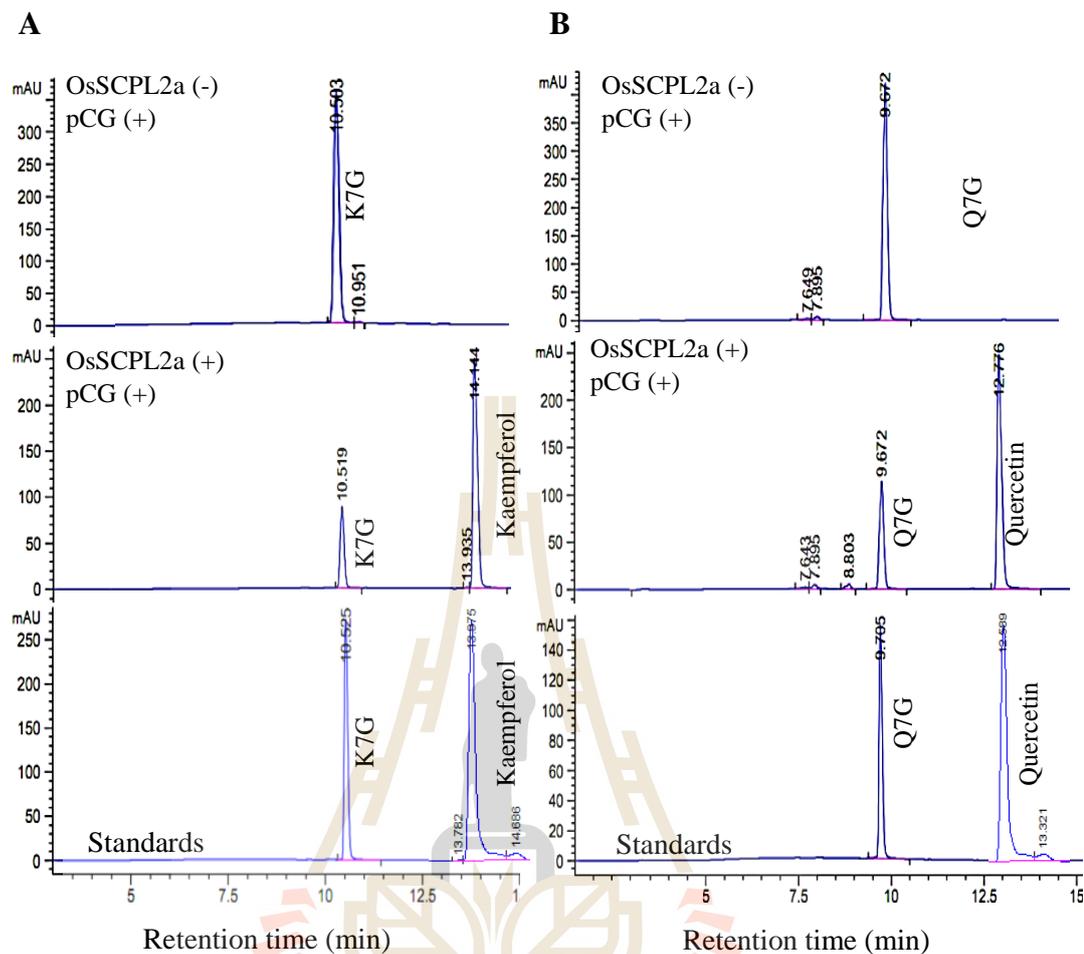
Figure 4.24 Detection of OsSCPL2a activity with the flavonoid compounds kaempferol-7-O-glucoside (K7G) (A) and quercetin-7-O-glucoside (Q7G) (B). UHPLC analysis of the product of the reaction performed on reaction products with the OsSCPL2a enzyme (+) and in its absence (-). Absorbance was measured at 360 nm.
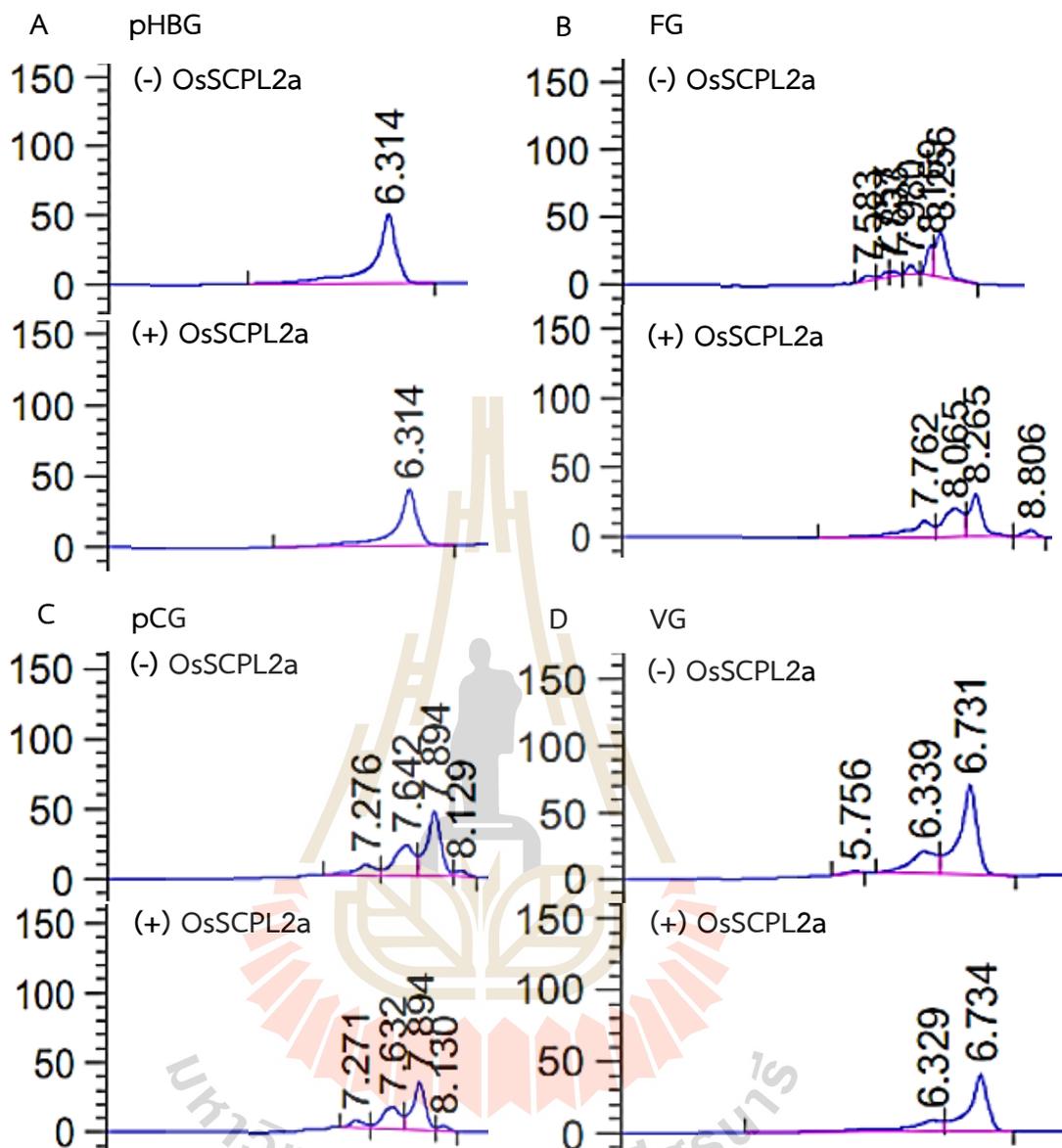
Figure 4.25 Detection of glucose ester contribution as acyl donor in OsSCPL2a activity. pHBG, p-hydroxybenzoyl-glucose (A) and VG, vanillyl-glucose (D), were detected by absorbance at 260 nm, while FG, feruloyl-glucose (B), and pCG, p-coumaroyl-glucose (C), were detected by absorbance at 280 nm. All the candidate acyl donors were tested with Q7G substrate.

The hydrolysis activity of OsSCPL2a protein was also confirmed by detection of glucose ester, corresponding to putative acyl donors in the UHPLC chromatograms. pHBG and VG have absorbance at 260 nm, while that of, FG and pCG was measured at the wavelength of 280 nm (**Figure 4.25**). These acyl donors showed no significant decrease of peak height in the enzymatic reaction **(Table 4.5)**, in which the shifting of absorbance and/or decreasing of the concentration the glucose ester could indicate that the donor acyl group (e.g. pHBA) was bonded to the acceptor substrate (Nishizaki *et al*., 2013).

Table 4.5 Peak height of acyl donor substrates in the reaction with quercetin-7-O-glucoside (Q7G) as acyl acceptor. Absorbance was measured in 360 nm.

| Acyl donor	Retention time (min)	Peak Height (mAU)	
		(-) OsSCPL2a	(+) OsSCPL2a
pHBG	6.3	49.7	39.3
FG	8.2	33.5	30.9
pCG	7.8	45.7	33.8
VG	6.7	66.7	59.3

Despite the apparent hydrolysis of flavonoid glucosides by OsSCPL2a, the decrease in the peak area observed in the UHPLC chromatogram varied, depending on what kind of glucose ester was included in the reaction mixture. For instance, the presence of pHBG and pCG in the reaction resulted in larger decreases in the peak area of all flavonoid 7-O-glucosides. Specifically, the A7G and L7G peaks showed the largest decrease when pHBG was included in the reaction, while the inclusion of pCG had the largest effect on K7G and Q7G **(Figure 4.26)**. Additionally, the peak absorbance area eluting at the position of flavonoid aglycones showed that there is no significant difference compare to the decreasing of acceptor substrates **(Table 4.6)**. This result could not supported the possibility of OsSCPL2a for having another activity such as acyltransferase activity.
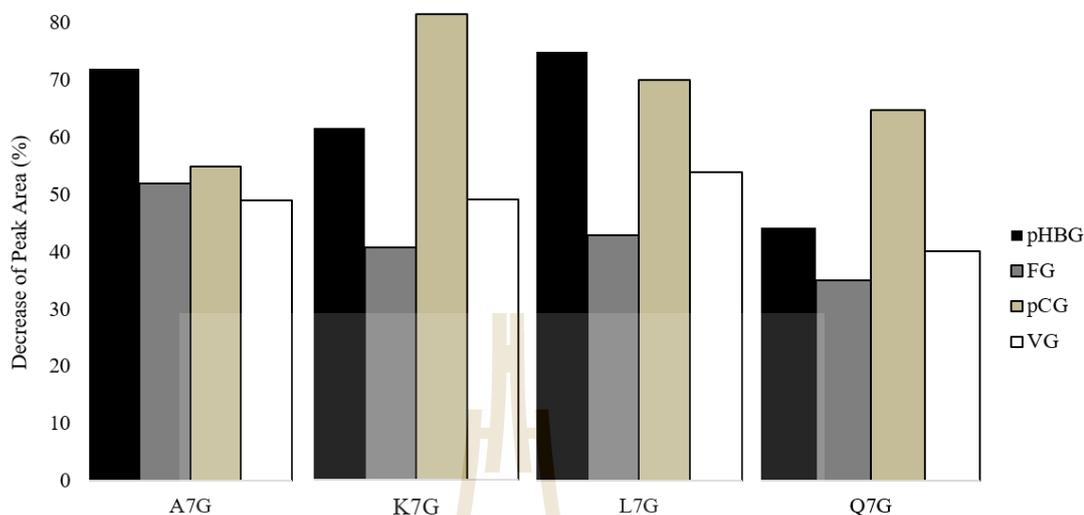
Figure 4.26 The decreasing of peak area of glycosylated flavonoid against glucose ester that was added as acyl donors in the reaction. Generally, the reactions with the addition of pHBG and pCG resulted in a larger decrease in glycosylated flavonoids, which suggests the possibility of acyl donor preferences in the reaction.

From the beginning, based on phylogenetic analysis of the amino acid sequences of both putative enzymes, OsSCPL2a and OsSCPL7 appeared likely to have acyltransferase activity. Despite those data, the conducted experiment showed that OsSCPL2a, which is closely related to the acyltransferases from other plants, appears to release 7-O-linked glucose from flavonoids instead of being an acyltransferase that acylates those flavonoid glucosides. In addition, the absence of new peaks in reactions with flavonoid glycosylated at the 3-O-position suggested the specificity of the enzyme to hydrolyze a substrate.

The hydrolysis activity found in a putative protein from SCPL is not impossible to occur, since SCPL proteins belong to the α/β hydrolases superfamily along with serine carboxypeptidases (SCP). This family shares a common motif composed of three non-successive amino acids: serine (Ser), histidine (His), and aspartic acid (Asp). This catalytic triad is a recurrent motif necessary for hydrolysis and other related functions, such as transfer reactions, and could explain the recruitment of SCPL acyltransferase during the evolution of secondary metabolite acylation (Milkowski and Strack, 2004). SCPL proteins are distinguished from SCP proteins, because they lost their activity as peptidases. This discovery was initiated by Li and Steffens (2000), whose reported the identification of a cDNA encoding a 1-O-β-acylglycose-dependent-acyltransferase

functioning in glucose polyester biosynthesis by *Lycopersicon pennellii.* After that, 19 genes in clade 1A of the SCP family, including a characterized sinapoyltransferase, were predicted as SCPL acyltransferase in *Arabidopsis thaliana* (Fraser *et al.*, 2005).

Table 4.6 Peak area of acyl acceptor substrates and products in the enzymatic reaction with pHBG and pCG as donor. Absorbance was measured at 360 nm.

| Acyl donor	Acyl acceptor	Peak area of acceptor (%)		Peak area of product (%)
		(-) OsSCPL2a	(+) OsSCPL2a	
pHBG	A7G	86.8	14.4	72.5
	L7G	89.2	14.8	72.9
	K7G	79.2	20.2	67.5
	Q7G	66.1	23.1	45.8
pCG	A7G	82.7	27.0	59.1
	L7G	84.9	14.5	71.0
	K7G	87.8	9.9	74.6
	Q7G	89.1	26.5	61.2

CHAPTER V
CONCLUSION

At the beginning, we had the hypothesis regarding rice serine carboxypeptidase-like (OsSCPL) enzymes, in which we postulated that these proteins can be characterized as acyltransferase proteins and that they might play a role to transfer acyl group from glucose ester conjugates to anthocyanins to form more complex anthocyanins. However, based on the experiment that we have done, we discovered unexpected protein functions.

First, we could not detect either rice serine carboxypeptidase-like proteins OsSCPL2a or OsSCPL7 in a *P. pastoris* yeast expression system, but OsSCPL7 expression resulted in the release of alcohol oxidase 1 (AOX1) protein, which usually cannot be found in the medium. This is likely due to some particular stress caused by expression of the OsSCPL7 gene, although we could not detect the protein. The AOX1 protein could bind to immobilized metal affinity chromatography (IMAC) resin, thereby further confounding the search for the protein of interest (OsSCPL7 in this case). Although we did not investigate further, the AOX1 protein itself affected the structure of cyanidin-3-*O*-glucoside, in a way that was more consistent with oxidation than acylation, since the reaction product did not absorb at 520 nm.

Second, rice serine carboxypeptidase-like proteins were mainly found in the insoluble protein fraction when expressed with an *E. coli* bacterial system. Although we identified the purified protein to contain OsSCPL2a in the soluble fraction, the protein hit is considered to have a low score compared to some background proteins from *E. coli*, suggesting that the bacterial contaminants were more abundant than the desired protein. Although OsSCPL2 is closely related to the acyltransferases from other plants, it appears to act as a hydrolase to release 7-O-linked glucose from flavonoids instead of an acyltransferase that acylates those flavonoid glucosides. This conclusion is only tentative, due to the low purity of the protein that was assayed and the lack of ability to detect the activity in crude extracts for comparison to extracts from cells not producing the enzyme. In the future, it may be necessary to increase the solubility of protein, so that it can be better purified, in order to determine the reaction and substrate specificity more certainly.
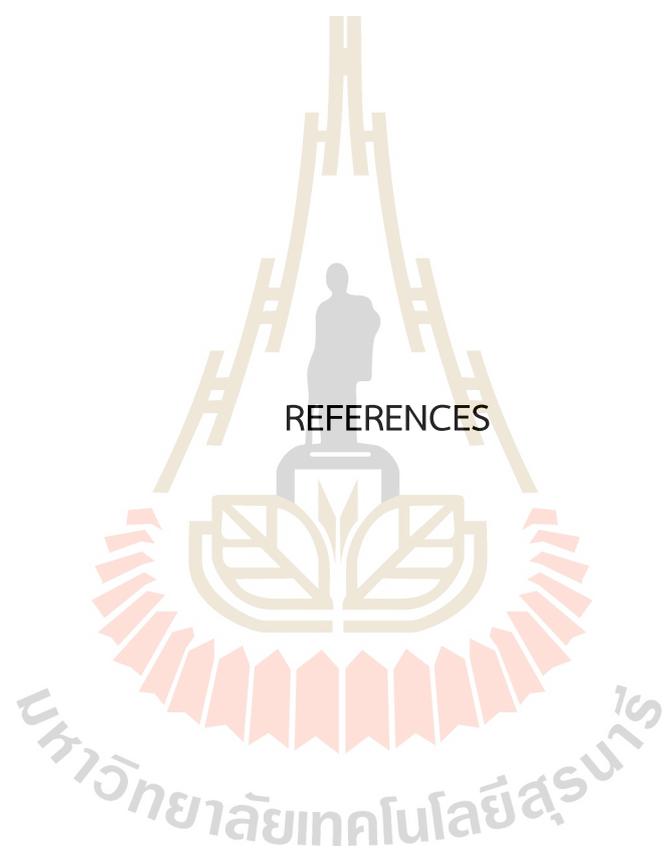
REFERENCES

REFERENCES

Al-Farsi, M.A., and Lee, C.Y. (2008). Optimization of phenolics and dietary fibre extraction from date seeds. *Food Chemistry. 108*(3), 977-985.

Armenteros, J.J.A., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology. 37*(4), 420-423.

Barnett, J. P., Scanlan, D. J., and Blindauer, C. A. 2012. Fractionation and identification of metalloproteins from a marine cyanobacterium. *Analytical and Bioanalytical Chemistry. 402*, 3371-3377.

Bontpart, T., Ferrero, M., Khater, F., Marlin, T., Vialet, S., Vallverdù-Queralt, A., Pinasseau, L., Ageorges, A., Cheynier, V. and Terrier, N. (2018). Focus on putative serine carboxypeptidase-like acyltransferases in grapevine. *Plant Physiology and Biochemistry. 130*, 356-366.

Bontpart, Tribaut, V. Cheynier, A. Ageorges and N. Terrier. (2015). BAHD or SCPL acyltransferasde? What a dilemma for acylation in the world of plant phenolic compounds. *New Phytologist. 208,* 695-707.

Bowen-Forbes, C.S., Zhang, Y., and Nair, M.G. (2010). Anthocyanin content, antioxidant, anti-inflammatory and anticancer properties of blackberry and raspberry fruits. *Journal of Food Composition and Analysis. 23,* 554-560.

Cairns, J.R.K., Mahong, B., Baiya, S. and Jeon, J.S. (2015). β-Glucosidases: multitasking, moonlighting or simply misunderstood? *Plant Science. 241*, 246-259.

Cereghino, J.L. and Cregg, J.M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews. 24*(1), 45-66.

Chan, M.K., Lim, S.K., Miswan, N., Chew, A.L., Noordin, R., and Khoo, B.Y. (2018). Expression of stable and active human DNA topoisomerase I in *Pichia pastoris*. *Protein Expression and Purification. 141*, 52-62.

Chang, A.Y., Chau, V.W., Landas, J.A. and Pang, Y. (2017). Preparation of calcium competent *Escherichia coli* and heat-shock transformation. *JEMI Methods. 1*, 22-25.

Choi, Y.J., Bourque, D., Morel, L., Groleau, D., and Míguez, C.B. (2006). Multicopy integration and expression of heterologous genes in *Methylobacterium*

extorquens ATCC 55366. *Applied and Environmental Microbiology. 72*(1), 753-759.

Ciarkowska, A., Ostrowski, M., and Jakubowska, A. (2018). A serine carboxypeptidase-like acyltransferase catalyzes synthesis of indole-3-acetic (IAA) ester conjugate in rice (*Oryza sativa*). *Plant Physiology and Biochemistry. 125*, 126-135.

Ciulu, M., Cádiz-Gurrea, M.d.l.L., and Segura-Carretero, A. (2018). Extraction and analysis of phenolic compounds in rice: a review. *Molecules. 23*, 2890.

D'Auria, J.C. (2006). Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology. 9,* 331-340.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research. 36*, 465-469

Duthie, G.G., Gardner, P.T., and Kyle, J.A. (2003). Plant polyphenols: are they the new magic bullet? *Proceedings of the Nutrition Society. 62*: 599-603.

Edgar, R .C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research. 32*(5), 1792-1797.

Francis, D.M. and Page, R. (2010). Strategies to optimize protein expression in *E. coli. Current Protocols in Protein Science. 61*(1), 5-24.

Fraser, C.M., Thompson, M.G., Shirley, A.M., Ralph, J., Schoenherr, J.A., Sinlapadech, T., Hall, M.C., and Chapple, C. (2007). Related Arabidopsis serine carboxypeptidase-like sinapoylglucose acyltransferases display distinct but overlapping substrate specificities. *Plant Physiology. 144*, 1986-1999.

Goffman, F. and Bergman, C. (2004). Rice kernel phenolic content and its relationship with antiradical efficiency. *Journal of the Science of Food and Agriculture. 84*, 1235-1240.

Gomes, L., Monteiro, G. and Mergulhão, F. (2020). The impact of IPTG induction on plasmid stability and heterologous protein expression by *Escherichia coli* biofilms. *International Journal of Molecular Sciences. 21*(2), 576.

Goufo, P. and Trindade, H. (2014). Rice antioxidants: phenolic acids, flavonoids, anthocyanins, proanthocyanidins, tocopherols, tocotrienols, γ-oryzanol, and phytic acid. *Food Science and Nutrition. 2*, 75-104.

Gupta, R. and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symposium on Biocomputing. 7*, 310-322.

Hashimoto, F., Tanaka, M., Maeda, H., Fukuda, S., Shimizu, K., and Sakata, Y. (2002). Changes in flower coloration and sepal anthocyanins of cyanic *Delphinium* cultivars during flowering. *Bioscience, Biotechnology, and Biochemistry. 66*, 1652-1659.

Hays, F. A., Roe-Zurz, Z. and Stroud, R. M. (2010). Overexpression and purification of integral membrane proteins in yeast. *Methods in Enzymology. 470*, 695-707.

He, K., Li, X., Chen, X., Ye, X., Huang, J., Jin, Y., Li, P., Deng, Y., Jin, Q., and Shi, Q. (2011). Evaluation of antidiabetic potential of selected traditional Chinese medicines in STZ-induced diabetic mice. *Journal of Ethnopharmacology. 137*, 1135-1142.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany. 62*, 2465-2483.

Hu, C., Zawistowski, J., Ling, W., and Kitts, D.D. (2003). Black rice (Oryza sativa L. indica) pigmented fraction suppresses both reactive oxygen species and nitric oxide in chemical and biological model systems. *Journal of Agricultural and Food Chemistry. 51*, 5271-5277.

Inan, M., Aryasomayajula, D., Sinha, J. and Meagher, M. M. (2005). Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. *Biotechnology and Bioengineering. 93*(4), 771-778.

Iqbal, S., Bhanger, M., and Anwar, F. (2005). Antioxidant properties and components of some commercially available varieties of rice bran in Pakistan. *Food Chemistry. 93*, 265-272.

Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends in Plant Science. 18*, 477-483.

Jovanovic, S.V., Steenken, S., Tosic, M., Marjanovic, B., and Simic, M.G. (1994). Flavonoids as antioxidants. *Journal of the American Chemical Society. 116*, 4846-4851.

Khoo, H.E., Azlan, A., Tang, S.T., and Lim, S.M. (2017). Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food and Nutrition Research. 61*, 1361779.

Kiess, M., Hecht, H. J. and Kalisz, H. M. (1998). Glucose oxidase from *Penicillium amagasakiense*. Primary structure and comparison with other glucose-methanol-choline (GMC) oxidoreductases. *European Journal of Biochemistry. 252*(1), 90-99.

Koch, C., Neumann, P., Valerius, O., Feussner, I. and Ficner, R. (2016). Crystal structure of alcohol oxidase from *Pichia pastoris*. *PLoS ONE. 11*(2). e0149846.

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends in Plant Science. 10*, 236-242.

Kogawa, K., Kazuma, K., Kato, N., Noda, N., and Suzuki, M. (2007). Biosynthesis of malonylated flavonoid glycosides on the basis of malonyltransferase activity in the petals of Clitoria ternatea. *Journal of Plant Physiology. 164*, 886-894.

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. *Journal of Agricultural and Food Chemistry. 63*, 9764-9769.

Kotchoni, S.O., Gachomo, E.W., Betiku, E. and Shonukan, O.O. (2003). A home-made kit for plasmid DNA mini-preparation. *African Journal of Biotechnology. 2*(4), 88-90.

Krygier, K., Sosulski, F., and Hogge, L. (1982). Free, esterified, and insoluble-bound phenolic acids. 1. Extraction and purification procedure. *Journal of Agricultural and Food Chemistry. 30*, 330-334.

Kumar, R. (2019). Simplified protocol for faster transformation of (a large number of) *Pichia pastoris* strains. *Yeast. 36*, 399-410.

Laokuldilok, T., Shoemaker, C.F., Jongkaewwattana, S. and Tulyathan, V. (2011). Antioxidants and antioxidant activity of several pigmented rice brans. *Journal of Agricultural and Food Chemistry. 59*(1), 193-199.

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. *The Plant Cell. 12*, 1295-1306.

Li, A.X. and Steffens, J.C. (2000). An acyltransferase catalyzing the formation of diacylglucose is a serine carboxypeptidase-like protein. *Proceedings of the National Academy of Sciences of the United States of America. 97*(12), 6902-6907.

Li, D., Zhang, Y., Liu, Y., Sun, R. and Xia, M. (2015). Purified anthocyanin supplementation reduces dyslipidemia, enhances antioxidant capacity, and prevents insulin resistance in diabetic patients. *The Journal of Nutrition. 145*(4), 742-748.

Li, W., Friel, J., and Beta, T. (2010). An evaluation of the antioxidant properties and aroma quality of infant cereals. *Food Chemistry. 121*, 1095-1102.

Liaqat, I., Ahmed, D., Saleem, A., Masih, R. and Chaudhery, R. (2018). Characterization of natural products from the peel of Lagenaria siceraria fruit using chromatographic techniques. *Acta Chromatographica. 30*(3), 191-194.

Lin, P. Y., and Lai, H.M. (2011). Bioactive compounds in rice during grain development. *Food chemistry. 127*, 86-93.

Lu, C., Li, Y., Cui, Y., Ren, J., Qi, F., Qu, J., Huang, H., and Dai, S. (2021). Isolation and functional analysis of genes involved in polyacylated anthocyanin biosynthesis in blue *Senecio cruentus*. *Frontiers in Plant Science. 12*, 253.

Luang, S., Cho, J.-I., Mahong, B., Opassiri, R., Akiyama, T., Phasai, K., Komvongsa, J., Sasaki, N., Hua, Y.-l., and Matsuba, Y. (2013). Rice Os9BGlu31 is a transglucosidase with the capacity to equilibrate phenylpropanoid, flavonoid, and phytohormone glycoconjugates. *Journal of Biological Chemistry. 288*, 10111-10123.

Matsuba, Y., Sasaki, N., Tera, M., Okamura, M., Abe, Y., Okamoto, E., Nakamura, H., Funabashi, H., Takatsu, M., and Saito, M. (2010). A novel glucosylation reaction on anthocyanins catalyzed by acyl-glucose–dependent glucosyltransferase in the petals of carnation and delphinium. *The Plant Cell. 22*, 3374-3389.

Mekapogu, M., Vasamsetti, B.M.K., Kwon, O.K., Ahn, M.S., Lim, S.H. and Jung, J.A. (2020). Anthocyanins in floral colors: Biosynthesis and regulation in chrysanthemum flowers. *International Journal of Molecular Sciences. 21*(18), p.6537.

Milkowski, C., Baumert, A., Schmidt, D., Nehlin, L., and Strack, D. (2004). Molecular regulation of sinapate ester metabolism in *Brassica napus*: expression of genes, properties of the encoded proteins and correlation of enzyme activities with metabolite accumulation. *The Plant Journal. 38*, 80-92.

Miyahara, T., Sakiyama, R., Ozeki, Y., and Sasaki, N. (2013). Acyl-glucose-dependent glucosyltransferase catalyzes the final step of anthocyanin formation in Arabidopsis. *Journal of Plant Physiology. 170*, 619-624.

Morton, L.W., Caccetta, R.A.A., Puddey, I.B., and Croft, K.D. (2000). Chemistry and biological effects of dietary phenolic compounds: relevance to cardiovascular disease. *Clinical and Experimental Pharmacology and Physiology. 27*, 152-159.

Mugford, S.T. and Milkowski, C. (2012). Serine carboxypeptidase-like acyltransferases from plants. *Methods in Enzymology. 516*, 279-297.

Mugford, S.T. and Osbourn, A. (2010). Evolution of serine carboxypeptidase-like acyltransferases in the monocots. *Plant Signaling and Behavior. 5*(2), 193-195.

Mugford, S.T., Qi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R.K., Papadopoulou, K., Melton, R., Philo, M., and Sainsbury, F. (2009). A serine carboxypeptidase-like

acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *The Plant Cell. 21*, 2473-2484.

Nishizaki, Y., Yasunaga, M., Okamoto, E., Okamoto, M., Hirose, Y., Yamaguchi, M., Ozeki, Y., and Sasaki, N. (2013). *p*-Hydroxybenzoyl-glucose is a zwitter donor for the biosynthesis of 7-polyacylated anthocyanin in *Delphinium*. *The Plant Cell. 25*, 4150-4165.

Noda, N. (2006). Molecular cloning of 1-O-acetylglucose dependent anthocyanin aromatic acyltransferase in ternatin biosynthesis of butterfly pea *(Clitoria ternatea)*. *Plant Cell Physiol. 47*, s109.

Oberg, K., Chrunyk, B.A., Wetzel, R. and Fink, A.L. (1994). Native-like secondary structure in interleukin-1. beta. Inclusion bodies by attenuated total reflectance FTIR. *Biochemistry. 33*(9), 2628-2634.

OECD/FAO (2018). OECD/FAO Agricultural Outlook 2018-2027. http://www.fao.org/publications/oecd-fao-agricultural-outlook/2018-2027/en/. Accessed on 15-07-2021.

Parr, A.J., and Bolwell, G.P. (2000). Phenols in the plant and in man. The potential for possible nutritional enhancement of the diet by modifying the phenols content or profile. *Journal of the Science of Food and Agriculture. 80*, 985-1012.

Patras, A., Brunton, N.P., O'Donnell, C., and Tiwari, B.K. (2010). Effect of thermal processing on anthocyanin stability in foods; mechanisms and kinetics of degradation. *Trends in Food Science and Technology. 21*, 3-11.

Qiu, Y., Liu, Q., and Beta, T. (2010). Antioxidant properties of commercial wild rice and analysis of soluble and insoluble phenolic acids. *Food Chemistry. 121*, 140-147.

Rechner, A.R. and Kroner, C. (2005). Anthocyanins and colonic metabolites of dietary polyphenols inhibit platelet function. *Thrombosis Research. 116*, 327-334.

Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology. 5*, 172.

Ross, J.A. and Kasum, C.M. (2002). Dietary flavonoids: bioavailability, metabolic effects, and safety. *Annual Review of Nutrition. 22*, 19-34.

Samad, N. (2015). Rice bran oil prevents neuroleptic-induced extrapyramidal symptoms in rats: Possible antioxidant mechanisms. *Journal of Food and Drug Analysis. 23*, 370-375.

Sasaki, N., Nishizaki, Y., Ozeki, Y., and Miyahara, T. (2014). The role of acyl-glucose in anthocyanin modifications. *Molecules. 19*, 18747-18766.

Sato, Y., Takehisa, H., Kamatsuki, K., Minami, H., Namiki, N., Ikawa, H., and Nagamura, Y. (2013). RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. *Nucleic Acids Research. 41*(1), 1206-1213.

Scalbert, A. and Williamson, G. (2000). Dietary intake and bioavailability of polyphenols. *The Journal of Nutrition. 130*, 2073S-2085S.

Setyaningsih, W., Saputro, I.E., Palma, M., Barroso, C.G. (2016). Profile of individual phenolic compounds in rice (*Oryza sativa*) grains during cooking processes. AIP Conference Proceedings. *AIP Publishing LLC.* p. 130012.

Sevchenko, A., Wilm, A., Vorm, O. and Mann, M. (1996). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical Chemistry. 68*, 850-858.

Shental-Bechor, D. and Levy, Y. (2009). Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Current Opinion in Structural Biology. 19*(5), 524-533.

Shirley, A.M., McMichael, C.M., and Chapple, C. (2001). The sng2 mutant of Arabidopsis is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose: choline sinapoyltransferase. *The Plant Journal. 28*, 83-94.

Silaban, S., Gaffar, S., Simorangkir, M., Maksum, I.P. and Subroto, T. (2018). Effect of IPTG concentration on recombinant human prethrombin-2 expression in *Escherichia coli* BL21(DE3) ArcticExpress. In *IOP Conference Series: Earth and Environmental Science. 217*(1), 012039. IOP Publishing.

Statistica. (2021). Top countries of rice export. Accessed on 28 July 2021. https://www.statista.com/statistics/255945/top-countries-of-destination-for-us-rice-exports-2011/. Accessed on 11-07-2021.

Stehle, F., Stubbs, M.T., Strack, D. and Milkowski, C. (2008). Heterologous expression of a serine carboxypeptidase-like acyltransferase and characterization of the kinetic mechanism. *The FEBS Journal. 275*(4), 775-787.

Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification. 41(1)*, 207-234.

Tian, S., Nakamura, K., and Kayahara, H. (2004). Analysis of phenolic compounds in white rice, brown rice, and germinated brown rice. *Journal of Agricultural and Food Chemistry. 52*, 4808-4813.

Vichapong, J., Sookserm, M., Srijesdaruk, V., Swatsitang, P., and Srijaranai, S. (2010). High performance liquid chromatographic analysis of phenolic compounds and their antioxidant activities in rice varieties. *LWT-Food Science and Technology. 43*, 1325-1330.

Wall, M. J., Crowell, A. M. J., Simms, G. A., Liu, F. and Doucette, A. A. (2011). Implications of partial tryptic digestion in organic-aqueous solvent systems for bottom up proteome analysis. *Analytica Chimica Acta. 703*, 194-203.

Walter, M., and Marchesan, E. (2011). Phenolic compounds and antioxidant activity of rice. *Brazilian Archives of Biology and Technology. 54*, 371-377.

Wang, F., Wang, X., Yu, X., Fu, L., Liu, Y., Ma, L., and Zhai, C. (2015). High-level expression of endo-β-N-acetylglucosaminidase H from Streptomyces plicatus in *Pichia pastoris* and its application for the deglycosylation of glycoproteins. *PLoS One. 10(3),* p.e0120458.

Weier, D., Mittasch, J., Strack, D., and Milkowski, C. (2008). The genes BnSCT1 and BnSCT2 from *Brassica napus* encoding the final enzyme of sinapine biosynthesis: molecular characterization and suppression. *Planta. 227*, 375-385.

Yawadio, R., Tanimori, S., and Morita, N. (2007). Identification of phenolic compounds isolated from pigmented rices and their aldose reductase inhibitory activities. *Food Chemistry. 101*, 1616-1625.

Yonekura-Sakakibara, K. and Hanada, K. (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. *The Plant Journal. 66*, 182-193.

Yoshida, K., Mori, M., and Kondo, T. (2009). Blue flower color development by anthocyanins: from chemical structure to cell physiology. *Natural Product Reports. 26*, 884-915.

Zawistowski, J., Kopec, A., and Kitts, D.D. (2009). Effects of a black rice extract (Oryza sativa L. indica) on cholesterol levels and plasma lipid parameters in Wistar Kyoto rats. *Journal of Functional Foods. 1*, 50-56.

Zhou, A., Jiang, X. and Xu, X. (1997). Improved alkaline lysis method for rapid isolation of plasmid DNA. *Biotechniques. 23*, 592-594.

Zhou, Z., Robards, K., Helliwell, S., and Blanchard, C. (2004). The distribution of phenolic acids in rice. *Food Chemistry. 87*, 401-404.

APPENDICES
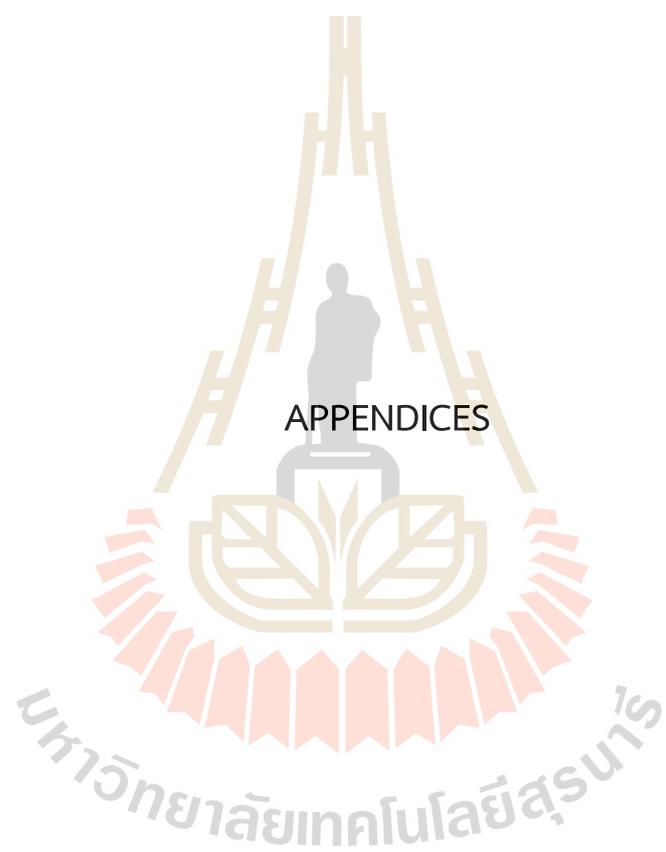
APPENDIX A

RESULTS OF AOX1 AND OSSCPL2A PROTEIN ACTIVITY TOWARD ANTHOCYANIN AND FLAVONOID

1.   Result of AOX1 activity by UHPLC chromatogram

Table 1. UHPLC result of AOX1 activity towards Cy3G with the presence of acyl donors.

Abs (nm)	Acyl donor	Peak area of acceptor (%)		Peak area of product (%)
		(-) AOX1	(+) AOX1	
360	pHBG	29.7	9.5	11.6
	FG	18.1	7.0	9.6
	pCG	24.0	7.2	11.3
	VG	29.7	10.0	15.5
	-	30.9	11.3	11.7
520	pHBG	92.6	67.8	ND
	FG	91.9	65.3	ND
	pCG	90.3	63.7	ND
	VG	89.9	66.2	ND
	-	95.2	77.7	ND

Note: pHBG, *p*-hydroxybenzoyl-glucose; FG, feruloyl-glucose; pCG, *p*-coumaroyl-glucose; VG, vanillyl-glucose; ND, not detected.

Table 2. UHPLC result of AOX1 activity towards P3G with the presence of acyl donors.

Abs (nm)	Acyl donor	Peak area of acceptor (%)		Peak area of product (%)
		(-) AOX1	(+) AOX1	
360	pHBG	41.3	44.0	ND
	FG	37.0	35.7	1.4
	pCG	45.0	44.5	ND
	VG	47.3	47.2	1.7
520	pHBG	95.9	95.5	ND
	FG	95.7	95.2	ND
	pCG	95.9	95.5	ND
	VG	95.8	95.3	ND

Note: pHBG, *p*-hydroxybenzoyl-glucose; FG, feruloyl-glucose; pCG, *p*-coumaroyl-glucose; VG, vanillyl-glucose; ND, not detected.

2.   UHPLC specification

Table 3. UHPLC parameter condition for AOX1 and OsSCPL2a enzymatic assay.

Specification	Description
Column	2.1x150 mm, SC-C18, Agilent-USA
Solvent	A: 0.2 % formic acid
	B: 100 % acetonitrile
	Wash solution: 100 % methanol
Standard concentration	300 µM
Filter	0.2 µm syringe filter CA-CN 13 mm
Volume injections	Cy3G: 10 µL
	P3f: 2 µL
Pressure	540 barr
Column temperature	40 °C
Signal and absorbance	LCDAD, wavelength: 260, 280, 360, 520, 530 nm

Table 4. Solvent gradient of UHPLC.

Time (min)	Solvent (%)		Flow (mL/min)
	A	B	
0	95	5	
2	95	5	
13	50	50	
14	30	70	0.2
16	0	100	
20	95	5	
25	95	5	

3.　　Result of OsSCPL2a activity by UHPLC chromatogram

Table 5. UHPLC result of OsSCPL2a activity toward 7-glycosylated flavonoid with the presence of acyl donors and was measured at 360 nm.

Acyl donor	Acyl acceptor	Peak area of acceptor (%)		Peak area of product (%)
		(-) OsSCPL2a	(+) OsSCPL2a	
FG	A7G	75.3	23.4	46.5
	L7G	75.2	32.1	41.7
	K7G	73.3	34.2	40.4
	Q7G	75.3	41.8	34.3
VG	A7G	82.9	33.2	51.9
	L7G	85.9	31.9	54.4
	K7G	88.2	41.8	45.2
	Q7G	ND	ND	ND

Note: FG, feruloyl-glucose; VG, vanillyl-glucose, A7G, apigenin-7-*O*-glucoside; L7G, luteolin-7-*O*-glucoside; K7G, kaempferol-7-*O*-glucoside; Q7G, quercetin-7-*O*-glucoside; ND, not detected.

APPENDIX B

SCPL EXPRESSION IN *E. COLI* WITH VARIOUS TREATMENT
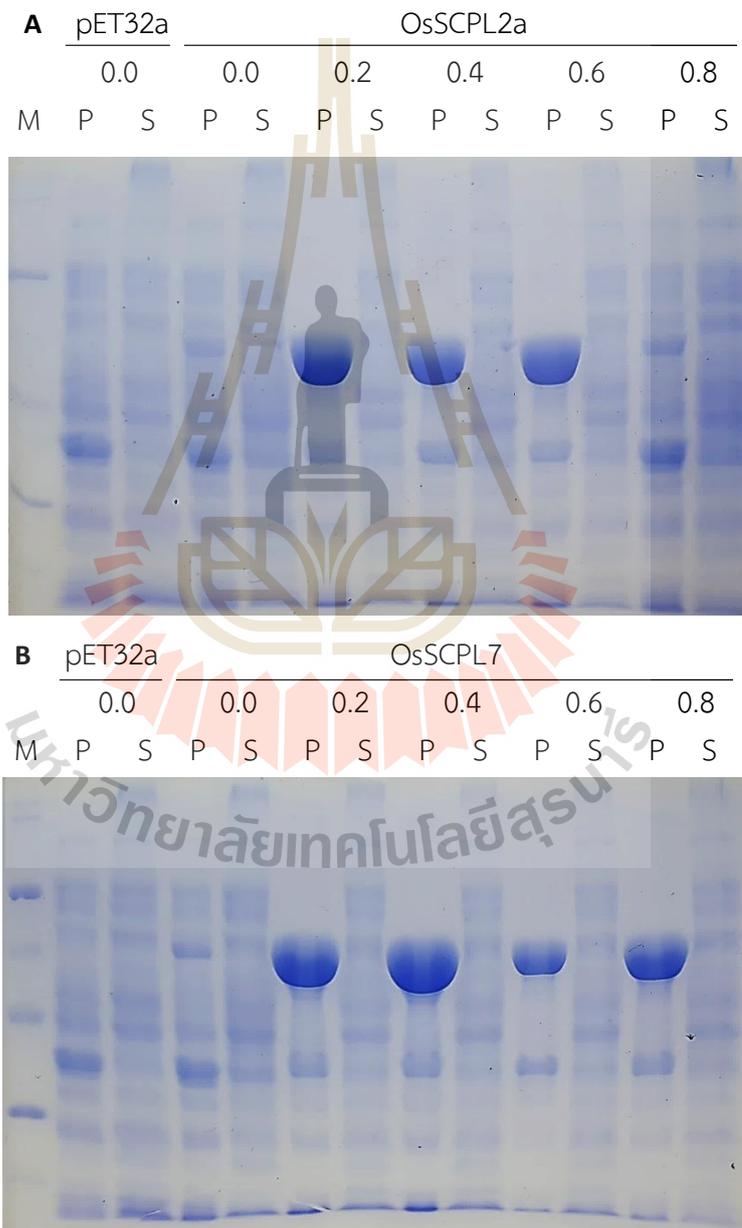
1. SCPLs expression in BL21(DE3)

Figure 1.OsSCPLs expression in E. coli strain BL21(DE3) competent cells with various concentration of IPTG. M, marker; P, pellet; S, supernatant.

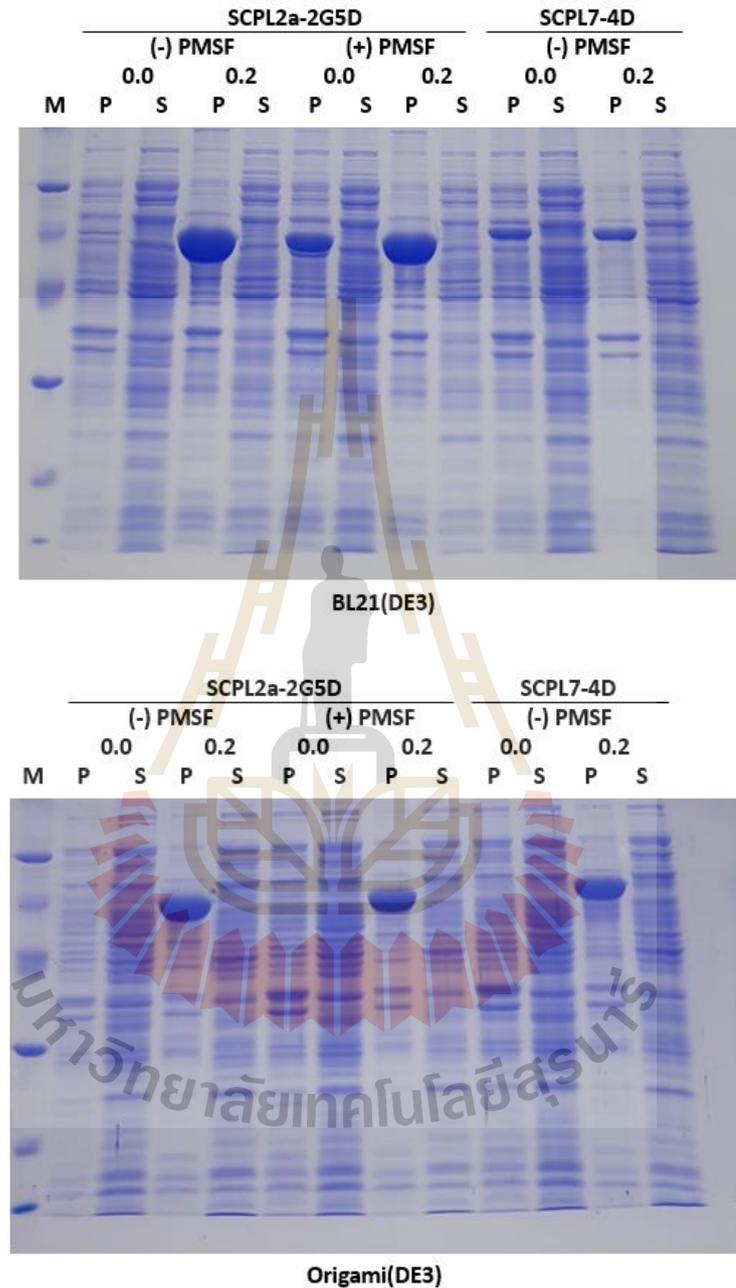## 3. Expression of OsSCPL2A_2G5D and OsSCPL7_4D

Figure 3 The expression of OsSCPL2a_2G5D (top) and OsSCPL7_4D (bottom) in competent cell BL21(DE3) and Origami(DE3), treated with and without PMSF in extraction buffer.

CURRICULUM VITAE

Name   Ms. Istiftakhun Nikmah
Date of Birth   07 February 1995
Place of Birth   Tuban, East Java, Indonesia
Education   Bachelor of Science (2013-2017)
                       Department of Biology, Faculty of Mathematics and
                       Natural Science, University of Brawijaya, Malang,
                       Indonesia.

Publications

1. Nikmah, Istiftakhun and F. Fatchiyah. 2017. Identification of Fingerprints Pattern in Mixed Family of Chinese-Javanese Ethnic. *Journal of Tropical Life Science*. 7(3): 263-267.

2. Nikmah, Istiftakhun, Kongdin, M., Hua, Y., Roytrakul, S., Fatchiyah, F., and Ketudat-Cairns, J.K. 2021. Release of alcohol oxidase 1 (AOX1) from *Pichia pastoris* during heterologous expression of rice enzymes. *Asia-Pacific Journal of Science and Technology*. 27(2): 1-10.

Grants and Fellowships   Thailand Research Fund International Research Network
                             IRN62W0004