# THE EXPRESSION AND DETECTION OF RICE SFR2 IN *ESCHERICHIA COLI* AND *PICHIA PASTORIS*

**Sasiprapa Kanjanawattana**



**A Thesis Submitted in Partial Fulfillment of the Requirements for the**

**Degree of Master of Science in Biotechnology**

**Suranaree University of Technology**

**Academic Year 2010**

# การศึกษาการแสดงออกและการตรวจสอบโปรตีน SFR2 จากข้าวใน *Escherichia coli* และ *Pichia pastoris*

นางสาวศศิประภา กาญจนวัฒนา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีชีวภาพ
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2553

# THE EXPRESSION AND DETECTION OF RICE SFR2 IN *ESCHERICHIA COLI* AND *PICHIA PASTORIS*

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

Thesis Examining Committee

_____

(Assoc. Prof. Dr. Montarop Yamabhai)

Chairperson

_____

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

Member (Thesis Advisor)

_____

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

Member

_____ _____

(Dr. Wut Dankittikul) (Asst. Prof. Dr. Suwayd Ningsanond)

Acting Vice Rector for Academic Affairs Dean of Institute of Agricultural Technology

การศึกษาการแสดงออกและการตรวจสอบโปรตีน SFR2 จากข้าวใน *Escherichia coli* และ *Pichia pastoris* (THE EXPRESSION AND DETECTION OF RICE SFR2 IN *ESCHERICHIA COLI* AND *PICHIA PASTORIS*) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร. มารินา เกตุทัต-คาร์นส์, 126 หน้า.

งานวิจัยนี้เป็นการศึกษาการแสดงออกของโปรตีน SFR2 ของข้าวใน *E. coli* และ *P. pastoris* ส่วนทางด้านปลายเอ็นและปลายซีของโปรตีน SFR2 ถูกสร้างใน *E. coli* ด้วยระบบของ pET32a ผลการทดลองพบว่ามีเพียงกรดอะมิโน 260 ตัวทางด้านปลายซีและกรดอะมิโน 174 ตัวทางด้านปลายเอ็นเท่านั้นสามารถแสดงออกใน *E. coli* ได้ส่วนของ PEST sequence ที่พบในโปรตีน SFR2 ซึ่งเป็นลำดับของกรดอะมิโนที่มีการเหนี่ยวนำให้เกิดการย่อยสลายของโปรตีนโดยโปรติเอสถูกทำให้เปลี่ยนแปลงโดยเปลี่ยนกรดอะมิโนกลูตามิกเป็นกลูตามีนโดยวิธีพีซีอาร์มิวตาเจนเนซิส อย่างไรก็ตามเมื่อทำให้มีการแสดงออกของโปรตีนพบว่า SFR2 ที่มีการมิวเทตไม่สามารถสร้างโปรตีนได้ ดังนั้นจึงเปลี่ยนระบบจาก pET32a ไปเป็น pCold I และทำให้มีการแสดงออกใน *E. coli* เช่นเดิม ผลการทดลองพบว่าโปรตีน SFR2 ไม่สามารถแสดงออกได้โดยใช้ pCold I เช่นกัน เนื่องจากข้อจำกัดของการย้อมสี Coomassie blue ในการตรวจสอบโปรตีนบน SDS-PAGE จึงได้ทำการตรวจสอบโปรตีนโดยเทคนิค western blot เพื่อตรวจสอบหาโปรตีน SFR2 โดยการใช้โปรตีนส่วนปลายซีของ SFR2 เป็นแอนติเจนเพื่อสร้างแอนติเซรัม ผลการทดลองพบการแสดงออกของโปรตีน SFR2 แต่มีในปริมาณเพียงเล็กน้อย ดังนั้นระดับอาร์เอ็นเอของ SFR2 จึงถูกตรวจสอบโดยใช้เทคนิค Northern blot พบว่า ระดับอาร์เอ็นเอของส่วนทางด้านปลาย 5' (ปลายเอ็นของโปรตีน) แสดงความสัมพันธ์กับผลการทดลองที่แสดงบน western blot คือ ระดับอาร์เอ็นเอน้อยจึงส่งผลให้มีการแสดงออกของโปรตีนที่น้อยด้วย อย่างไรก็ตามผลการทดลองของระดับอาร์เอ็นเอทางปลาย 3' (ปลายซีของโปรตีน) ไม่ได้แสดงออกในทิศทางเดียวกัน *P. pastoris* ถูกใช้เป็นเซลล์เจ้าบ้านเพื่อให้มีการแสดงออกของโปรตีน SFR2 อย่างไรก็ตามผลการทดลองใน SDS-PAGE ไม่ได้บ่งบอกอย่างชัดเจนว่ามีการแสดงออกของ SFR2 หรือไม่

สาขาวิชาเทคโนโลยีชีวภาพ ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2553 ลายมือชื่ออาจารย์ที่ปรึกษา_____

SASIPRAPA KANJANAWATTANA : THE EXPRESSION AND DETECTION OF RICE SFR2 IN *ESCHERICHIA COLI* AND *PICHIA PASTORIS*. THESIS ADVISOR : ASSOC. PROF. MARIENA KETUDAT-CAIRNS, Ph.D., 126 PP.

SFR2/PROTEIN EXPRESSION/PROTEIN DETECTION

The rice sensitive to freezing2 (*SFR2*) gene was cloned and expressed in *E. coli* and *P. pastoris*. The SFR2 proteins with N-terminal and C-terminal deletions were expressed in *E. coli* with the pET32a system. Results indicated that only 260 amino acids of the C-terminal and 174 amino acids of the N-terminal regions could be produced in *E. coli*. A PEST sequence, a region known to target its protein to proteolysis was found in the rice SFR2 protein. The PEST sequence in SFR2 was mutated by changing glutamic acids to glutamine. Then the mutated gene was expressed in *E. coli*. However, Coomassie blue stained SDS-PAGE could not detect any expressed SFR2 protein. The expression vector was changed to pCold I and again the *SFR2* gene deletion were expressed in *E. coli*. The results still showed no detectable protein on SDS-PAGE. Because of the limitation of Coomassie blue staining, western blot analysis was done. The C-terminal region of SFR2 protein (28_SFR2) was used as antigen to produced rabbit antisera. With this antisera, it was shown that the SFR2 protein could be expressed in *E. coli*, but only in very small amounts which were not detected by Coomassie blue stained SDS-PAGE. To determine the reasons, the RNA levels were investigated. The amount of RNA of the 5's region, encoding the N-terminal region, correlated with the western blot results that the lower RNA levels showed low protein levels. In contrast, the RNA encoding the C-terminal region did not show the same correlation.

*P. pastoris* was also used as expression host but no detectable protein can be found on SDS-PAGE even with the SFR2 with codons optimized for *P. pastoris* expression.



School of Biotechnology Student's Signature_____

Academic Year 2010 Advisor's Signature_____

# ACKNOWLEDGEMENT

I am heartily thankful to my advisor, Assoc. Prof. Dr. Mariena Ketudat-Cairns, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. I would like to thank to Assoc. Prof. Dr. Aoyagi Hideki to provide ArcticExpress(DE3) RIL *E. coli* host strain.

Thanks to all members of SUT School of Biotechnology for their help. Special thanks go to Mr. Surachai Rattanasuk, who prepare antisera, and Miss Chanida Kupradit, who taught me about northern blot technique. Miss Darawan Ruamkuson, Miss Wanwisa Suphap, Mr. Sumeth Imsoonthornruksa, Dr. Wipaporn Wanthanalert and Miss Pongpan Songwattana, are thank for their consultations and encouragements.

This thesis work and M.Sc. scholarship were supported by SUT-NRCT grant.

Finally, I would like to extend my gratitude to my family for their love and understanding.

Sasiprapa Kanjanawattana

# CONTENTS

# CONTENTS (Continued)

**Page**

# CONTENTS (Continued)

**Page**

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF FIGURES (Continued)

**Figure** **Page**

# LIST OF FIGURES (Continued)

**Figure** **Page**

# LIST OF FIGURES (Continued)

**Figure** **Page**

# CHAPTER I

# INTRODUCTION

## 1.1 Introduction

Rice (*Oryza sativa*) is one of the most important food crops for humans. More than 500 million tons are produced annually worldwide (Goff *et al.*, 2002). Rice has important relationships with other cereal species and is a model plant for the grasses. Rice was the first crop plant to have its genome sequenced (Jung *et al.*, 2007). Rice is an ideal system to study factors affecting growth and response to stress of grain crops, since genes of interest can be identified and cloned directly from their sequences.

The human population is increasing at an alarming rate, whereas at the same time agricultural productivity is decreasing due to the effect of various environmental problems. In particular, cold stress is a serious threat to the sustainability of crop yields. Indeed, cold stress can lead to major crop losses. Phenotypic symptoms in response to cold stress include poor germination, stunted seedlings, yellowing of leaves (chlorosis), reduced leaf expansion and wilting, which may lead to death of tissue (necrosis). Cold stress also severely hampers the reproductive development of plants. The major negative effect of cold stress is that it induces severe membrane damage. This damage is largely due to the acute dehydration associated with freezing during cold stress (Yadav, 2010).

The *sensitive to freezing2* gene (*SFR2*) belongs to β-glycosidase family 1 (GH1). It has been found in many plant species, such as *Arabidopsis* (At), maize, wheat, *Glycin Max* (GM), *Lycopersv esculentum* (Le), *Pinus taeda* (Pt), sorghum and

barley (Opassiri *et al.*, 2006). The highly conserve peptide active site motifs, TFNEP and I/VTENG of family 1 β-glycosidase have been found in AtSFR2, LeSFR2, GmSFR2 and PtSFR2. SFR2 is more closely related to bacterial glucosidases than other members of GH1 family from plants (Opassiri *et al.*, 2006; Fourrier *et al.*, 2008). This protein is essential for freezing tolerance in *Arabidopsis* (Thorlby *et al.*, 2004) and localized to the chloroplast outer envelope membrane (Fourrier *et al.*, 2008). The *SFR2* gene encodes a galactolipid remodeling enzyme to transfers galactosyl residues from the abundant monogalactolipid to different galactolipid acceptors forming oligogalactolipids and diacylglycerol, which is further converted to triacylglycerol to change ratio of bilayer forming to non-bilayer forming membrane lipids to stabilize the membrane during freezing (Moellering *et al.*, 2010).

*Escherichia coli* is the most frequently used host for heterologous gene expression. However, high-level production of functional proteins in *E. coli,* especially those from eukaryotic sources, has often been difficult (Weickert *et al.*, 1996). There have been many proteins from eukaryotic species that express at a low level or are not expressed at all in *E. coli*. Many reasons for this phenomenon have been explained, for example not proper codon usage (Wu *et al.*, 2004; Zhou *et al.*, 2004), codon position at the initiation site (Nishikubo *et al.*, 2005; Pai and Murugan, 2008), AT/GC ratio, mRNA secondary structure and mRNA stability (Paulus *et al.*, 2004; Zhang *et al.*, 2006). SFR2 is one of the protein that has not been able to be expressed in *E. coli*. One or more of these reasons may be the cause of lack of SFR2 expression in *E. coli*.

In this work, we systemically delete the N-terminal parts and portion of the C-terminal parts of the *SFR2* protein coding gene and tried to produce the truncated recombinant rice SFR2 proteins in *E. coli*. The PEST sequence a kind of sequence

known to target protein to degradation, found in this protein was also mutated to test whether this would allow the mutated protein to express in *E. coli*. An *SFR2* gene synthesized with codons optimized for *Pichia pastoris* was also expressed in *E. coli* and *P. pastoris*. Western blot analysis using a rabbit polyclonal antibody against the C-terminal region of this protein as antisera was performed and small amount of the SFR2 proteins were detected. Northern blot analysis using two different probes to detect the RNA expression showed that small amounts of RNA could be detected in *E. coli* in which no expressed SFR2 proteins were detected on Coomassie blue stained SDS-PAGE.

## 1.2 Research objectives

1. To determine the parts of the *SFR2* gene or protein that results in negligible expression in *E. coli* and *P. pastoris*.
2. To mutate the PEST sequence region in the *SFR2* gene.
3. To express and detect the deleted and mutated SFR2 protein and RNA in *E. coli*.
4. To clone and express the codon optimized SFR2 gene in *E. coli* and *P. pastoris*.

# CHAPTER II

# LITERATURE REVIEWS

## 2.1 RICE

Cereal crops constitute more than 60% of the total worldwide agricultural production. Rice, wheat, and maize are the three most important cereals. More than 500 million tons of each are produced annually worldwide (Goff *et al.*, 2002). The total production of rice paddy area is about 154 million hectares and the annual production of rice is approximately 731 million tons (Zhao *et al.*, 2010). Thailand is a world class rice producer and exporter, which cultivates approximately 66 million tons per year (Thai rice exporter association., 2010). Most rice grown is consumed directly by humans, and about one-third of the world population depends on rice for more than 50% of caloric intake (Goff *et al.*, 2002). Most rice cultivars can be placed within two subspecies of rice: *Oryza sativa* ssp. *japonica* and *Oryza sativa* ssp. *indica*, which differ in physiological and morphological traits (Jung *et al.*, 2007). Rice serves as a model species in cereal genome research (Zhao *et al.*, 2004). The genome of *Oryza sativa L.* is 389 MB, 99% of which has been sequenced (IRGSP, 2005; Jung *et al.*, 2007). A total of 37,544 non-transposable-element-related protein-coding genes have been identified, of which 71% had a putative homologue in *Arabidopsis*. In a reciprocal analysis, 90% of the *Arabidopsis* proteins had a putative homologue in the predicted rice proteome. Twenty-nine percent of the 37,544 predicted genes appear to all in clustered gene families. Genome-wide gene-indexed mutant collections, rice gene expression microarrays and genome tilling array platforms are new methods

used to study rice gene function. An intensive effort to analyze the function of the ten of thousands of predicted rice genes is now being undertaken. Integration of data from diverse transcriptomic, proteomic and computational approaches is needed to understand the function of these genes and of regulatory regions in non-coding RNAs and other newly or poorly annotated regions of the genome. Cloning of the genes and recombinantly expressing and characterizing their gene products can also be done to study some of the interesting gene in the rice genome.

## 2.2 β-glucosidase

β-glucosidases (EC 3.2.1.21) are enzymes widely found in various living organisms such as animals, plants, fungi, eubacteria and archaea (Opassiri *et al.*, 2003; Byeon *et al.*, 2005). These enzymes hydrolyze β-O-glycosidic bond between the reducing side of glucose and aryl or akyl aglycone or oligosaccharide to release glucose and an aglycone. Nakano *et al*. (1998) reported that β-glucosidase from *Clavibacter michiganense* can also hydrolyze the ester linkage in steviol glycoside, but it was not as fast as hydrolyzing a glucosidic linkage. The β-glucosidases have been characterized to date fall predominantly in glycosyl hydrolase (GH) families 1 and 3. They also found in GH5, 9, 30 and 116 (Ketudat-Cairns and Esen., 2010). Most plant β-glucosidases characterized belong to glycosyl hydrolase family 1 (Opassiri *et al.*, 2006). β-glucosidases play roles in many biological processes, such as degradation of cellulosic biomass, hydrolysis of glycolipids, cyanogenesis and modification of secondary metabolites (Hsieh and Graham, 2001). In plants, β-glucosidases play roles in phytohormone regulation such as ABA metabolism, the hydrolysis of conjugated gibberellins, the conversion of storage forms of cytokinin to

active forms (Dietz *et al*., 2000). Moreover this enzyme involve in stress metabolism, resistance to some phytopathogens and herbivores to cyanogenesis (Esen, 1992) and defense mechanism against microbes, insects or parasitic plants (Hsieh and Graham, 2001). Furthermore, β-glucosidase has a function in lignification, for example, the *BGLU45* and *BGLU46* gene products in *Arabidopsis thaliana* hydrolyze monolignol glucosides to release monolignols that polymerize into lignin (Escamilla-Trevino *et al.*, 2006). Other functions of β-glucosidases in plants are production of aromatic volatiles, activation of metabolic intermediate by releasing glucose blocking group from the inactive glucoside (Opassiri *et al.*, 2006), floral development, and pigmentation (Hsieh and Graham, 2001). To achieve specificity for these various functions, β-glucosidases must differentially bind to a wide variety of aglycones, in addition to the glucose of the substrate (Opassiri *et al.*, 2006).

β-glucosidase can also be used in industry. The immobilized β-glucosidase from *Candida molischiana* 35M5N was used to hydrolyze flavor compounds to enhance wine and fruit juices flavor (Gueguen *et al.*, 1996). Recently, there were some reports that β-glucosidase can also be used to reduce the cost to produce ethanol from plant material or lignocellulose by the production of the cellulolytic enzyme via microoraganism or consolidated bioprocessing (CBP). The conversion of the resulting sugar to the desired product is done in single process step. β-glucosidase genes of fungal origin were isolated and heterologously expressed in *Saccharomyces cerevisciae* to enable growth on the disaccharide cellobiose as a sole carbon source. β-glucosidase genes were fused to the secretion signal of the *Trichoderma reesei xyn*2 gene and a *S. cereviciae* PGK1 promoter and terminator to control transcription. They found the recombinant *S. cerevisiae* strian was able to effectively utilize cellobiose as

its sole carbon source (Rooyen *et al.*, 2005).

## 2.3 The *Sensitive to Freezing2* gene (*SFR2* gene)

Many plants increase freezing tolerance upon exposure to low non-freezing temperatures, a phenomenon known as cold acclimatization (Thomashow, 1999). The freezing tolerance in plant is not a "constitutive" but is induced in response to low temperature (Thomashow, 1998). The details of the mechanism that permit freezing tolerance are not well understood. Several processes that can occur in the development of freezing are (a) stabilization of membranes (b) accumulation of sugars, other osmolytes and antifreeze proteins and (c) multiple changes in gene expression (Buchanan *et al.*, 2000). However, one important conclusion is that cold acclimation includes the expression of certain cold induced genes that function to stabilize membranes against freeze-induced injury (Thomashow, 1999). Freezing stress and dehydration stress are closely related. Freezing causes the ice crystals in apoplast (the space within the plant tissue that is exterior to the living cells). The plant cell walls and the external membranes of the protoplasts form a barrier that prevents ice from growing into the interiors of the cells. The apoplastic ice attracts water from the living cells, in effect acting as a desiccant distributed throughout the plant tissues. It is this loss of water which is the main injurious factor of frost, as it is in drought. As water is lost from the protoplast, whether due to freezing or dehydration, the solutes in the protoplast become concentrated, as do the solutes in the residual liquid phase of the apoplast. Elevated levels of certain solutes, particularly inorganic salts, can be injurious to the cell. The importance of understanding and controlling tolerance to abiotic stresses in plants have been highlighted in agriculture, where significant

worldwide economic losses are suffered as a result of unpredictable frosts (Warren *et al.*, 2003).

The *Sensitive to freezing2* gene belongs to glycoside hydrolase family 1 β-glycosidases. This protein is essential for freezing tolerance in *Arabidopsis*. The SFR2 protein contains the highly conserve peptide active site motifs, TFNEP and I/VTENG (Thorlby *et al.*, 2004). Thorlby *et al*. (2004) compared the wild type gene and the mutant gene (*sfr2-1*), the transcription level of *SFR2* and *sfr2-1* when exposed to the low temperature for 24 hours were not different. They concluded that *SFR2* is constitutively expressed rather than inducible. They also determined the localization by using a translation fusion of a β-glucuronidase (GUS) reporter to the 31$^{st}$ codon of *SFR2* gene. The results indicated that the reporter protein are in the most aerial organ types (hypocotyls, cotyledons, stems, leaves, pedicels, sepals, anthers and pistils) but not in petal or filament. Transcription of *SFR2* was detected in all green tissues and at a much lower level in roots, which implied that SFR2 protein plays a role in freezing protection of photosynthetic tissues or possibly a more specific role in a plastid protection. In 2008, Fourrier *et al*. found that SFR2 in *Arabidopsis* was localized to chloroplast outer envelope membrane by analysis of transgenic plants expressing SFR2 fusions to GFP by confocal microscopy and by immunological analysis of isolated chloroplasts treated with thermolysin protease. They suggested a role for this protein as protecting the chloroplast because the chloroplasts of the *sfr2* mutant showed clear evidence of rapid damage after a freezing episode under electron microscope. The mechanism of SFR2 studied in *Arabidopsis* by Moellering et al., 2010 suggested that *SFR2* gene encodes a galactolipid remodeling enzyme of the outer chloroplast envelope membrane. SFR2 processively transfers galactosyl residues from the abundant monogalactolipid to different galactolipid acceptors forming

oligogalactolipids and diacylglycerol, which is further convert to triacylglycerol. The combined activity of SFR2 and triaclyglycerol-biosynthetic enzymes leads to the removal of monogalactolipids from envelope membrane, changing the ratio of bilayer to non-bilayer forming membrane lipids to stabilize the membrane during freezing.

*SFR2* gene in rice has been studied by Nakphaijit *et al*., 2006. Northern blot analysis using 3' UTR of rice *SFR2* gene as probe was done by Rodjana Opassiri and Takashi Akiyama to detect mRNA level of rice *SFR2* in cold stress conditions. They found that the transcript of *SFR2* was induced and maintained a similar high level from 6 to 48 hours when rice seedling were treated at 5 °C, whereas SFR2 mRNA were not detected when grown at 28 °C, which was not consistant with the *Arabidopsis SFR2* report (Thorlby *et al.*, 2004). Nakphaijit *et al*., 2006 concluded that the rice *SFR2* transcript level was upregulated in response to cold stress at 5 °C in rice seedling. The protein localization using GFP and GUS as reporter proteins in hydrilla and onion cells indicated that the SFR2 localized to chloroplasts.

In *Arabidopsis thaliana*, there are 48 members in glycosyl hydrolase (GH) family 1 (Xu *et al*., 2004). Of these 47 shared a common evolutionary origin and were assigned a new gene names (*BGLU1-BGLU47*), whereas the remaining family member (*sfr2*) belong to a distinct lineage (Xu *et al.*, 2004). In rice, 40 members of GH1 were identified in the genome sequences (Opassiri *et al*., 2006). The thirty-seven gene-derived cDNA showed a high level of shared deduced amino acids sequence identity to each other and other known plant β-glucosidase sequences. The 3 remaining family members, *Os11bglu36 (sfr2), Osbglu39* and *Osbglu40* are more closely related to bacterial GH1 genes (Opassiri *et al.*, 2006). Analysis of the primary structure of SFR2 revealed, intriguingly, that the protein is most closely related to β-.

glycosidases from thermophillic and halophillic archaebacteria (Martin and Jones, 2004) The amino acids sequence of SFR2 from rice showed that this gene is more divergent from every other glycosyl hydrolase family 1 enzyme from rice and other plants except for *SFR2* orthologues (Opassiri *et al.*, 2006). The cDNA corresponding to the rice ortholog shares about 60% sequence identity with the *Arabidopsis* ortholog. Moreover, BLAST searches using the rice *sfr2* cDNA as the query sequence revealed a number of maize, sorghum, and barley ESTs, indicating that *sfr2* orthologs occur in other plant taxa (Xu *et al.*, 2004).

## 2.4 *Escherichia coli* expression system

Among many systems available for heterologous protein production, the Gram negative bacterium *Escherichia coli* remain one of the most attractive, because of its ability to grow rapidly and at high density on inexpensive substrate, its well-characterized genetics and availability of an increasingly large number of cloning vectors and mutant host strains (Baneyx, 1999). The majority of proteins synthesized naturally by *E. coli* are intracellular. Few of the cell's native protein are exported to the periplasmic space and fewer still are exported into extracellular medium. Thus the majority of heterologous proteins expressed in *E. coli* accumulate in the cytoplasm. However, *E. coli* cannot perform post-translational modifications of recombinant eukaryotic proteins such as glycosylation and acetylation of proteins. Lack of post-translational modification may or may not have a significant effect upon the biological activity of the heterologous protein (Walsh, 2000). Moreover, the high-level production of functional proteins in *Escherichia coli,* especially those from eukaryotic sources, have often been difficult with many reasons (Weickert *et al.*, 1996).

### 2.4.1 Codon

*E. coli* is a commonly used expression host for high-level production of heterologous protein, but has a marked bias in codon usage. Differences in codon usage between prokaryotes and eukaryotes can have a significant impact on heterologous protein production. A gene with a codon usage different from the targeted host organism often has poor expression in that particular host. The presence of rare codons in cloned genes affects the mRNA and protein expression levels (Kurland and Gallant, 1996). The excessive presence of rare codons lead to translational errors as a result of ribosomal stalling at a position requiring incorporation of amino acids coupled to minor tRNAs, or even at sites requiring major tRNAs, but which are depleted because of overutilization of a particular amino acid. The mistranslational events related to rare tRNAs are observed as codon misreadings and as processing errors and they manifest themselves as amino acid substitutions or frameshift events (Wu *et al.*, 2004). The rare arginine (AGG, AGA, CGG, and CGA), leucine (CUA), isoleucine (AUA), and proline codons (CCC) often lead to frameshift errors and ultimately to undesired products. In addition, it is likely that an excess of any of these codons, even without clusters, could create translational problems (Kane, 1995). In some cases, rare codons inhibit protein synthesis and cell growth. Eukaryotic organisms often use codons that are rarely used in the highly expressed genes of *E. coli*. Currently there are two alternative strategies to circumvent this codon bias. One is to alter rare codons in the target gene to the preferred codons of *E. coli* without affecting the encoded amino acid sequence. Another is to expand the intracellular tRNA pool by introduction of a plasmid encoding extra copies of tRNAs for codons rarely used in *E. coli* (Zhou *et al.*, 2004).

### 2.4.2 mRNA secondary structure

Translational efficiency in *E. coli* is known to be strongly influenced by the secondary structure around the ribosome-binding site and the initiation codon in the translational-initiation region of the mRNA. Several quantitative studies have reported that translational efficiency is attributable to effects on ribosome accessibility predominantly caused by the secondary structure surrounding the ribosome-binding site (Seo *et al.*, 2009). The intramolecular mRNA fold involving the Shine-Dalgarno (SD) sequence and/or initiation codon, if they are stable enough, can completely block the access of the ribosomal particle to the initiation site on mRNA. Therefore, the existence of secondary or tertiary structure in the ribosome binding site region seems to always reduce the initiation rate (Spirin., 1999). Recently, researchers tried to improve the translation initiation by adding an RNA stem loop after the start codon, insertion of the stem-loop allowed *in vitro* expression of previously non-expressible genes and enhanced the expression of all other genes investigated and demonstrated that stabilization of the ribosome binding site is due to stem-loop introduction. Thus, the inserted RNA stem-loop supports the formation of a separate translational initiation domain, which is more accessible to the ribosome binding site (Paulus *et al.*, 2004; Seo *et al.*, 2009).

### 2.4.3 PEST sequences

PEST sequences are sequences that have been associated with rapidly degraded proteins. These sequences have been suggested to serve as signals for proteolytic degradation (Chen and Clarke, 2002). In 1986, Rogers *et al*. chose 10 proteins with intracellular half-lives less than 2 hours. They found that these proteins contain one or more regions rich in proline (P), glutamic acid (E), serine (S) and

threonine (T). These region are generally, but not always, flanked by clusters containing several positively charged amino acids. They named these regions PEST regions. Based on their observation, they developed an algorithm that would search for such regions in a given protein sequence called PEST-FIND. And they defined PEST sequence as hydrophilic stretches of amino acids greater than or equal to 12 residues in length. PEST-FIND produces a score ranging from about -50 to +50. By definition, a score above zero denotes a possible PEST region, but a value greater than +5 sparks real interest (Rechsteiner and Rogers, 1996). Salama *et al*. added the carboxyl terminus of CLN2 that has PEST sequence properties onto thymidine kinase and found the chimera was rapidly degraded throughout the yeast cell cycle. Moreover, they also showed that the removal of the most prominent PEST region in the CLN2 tail stabilized the fusion protein (Salama *et al.*, 1994). However, there are stable proteins that contain PEST regions and there are other factors that influence the degradation of intracellular protein, such as intracellular location or masking due to association with other macromolecules. There are many properties of proteins involved in PEST protein degradation like larger proteins are degrade faster than smaller ones and negatively charged protein are degraded faster than the positively charged proteins (Rogers *et al.*, 1986).

For the pathway for degrading PEST protein, it is still not clear. The researcher speculated that proline residues within PEST sequences might induce degradation by proline endopeptidases called calpains. Calpain (EC 3.4.22.17) is a $Ca^{2+}$-dependent cysteine protease first isolated in 1978, with a pH optimum between 7.0 and 8.0. If the negatively charged PEST regions bind $Ca^{2+}$, they could provide both peptide bonds and a necessary co-factor for proteolysis by $Ca^{2+}$-activated protease (Rechsteiner and Rogers, 1996; Mitchell and Bell, 2003). However, some

studies denied this hypothesis, site-directed mutagenesis of plasma membrane $Ca^{2+}$ ATPase, an excellent calpain substrate, demonstrated that a calmodulin-binding site is crucial for proteolysis, but two nearby PEST regions could be changed markedly with little effect on rates of proteolysis by calpain (Molinari *et al.*, 1995). Some evidence supported the idea that PEST sequences target proteins for degradation by the 26S proteosome in eucaryotes, because the ubiquitin-26S proteosome-mediated degradation of the PEST containing proteins, phytochrome, Matα2 repressor and p53 was discovered (Rechsteiner, 1991).

## 2.5 Protein expression in *P. pastoris*

The use of the methylotrophic yeast, *P. pastoris* as a cellular host for the expression of recombinant proteins has become increasing popular in recent times. *P. pastoris* is easier to genetically manipulate and culture than mammalian cells and can be grown to high cell densities. Equally important, *P. pastoris* is also a eukaryote, and thereby provides the potential for producing soluble, correctly folded recombinant proteins that have undergone all the post-translational modifications required for functionality (Daly and Hearn, 2005). In addition, other reports described recombinant proteins that do not fold correctly in bacterial systems, but are expressed successfully in yeast systems (Fan *et al.*, 2007; Tarahomjoo *et al.*, 2008).

The conceptual basis for the *P. pastoris* expression system involves methanol utilization which requires a novel metabolic pathway involving several unique enzymes. The enzyme alcohol oxidase (AOX) catalyzes the first step in the methanol utilization pathway, the oxidation of methanol to formaldehyde and hydrogen peroxide. AOX is sequestered within the peroxisome along with catalase, which degrades hydrogen peroxide to oxygen and water. A portion of the formaldehyde

generated by AOX leaves the peroxisome and is further oxidized to formate and carbon dioxide by two cytoplasmic dehydrogenases, reactions that are a source of energy for cells growing on methanol (Cereghino and Cregg, 2000). The *P. pastoris* system has strong promoters available to drive the expression of foreign genes of interest, thus enabling production of large amounts of the target proteins with relative technical ease and at a lower cost than most other eukaryotic systems (Daly and Hearn, 2005).

Numerous strains of *P. pastoris* with a wide range of genotypes are now available. The SMD1168 and SMD1168H strains, for example, are defective in the vacuole peptidase A (pep4). This enzyme is responsible for activating carboxypeptidase Y and protease B1 and, hence, these SMD1168 and SMD1168H strains are also defective in these proteases but these protease-deficient cells are not as vigorous as wild-type strains. In addition to lower viability, they possess a slower growth rate and are more difficult to transform. Therefore, the use of protease-deficient strains is only recommended in situations where other measures to reduce proteolysis have yielded unsatisfactory results (Cereghino and Cregg, 2000; Daly and Hearn, 2005). However, there are many reports of satisfied results with high level expression of secreted proteins (Chen *et al.*, 2009; Sanchez-Venegas *et al.*, 2009; Luang *et al.*, 2010).

So, in this research *P. pastoris* strain SMD1168H was used as a host strain for the expression of SFR2 from a gene with codon optimized for *P. pastoris*.

# CHAPTER III

# MATERIALS AND METHODS

## 3.1 Cloning

### 3.1.1 Materials

#### 3.1.1.1 Rice SFR2 cDNA

The rice *SFR*2 cDNA (DDBJ Accession AK119461) in the pME18SFL3 vector from clone 001-133-E10 provided by the Rice Genome Resource Center (National Institute for Agrobiological Sciences, Japan) was used as template to amplify the cDNA for expression.

#### 3.1.1.2 Rice SFR2 gene synthesized

The synthetic rice *SFR2* gene with optimized codon for *P. pastoris* ordered from Genescript was used as template to amplify DNA for expression (Appendix I).

### 3.1.1.3 Primers

**Table 3.1** Oligonucleotide primers for amplification and sequencing of the *SFR2* gene in *E. coli* and *P. pastoris* vector (continued).

| **Primers** | **Sequence (5' → 3')** |
|---|---|
| AK_SFR2_F | CACCATGCCACTACCGGCGTTC |
| 26_SFR2_F | CACCGACTTCCGCGCCTTTCC |
| 28_SFR2_F | CACCGTGAGGCTTGTTGTTGATTGT |
| 29_SFR2_F | CACCCCTGGTCTAAAGCTTGTG |
| 30_SFR2_R | GGAGAATAAAAAGTATGAAGGCCGAGG |
| 33_mutPEST_SFR2_F | TGGGCAAGGTTAATGCCTCAGCAACCAACCCAACAATTGAAG TGAGTT |
| 34_mutPEST_SFR2_R | GCAAAAATAACTGAGCTCTTCAATTGTTGGGTTGGTTGCTGA GGCATTAACCTTGCCCA |
| 28_SFR2_R | ACAATCAACAACAAGCCTCAC |
| 37_SFR2_KpnI_F | GCCGGGTACCCTAGCAACAGCGCCTGCGCAT |
| 32_SFR2_SalI_R | GCCGGTCGACGGAGAATAAAAAGTATGAAGGCCGAGG |
| 26_SFR2_Syn_F | CACCGATTTCAGAGCTTTTCCTTC |
| 30_SFR2_Syn_R | AAGACCTTCATATTTCTTGTTTTCT |
| pET393_seq_F | GAAAGAGTTCCTCGACGC |
| T7 promoter | TAATACGACTCACTATAGGG |
| T7 terminator | GCTAGTTATTGCTCAGCGG |
| 35_pCold_seq_F | ACGCCATATCGCCGAAAGA |
| 36_pCold_seq_R | CCAAATGGCAGGGATCTTAGATTC |
| M13_F | GTAAAACGACGGCCAG |
| M13_pUC_R | CAGGAAACAGCTATGAC |

AK_SFR2_F

atgccactaccggcgttcgtggcggcggcggcgaggctcgccgtcctcgtggccgcggcg 60

**M P L P A F V A A A A R L A V L V A A A** 20

gcgacggcggccaatgccgcctcctacgcgcggtaccggcggcgccacctccgccgcatc 120

**A T A A N A A** S Y A R Y R R R H L R R I 40

26_SFR2_F

cccagccccatcgacgagtccgccgatcccctcgccgacttccgcgcctttccctcctcc 180

P S P I D E S A D P L A D F R A F P S S 60

37_SFR2_KpnI_F

gacgctgacgactcagaggaggataatttcttctttgggctagcaacagcgcctgcgcat 240

D A D D S E E D N F F F G L A T A P A H 80

gttgaggacaggctagaagatgcttggctccagtttgcaactgaaacttcctgcgatgac 300

V E D R L E D A W L Q F A T E T S C D D 100

aatggaaacgtgcgcgaccagaggccagtagatgcactgatggcatctgctgctggtgat 360

N G N V R D Q R P V D A L M A S A A G D 120

ggaggctcccagcaatcttggaggtcaacaggtggggaaaatattggtgatagagagcag 420

G G S Q Q S W R S T G G E N I G D R E Q 140

aggaaaccacttagggtagccatggaggctatgctcaggggtttgaaattcttgctgag 480

R K P L R V A M E A M L R G F E I L A E 160

27_SFR2_F

agtggagaatctgctggcggcgataactgcagccacaatgttgctgcttggcataatgtt 540

S G E S A G G D N C S H N V A A W H N V 180

ccatgcccgcaagaaaggcttagattttggtctgatcctgatgctgagttgaaacttgct 600

P C P Q E R L R F W S D P D A E L K L A 200

33_mutPEST_SFR2_F and 34_mutPEST_SFR2_R

aaggagaccggcatcagcgttttccgcatgggggtagattgggcaaggttaatgcctgag 660

K E T G I S V F R M G V D W A R L M P E 220

gaaccaaccgaagaattgaagagctcagttaattttgcagcacttgagcggtatagatgg 720

E P T E E L K S S V N F A A L E R Y R W 240

atcattcaaagggttcgcgaatatggaatgaaagtaatgcttactctgtttcatcactca 780

I I Q R V R E Y G M K V M L T L F H H S 260

cttccaccttgggctggaaaatatggcgggtggaagatggaaaaaactgtcacctatttc 840

L P P W A G K Y G G W K M E K T V T Y F 280

28_SFR2_F and 28_SFR2_R

atggatttgtgaggcttgttgttgatcgtgtatccaatttagtggactactgggtgatt 900

M D F V R L V V D R V S N L V D Y W V I 300

ttcaatgaacctcacgttttgtgatgctaacttattgtgccggtgcttggcctggtgga 960

F N E P H V F V M L T Y C A G A W P G G 320

gaccctaatgcaattgaagtagcaacatctactctgccaactggtgtatacaatcaggct 1020

D P N A I E V A T S T L P T G V Y N Q A 340

ttgcattggatggctattgcacattctgaagcctatgactacatacattcgaaaagcaag 1080

L H W M A I A H S E A Y D Y I H S K S K 360

aacgaaaggaagccaatagttggtgttgcacaccatgtatcgtttacaaggccatatggg 1140

N E R K P I V G V A H H V S F T R P Y G 380

ctatttgatgttgctgctgtcgcgctagctaactcattgacccttttccttacgtggat 1200

L F D V A A V A L A N S L T L F P Y V D 400

agcatatgtgataaattggacttcattggaatcaactactatggacaggaggttatatca 1260

S I C D K L D F I G I N Y Y G Q E V I S 420

29_SFR2_R

ggacctggtctaaagcttgtggataatgatgagtatagtgaatctggtcgtggggtttat 1320

G P G L K L V D N D E Y S E S G R G V Y 440

cctgatgggctgttccgcatcctgattcaattcaatgaacgatataagagattaaatata 1380

P D G L F R I L I Q F N E R Y K R L N I 460

ccttttgtcattactgaaaatggagtttctgatgagactgatctgatacggaaaccatat 1440

P F V I T E N G V S D E T D L I R K P Y 480

atactggaacacttgttagccacatacgctgccatcattatgggtgtccgtgtacttggt 1500

I L E H L L A T Y A A I I M G V R V L G 500

tatttgttttggacaacatcagataattgggaatgggcggatggctatggtcccaagttt 1560

Y L F W T T S D N W E W A D G Y G P K F 520

30SFR2_R and 32_SFR2_SalI_R

gggcttgttgctgttgaccgtgctaataacctagcacggaaacctcggccttcatacttt 1620

G L V A V D R A N N L A R K P R P S Y F 540

ttattctccagggttgttacaactggaaaaattacaagacaggacagaatgtctgcttgg 1740

L F S R V V T T G K I T R Q D R M S A W 560

agggagctgcaacaagctgcagttcaaaagaaaacacgtccatttttcagggcagtggat 1800

R E L Q Q A A V Q K K T R P F F R A V D 580

aagcatggtcggatgtatgcaggtggtctagatcggcctattcagaggcccttcatattg 1860

K H G R M Y A G G L D R P I Q R P F I L 600

```
cgggattggagatttggtcactataaaatggaaggcttgcaggatcctttgagttgcttt 1920
 R  D  W  R  F  G  H  Y  K  M  E  G  L  Q  D  P  L  S  C  F  620
ataagatgcatttttgcaccattttcacgccaaaagaagattcactacatagaagatgat 1980
 I  R  C  I  F  A  P  F  S  R  Q  K  K  I  H  Y  I  E  D  D  640
gttatttcttattctattaactga                                     2004
 V  I  S  Y  S  I  N  -                                      648
```

**Figure 3.1** *SFR2* gene sequence and primers used for SFR2 gene amplification. Nucleotides in grey highlight indicated the positions for primer to amplify the SFR2 gene fragments as Table 3.2. A presicted signal peptide is indicated in bold letters. The family1 glycosyl hydrolase motif (IFNEP and VTENG) were indicated in oval and PEST sequence was indicated in box.

#### 3.1.1.4 Chemicals

Reagents for PCR amplification consisted of 10X *Taq* buffers (Promega), 25 mM $MgCl_2$ (Promega), 2.5 mM dNTPs mix (Promega), 10 μM forward and reverse primers and *Taq* polymerase (Homemade). Reagents for agarose gel electrophoresis included 1X TAE buffer, agarose low EEO, 1 kb and 100 bp DNA Ladder Markers and loading dye (Fermentas). PCR purification kit, gel extraction kit and plasmid miniprep kit were purchased from Qiagen. The *Kpn*I and *Sal*I were purchased from Promega. The LR clonase was purchased from Invitrogen. All chemicals used were molecular grade or analytical grade.

#### 3.1.1.5 Bioinformatics

The National Center for Biotechnology Information website or NCBI (http://www.ncbi.nlm.nih.gov/) was used for the public sequence database

search for AK119461. The Expasy translation tool was used to translate SFR2 gene to protein sequences (http://au.expasy.org/tools/dna.html). The multiple sequence alignment (ClustalW2) was used to make the alignment of rice SFR2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). PESTfind analysis web tool was used to determine the PEST sequence within the SFR2 protein sequence (http://www.es.embnet.org/Services/EMBnetAT/htdoc/PESTfind/).

**3.1.1.6 Plasmids**

(a) pENTR-D/TOPO

The pENTR Directional TOPO® Cloning vector (Invitrogen) was used to clone blunt-end PCR products containing the gene of interest in a 5′ to 3′ orientation. The PCR product includes a 4 base pair sequences (CACC) at the 5' end of the forward primer. pENTR-D/TOPO provides rapid, efficient cloning of PCR products in the correct orientation. This vector has flanking universal M13 sequencing sites for simplified sequence verification of the cloned insert. The pENTR-D/TOPO plasmid has kanamycin resistance marker for selection in *E. coli*. This vector was used to allow for entry into the Gateway cloning system.

(b) pET32a/DEST

The pET32a/DEST (provided by J. R. Ketudat-Cairns and described in Opassiri *et al*., 2006) was constructed from pET32a (Novagen) by inserting Gateway conversion cassette reading frame A into the *EcoR*V (206) site of pET32a(+). The pET-32a(+)DEST vector is designed for high-level expression of peptide sequences fused with the 109 amino acids Trx•Tag™ thioredoxin protein. The

fusion proteins also contained cleavable His•Tag and S•Tag sequences for detection and purification. It can be used as a destination plasmid to receive the insert fragment from the pENTR-D/TOPO cloning plasmid.

(c) pCold I expression vector

The pCold I (Takara) is a cold-shock expression vector designed to perform efficient protein expression utilizing the promoter derived from the *cspA* gene, which is one of the cold-shock genes. At the downstream of the *cspA* promoter, the lac operator was inserted so that the expression is strictly controlled. In addition, a 5' untranslated region (5' UTR), translation enhancing element (TEE), His-Tag sequence, Factor Xa cleavage site, and multiple cloning site (MCS) are located at the downstream of the *cspA* promoter. The advantages of pCold I are increased levels of expression, ability to express otherwise inexpressible proteins and increased levels of soluble protein.

(d) pPICZαBNH8/DEST

The pPICZαBNH8/DEST (provided by J. R. Ketudat-Cairns) was derived from pPICZαB (Invitrogen) by inserting an 8-histidine tag after the signal peptide cleavage the Gateway conversion cassette C into the *SnaB*I of pPICZαB (Luang *et al.*, 2010). This expression vector contains N-terminal α-factor prepropeptide sequence for protein secretion and 8x His-tag for purification. It also contains an AOX1 promoter which can be induced with methanol. The pPicZαBNH8/DEST plasmid has a zeocin resistance marker for selection both in *P. pastoris* and *E. coli*.

### 3.1.2 Methodologies

#### 3.1.2.1 Cloning of N-terminal and C-terminal SFR2 deletion

The PCR technique was used to amplify different parts of *SFR2* gene using 9 different primer pairs designed from the full length rice *SFR*2 cDNA sequence from the Genbank accession AK1194661 sequence. The forward primers for N-terminal deletion were AK_SFR2_F, 26_SFR2_F, 27_SFR2_F, 28_SFR2_F and 29_SFR2_F (Table 3.1). The reverse primer was 30_SFR2_R (Table 3.1). The *SFR2* N-terminal deletion products were AK_SFR2, 26_SFR2, 27_SFR2, 28_SFR2 and 29_SFR2. The PCR conditions are shown in Table 3.2.

The forward primers for C-terminal deletion were 26_SFR2_F and 27_SFR2_F (Table 3.1). The reverse primers were 34_mutPEST_SFR2_R and 28_SFR2_R (Table 3.1). The C-terminal SFR2 deletion products were 31_SFR2_delC(mP), 32_SFR2_delC(mP), 33_SFR2_delC and 34_SFR2_delC. The PCR conditions are shown in Table 3.2.

The PCR products were run on 1% agarose gel then excised and purified with a gel extraction kit (QIAGEN). The purified PCR products were cloned into pENTR/D-TOPO cloning vector (Invitrogen). Each reaction consisted of 1 μl of 15-20 ng of pENTR/D-TOPO vector, 1 μl of salt solution and 20 ng of PCR products. The solution mixtures were incubated at 25 °C overnight and transformed into *E. coli* TOP10 by heat shock. The *E. coli* cells were spread on LB agar containing 100 μl/ml kanamycin. The pENTR/D-TOPO insert was extracted from the positive clones by plasmid prep kits and DNA sequencing were done with the M13 primers.

Gateway technology was used to exchange the insert from the entry

clones (pENTR-D/TOPO to the destination vector (pET32a/DEST expression vector). The LR clonase exchanged the insert between the attL1 and attL2 sites of pENTR/D-TOPO into the attR1 and attR2 sites (Katzen, 2007) in pET32a/DEST. The products of LR reaction are expression clones, with the DNA of interest. The LR reaction consisted of 1 μl of 300 ng entry clones, 2 μl of 100 ng/μl pET32a/DEST and 2 μl of LR clonase enzyme. The reaction mixtures were incubated at 25 °C overnight, then 2 μl of proteinase K was added and incubated at 37 °C for 10 min to stop the reaction. Four microliters of each reaction were transformed into *E. coli* DH5α by electroporation. The *E. coli* cells were spread on LB agar containing 100 μl/ml ampicillin. Plasmids were extracted by plasmid prep kits and DNA sequencing was done with the pET32_393F and T7 terminator primer.

#### 3.1.2.2 Cloning of 26_SFR2_Syn

pJET_SFR2_Syn ordered from Genescript was used as template to amplify 26_SFR2_Syn (Appendix I). Primers 26_SFR2_Syn_F and 30_SFR2_Syn_R were used. The PCR products were purified with a gel extraction kit and cloned into pENTR-D/TOPO and pET32a/DEST, respectively. The PCR condition consisted of 35 cycles, 94 °C 2 min for initiation denature then 94 °C 30 s for denaturation, 55 °C 30 s for annealing and 72 °C 2 min for extension in each cycle and 72 °C 10 min for final extension.

**Table 3.2** PCR primers and annealing conditions for amplification of N- and C-terminus deletions of *SFR2* gene.

| PCR | Primer pair | Annealing | Extension times |
|---|---|---|---|
| AK_SFR2 | AK_SFR2_F | 55 °C 30 s | 2 min |
| | 30_SFR2_R | | |
| 26_SFR2 | 26_SFR2_F | 55 °C 30 s | 2 min |
| | 30_SFR2_R | | |
| 27_SFR2 | 27_SFR2_F | 55 °C 30 s | 1.5 min |
| | 30_SFR2_R | | |
| 28_SFR2 | 28_SFR2_F | 55 °C 30 s | 1 min |
| | 30_SFR2_R | | |
| 29_SFR2 | 29_SFR2_F | 50 °C 30 s | 30 s |
| | 30_SFR2_R | | |
| 31_SFR2 _delC(mP) | 26_SFR2_F | 58 °C 30 s | 30 s |
| | 34_mutPEST_SFR2_R | | |
| 32_SFR2 _delC(mP) | 27_SFR2_F | 55 °C 30 s | 30 s |
| | 34_mutPEST_SFR2_R | | |
| 33_SFR2 _delC | 26_SFR_F | 50 °C 30 s | 30 s |
| | 28_SFR2_R | | |
| 34_SFR2 _delC | 27_SFR2_F | 50 °C 30 s | 30 s |
| | 28_SFR2_R | | |

Note: The PCR were done with 94 °C initial denaturation for 2 min and 35 cycles of 30 s 94 °C denaturation, 30 s annealing time at temperature as shown, 72 °C extension for the stated time followed by final extension at 72 °C for 10 min.

### 3.1.2.3 Mutation of PEST sequence in *SFR2* gene by PCR mutagenesis for cloned into pET32a/DEST

The codons encoding glutamic acids (E) in the PEST sequence were mutated to codons for glutamine (Q) by PCR mutagenesis. Briefly, two primer pairs were used to amplify the *SFR2* gene in 2 parts. The first primer pairs were 26_SFR2_F and 34_mutPEST_SFR2_R (Table 3.1) used to generate the 5' end part of the *SFR2* gene. The second primer pair, 33_mutPEST_SFR2_F and 30_SFR2_R (Table 3.1) were used to generate the 3' end part of the mutated *SFR2*. Note that the 34_mutPEST_SFR2_R and 33_mutPEST_SFR2_F primer had complementary sequences containing the mutated codons changing glutamic acid to glutamine. Both regions of the gene were amplified and excised from a 1% agarose gel and purified with the gel extraction kit. In the next step, the above PCR products were used as templates to anneal to each other and the 26_SFR2_F and 30_SFR2_R primers were used to generate full length of 26_mutPEST_SFR2 by two step PCR (Fig. 3.2). The PCR mixture contained 10 ng of the both parts of *SFR2* template, 1X *Taq* buffer (Promega), 2.5 mM $MgCl_2$, 2.5 mM dNTP, 0.2 μM of each primer (26_SFR2_F and 30_SFR2_R) and 1 μl *Taq* polymerase (Homemade). The PCR program consisted of 20 repetitive cycles with a strand separation step at 94 °C for 30 s, an annealing step at 69 °C for 10 min and elongation step at 72 °C for 5 min. This step is to anneal both parts together (modified from Wurch et al., 1998). The second step for increasing the number of 26_mutPEST_SFR2 DNA strands used the 26_SFR2_F and 30_SFR2_R primers. Denaturing step at at 94 °C for 30 s, an annealing step at 55 °C for 30 s and an elongation step at 72 °C for 2 min were used. The 1464 bp band was confirmed on 1% agarose gel and purified with the gel extraction kit. The 26_mutPEST_SFR2

cDNA was then cloned into pENTR-D/TOPO and transferred to the pET32a expression vector by LR clonase reaction.

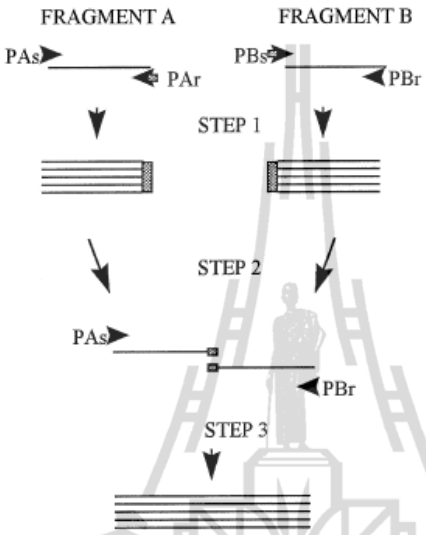



**Figure 3.2** Strategy to create chimeric gene using overlap extension PCR technique in the absence of restriction enzyme (Wurch *et al.*, 1998).

**3.1.2.4 Cloning of 37_SFR2 and 37_mutPEST_SFR2 into pCold I**

The 37_SFR2 and the 37_mutPEST_SFR2 cDNA were cloned into pCold I by standard restriction enzyme digestion and ligation. The 37_SFR2 and the 37_mutPEST_SFR2 were amplified with the 37_SFR2_KpnI_F and the 32_SFR2 SalI_R primers which containing *Kpn*I and *Sal*I restriction sites, respectively, using the 26_SFR2 and 26_mutPEST_SFR2 cDNA as templates, respectively. The PCR mixture (50 µl) contained 10 ng of template, 1X *Taq* buffer (Promega), 2.5 mM $MgCl_2$, 2.5 mM dNTP, 0.2 µM of each primer and 1 µl of *Taq* polymerase (Homemade). The PCR products were loaded onto a 1% agarose and purified with a gel extraction kit. Both inserts were ligated into pGEM-T easy vector. The ligation reaction contained 50 ng of inserts, 1X rapid ligation buffer, 50 ng of pGEM-T easy

vector and 1 µl of T4 DNA ligase. The reaction mixtures were incubated at 4 °C overnight, then transformed in to *E. coli* DH5α by electroporation. The LB agar plates containing 100 µg/ml ampicillin was prepared, then 100 µl of 100 mM IPTG and 50 µl of 20 mg/ml X-gal was spread over the LB ampicillin plate and allowed to dry for 30 min at 37 °C prior to use. The *E. coli* cells were plated on LB/ampicillin/X-gal/IPTG agar. The positive clones were chosen to do colony PCR. Plasmids were extracted with a plasmid extraction kit and designated as pGEM_37_SFR2 and pGEM_37_mutPEST_SFR2. Ten nanograms of pCold I vector and pGEM containing 37_SFR2 and 37_mutPEST_SFR2 cDNA inserts were cut with *Kpn*I for 2 h at 37 °C then purified by DNA precipitation and cut with *Sal*I for 2 h at 37 °C. The digested inserts and pCold I vector were purified again with a gel extraction kit. The linear pCold I vector and the 37_SFR2 or 37_mutPEST_SFR2 cDNA were ligated together by T4 DNA ligase (NEB) overnight at 16 °C. The ligation reactions were transformed into *E. coli* DH5α and spread on LB agar plates containing 100 µg/ml ampicillin. The transformants were incubated for 18 h at 37 °C. The positive clones were collected and tested by colony PCR and plasmids were extracted with a plasmid prep kit (QIAGEN).

#### 3.1.2.5 Cloning of the 26_SFR2_Syn cDNA into pPICZαBNH8/DEST

The TOPO_26_SFR2_Syn plasmid was used as the entry clone to transfer the insert to the destination vector, pPICZαBNH8/DEST (Luang *et al*., 2010). One microliter of 300 ng entry clones, 2 µl of 100 ng/µl pPICZαBNH8/DEST and 2 µl of LR clonase were mixed. The reaction mixtures were incubated at 25 °C overnight then 1 µl of proteinase K was added and incubated at 37 °C for 10 min to

stop the reaction. Four microliters of ligation reaction were transformed into *E. coli* DH5α by electroporation. The *E. coli* cells were spread on an LB agar containing 100 μl/ml Zeocin and incubated for 18 h at 37 °C. The pPICZ_26_SFR2_Syn plasmids were extracted from the positive clones for further use. Plasmids were sequenced from the 5' AOX and 3' AOX primers to ensure the correct sequence and reading frame.

## 3.2 Protein Expression

### 3.2.1 Protein Expression in *E. coli*

#### 3.2.1.1 Materials

##### 3.2.1.1.1 *E. coli* strains

(a) BL21(DE3) p*lys*S

BL21(DE3) p*Lys*S cells allowed high-efficiency protein expression of any gene that is under the control of a T7 promoter and has a ribosome binding site. BL21(DE3) p*Lys*S is lysogenic for λ-DE3, which contains the T7 bacteriophage gene I, encoding T7 RNA polymerase under the control of the *lac* UV5 promoter. This strain also contains a plasmid, p*Lys*S, which carries the gene encoding T7 lysozyme. T7 lysozyme lowers the background expression level of target genes under the control of the T7 promoter but does not interfere with the level of expression achieved following induction by IPTG.

Genotype: F–, *omp*T, *hsd*$S_B$ ($r_B$–, $m_B$–), *dcm*, *gal*, λ(DE3), p*Lys*S, $Cm^r$.

(b) Origami(DE3)

This host strain is a K-12 derivative that contains mutations in both the thioredoxin reductase (trxB) and glutathione reductase (*gor*) genes, which greatly enhance disulfide bond formation in the cytoplasm. Origami hosts are compatible with ampicillin resistant plasmids and are ideal for use with pET-32 vectors, since the thioredoxin fusion tag further enhances the formation of disulfide bonds in the cytoplasm. The *trxB* and *gor* mutations are selectable on kanamycin and tetracycline, respectively; therefore this strain cannot be used with plasmids carrying kanamycin- or tetracycline-resistance genes.

Genotype: Δ(*ara-leu*)*7697* Δ*lacX74* Δ*phoA Pvu*II *phoR araD139 ahpC galE galK rps*L F′[$lac^+$ $lacI^q$ *pro*] (DE3) *gor522*::Tn*10 trxB* pLacI ($Cam^R$, $Str^R$, $Tet^R$)

(c) Origami B(DE3)

This host strains carries the same mutations in *trxB* and *gor* as the original Origami strains, except that they are derived from a *lacZY* mutant of BL21 to enable precise control of expression levels by adjusting the concentration of IPTG. Thus the Origami B strains combine the desirable characteristics of BL21, Tuner™, and Origami strains in one strain background. These strains also include the *lon* and *ompT* deficiences of BL21, which increase protein stability. Such strains are suitable for production of protein from target genes cloned in pET vectors by induction with IPTG.

Genotype: $F^-$ *ompT hsd*$S_B$($r_B^-$ $m_B^-$) *gal dcm lacY1 ahpC* (DE3) *gor522*:: Tn*10 trxB* ($Kan^R$, $Tet^R$)

(d) Rosetta-gami(DE3)

This host strain is an Origami derivative that combines the enhanced disulfide bond formation resulting from *trxB/gor* mutations with enhanced expression of eukaryotic protein that contain codons rarely used in *E. coli*. These strains supply tRNAs for AGG, AGA, AUA, CUA, CCC, and GGA codons on a compatible chloramphenicol-resistant plasmid. The Rosetta-gami strains are resistant to kanamycin, tetracycline, streptomycin, and chloramphenicol. These strains are recommended for use with expression plasmids carrying the ampicillin resistance marker *bla*.

Genotype: Δ(*ara-leu*)*7697* Δ*lacX74* Δ*phoA Pvu*II *phoR araD139 ahpC galE galK rpsL* (DE3) F′[*lac*+ *lacI*$^{q}$ *pro*] *gor522*::Tn*10 trxB* pRARE (Cam$^{R}$, Kan$^{R}$, Str$^{R}$, Tet$^{R}$)

(e) ArcticExpress (DE3) RIL

ArcticExpress was engineered to address the common bacterial gene expression problem of protein insolubility. These cells are derived from the high-performance Stratagene BL21-Gold cells, enabling efficient high-level expression of heterologous proteins in *E. coli*. This strain contained the chaperonins Cpn60 and Cpn10 from the psychrophilic bacterium *Oleispira antarctica*, which help improve protein processing at lower temperatures, potentially increasing the yield of active and soluble recombinant protein. These cells have extra copies of the *argU*, *ileY,* and *leuW* tRNA genes. These genes encode tRNAs that recognize the arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA, respectively**.** The ArcticExpress RIL strains have available the tRNAs that most frequently restrict translation of heterologous proteins from organisms that have AT-rich genomes.

Genotype: *E. coli* B F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal* λ(DE3) *endA* Hte [*cpn10 cpn60* Gentr] [*argU ileY leuW* Strr]

#### 3.2.1.1.2 Chemicals

Reagents for *E. coli* competent cells preparation included 10% ice-cold glycerol and GYT medium. The *E. coli* lysis buffer contained 20 mM Tris-Cl pH 8.0, 150 mM NaCl, 0.1% Triton X-100, 1 mg/ml lyzozyme and 1 mM PMSF.

### 3.2.1.2 Methodologies

#### 3.2.1.2.1 pET system

Each pET that contained the AK_SFR2 or the truncated *SFR2* gene was transformed into *E. coli* strain BL21(DE3) p*Lys*S, Origami(DE3), Origami B(DE3) and Rosetta-gami(DE3) by electroporation. *E. coli* cells were spread on LB agar containing 100 μg/ml ampicillin and appropriate antibiotics, depending on the expression host. A single colony of each positive clone was cultured in 3 ml LB broth containing the appropriate antibiotics for 18 h at 37 °C, 200 rpm to use as a starter. Two hundred microlitter of starter cells were added into 10 ml of fresh LB medium containing 100 μl/ml ampicillin and cultured at 37 °C 200 rpm until the $OD_{600}$ reached to 0.6. The culture temperature was decreased to 20 °C and IPTG was added to a final concentration of 0.3 mM for induction. Proteins were expressed for 24 h at 20 °C, 200 rpm. At the end of the cultivation, cell pellets were harvested by centrifugation at 4,000 rpm, 4 °C for 10 min. The medium was removed and cell

pellets were resuspended in 500 µl of lysis buffer and cells were broken by sonication. The total cells were submitted to 15% gel SDS-PAGE and the gels were stained with Coomassie Brilliant Blue R250.

The pET_26_SFR2 and pET_26_SFR2_syn plasmids were transformed into Arctic Express (DE3) RIL cells by heat shock and spread on LB agar containing 100 µg/ml ampicillin. The transformants were incubated for 18 h at 37 °C. The positive clones were selected and cultured in 3 ml LB medium containing 20 µg/ml gentamycin and 100 µg/ml ampicillin for 18 h at 37 °C. Two hundred microliters of the starter were transferred to the 10 ml fresh LB medium containing antibiotics and incubated for 3 h at 30 °C then IPTG was added to a final concentration 1 mM. Protein was expressed for 24 h at 10 °C. Cell pellets were collected by centrifugation at 4000 rpm, 4 °C for 10 min. The supernatant was removed and cell pellets were resuspended in 500 µl of lysis buffer. Total cells were submitted to SDS-PAGE.

**3.2.1.2.1 pCold I system**

The pCold_37SFR2 and pCold_37mutPEST_SFR2 were transformed into *E. coli* strains BL21(DE3) p*Lys*S, Origami(DE3), Origami B(DE3) and Rosetta-gami(DE3) by electroporation and spread on LB agar containing 100 µg/ml ampicillin and host-appropriate antibiotics. Positive clones were inoculated into 3 ml LB medium containing 100 µg/ml ampicillin then transferred to the fresh medium and incubated at 37 °C until $OD_{600}$ reached to 0.6. The expression temperature was reduced to 16 °C to induce the *cspA* promoter. The protein was expressed for 24 h. Cell pellets were collected by centrifugation at 4000 rpm, 4 °C for

10 min. The supernatant was discarded, and the cell pellets were resuspended in 500 µl of lysis buffer, and the cells were broken by sonication. Total cell lysates were submitted to SDS-PAGE to visualize the protein profile.

### 3.2.2 Protein expression in *P. pastoris*

#### 3.2.2.1 Materials

##### 3.2.2.1.1 *P. pastoris* strains

(a) *P. pastoris* strain SMD1168H

SMD1168H is methylotrophic yeast strain with *Mut*$^+$ and *His*$^-$ genotype. It is a protease deficient strain with the protenase A (pep4) mutation to reduce protein proteolysis.

##### 3.2.2.2.2 Chemicals

Reagents for *P. pastoris* competent cells preparation included ice-cold DI water, 1 M sorbitol and YPD medium. The reagent for breaking cells contained 50 mM sodium phosphate, pH 7.4, 1 mM PMSF, 1 mM EDTA and 5% glycerol. Reagents for protein expression were BMGY medium and absolute methanol.

#### 3.2.2.2 Methodologies

##### 3.2.2.2.1 pPICZαBNH8 system

The pPICZ_26_SFR2_Syn plasmid was cut with *Sac*I at 37 °C for 2 h to generate the linear plasmids. The linear plasmid was transformed into

competent *P. pastoris* SMD1168H by elecroporation at 1500 V and cells were spread on YPD agar containing 100 μg/ml zeocin. The transformants was incubated for 48-72 h at 30 °C until colonies were formed. The positive clones were inoculated in 50 ml BMGY medium and grown at 28 °C, 200 rpm for 24 h. Methanol was added every 12 h to a final concentration of 3% until 72 h. Induced cultures were harvested and aliquotted to load on SDS-PAGE and analyzed by western blot. Supernatants and cell pellets were separated and collected by centrifugation at 4000 rpm for 10 min, then kept at -70 °C overnight. Cell pellets were thawed and 5 ml of breaking buffer was added. One milliliters glass beads (0.5 mm) were used for breaking cells then broken cells were centrifuged at 14,000 rpm for 10 min and the supernatant was transferred to fresh 1.5 ml microcentrifuge tubes. Both fractions were submitted to SDS-PAGE.

## 3.3 Protein Detection

### 3.3.1 Materials

#### 3.3.1.1 Chemicals

Reagents for western blot analysis included anti-rabbit IgG secondary antibody conjugated to alkaline phosphatase, which was purchased from Sigma. SIGMA *FAST*™ BCIP/NBT (5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium) purchased from Sigma was used as substrate.

### 3.3.2 Methodologies

#### 3.3.2.1 Polyclonal antibody production

##### 3.3.2.1.1 Antigen preparation

The pET28_SFR2 was transformed into the expression host, Origami B (DE3). The positive clones was chosen and incubated in 3 ml LB broth containing appropriate antibiotic for the starter culture. The starter culture was cultured at 37 °C overnight, then transferred to 50 ml fresh LB medium containing 100 µg/ml ampcillin, 12.5 µg/ml tetracycline and 15 µg/ml kanamycin and cultured at 37 °C for 3 h. The culture was transferred to 20 °C for 30 min. The IPTG was added 0.4 mM final concentration for protein induction for 24 h at 200 rpm. At the end of the cultivation, cells were harvested by centrifugation at 4000 rpm, 4 °C for 15 min. Supernatant was discarded and cells were kept at -70 °C for 3 h. Then cells were resuspended in Tris-HCl buffer, pH 7.0, containing 0.5 mM PMSF and cells were broken by sonication. The soluble and insoluble fractions were separated by centrifugation at 4000 rpm, 4 °C for 15 min. The soluble fraction was transferred to new tubes whereas the insoluble fraction was resuspended in Tris-HCl buffer pH 7.0. Both fractions were analyzed by SDS-PAGE to detect the recombinant protein in the protein profiles.

The 28_SFR2 protein band in the insoluble fraction was excised from the polyacrylamide gel. The band was soaked in elution buffer containing 250 mM Tris-HCl, pH 7.0, 100 mM NaCl and 0.1% SDS overnight, then 0.1 % SDS and NaCl were removed by dialysis against Tris-HCl, pH 7.0, overnight. The remaining solution in the dialysis bag was assessed for the concentration by

Bradford's method (1976) and kept at 4°C until further used as antigen for antibody production.

#### 3.3.2.1.2 Rabbit injection

Two New Zealand white rabbits were used for antisera production. The injection schedule was done as shown in Table 3.3. Each rabbit was injected (subcutaneously) 4 positions in the back. The sera from the final day were used to confirm the specificity and the antibody concentration by ELISA.

**Table 3.3** Antigen injection and polyclonal antibody collection schedule.

| Days | Activities |
|---|---|
| 0 | Pre-immune bleed and immunized with 100 - 200 µg/rabbit antigen in Complete Freund's Adjuvant (CFA) |
| 14 | Boost with 50-100 µg/rabbit antigen in Incomplete Freund's Adjuvant (IFA) |
| 28 | Boost with 50-100 µg/rabbit antigen in IFA |
| 35 | Test bleed with ELISA |
| 42 | Boost with 50-100 µg/rabbit antigen in IFA |
| 49 | Production bleed and ELISA |
| 52 | Final bleed |

### 3.3.2.2 Antisera preparation

Antisera used for western blot analysis was produced by rabbit injection. The C-terminal part of the recombinant SFR2 was used as antigen to produced antisera as described in the 3.3.2.1.2. Before use in western blots, the SFR2

antisera were partially purified. The empty vector, pET32a was transformed into BL21(DE3) p*lys*S and 10 ml cultures were induced and express thioredoxin protein. Then, cell pellets were harvested by centrifugation at 4000 rpm 10 min 4 °C. The lysis buffer was added and the cell suspension was sonicated. The soluble fraction was collected and 20 µl of SFR2 antisera were added into 1 ml cell lysates. The reaction was incubated overnight at 37 °C. The SFR2 antisera were separated by centrifugation at 14000 rpm for 5 min and supernatant was collected and used as antisera.

**3.3.2.3 Western blot analysis**

Induced *E. coli* cells carrying SFR2 expressing plasmids were harvested and loaded on SDS-PAGE. The polyacrylamide gel was soaked in transfer buffer containing 10% methanol and 1X Tris-glycine for 15 min. The PVDF membrane was soaked in absolute methanol before used and soaked with transfer buffer. Six pieces of 3M Whatman paper were soaked in transfer buffer for 15 min. The proteins were transferred to PVDF membrane by Semi-Phor (Hoefer Science) for 1 h at 400 mA, according to manufacture's protocol. The membrane was blocked with 3% BSA in phosphate buffered saline (PBS) with 0.05% Tween-20 overnight then 1 ml of partially purified antisera was added to the membrane and incubated for 2-3 h. The membrane was washed with PBST (PBS with 0.05% Tween-20) 3 x 15 min and then 1:20,000 diluted secondary antibody (anti-rabbit IgG) in PBST was added to the membrane. Then, the membrane was washed with PBST 3 x 15 min. One milliliter of 1X alkaline phosphatase substrate (NBT/BCIP) was added and incubated in the dark for 5 min to develop the color then the membrane was wash with DI water to stop a reaction and allowed to dry.

## 3.4 RNA detection

### 3.4.1 Materials

#### 3.4.1.1 Chemicals

*E. coli* RNA was extracted using the RNeasy Mini Kit (QIAGEN) and RNA was detected by DIG High Prime DNA labeling and Detection Starter Kit I (Roche). Diethyl pyrocarbonate (DEPC) used as an RNase inactivator was purchased from Bio-Basics. Nylon membrane, positively charged for used in northern blots was purchased from Roche.

### 3.4.2 Methodologies

#### 3.4.2.1 RNA extraction

The pET plamids containing *SFR2* gene were transformed into *E. coli* strain Origami(DE3) by electroporation. The positive clone was selected and inoculated into 3 ml LB broth containing 15 µg/ml of kanamycin, 12.5 µg/ml of tetracycline and 100 µg/ml of ampicillin and cultured at 37 °C overnight. Then 200 µl of starter was transferred to 10 ml fresh LB medium containing antibiotics, cultured at 37 °C for 3 h, and transferred to 20 °C for 30 min. The IPTG was added 0.3 mM final concentration for protein induction for 24 h.

At the end of the cultivation, Total RNA was extracted from these *E. coli* cells by the RNeasy Mini Kit procedure (QIAGEN). Briefly, both induced and uninduced *E. coli* cells were diluted to $OD_{600}$ of 0.6 with LB medium and cells were harvested by centrifugation at 4000 rpm 4 °C for 10 min. Then, cells were resuspended in 600 µl RLT buffer from the kit containing 1% β-mercaptoethanol

(ME). The cells were broken by sonication. The suspension was centrifuged again at 14,000 rpm for 2 min at room temperature. The supernatant was transferred to a new microcentrifuge tube. Five hundred microliters of absolute ethanol was added to the supernatant and mixed well by pipetting. The samples, including any precipitated form, were transferred to RNeasy spin columns and centrifuged for 30 s at 12,000 rpm. The flow-through was discarded and 700 µl of RW1 buffer was added to the column, which was centrifuged for 30 s at 12,000 rpm. The flow-through was discarded then 500 µl of RPE buffer was added to each column and it was centrifuged for 30 s to wash the membrane. The flow-through was discarded and 500 µl of RPE buffer was added to each column and it was centrifuged for 2 min at 12,000 rpm. The flow-through was discarded and the columns were centrifuged again to remove ethanol residue. In the final step, the columns were placed in new 1.5 ml collection tubes, then 50 µl of RNase-free water was added directly to the membrane, and the RNeasy spin columns were centrifuged for 1 min at 12,000 rpm to elute RNA. The total RNA was detected by electrophoresis on a 1% formaldehyde agarose gel (FA gel) in 1X FA gel buffer (20 mM MOPS, 5 mM sodium acetate and 1 mM EDTA with the pH adjusted to 7.0 with NaOH).

#### 3.4.2.2 Northern blot analysis

##### 3.4.2.2.1 RNA transfer

After total RNA was extracted from *E. coli* cells. The RNA was electrophoresed on a 1% FA gel in 1X FA gel buffer at 50 V for 2 h. Then, the FA gel was soaked in 20X SSC 2 times, 15 min each, to remove the formaldehyde. A positively charged nylon membrane (Roche) was cut to the size of

the gel and equilibrated in 2X SSC for 15 min. Three pieces of 3M Whatman paper were soaked with 20X SSC for 15 min for used as the wick. Another 3 pieces of 3M Whatmam paper were soaked with 2X SSC for 15 min. The capillary transfer method was set up to transfer RNA (Fig. 3.3). Briefly, the reservoir containing 20X SSC was prepared then a tray was placed in the middle of the reservoir. The wick was laid on the top of the tray with the ends dipping into the reservoir. The gel was laid with top side facing down and nylon membrane was placed on top of the gel and the bubbles between the gel and the membrane removed. The 3 pieces of 3M Whatman that were soaked in 2X SSC were laid on the top of the membrane. Three inches of cut-to-size paper towel were stacked on top of the 3M Whatman paper. A glass plate was placed on the top and a weight placed on top of of the whole setup (Streit *et al.*, 2008). The RNA was allowed to transfer overnight.

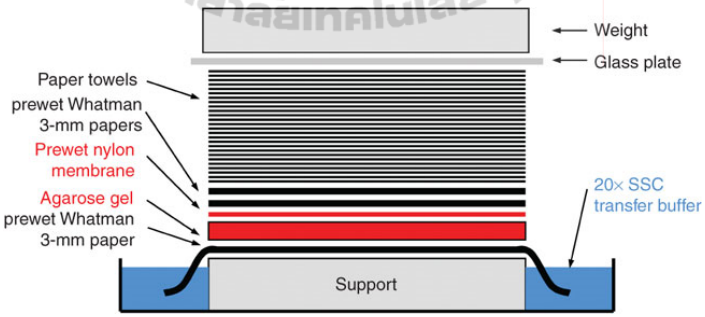



**Figure 3.3** Upward capillary transfer. Capillary transfer of nucleic acids from an agarose gel to solid supports is achieved by drawing the transfer buffer from the reservoir upward through the gel into the stack of paper towels (Streit *et al.*, 2008).

At the end of the transfer, the RNA was crosslinked to the membrane with UV light (365 nm) for 3 min and the membrane was rinsed briefly with distilled water treated with DEPC, then allowed to air dry.

**3.4.2.2.2 Probe preparation**

The purified PCR products of 28_SFR2 and 34_SFR2_delC were used as templates to synthesize DNA-labeling probe following the protocol of DIG High Prime DNA labeling and Detection Starter Kit I (Roche). The 28_SFR2 cDNA fragment was used as DNA probe for N-terminal deletion RNA and 34_SFR2_delC cDNA fragment was used as DNA probe for C-terminal deletion RNA. The 28_SFR2 cDNA fragment was amplified with the 28_SFR2_F and 30_SFR2_R primer pair. The 34_SFR2_delC was amplified with the 27_SFR2_F and 28_SFR2_R primers. The PCR conditions were as shown in Table 3.2. Both PCR products were purified with the gel extraction kit (QIAGEN). The concentrations of the purified PCR products were measured with a Nano Drop spectrophotometry.

The DNA probe was labeled *in vitro* with Digoxigenin-11-UTP. The probe was synthesized by adding 100 ng of purified PCR products and distilled water to final volume of 8 µl. DNA was denatured by boiling for 10 min and quickly chilling on ice. Dig-High Prime was mixed thoroughly and 2 µl were added to the denatured DNA then the reaction was incubated at 37 °C overnight. The reaction was stopped by heating to 65 °C for 10 min. The DNA probe was used immediately for hybridization.

**3.4.2.2.3 Hybridization**

The DIG easy hyb granule was reconstituted by adding 64 ml of MQ water treated with DEPC and incubated at 37 °C with shaking to dissolve. Appropriated volumes of 10 ml for prehybridization and 3 ml for hybridization of DIG easy hyb were pre-heat at 50 °C for 30 min. The membrane was placed into a 50 ml conical tube and 10 ml of pre-heated DIG easy hyb was added for prehybridization with gentle agitation at 50 °C for 30 min. DIG-labeled DNA probe (3.4.2.2.2) was denatured by boiling for 5 min and rapidly cooled in ice water. The 10 µl of denatured probe was added directly to 3 ml pre-heated DIG easy hyb and mixed well (avoiding foaming, bubbles). Then the prehybridization mixture was poured off and the probe/hybridization mixture was added to the membrane and incubated at 50 °C overnight with gentle agitation.

**3.4.2.2.4 Immunological detection**

The hybridization mixture was removed, then membrane was washed with 2X SSC with 0.1% SDS 2 times, 10 min each, at 25 °C under constant agitation, followed by 0.5X SSC with 0.1% SDS 2 times, 30 min each, at 60 °C with gentle agitation. After hybridization and wash, the membrane was removed from the conical tube and placed on a sterile petri dish, then rinsed briefly with 1X washing buffer 2 times, 10 min each. Next, the membrane was blocked with 20 ml of 1X blocking solution for 2 h. The membrane was then incubated in 5 ml antibody solution for 1 h. The antibody solution was prepared by centrifuging Anti-Digoxigenin-AP solution for 5 min at 10,000 rpm in the original vial and pipeting 1 µl from the surface with Anti-Digoxigenin-AP, diluted 1:5,000 in blocking solution. The

membrane was washed twice with 1X washing buffer for 30 min and the membrane was equilibrated in 1X detection buffer for 5 min. Then 1 ml of 1X color substrate solution was added to the membrane in the dark and it was left for the color to develop for 1 h. Then the membrane was wash with MQ water to stop the reaction.

# CHAPTER IV

# RESULTS AND DISCUSSION

## 4.1 Cloning

### 4.1.1 Prokaryotic expression vectors

#### 4.1.1.1 pET System

##### 4.1.1.1.1 Cloning of N-terminal SFR2 deletion

Five different primer pairs (Table 3.2) were used to amplify different truncations of the SFR2 gene to generate the AK_SFR2, 26_SFR2, 27_SFR2, 28_SFR2 and 29_SFR2 cDNA with the expected sizes, which were 1629 bp, 1464 bp, 1164 bp, 780 bp and 360 bp, respectively (Fig 4.1). The AK_SFR2, 26_SFR2 and 27_SFR2 fragments contained the coding sequence for the highly conserve SFR2 active site motif, TFNEP and I/VTENG (Thorlby *et al.*, 2004) and the PEST sequence. The 28_SFR2 fragment contained the coding sequence for the conserve peptide active site motif but not the PEST sequence. The 29_SFR2 contained the SFR2 active site motif I/VTENG but not PEST sequence.

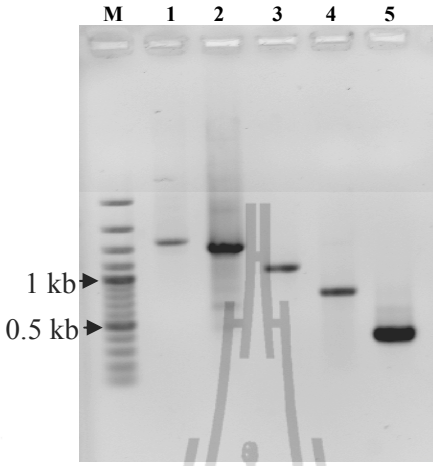



**Figure 4.1** Agarose gel electrophoresis of PCR products of lane 1: AK_SFR2, lane 2: 26_SFR2, lane 3: 27SFR2, lane 4: 28_SFR2 and lane 5: 29_SFR2 and lane M: 100 bp marker.

All purified PCR products were cloned into pENTR-D/TOPO and transformed into TOP10 by the heat-shock method. The positive clones of each construct were selected for plasmid preparation. Doubles restriction enzymes digests were performed on TOPO_AK_SFR2, TOPO_26_SFR2, TOPO_27_SFR2 and TOPO_28_SFR2 with *Not*I and *Hind*III to generate 2 DNA fragments with the expected sizes of 2936/1273 bp, 2927/1117 bp, 2924/820 bp and 2936/424 bp, respectively (Fig. 4.2 lane 1-4). The TOPO_29_SFR2 was cut with *Not*I and *Asc*I to give the products size of 2580 bp and 360 bp (Fig. 4.2 lane 5).

The results indicated that different size of the *SFR2* gene were amplified and cloned into pENTR-D/TOPO. Next, LR clonase enzyme (Invitrogen) was used to transfer SFR2 deletion fragments to pET32a/DEST. The LR reactions were transformed into DH5α, then colonies containing recombinant plasmids were screened using colony PCR technique with gene specific primers. The positive clones were selected for plasmid extraction and sequencing to verify the

sequence. The analysis of the nucleotide sequences were done and correct nucleotide sequences with in-frame translated products were obtained in all SFR2 deletions (Appendix I)

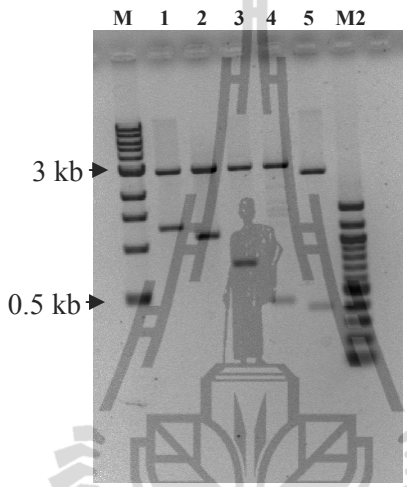



**Figure 4.2** Agarose gel electrophoresis of TOPO containing N-terminus deletion fragments of *SFR2* digested with *Not*I and *Hind*III (lane 1-4) and *Not*I and *Asc*I (lane 5). Lane M: 1kb marker, Lane 1: TOPO_AK_SFR2, Lane 2: TOPO26_SFR2, Lane 3: TOPO27_SFR2, Lane 4: TOPO28_SFR2, Lane 5: TOPO29_SFR2 and Lane M2: 100 bp marker.

#### 4.1.1.1.2 Cloning of C-terminal SFR2 deletion

Another four primer pairs (Table 3.2) were used to amplify the *SFR2* gene fragments to generate 31_SFR2_delC(mP), 32_SFR2_delC(mP), 33_SFR2_delC and 34_SFR2_delC (Fig. 4.15). The expected sizes of the above fragments were 522 bp, 165 bp, 714 bp and 357 bp, respectively (Fig. 4.3). In both 31_SFR2_delC(mP) and the 32_SFR2_delC(mP), the PEST sequence at the C-terminal were mutated (mP).

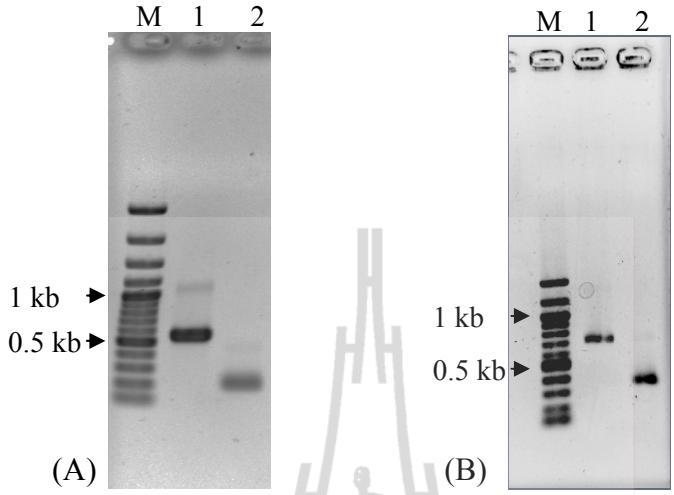



**Figure 4.3** Agarose gel electrophoresis of PCR products (A) Lane M: 100 bp marker**,** Lane 1: 31_SFR2_delC(mP), Lane 2: 32_SFR2_delC(mP). (B) Lane M: 100 bp marker**,** Lane 1: 33_SFR2_delC, Lane 2: 34_SFR2_delC.

The PCR products were purified and cloned into pENTR-D/TOPO to obtain TOPO_31_SFR2_delC(mP), TOPO_32_SFR2_delC(mP), TOPO_33_SFR2_delC and TOPO_34_SFR2_delC. The first two were cut with *Pst*I to generate the linear plasmids with the expected sizes of 3102 bp and 2745 bp, respectively (Fig. 4.4A). The later two were cut with *Nco*I and *Sac*I to give 2 DNA fragments with the expected sizes of 2765/529 bp and 2765/172 bp, respectively (Fig. 4.4B).

The results indicated that the gene encoding the N-terminal region of SFR2 could be amplified and cloned into the cloning vector, pENTR-D/TOPO. The plasmids were extracted from the positive clone and DNA sequencing were done. The nucleotide sequences showed no mutation. The LR reactions were performed to transferred the inserts into the destination vector, pET32a/DEST. The

LR reaction products were transformed into *E. coli* DH5α. The positive clones were selected by the colony PCR technique. Eighteen colonies of ampicillin resistance were screened, almost all of them showed positive bands.

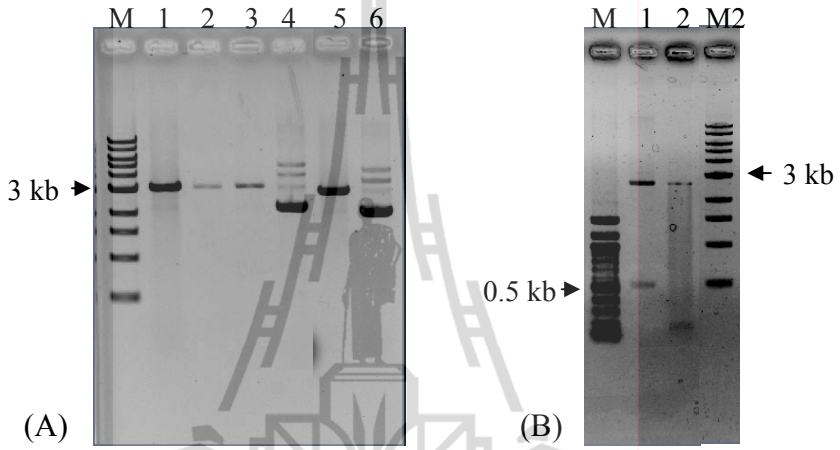



**Figure 4.4** Agarose gel electrophoresis of pENTR-D/TOPO containing 31_SFR2_ delC(mP), 32_SFR2_delC(mP) cut with *Pst*I, 33_SFR2_delC and 34_SFR2_delC cut with *Nco*I and *Sac*I. (A) Lane M: 100 bp marker**,** Lane 1-3: digested TOPO_31_SFR2_delC(mP), Lane 4: uncut TOPO_31_ SFR2_delC(mP), Lane 5: digested TOPO_32_SFR2_ delC9(mP) and Lane 6: uncut TOPO_32_SFR2_delC(mP). (B) Lane M: 100 bp marker**,** Lane 1: Digested TOPO_33_SFR2_delC, Lane 2: digested TOPO_34_SFR2_delC and Lane M2: 1 kb marker.

#### 4.1.1.1.3 Cloning of 26_SFR2_Syn

The 26_SFR2_Syn codons were optimized for expression in *P. pastoris* without changing the amino acids sequences. The PCR method was used to amplify the 26_SFR2_Syn gene (Fig. 4.15) with the 26_SFR2_Syn_F and 30_SFR2_Syn_R primers. The resulting product was 1464 bp (Fig. 4.5). The PCR

product was purified and cloned into pENTR-D/TOPO and designate as TOPO_26_SFR2_Syn. The plasmids were cut with *Not*I and *Hind*III to generate 2 DNA fragments with the expected sizes of 3151 bp and 890 bp (Fig 4.6).

The nucletotide sequences in pENTR-D/TOPO containing 26_SFR2_Syn was analyzed and no mutations were found. A pET_26_SFR2_Syn was constructed by transferring the 26_SFR2_syn from TOPO_26_SFR2_Syn with LR clonase enzyme and the LR reaction was transformed into *E. coli* DH5α. Colony PCR screening twenty-four ampicillin resistance clones showed all of them had the positive band. The pET_26_SFR2_Syn plasmid was extracted from one positive clone and used for protein expression. The TOPO_26_SFR2_Syn plasmid was also used as entry clone for cloning 26_SFR2_syn into pPICZαBNH8/DEST with a method similar to that described.

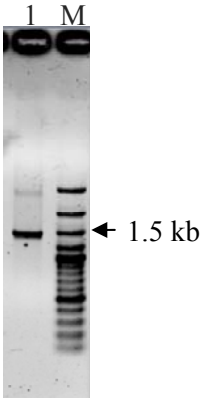



**Figure 4.5** Agarose gel electrophoresis of PCR product of the 26_SFR2_Syn. Lane 1: purified 26_SFR2_Syn PCR product and Lane M: 100 bp marker.

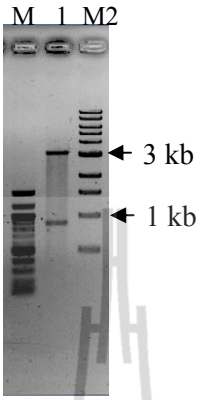



**Figure 4.6** Agarose gel electrophoresis of TOPO_26_SFR2_Syn plasmid cut with *Not*I and *Hind*III. Lane M: 100 bp marker, Lane 1: digested TOPO_SFR2 _syn and Lane M2: 1 kb marker.

#### 4.1.1.1.4 Cloning of 26_ mutPEST_SFR2

**(a) PEST sequence of SFR2 gene**

The algorithm, PESTfind (http://www.es.embnet.org/ Services/EMBnetAT/htdoc/PESTfind/) searches for hydrophilic regions of 12 or greater amino acids that contain at least one P (proline), one E (glutamic acid) or D (aspartic acid), and one S (serine) or T (threonine), flanked by K (lysine), R (arginine), or H (histidine) residues. The algorithm assigns a score to each possible PEST sequence found. The score ranges from -50 to +50, with a score above zero denoting a possible PEST region while a value greater than + 5 being of particular interest (Chen et al., 2002). In this study, a PEST sequence with the PEST score of +5.01 (amino acid residues 216 - 227) was found with the hydrophobicity index of 35.57. Within 11 residues PEST region, there are 5 charged amino acids (1 arginine and 4 glutamic acids), 1 polar amino acid (1 thereonine) and 5 non-polar amino acids (2 leucine, 1 methionine and 2 prolines). The low hydrophobicity suggested that the

region may be surface accessible to proteases or for protein-protein interaction with other proteins such as molecular chaperones, trafficking proteins, or components of proteolytic systems (Chen et al., 2002).

**(b) Cloning of 26_ mutPEST_SFR2**

The expression of N-terminal deletions of the *SFR2* gene in *E. coli* showed that only recombinant 28_SFR2 and 29_SFR2 were able to produce SFR2 protein in sufficient quantity to be detected on Coomassie brilliant blue stained SDS-PAGE (Fig. 4.18). Both gene deletions did not contain the coding sequence for the PEST sequence (Fig. 4.15). Thus, a hypothesis proposed from this result was that the PEST sequence in SFR2 protein had some effected on the protein expression in *E. coli*.

The four glutamic acids (E) in the PEST sequence were mutated to glutamine (Q) by PCR mutagenesis with TOPO_AK_SFR2 as template. The 26_SFR2_F and 34_mutPEST_SFR2_R primers were used to amplify the first part (5' region) of the SFR2 of 540 bp (Fig. 4.7A). For the second part (3' region), the 33_mutPEST_SFR2_F and 30_SFR2_R primers were used and the expected size, 981 bp (Fig. 4.7A), PCR product was obtained. Then, the full length 26_mutPEST_SFR2 was generated by annealing the two parts together and amplifying with the 26_SFR2_F and 30_SFR2_R primers. The PCR product of 1464 bp was obtained from two-steps PCR (Fig. 4.7B). On the other hand, separating the annealing step of the two fractions and amplification with the outer primers (26_SFR2 and 30_SFR2) resulting in a smeared product with the size similar to the full length 26_SFR2 (Fig. 4.8).

Although this technique is elegantly simple in principle, there are often problems with amplifying a final PCR product of a specific size. In many cases, the amplified products are found to be heterogeneous in size and migrate as smears on agarose gels. Even increasing the amount of template DNA does not reliably improve the yield of the final PCR product (Aiyar *et al.*, 1996). As alternative method (Ho *et al.*, 1989; Wurch *et al.*, 1998) has been described, in this method the overlap fragment are mixed, and a full-length DNA containing the desired mutation is amplified, using 2 outer primers. The result showed a sharper band (Fig. 4.7B).

The mutated PCR products were cloned into pENTR-D/TOPO to obtain TOPO_26_mutPEST_SFR2. The plasmids were cut with *Hind*III to check. They generated linear plasmids with the expected size of 4044 bp (Fig. 4.9). The analysis of nucleotides sequence indicated that the nucleotides encoding the amino acid position 220 and 221 glutamic acids were mutated to glutamine. However, the codons for the amino acids at positions 224, 225 and 226 (counted from the ATG start codon) were missing with unknown reason (Fig. 4.10). However, the sequence still showed in frame protein translation. This new mutated sequence was analyzed again with the PESTfind program and the result showed no PEST sequence property was found in this new mutated SFR2 protein.

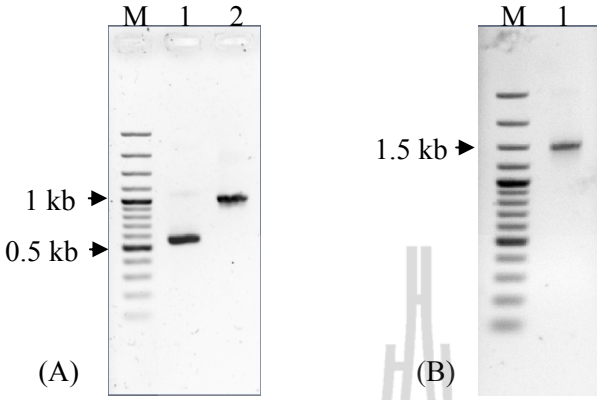



**Figure 4.7** Agarose gel electrophoresis of purified mutagenesis PCR products (A) the first part and the second part amplified for the PEST sequence mutation. (B) full length 26_mutPEST_SFR2 PCR product. (A) Lane M: 100 bp marker, Lane 1: purified first (5') part of the *SFR2* gene (N-terminal), Lane 2: purified second (3') part of the *SFR2* gene (C-terminal). (B) Lane M: 100 bp marker, Lane 1: purified full-length 26_mutPEST_SFR2 PCR product.

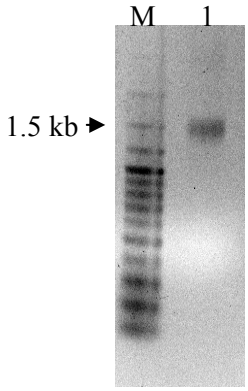



**Figure 4.8** Agarose gel electrophoresis of PCR product of full-length 26_mutPEST _SFR2 PCR product without outer primers (26_SFR2 and 30_SFR2) in the PCR reaction. Lane M: 100 bp marker, Lane 1: full length of the 26_mutPEST_SFR2 annealing reaction.

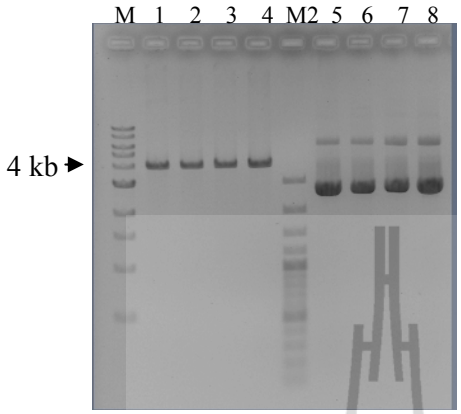



**Figure 4.9** Agarose gel electrophoresis of TOPO_26_mutPEST_SFR2 plasmid. Lane M: 1 kb marker, Lane 1-4: TOPO_26_ mutPEST_SFR2 cut with *Hind*III (different clones), Lane M2: 100 bp marker, Lane 5-8: uncut plasmids (different clones).

```
26_SFR2             DFRAFPSSDADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDA
26_mutPEST_SFR2     DFRAFPSSDADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDA
                    ************************************************************

26_SFR2             LMASAAGDGGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSH
26_mutPEST_SFR2     LMASAAGDGGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSH
                    ************************************************************

26_SFR2             NVAAWHNVPCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNF
26_mutPEST_SFR2     NVAAWHNVPCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPQQPT---KSSVNF
                    *********************************************::**    ******

26_SFR2             AALERYRWIIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVS
26_mutPEST_SFR2     AALERYRWIIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVS
                    ************************************************************

26_SFR2             NLVDYWVIFNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAY
26_mutPEST_SFR2     NLVDYWVIFNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAY
                    ************************************************************

26_SFR2             DYIHSKSKNERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGIN
26_mutPEST_SFR2     DYIHSKSKNERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGIN
                    ************************************************************

26_SFR2             YYGQEVISGPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDE
26_mutPEST_SFR2     YYGQEVISGPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDE
                    **********************************************************

26_SFR2             TDLIRKPYILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLA
26_mutPEST_SFR2     TDLIRKPYILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLA
                    ************************************************************

26_SFR2             RKPRPSYFLFS
26_mutPEST_SFR2     RKPRPSYFLFS
                    ***********
```

**Figure 4.10** The amino acids sequence alignment of the 26_SFR2 and 26_ mutPEST _SFR2 cDNA. The PEST region is highlight in grey. The mutated amino acids are shown in bold letters. The amino acid sequences were aligned by ClustalW2. The conserved sequence VITENG and VIFNEP are underlined.

The results indicated that the 26_mutPEST_SFR2 cDNA was amplified using 26_SFR2_F and 30_SFR2_R primers and successfully cloned into pENTR-D/TOPO. Although the sequencing result showed deletion of some amino acids in the PEST sequence, the resulting 26_mutPEST_SFR2 did not have a PEST sequence property and the downstream sequences were in-frame. The TOPO_26_mutPEST_SFR2 was used as an entry clone to transfer the insert into pET32a/DEST. The LR reaction was transformed into DH5α by electrophoration. The

positive clones were detected by the colony PCR technique and the correct clone was designated as pET_26_mutPEST_SFR2.

#### 4.1.1.2 pCold I System

##### 4.1.1.2.1 Cloning of 37_SFR2

The 37_SFR2 cDNA was amplified with the 37_SFR2_KpnI_F and 32_SFR2_SalI_R primers using TOPO_26_SFR2 as template to generate the PCR products with the expected size of 1410 bp (Fig. 4.11A).

The PCR products were purified and cloned into the pGEM-T Easy vector, then the pGEM containing 37 _SFR2 was cut with *Kpn*I and *Sal*I, the insert purified and ligated into the precut pCold I vector between the *Kpn*I and *Sal*I cloning sites (Fig. 4.12). The ligation reaction was transformed into *E. coli* DH5α. The resulting plasmid was named pCold_37_SFR2. The plasmids were extracted from the positive clones that were confirmed by the colony PCR technique. The pCold_37_SFR2 was cut with *Pst*I to generate two DNA fragments to verify the direction of the insert with the expected sizes of 4691 bp and 1149 bp (Fig. 4.13). DNA sequencing was used to confirm the proper construct and the translation product was shown to be in-frame and no mutations were found.

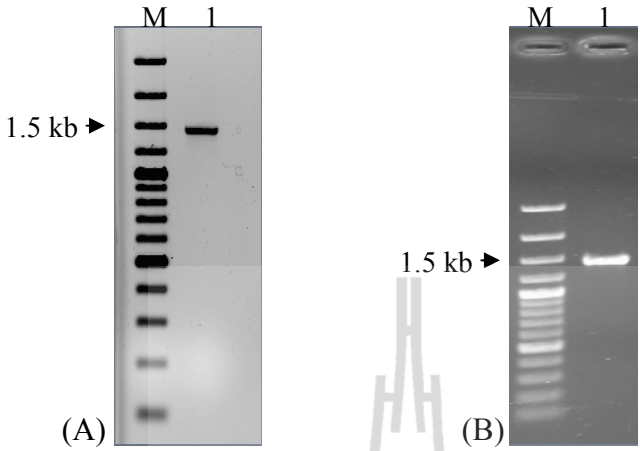



**Figure 4.11** Agarose gel electrophoresis of the purified 37_SFR2 and 37_mutPEST _SFR2 cDNA for cloning into pGEM-T Easy vector. (A) Lane M: 100 bp marker and Lane 1: purified 37_SFR2. (B) Lane M: 100 bp marker and Lane 1: purified 37_mutPEST_SFR2.

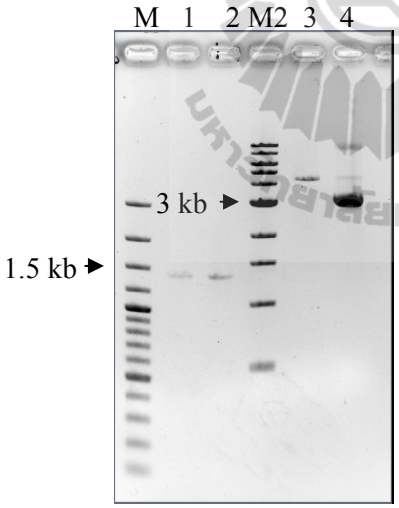



**Figure 4.12** Agarose gel electrophoresis of purified 37_SFR2 and 37_mutPEST _SFR2 and pCold I expression vector cut with *Kpn*I and *Sal*I. Lane M: 100 bp marker, Lane 1: *Kpn*I-*Sal*I excised 37_SFR2 cDNA, Lane 2: *Kpn*I-*Sal*I excised 37_mutPEST_SFR2 cDNA, Lame M2: 1 kb marker, Lane 3: *Kpn*I-*Sal*I digested pCold I vector and Lane 4: uncut pCold I vector.

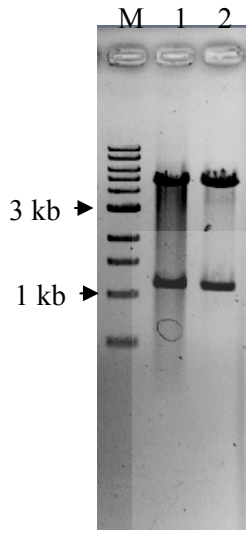



**Figure 4.13** Agarose gel electrophoresis of pCold_37_SFR2 and pCold_37_ mutPEST_SFR2 cut with *Pst*I. Lane M: 1 kb marker, Lane 1: Digested pCold_37_SFR2 and Lane 2: Digested pCold_37_ mutPEST_SFR2.

**4.1.1.2.2 Cloning of 37_mutPEST_SFR2**

The 37_mutPEST_SFR2 fragment was amplified from the 26_mutPEST_SFR2 PCR product, rather than not the TOPO_26_mutPEST_SFR2. The 37_SFR2_KpnI_F and the 32_SFR2_SalI_R were used as primers. The expected size of this fragment was 1410 bp (Fig. 4.11B)

The PCR products were purified and cloned into the pGEM-T Easy vector, then the pGEM_37_mutPEST_SFR2 was cut with *Kpn*I and *Sal*I to release the 37_mutPEST_SFR2 insert, which was ligated into pCold I between the *Kpn*I and *Sal*I cloning sites (Fig. 4.12). The obtained plasmid was named pCold_37_mutPEST _SFR2. The pCold_37_mutPEST_SFR2 was cut with *Pst*I to generate two DNA fragments of 4691 bp and 1149 bp to verify the direction of the insert (Fig. 4.13). Nucleotide sequences confirmed that the PEST sequences were mutated. The sequencing results indicated that the glutamic acid (E) at position 220,

221 and 224 were mutated to glutamine (Q) but the glutamic acid at position 225 was not changed when (Fig. 4.14 arrow). The 37_mutPEST_SFR2 was analyzed again with PESTfind program and the result did not predict it to be a PEST sequence.

```
37_SFR2             LATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGDGGSQQSWRSTGGE
37_mutPEST_SFR2     LATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGDGGSQQSWRSTGGE
                    ************************************************************

37_SFR2             NIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNVPCPQERLRFWSDP
37_mutPEST_SFR2     NIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNVPCPQERLRFWSDP
                    ************************************************************

37_SFR2             DAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRWIIQRVREYGMKVM
37_mutPEST_SFR2     DAELKLAKETGISVFRMGVDWARLMPQQPTQELKSSVNFAALERYRWIIQRVREYGMKVM
                    ************************::**:****************************::

37_SFR2             LTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVIFNEPHVFVMLTYC
37_mutPEST_SFR2     LTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVIFNEPHVFVMLTYC
                    ************************************************************

37_SFR2             AGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSKNERKPIVGVAHHV
37_mutPEST_SFR2     AGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSKNERKPIVGVAHHV
                    ************************************************************

37_SFR2             SFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVISGPGLKLVDNDEYS
37_mutPEST_SFR2     SFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVISGPGLKLVDNDEYS
                    ************************************************************

37_SFR2             ESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPYILEHLLATYAAII
37_mutPEST_SFR2     ESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPYILEHLLATYAAII
                    ************************************************************

37_SFR2             MGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYFLFS
37_mutPEST_SFR2     MGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYFLFS
                    ***************************************************
```

**Figure 4.14** The amino acids sequence alignment of 37_SFR2 and 37_mutPEST _SFR2. The PEST region is highlight in grey. The mutated amino acids were shown in bold letters. The E225 that was not changed is indicated with an arrow. The amino acid sequences were aligned by ClustalW2.

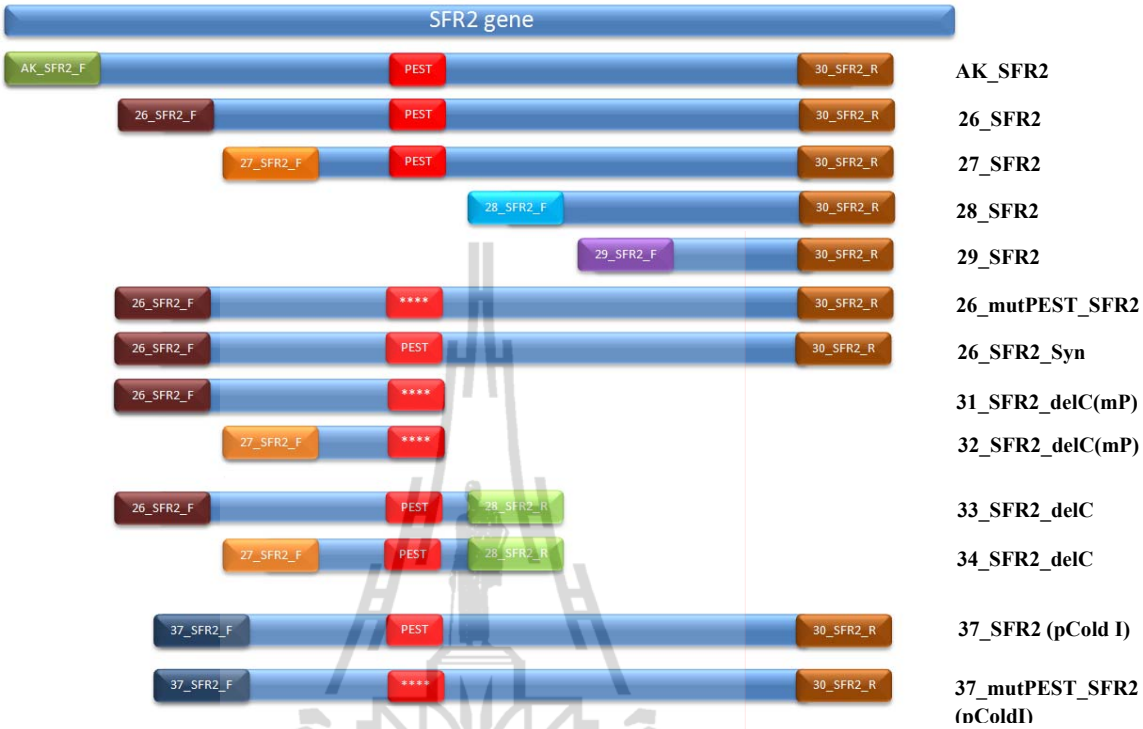



**Figure 4.15** Summary of all SFR2 fragments cloned into pET32a and pCold I to test their protein expression in *E. coli*.

### 4.1.2 Eukaryotic expression vector

#### 4.1.2.1 pPICZαBNH8 System

##### 4.1.2.1.1 Cloning of 26_SFR2_syn into pPICZαBNH8

The TOPO_26_SFR2_syn was used as an entry clone to transfer the 26_SFR2_Syn to the pPICZαBNH8/DEST (Luang *et al*., 2010) with the LR clonase reaction. The LR reaction was transformed into *E. coli* DH5α. The positive clones were selected by colony PCR techniques. The pPICZαBNH8_ 26_SFR2_Syn was extracted and nucleotide sequence analysis was done to identify the 26_SFR2_Syn Sequencing indicated that the 26_SFR2_Syn cDNA was inserted in

the correct reading frame and no mutation was found (Fig. 4.16). Then the pPICZαBNH8_ 26_SFR2_Syn was linearized and transformed into *P. pastoris* SMD1168H by the electroporation method. The yeast colony PCR method was used to screen for positive clones with the 5' AOX and 3' AOX primers. Screening showed the positive band of 2044 bp and a 500 bp band of the endogenous AOX1 gene (Fig. 4.17).


**Figure 4.16** Translated DNA sequence of pPICZαBNH8_ 26_SFR2_Syn. The DNA sequence was translated by Expasy translation tool.

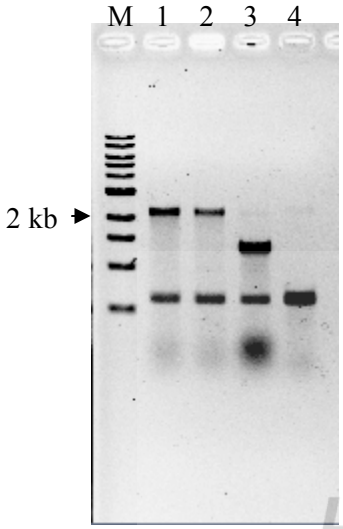



**Figure 4.17** Agarose gel electrophoresis of pPICZαBNH8_ 26_SFR2_Syn colonies screened by colony PCR with the 5' AOX and 3' AOX primers. Lane M: 1 kb marker, Lane 1: colony 1 of pPICZαBNH8_ 26_SFR2_Syn, Lane 2: Colony 2 of pPICZαBNH8_ 26_SFR2_Syn, Lane 3: Positive control (pPICZαBNH8_$Ek_L$), Lane 4: Negative control of untransformed yeast cells.

## 4.2 Protein expression

### 4.2.1 Protein expression in *E. coli*

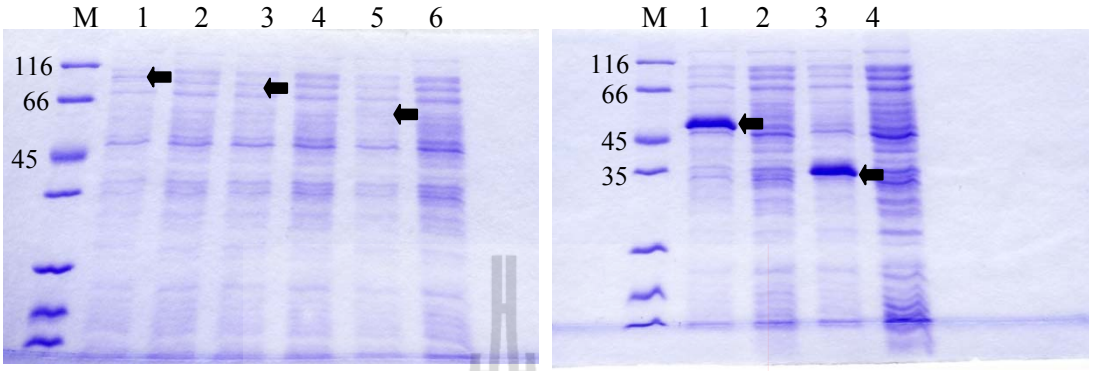
#### 4.2.1.1 pET System

##### 4.2.1.1.1 Expression of N-terminal SFR2 deletion

Recombinant pET plasmid that contained the AK_SFR2, 26_SFR2, 27_SFR2, 28_SFR2 and 29_SFR2 cDNA fragments were transformed into BL21(DE3) p*lys*S cells by electroporation method. Then, positive clones were selected by colony PCR. Two hundred microliters of starter cells were inoculated in

10 ml LB broth containing ampicillin. The cultures were incubated at 37 °C 200 rpm for 3 h at $OD_{600}$ reached to 0.6. After 3 h of cultivation, cultures were transferred to 20 °C 200 rpm for 30 min and IPTG was added to 0.3 mM final concentration and the cultures were further incubated at 20 °C for 16 h induction. At the end of the cultivation, cells were broken by sonication and total proteins were detected on SDS-PAGE. The protein expression results indicated that the cells that carried pET plasmids that contained the AK_SFR2, 26_SFR2, 27_SFR2 cDNA fragment did not show any intense band for SFR2 protein of the expected molecular weights of 83 kDa, 78 kDa and 65 kDa (include thioredoxin fusion tag), respectively (Fig. 4.18A). In contrast, pET plasmids containing 28_SFR2 and 29_SFR2 cDNA showed intense bands with molecular weights of 55 kDa and 40 kDa (including thioredoxin fusion tag), respectively (Fig. 4.18B).

These results indicated that only 260 amino acids at the C-terminal part (amino acids 284-544) of the SFR2 can be produced in *E. coli* in the pET32a system. However, the recombinant 28_SFR2 and 29_SFR2 proteins were found only in the insoluble fractions and no enzyme activities were detected.




(A) (B)

**Figure 4.18** SDS-PAGE analysis of total protein of BL21(DE3) p*lys*S cells induced to produce recombinant AK_SFR2, 26_SFR2, 27_SFR2, 28_SFR2 and 29_SFR2 proteins in for 16 h at 20 °C. (A) Lane M: protein marker, Lane 1: induced cells containing AK_SFR2, Lane 2: uninduced cells containing AK_SFR2, Lane 3: induced cells containing 26_SFR2, Lane 4: uninduced cells containing 26_SFR2, Lane 5: induced cells containing 27_SFR2, Lane 6: uninduced cells containing 27_SFR2. (B) Lane 1: induced cells containing 28_SFR2, Lane 2: uninduced cells containing 28_SFR2, Lane 3: induced cells containing 29_SFR2, Lane 4: uninduced cells containing 29_SFR2. The arrows indicated the expected size of recombinant SFR2 protein fragments.

#### 4.2.1.1.2 Expression of 26_mutPEST_SFR2

From the N-terminal deletion protein profile showed that only the C-terminal one third tail of the SFR2 that did not contain the PEST sequence region can be expressed in *E. coli*, while the proteins with the N-terminal part that contained the PEST sequence could not. Therefore, the PEST sequence in SFR2 was

investigated, because PEST sequences have been shown to be signals for proteolytic degradation of proteins (Rechstener et al., 1996).

To examine whether the PEST sequence directly affects the processing or proteolytic degradation of SFR2, pET plasmid containing 26_mutPEST_SFR2 (Fig. 4.15) was transformed into 4 different expression hosts, BL21(DE3) p*lys*S, Rosetta-gami(DE3), Origami(DE3) and Origami B(DE3). The positive clones of each expression host were chosen and inoculated into 10 ml LB broth containing appropriate antibiotic. The cells were cultured for 3 h at 37 °C and transferred to 20 °C for 30 min. IPTG was added to the final concentration of 0.3 mM to induce the protein expression for 16 h. At the end of the cultivation, cells were harvested and broken by sonication. Total protein from each cell type was load on SDS-PAGE to visualize the protein profile. The protein expression results did not show any intense protein band of 26_mutPEST_SFR2 with the expected molecular weight of 78 kDa (including thioredoxin fusion tag), which was the same result with the wild type 26_SFR2 in all expression hosts. However, the positive control, pET32a, expressed thioredoxin fusion tag with the molecular weight of 22 kDa as clearly be seen in fig 4.19 lanes 3, 6, 9 and 12.

The protein expression results suggested that SFR2 still could not express in *E. coli* even though the PEST sequence was partially mutated. Therefore, the protein degradation might not be the reason why the SFR2 protein were not expressed in *E. coli*, even though the PEST sequences have been shown to target proteins for rapid intracellular degradation (Chen *et al*., 2002). Rogers *et al*., 1986 have hypothesized that these negatively charged motifs could sequester calcium for activating calpains in the close vicinity of substrates. But some studies disagree with this hypothesis. Some researchers found that the PEST sequence was not a target site

for calpain protease (Molinari *et al.*, 1995; Carillo *et al.*, 1996). There are PEST sequence containing proteins that are stable. There are other factors that influence the degradation of intracellular proteins such as intracellular location or masking due to association with other macromolecules. There are many properties of proteins involved in PEST protein degradation, for example larger proteins are degraded faster than smaller ones and negatively charged proteins are degraded faster than positively charged proteins. Many rapidly degraded proteins are found in prokaryotic cells, such as Lambda proteins N and CII, which have half-lives of less than 5 minutes, and the *E. coli* heat shock σ factor is also highly unstable. However, none of these proteins contain a PEST sequence. Therefore, other structural features must affect protein proteolysis (Rogers *et al.*, 1986). The observation that the E1A and myc protein that contain many sites of PEST sequences can be overproduced in *E. coli* (Ferguson *et al.*, 1984) suggests that a PEST sequence does not necessarily signal rapid proteolysis in bacteria (Rogers *et al.*, 1986).

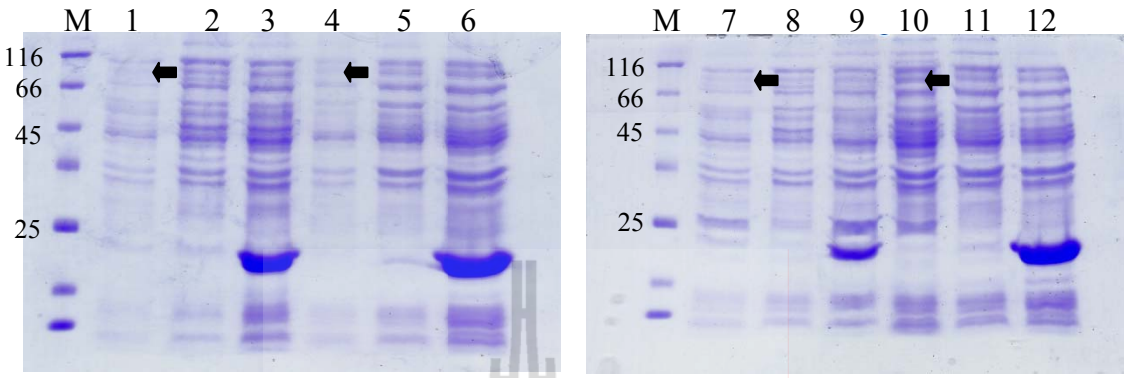



**Figure 4.19** SDS-PAGE analysis of total protein of recombinant 26_mutPEST _ SFR2 expressed in BL21(DE3) p*lys*S, Rosetta-gami(DE3), Origami(DE3) and Origami B(DE3) for 16 h at 20 °C. Lane M: protein marker, Lanes 1, 4, 7 and 10: induced 26_mutPEST_SFR2 in Origami(DE3), Origami B(DE3), Rosetta-gami(DE3) and BL21(DE3) plysS, respectively, Lanes 2, 5, 8, 11: non-induced 26_mutPEST_SFR2 in Origami(DE3), Origami B(DE3), Rosetta-gami(DE3) and BL2(DE3) p*lys*S, respectively, Lanes 3, 6, 9, 12: thioredoxin fusion tag as positive control expressed in Origami(DE3), Origami B(DE3), Rosetta-gami(DE3) and BL21(DE3) p*lys*S, respectively, from the same expression vector (pET32a). The arrows indicated the expected size of recombinant 26_muPEST_SFR2 protein.

#### 4.2.1.1.3 Expression of C-terminal SFR2 deletion

The pET plasmids containing 31_SFR2_delC(mP) and 32_SFR2_delC(mP) (Fig. 4.15) were transformed into Origami(DE3), Origami B(DE3) and Rosetta-gami(DE3). The pET plasmids containing 33_SFR2_delC and 34_SFR2_delC were transformed into Origami(DE3) and BL21(DE3) p*lys*S. A

positive clone of each expression host was selected and inoculated in 10 ml LB broth containing appropriate antibiotics. The cultures were incubated at 37 °C 200 rpm for 3 h. Cultures were transferred to 20 °C 200 rpm for 30 min then the cultures were induced to express protein with 0.3 mM IPTG for 16 h. At the end of the cultivation, cells were harvested and broken by sonication and total proteins were detected on SDS-PAGE. The protein expression results of recombinant 31_SFR2_delC(mP) and 32_SFR2_delC(mP) showed intense bands of both recombinant protein with the expected molecular weights of 42 kDa and 29 kDa, respectively (including thioredoxin fusion tag) in all expression hosts. But 32_SFR2_delC(mP) showed lower level of expression in Rosetta-gami(DE3) when compared to the other expression hosts (Fig. 4.20 and Fig. 4.21).

In contrast, the recombinant 33_SFR2_delC and 34_SFR2_delC protein did not show any intense band of target protein of the expected molecular weights 50 kDa and 37 kDa, respectively (which includes thioredoxin fusion tag) either in Origami(DE3) and BL21(DE3) p*lys*S when compare with the uninduced condition (Fig. 4.22 and Fig. 4.23).

These results suggested that only the first 174 amino acids at the N-terminal (amino acids 59-232 of the precursor protein) of the SFR2 protein can be expressed in *E. coli* in the pET32a system.

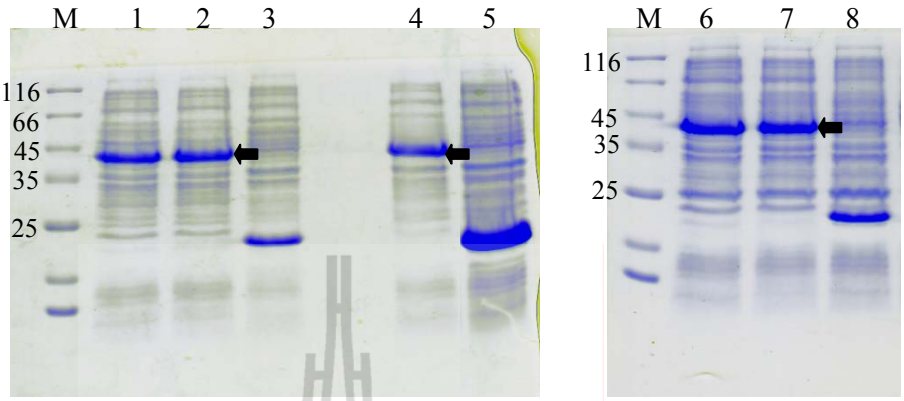



**Figure 4.20** SDS-PAGE analysis of total protein of cells with recombinant 31_SFR2_delC(mP) protein expressed in Origami(DE3), Origami B(DE3) and Rosetta-gami(DE3) for 16 h at 20 °C. Lane M: protein marker, Lanes 1 and 2: induced 31_SFR2_delC(mP) in Origami(DE3), Lanes 4: induced 31_SFR2_delC(mP) in Origami B(DE3), Lanes 6 and 7: induced 31_SFR2_delC(mP) in Rosetta-gami(DE3), Lanes 3, 5 and 8: thioredoxin fusion tag as positive control expressed in Origami(DE3), Origami B(DE3) and Rosetta-gami(DE3), respectively. The arrows indicated the expected size of recombinant 31_SFR2_delC (mP) protein.

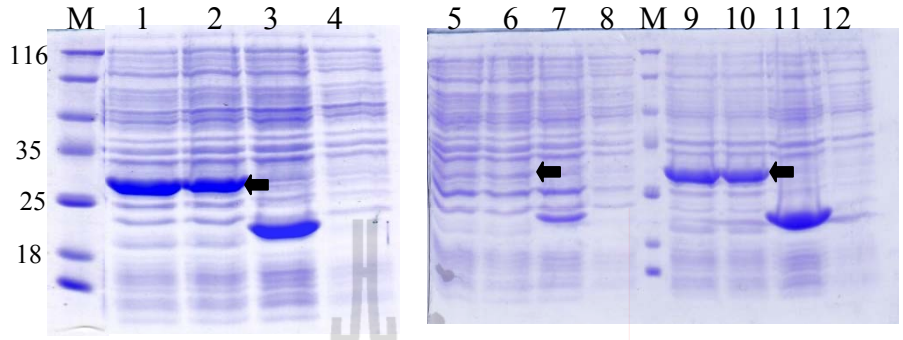



**Figure 4.21** SDS-PAGE analysis of total protein of cells with recombinant 32_SFR2 _delC(mP) protein expressed in Origami(DE3), Rosetta-gami(DE3) and Origami B(DE3) for 16 h at 20 °C. Lane M: protein marker, Lanes 1 and 2: induced 32_SFR2_delC(mP) in Origami(DE3), Lanes 5 and 6: induced 32_SFR2_delC(mP) in Rosetta-gami(DE3), Lanes 9 and 10: induced 32_SFR2_delC(mP) in Origami B(DE3), Lanes 3, 7 and 11: thioredoxin fusion tag as positive control expressed in Origami(DE3), Rosetta-gami(DE3) and Origami B(DE3), respectively, Lanes 4, 8 and 12: uninduced 32_SFR2_delC(mP) in Origami(DE3), Rosetta-gami(DE3) and Origami B(DE3), respectively. The arrows indicated the expected size of recombinant 32_SFR2_delC (mP) protein.

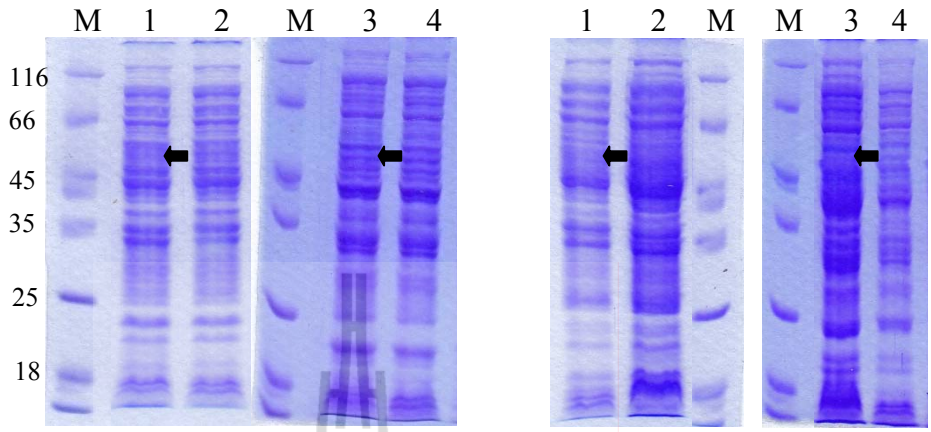



**Figure 4.22** SDS-PAGE analysis of total protein of cells with recombinant 33_SFR2_delC protein expressed in Origami(DE3) and BL21(DE3) p*lys*S for 16 h at 20 °C. (A) Lane M: protein marker, Lanes 1, 3: Origami(DE3) induced to express 33_SFR2_delC, Lanes 2, 4: uninduced Origami(DE3) with 33_SFR2_delC expreesion vector. (B) Lanes 1, 3: BL21(DE3) p*lys*S induced to express 33_SFR2_delC, Lanes 2, 4: uninduced BL21(DE3) p*lys*S 33_SFR2_delC expression vector. The arrows indicated the expected size of recombinant 33_SFR2_delC protein.

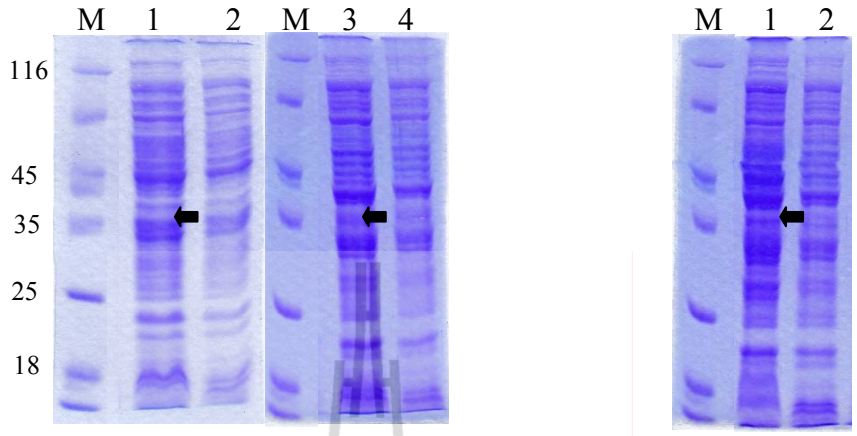



**Figure 4.23** SDS-PAGE analysis of total protein of cells with recombinant 34_SFR2_delC protein expressed in Origami(DE3) and BL21(DE3) p*lys*S for 16 h at 20 °C. (A) Lane M: protein marker, Lanes 1, 3: Origami(DE3) induced to express 34_SFR2_delC, Lanes 2, 4: uninduced Origami(DE3) with 34_SFR2_delC expression vector. (B) Lane 1: BL21(DE3) p*lys*S induced to express 34_SFR2_delC, Lane 2: uninduced BL21(DE3) p*lys*S with 33_SFR2_delC expression vector. The arrows indicated the expected size of recombinant 34_SFR2_delC protein.

#### 4.2.1.1.4 Expression of pET_26_SFR2_Syn

The pET_26_SFR2_Syn expression plasmid was transformed into Origami(DE3), Origami B(DE3) and Rosetta-gami(DE3). The positive clones were selected and inoculated in 10 ml LB broth containing appropriate antibiotics. The cultures were incubated at 37 °C, 200 rpm for 3 h, then transferred to 20 °C, 200 rpm for 30 min. The IPTG was added at the final concentration of 0.3 mM to the cultures for 16 h protein induction. At the end of the cultivation, cells were

broken by sonication and total proteins were loaded on SDS-PAGE to detect the protein profiles. The results indicated that no intense band of recombinant protein was detected on SDS-PAGE in all expression hosts (Fig. 4.24A). The expected size of 26_SFR2_Syn was 78 kDa. In *E. coli* strain Rosetta-gami(DE3) a protein band with molecular weight of 25 kDa was seen. This protein band in bigger than the thioredoxin fusion protein in the control. This band might be the fusion tag and a small region of SFR2 with premature stop translation. Western blot analysis using 28_SFR2 with an anti-fusion tag antibody was performed to investigate this protein band. The results did not show the target band on the PVDF membrane (Fig. 4.2B). This indicated that the anti- fusion tag antibody cannot bind to this protein, whereas the positive control, the pET32a that expressed in Rosetta-gami(DE3) showed a positive band of fusion tag with molecular weight of 23 kDa (Fig. 4.24B). Note that, the antibody was raised against 28_SFR2 that contained thioredoxin fusion tag. Therefore, the 25 kDa of protein was not the small part of the thioredoxin or SFR2 protein. Nakphaijit, 2006, also found the 25 kDa of protein band when she expressed SFR2 protein in the Rosetta-gami(DE3) strain. This protein was investigated by attempting to purify via Ni-column (QIAGEN) and it was found that this band was in the flow through and elution fractions. She concluded that this band is likely not a small region of SFR2. It might be the host protein expressed from the Rosetta-gami(DE3) cells.

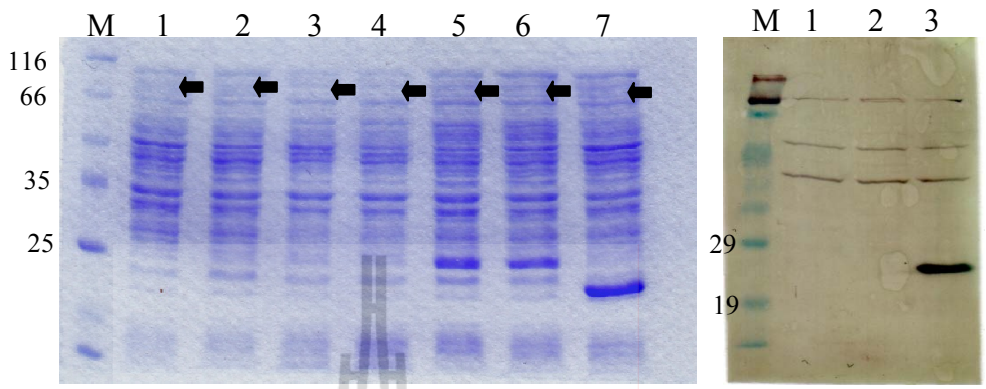



**Figure 4.24** SDS-PAGE analysis of total protein of cells with (A) recombinant 26_SFR2_Syn protein expressed in Origami(DE3), Origami B(DE3) and Rosetta-gami(DE3) for 16 h at 20 °C. Lane M: protein marker, Lanes 1, 2: Origami(DE3) induced to express 26_SFR2_Syn, Lanes 3, 4: Origami B(DE3) induced to express 26_SFR2_Syn, Lanes 5, 6: Rosetta-gami(DE3) induced 26_SFR2_Syn, Lane 7: Origami(DE3) induced to express thioredoxin fusion tag as positive control. (B) Western blot analysis using anti-28_SFR2 fusion tag antisera to detect the suspected protein band (25 kDa) that expressed from Rosetta-gami(DE3). Lane M: prestained protein marker, Lanes 1 and 2: Rosetta-gami(DE3) induced to express 26_SFR2_Syn, Lane 3: thioredoxin fusion tag. The arrows indicated the expected size of recombinant 26_SFR2-Syn protein.

#### 4.2.1.1.5 Expression of pET26_SFR2 and pET26_SFR2_Syn in ArcticExpress RIL(DE3)

The pET32a plasmid containing 26_SFR2 and 26_SFR2_Syn were transformed into Arctic Express RIL(DE3) cells by heat shock technique. The positive clones were selected and inoculated in 10 ml LB medium containing gentamycin and ampicillin. The cultures were incubated at 30 °C, 200 rpm for 3 h then transferred to 10 °C, 200 rpm for 30 min. The IPTG was added to final concentration of 1 mM to the cultures for protein induction, 24 h. At the end of the cultivation, cells were broken by lyzozyme and total proteins were loaded on SDS-PAGE to detect the protein profiles. The protein profiles between induced and uninduced cells with constructs to express each protein were nearly the same. The results did not show any intense band of 26_SFR2 and 26_SFR2_Syn at molecular weight 78 kDa but showed only the protein band of chaporonin, Cpn60 that was co-expressed and constitutively expressed protein in this *E. coli* host strain with molecular weight 57 kDa (Fig. 4.25).

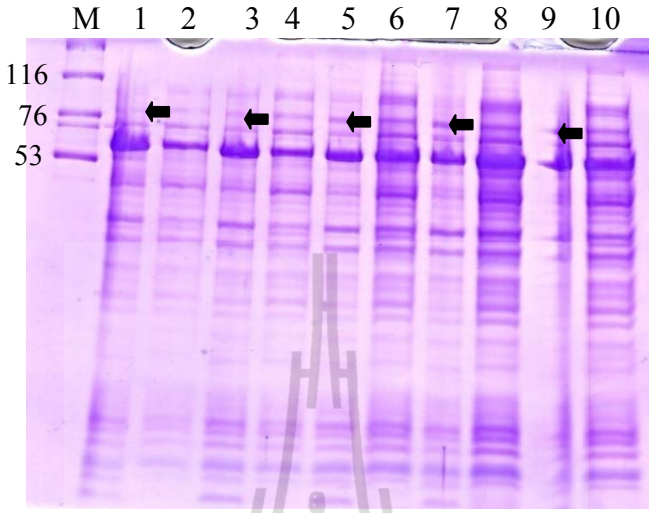



**Figure 4.25** SDS-PAGE of total protein of cells with recombinant 26_SFR2 and 26_SFR2_syn protein expressed in Arctic Express RIL(DE3) for 24 h at 10 °C. Lane M: protein marker, Lanes 1, 3 and 5: cells induced to express 26_SFR2, Lane 2, 4 and 6: uninduced cells for 26_SFR2, Lane 7 and 9: cells induced to express 26_SFR2_Syn, Lane 8 and10: uninduced cells for 26_SFR2_Syn. The arrows indicated the expected size of recombinant 26_SFR2 and 26_SFR2_Syn protein.

#### 4.2.1.2 pCold I system

##### 4.2.1.2.1 Expression of pCold_37_SFR2 and pCold_37 _mutPEST_SFR2

The pCold plasmid containing 37_SFR2 and 37_mutPEST_SFR2 were transformed into BL21 (DE3) p*lys*S, Origami (DE3) and Rosetta-gami (DE3). The positive clones of each expression hosts were selected by colony PCR techniques and inoculated in 10 ml LB medium containing appropriate antibiotics. The cultures were incubated at 37 °C 200 rpm for 3 h and then transferred

to 16 °C 200 rpm for 24 h to induce the protein expression. The total protein was loaded on SDS-PAGE to observe the protein profiles. The protein expression results indicated that no intense band of 37_SFR2 nor 37_mutPEST_SFR2 were not detected on SDS-PAGE in all expression hosts used for the pCold I expression system (Fig. 4.26 and Fig. 4.27). The recombinant 37_mutPEST_SFR2 showed the same protein expression result as the 26_mutPEST_SFR2, that SFR2 protein was not produce in *E. coli*, in both pET32a and the pCold I system, even though the PEST sequence has been mutated.

Hayashi and Kojima, 2008 reported that the expression of 10 proteins that could not be expressed in the pCold I system expressed using conventional *E. coli* expression methodologies. They modified the pCold I by adding GST gene before the multiple cloning site and compared the expression level with other expression vectors, including native pCold I, pGEX, pET and pMAL. Their results indicated that the target proteins could be observed in elute fraction when pCold-GST was used as the expression vector, whereas other expression vectors including the native pCold I expression vector could not detect any expressed protein (Hayashi and Kojima, 2008).

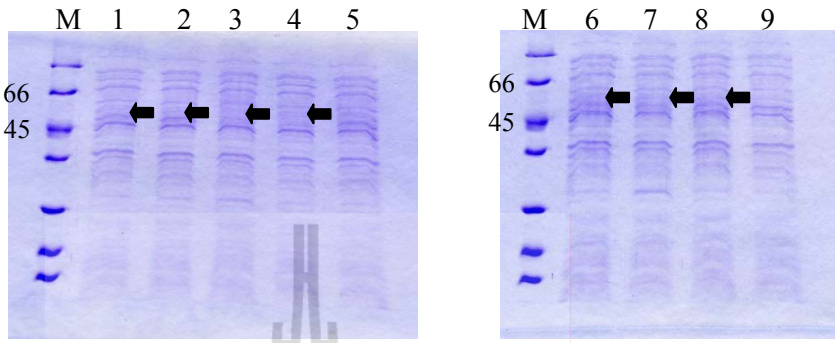



**Figure 4.26** SDS-PAGE of total protein of cells with recombinant 37_SFR2 and 37_mutPEST_SFR2 expressed in BL21(DE3) p*lys*S. Cells were induced for 24 h at 16 °C. Lane M: protein marker, Lanes 1-4: cells induced to express 37_SFR2, Lane 5: uninduced cells for 37_SFR2, Lane 6-8: cells induced to express 37_mutPEST_SFR2, Lane 9: uninduced cells for 37_mutPEST_SFR2. The arrows indicated the expected size of recombinant 37_SFR2 and 37_mutPEST_SFR2 protein.

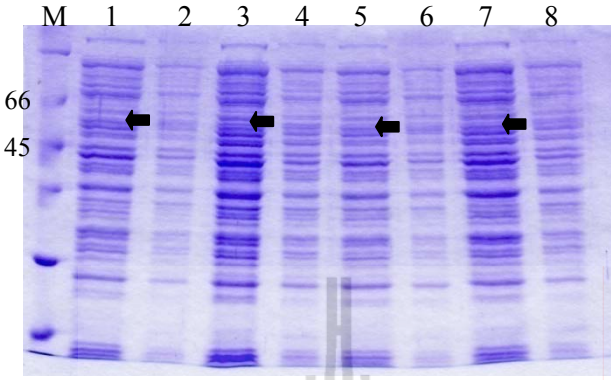



**Figure 4.27** SDS-PAGE of total protein of cells with recombinant 37_SFR2 and 37_mutPEST_SFR2 protein expressed in Rosetta-gami(DE3) and Origami(DE3). The cells were cultured for 24 h at 16 °C. Lane M: protein marker, Lanes 1 and 5: induced 37_SFR2 in Rosetta-gami(DE3) and Origami(DE3), respectively, Lanes 2 and 6: uninduced 37_SFR2 in Rosetta-gami(DE3) and Origami(DE3), respectively, Lanes 3 and 7: induced 37_mutPEST_SFR2 in Rosetta-gami(DE3) and Origami(DE3), respectively, Lanes 4 and 8: uninduced 37_mutPEST _SFR2 in Rosetta-gami(DE3) and Origami(DE3), respectively. The arrows indicated the expected size of recombinant 37_SFR2 and 37_mutPEST_SFR2 protein.

To summarize the expression of *SFR2* gene in *E. coli*, the results suggested that there are only 260 amino acid from the N-terminal and 174 amino acids from the C-terminal part of the protein that can be expressed and produce truncated SFR2 protein in *E. coli* using pET32a system. However, the full-length of SFR2 protein could not express in various host *E. coli* strains. Two hypotheses have be brought up to explain this problem. First, rare codon usage was analyzed. The N-terminal, the C-terminal SFR2 gene deletion and the *SFR2* gene codon optimization were analyzed for rare codon usage using *E. coli* Codon Usage Analyzer 2.0 developed by Morris Maduro (http://www.biology.ualberta.ca/pilgrim.hp/links/usage2.0c.html). The results showed similar percentage of rare codons in each *SFR2* gene fragment, which are approximately 19-23% for both the expressible and non-expressible SFR2 protein (Table 4.1). The bioinformatic analysis results indicated that rare codons were not the reason why SFR2 proteins were not able to be expressed in *E. coli*. This results agreed with an other report that improved codon usage did not enhance the protein expression (Klompus *et al.*, 2008). This bioinformatic results also agreed with the expression of SFR2 protein in Rosetta-gami(DE3) and Arctic Express(DE3) RIL from which no protein was produced, despite their containing additional rare codon tRNA.

**Table 4.1** Percentage of SFR2 rare codons in *E. coli*

| Gene | Rare codon percentage (%) | Expressible (detect band on SDS_PAGE) |
|---|---|---|
| AK_SFR2 | 20% | No |
| 26_SFR2 | 20% | No |
| 27_SFR2 | 20% | No |
| 28_SFR2 | 21% | Yes |
| 29_SFR2 | 23% | Yes |
| 31_SFR2_delC(mP) | 19% | Yes |
| 32_SFR2_delC(mP) | 16% | Yes |
| 33_SFR2_delC | 20% | No |
| 34_SFR2_delC | 19% | No |
| 26_SFR2_syn | 21% | No |

Second is the secondary the structure of RNA, different regions of the same RNA will fold together via base pairing interactions to make intricate secondary structures of RNA (Eccles, 1992). The unpredicted RNA secondary structure might block the expression of SFR2 protein in *E. coli*. The theoretical Gibb's free energy (ΔG value) was calculated using RNA secondary structure prediction (http://www.genebee.msu.su/services/rna2_reduced .html) to indicate the stability of RNA secondary structure. The ΔG value had some relation between expressible and non-expressible SFR2 protein (Table 4.2). The more negative ΔG value means higher stability of secondary structure of RNA, which may result in lower expression level.

The unexpressible proteins have a higher value of ΔG approximately between -335 to -170 kcal/mol while the expressible proteins have lower ΔG value of approximately -110 to -24 kcal/mol, except for 34_SFR2_delC which was not one of the expressible constructs, but have low ΔG value of -48 kcal/mol. From these analysis results suggest that the ΔG of the secondary structure might cause some effect on the expressibility of the SFR2 protein. The study of Klompus *et al.*, 2008 and Han *et al.*, 2010 have shown that genes synthesized with optimized codon usage for *E. coli* with significantly lower ΔG values can be expressed in *E. coli* at a much higher levels when compare to the wild-type genes.

**Table 4.2** Gibb's free energy (ΔG value) of SFR2 RNA secondary structure.

| **Gene** | **ΔG value (kcal/mol)** | **Expressible (detect band in SDS_PAGE)** |
|---|---|---|
| AK_SFR2 | -335.7 | No |
| 26_SFR2 | -272.5 | No |
| 27_SFR2 | -176.6 | No |
| 28_SFR2 | -111.7 | Yes |
| 29_SFR2 | -45.5 | Yes |
| 31_SFR2_delC(mP) | -108.7 | Yes |
| 32_SFR2_delC(mP) | -24.6 | Yes |
| 33_SFR2_delC | -148.7 | No |
| 34_SFR2_delC | -48.4 | No |
| 26_SFR2_syn | -287.8 | No |

## 4.3 Western blot analysis

Due to the limitation of the Coomassie brilliant blue staining detection of SDS-PAGE method, western blot analysis was performed using partially purified anti-28_SFR2 antisera. The pET plasmid harboring AK_SFR2, 26_SFR2, 27_SFR2, 28_SFR2, 29_SFR2, 26_mutPEST_SFR2, 26_SFR2_syn, 31_SFR2_delC(mP), 32_SFR2_delC (mP), 33_SFR2_delC and 34_SFR2_delC were transformed in Origami (DE3) and proteins were expressed at 20 °C for 24 h with 0.3 mM final IPTG concentration. The total proteins were submitted to SDS-PAGE and transferred to PVDF membrane. The western blot results indicated that the proteins which were not detected on SDS-PAGE (AK_SFR2, 26_SFR2, 27_SFR2, 26_mutPEST _SFR2, 26_SFR2_syn, 33_SFR2_delC and 34_SFR2_delC) showed small amounts of protein bands at the expected molecular weight. As expected the 28_SFR2, 29_SFR2, 31_SFR2_delC(mP) and 32_SFR2_delC(mP) were detected as high intensity bands on PVDF membrane (Fig. 4.28). From this result we can conclude that SFR2 proteins were able to be expressed in *E. coli*, however, with very small amounts so they were not detected on SDS-PAGE with Coomassie blue staining. For the bands of lower molecular weight than the target size might be the degradation products of the SFR2 protein. Also some cross contamination of *E. coli* protein were observed in every lane. The small amount of protein band can be detected only with the anti-28_SFR2 antisea but not with the anti 6-His probe with the same amount of protein.

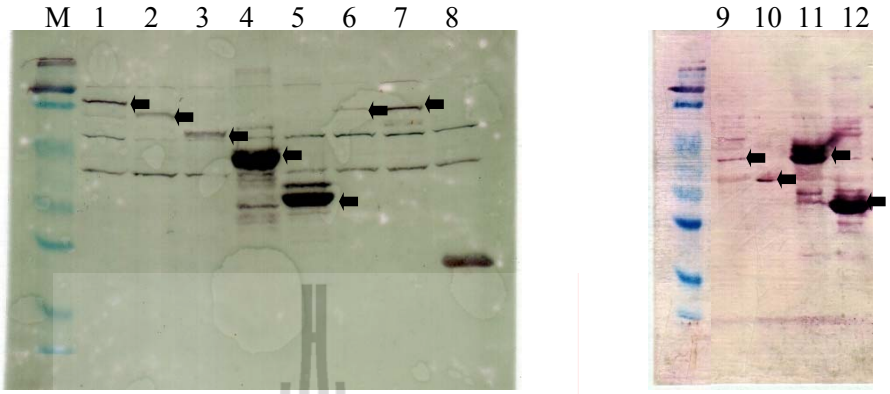



**Figure 4.28** Western blot analysis to detect the N- and C-terminal SFR2 deletions, 26_mutPEST_SFR2, 26_SFR2_Syn and thioredoxin fusion tag protein expressed in Origami(DE3). The cultures were induced for 24 h at 20 °C and partially purified anti-28_SFR2 antisera was used for detection. Lane M: pre-stained marker, Lane 1: AK_SFR2, Lane 2: 26_SFR2, Lane 3: 27_SFR2, Lane 4: 28_SFR2, Lane 5: 29_SFR2, Lane 6: 26_mutPEST_SFR2, Lane 7: 26_SFR2_syn, Lane 8: thioredoxin fusion tag, Lane 9: 33_SFR2_delC Lane 10: 34_SFR2_delC, Lane 11: 31_SFR2_delC(mP) and Lane 12: 32_SFR2_delC(mP). The arrows indicated the expected size of recombinant SFR2 protein fragments.

## 4.4 RNA detection

The RNA of SFR2 was detected by northern both analysis using DNA:RNA hybridization. The pET plasmid containing the N- and C-terminal deletion genes were transformed into Origami (DE3). The protein expression was induced at 20 °C, 200 rpm for 24 h with 0.3 mM IPTG final concentration. At the end of the cultivation, cell pellets were harvested and total RNA was extracted with the RNeasy mini kit (QIAGEN). Then samples of 20 µg of total RNA were loaded on a 1% FA gel and

transferred to the membrane. The northern blot results of the N-terminal deletions using the 28_SFR2 gene as probe showed small amounts of AK_SFR2, 26_SFR2 and 27_SFR2 RNA and a high amount of 28_SFR2 detected on the nitrocellulose membrane (Fig. 4.29A). The results suggested that the transcription level of AK_SFR2, 26_SFR2 and 27_SFR2 were low. Consequently, the low levels of proteins were detected. This result concluded that the reduced transcription rate results in decrease protein level (Wu *et al.*, 2004). Based on all the results there was a correlation between SDS-PAGE, western blot and northern blot analysis.

However, the RNA level of the C-terminal deletion showed conflicting results with the N-terminal SFR2 deletions (Fig. 4.29B). The northern blot results of the C-terminal deletion using the 34_SFR2_delC gene as probe showed intense bands of both 32_SFR2_delC(mP) and 33_SFR2_delC, in which the proteins were detected and non-detected by Coomassie blue, respectively. The 31_SFR2_delC(mP) and 34_SFR2_delC showed very thin detectable bands. These results did not a show correlation between RNA level and protein level. Some report on the relationship between RNA and protein levels in yeast and *E. coli* suggested that >70% of yeast and about only half of *E. coli* protein levels are determined by transcriptional regulation. The weaker correlation in *E. coli* may come from bacterial operon structure, in which genes are co-transcribed but differentially translated. Besides, the researchers also suggested that proteins present at high copies per mRNA are of low molecular weight; proteins present at low copies per mRNA show no such constraint. These results may indicate a ceiling on the capacity of the cell to maintain high ratios of protein to mRNA levels for large proteins, and possibly a limit on the capacities of the translational or degradative apparatus. However, it is consistent with the findings that ribosome density on mRNAs decreases with increasing gene lengths, and that

longer mRNAs have disproportionately lower ribosome initiation rates (Lu *et al.*, 2007). However, some studied indicated that the mRNA copy number and protein copy number for the same gene are not at all correlated based on the fact that mRNA is short-lived, at least in bacteria, whereas some proteins are long-lived. Because gene expression tends to be a stochastic process, mRNA showed up in bursts in the cell. And since mRNAs have relatively short lifetimes compared to proteins, for example, protein levels may be high after mRNA transcripts have been degraded (Taniguchi *et al.*, 2010).

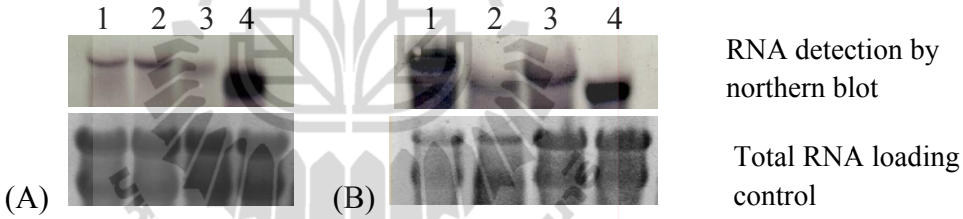



**Figure 4.29** Northern blot analysis to detect the RNA from N- and C-terminal gene deletions of SFR2 using 28_SFR2 and 34_SFR2_delC as probes. The recombinant proteins were expressed from Origami (DE3) at 20 °C for 24 h. (A) Total RNA from N-terminal SFR2 deletions. Lane 1: AK_SFR2, Lane 2: 26_SFR2, Lane 3: 27_SFR2, Lane 4: 28_SFR2. (B) RNA level from C-terminal SFR2 deletion. Lane 1: 33_SFR2_delC, Lane 2: 34_SFR2_delC, Lane 3: 31_SFR2_delC(mP), Lane 4: 32_SFR2_delC(mP).

**Table 4.3 Summarized all SFR2 fragment detection**

| Figures | Name of constructs | Detect on SDS-PAGE | Detect on WB | Detect on NB |
|---|---|---|---|---|
|  | AK_SFR2 | ✗ | ✓ | ✓ |
|  | 26_SFR2 | ✗ | ✓ | ✓ |
|  | 27_SFR2 | ✗ | ✓ | ✓ |
|  | 28_SFR2 | ✓✓ | ✓✓ | ✓✓ |
|  | 29_SFR2 | ✓✓ | ✓✓ | - |
|  | 26_mutPEST_ SFR2 | ✗ | ✓ | - |
|  | 26_SFR2_syn | ✗ | ✓ | - |
|  | 31_SFR2_delC (mP) | ✓ | ✓✓ | ✓ |
|  | 32_SFR2_delC (mP) | ✓ | ✓✓ | ✓✓ |
|  | 33_SFR2_delC | ✗ | ✓ | ✓✓ |
|  | 34_SFR2_delC | ✗ | ✓ | ✓ |
|  | 37_SFR2 (pCold) | ✗ | - | - |
|  | 37_mutPEST_ SFR2 (pCold) | ✗ | - | - |

Note: WB = western blot, NB = northern blot, ✓ = Detected, ✓✓ = Detected on high amount, ✗ = Not detected, - = Not used in experiments

## 4.5 Protein expression in eukaryotic cells

### 4.5.1 pPICZαBNH8 system

To investigate the expression of 26_SFR2_syn (codon optimized for *P. pastoris*) in eukaryotic system, the *P. pastoris* strain SMD1168H was selected as host strain. The pPicZαBNH8_ 26_SFR2_syn was transformed into *P. pastoris* by electroporation. The 15 colonies were grown in 50 ml BMGY medium for 24 h at 28 °C, 200 rpm and methanol induction was done for 72 h. At the end of the cultivation, the supernatant was used for enzyme activity test with *p*NP-β-D-galactoside, however the sample and negative control did not show any significant difference in activity values. The cell pellets were broken by adding breaking buffer and glass beads and loaded on SDS-PAGE to check for the protein profiles. The results did not show any intense band in all fractions (breaking cell pellet, supernatant and total protein) (Fig. 4.30). The result from protein expression in *P. pastoris* could not verify that SFR2 was produced because no detectable bands on SDS-PAGE nor any detectable activity was found. The SFR2 expression in *P. pastoris* in a previous study indicated that the SFR2 protein was not found by activity or protein assay both on SDS-PAGE and western blot using anti-6-His as antibody, but the RNA of SFR2 was detected (Nakphaijit., 2006).

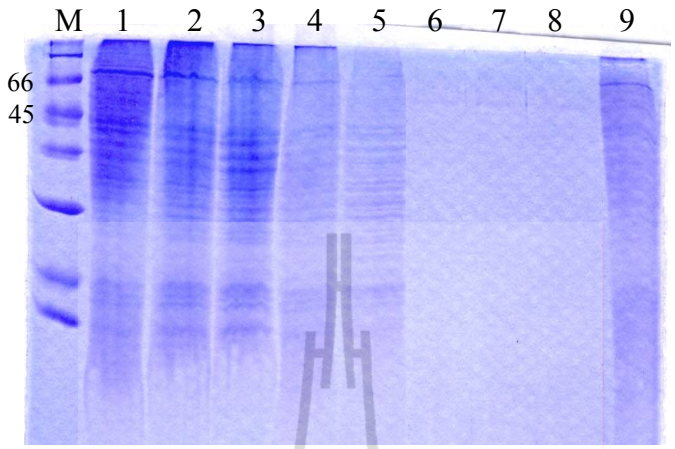



**Figure 4.30** SDS-PAGE analysis of recombinant 26_SFR2_Syn in *P. pastoris* SMD1168H. Lane M: protein marker, Lanes 1 and 2: cell pellets of cell induced to express 26_SFR2_Syn, Lane 3: cell pellets of $EK_L$, Lanes 4 and 5: Total protein contained in media of cells induced to express 26_SFR2_Syn, Lane 6 and 7: supernatant of 26_SFR2_Syn, Lane 8: supernatant of $EK_L$ and Lane 9: total protein of uninduced containing the 26_SFR2_Syn expression cassette.

# CHAPTER V

# CONCLUSION

The rice SFR2 gene were amplified from the AK119461 cDNA with different primers pairs to generate N-terminal SFR2 deletion, C-terminal SFR2 deletion, 26_mutPEST_ SFR2. The 26_SFR2_syn was amplified from pJET_26_SFR2_Syn. All of constructs were cloned into pET32a expression vector. The 37_SFR2 and 37_mutPEST_SFR2 truncated cDNA were also cloned into pCold I expression vector. Recombinant proteins were expressed in different expression hosts, such as BL21(DE3) plysS, Origami(DE3), Origami B(DE3), Rosetta-gami(DE3) and Arctic Express (RIL)(DE3). The results indicated that only the last 260 amino acids of the C-terminus and first 174 amino acids of the N-terminus of the conserved region of the SFR2 protein could be produced in *E. coli*.

The protein degradation signal, PEST sequence, in SFR2 protein was mutated using PCR mutagenesis. The glutamic acids in the PEST sequence were changed to glutamine and some were deleted. The 26_mutPEST_SFR2 was cloned and expressed, however, SFR2 with the mutated PEST sequence was not detected by Coomassie blue staining of SDS-PAGE gels in *E. coli*, even though the PEST sequence was mutated.

Western blot analysis of SFR2 proteins was done using a polyclonal anti-28_SFR2 antisera. The SFR2 protein was expressed in *E. coli* strain Origami (DE3) and total proteins were loaded on SDS-PAGE then transferred to PVDF membrane.

The result indicated that some SFR2 proteins which were not detected by Coomassie blue staining, can be detect on PVDF membrane but only very small amounts of the proteins were found.

The RNA level was assayed using northern blot analysis. The 28_SFR2 and 34_SFR2_delC cDNA were used as DNA probes. The total RNA was extracted and transferred to nitrocellulose membrane. The result showed a correlation between protein and RNA levels for the 5' end deletion genes, but for 3' end gene deletion, no such correlation was seen.

The pPICZαBNH8/DEST system was used to express a synthesized SFR2 gene with codon optimized for *P. pastoris* in *P. pastoris* protease deficient strain SMD1168H. No detectable SFR2 proteins were seen on SDS-PAGE.

All of the results together suggested that the SFR2 protein can be expressed in *E. coli*, but in very small amounts of protein, maybe because of low levels of RNA. High negative value of ΔG indicated the high stability of the secondary structure of RNA. The unpredicted RNA secondary structure might block the expression of the SFR2 protein in *E. coli*.

This protein may be produced from plant system (Nakphaijit, 2006), which offers the possibility to identify the function in further studies.

# REFERENCES


Aiyar, A., Xiang, Y., and Leis, J. (1996). **In vitro Mutagenesis Protocols**, Humana Press Inc., Hertfordshire.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. **Current Opinion in Biotechnology**. 10:(5), 411-421.

Buchanan, B. B., Gruissem, W., and Jones, R. L. (2000). **Biochemistry and Molecular Biology of Plants**, American Society of plant Physiologists Rockvile Maryland.

Byeon, G. M., Lee, K. S., Gui, Z. Z., Kim, I., Kang, P. D., Lee, S. M., Sohn, H. D., and Jin, B. R. (2005). A digestive β-glucosidase from the silkworm, *Bombyx mori*: cDNA cloning, expression and enzymatic characterization. **Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology**. 141:(4), 418-427.

Carillo, S., Pariat, M., Steff, A. m., Jariel-Encontre, I., Poulat, F., Berta, P., and Piechaczyk, M. (1996). PEST motifs are not required for rapid calpain-mediated proteolysis of c-fos protein. **Biochem Journal**. 313:(1), 245-251.

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

Chen, E., and Clarke, D. (2002). The PEST sequence does not contribute to the stability of the cystic fibrosis transmembrane conductance regulator. **BMC Biochemistry**. 3:(1), 29.

Chen, G.-H., Chen, W.-M., Huang, G.-T., Chen, Y.-W., and Jiang, S.-T. (2009). Expression of Recombinant Antibacterial Lactoferricin-Related Peptides from *Pichia pastoris* Expression System. **Journal of Agricultural and Food Chemistry**. 57:(20), 9509-9515.

Daly, R., and Hearn, M. T. W. (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. **Journal of Molecular Recognition**. 18:(2), 119-138.

Eccles, S. (1992). **Plant Protein Engineering**, Cambridge University Press, Cambridge.

Escamilla-Trevino, L. L., Chen, W., Card, M. L., Shih, M.-C., Cheng, C.-L., and Poulton, J. E. (2006). *Arabidopsis thaliana* β-Glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. **Phytochemistry**. 67:(15), 1651-1660.

Esen, A. (1992). Purification and Partial Characterization of Maize (*Zea mays* L.) β-glucosidase. **Phytochemistry**. 98:(1), 174-182.

Fan, Y., Zhang, Y., Yang, X., Pei, X., Guo, S., and Pei, Y. (2007). Expression of a Beauveria bassiana chitinase (Bbchit1) in *Escherichia coli* and *Pichia pastoris*. **Protein Expression and Purification**. 56:(1), 93-99.

Ferguson, B., Jones, N., Richter, J., and Rosenberg, M. (1984). Adenovirus E1a gene product expressed at high levels in *Escherichia coli* is functional. **Science**. 224:(4655), 1343-1346.

Fourrier, N., Bédard, J., Lopez-Juez, E., Barbrook, A., Bowyer, J., Jarvis, P., Warren, G., and Thorlby, G. (2008). A role for *SENSITIVE TO FREEZING2* in protecting chloroplasts against freeze-induced damage in *Arabidopsis*. **The Plant Journal**. 55:(5), 734-745.

Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.-l., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002). A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). **Science**. 296:(5565), 92-100.

Gueguen, Y., Chemardin, P., Janbon, G., Arnaud, A., and Galzy, P. (1996). A very efficient β-glucosidase catalyst for the hydrolysis of flavor precurcor of wines and fruit juices. **Journal of Agricultural and Food Chemistry**. 44, 2336-2340.

Han, J.-H., Choi, Y.-S., Kim, W.-J., Jeon, Y. H., Lee, S. K., Lee, B.-J., and Ryu, K.-S. (2010). Codon optimization enhances protein expression of human peptide deformylase in *E. coli*. **Protein Expression and Purification**. 70:(2), 224-230.

Hayashi, K., and Kojima, C. (2008). pCold-GST vector: A novel cold-shock vector containing GST tag for soluble protein production. **Protein Expression and Purification**. 62:(1), 120-127.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. **Gene**. 77:(1), 51-59.

Hsieh, M.-C., and Graham, T. L. (2001). Partial purification and characterization of a soybean β–glucosidase with high specific activity towards isoflavone conjugates. **Phytochemistry**. 58:(7), 995-1005.

IRGSP (2005). The map-based sequence of the rice genome. **Nature**. 436:(7052), 793-800.

Jung, K.-H., An, G., and Ronald, P. C. (2007). Towards a better bowl of rice: assigning function to tens of thousands of rice genes. **Nature Reviews Genetics** 9:(2), 91-101.

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. **Current Opinion in Biotechnology**. 6:(5), 494-500.

Katzen, F. (2007). Gateway recombinational cloning: a biological operating system. **Expert Opinion on Drug Discovery**. 2:(4), 571-589.

Ketudat-Cairns, J. and Esen. A. β-glucosidase. A review. Cellular and Molecular Life **Science**. 67:(20), 3389-3405.

Klompus, S., Solomon, G., and Gertler, A. (2008). A simple novel method for the preparation of noncovalent homodimeric, biologically active human interleukin 10 in *Escherichia coli*--Enhancing protein expression by degenerate PCR of 5' DNA in the open reading frame. **Protein Expression and Purification**. 62:(2), 199-205.

Kurland, C., and Gallant, J. (1996). Errors of heterologous protein expression. **Current Opinion in Biotechnology**. 7:(5), 489-493.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. **Nature Biotechnology**. 25:(1), 117-124.

Luang, S., Hrmova, M., and Ketudat Cairns, J. R. (2010). High-level expression of barley β-D-glucan exohydrolase HvExoI from a codon-optimized cDNA in *Pichia pastoris*. **Protein Expression and Purification**. 73:(1), 90-98.

Martin, C., and Jones, J. (2004). Cold tolerance, SFR2, and the legacy of Gary Warren. **The Plant Cell**. 16:(8), 1955-1957.

Mitchell, D., and Bell, A. (2003). PEST sequences in the malaria parasite *Plasmodium falciparum*: a genomic study. **Malaria Journal**. 2:(1), 16.

Moellering, E. R., Muthan, B., and Benning, C. (2010). Freezing Tolerance in Plants Requires Lipid Remodeling at the Outer Chloroplast Membrane. **Science**. 330:(6001), 226-228.

Molinari, M., Anagli, J., and Carafoli, E. (1995). PEST Sequences Do Not Influence Substrate Susceptibility to Calpain Proteolysis. **Journal of Biological Chemistry**. 270:(5), 2032-2035.

Nishikubo, T., Nakagawa, N., Kuramitsu, S., and Masui, R. (2005). Improved heterologous gene expression in *Escherichia coli* by optimization of the AT-content of codons immediately downstream of the initiation codon. **Journal of Biotechnology**. 120:(4), 341-346.

Opassiri, R., Ketudat Cairns, J. R., Akiyama, T., Wara-Aswapati, O., Svasti, J., and Esen, A. (2003). Characterization of a rice β-glucosidase highly expressed in flower and germinating shoot. **Plant Science**. 165:(3), 627-638.

Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A., and Ketudat Cairns, J. (2006). Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β-glucosidase. **BMC Plant Biology**. 6:(1), 33.

Pai, V. R., and Murugan, V. (2008). A synergistic effect of suppressive CGG codon in +2 position and downstream CAT repeats for efficient heterologous protein

expression in *Escherichia coli*. **Biochemical and Biophysical Research Communications**. 371:(3), 380-384.

Paulus, M., Haslbeck, M., and Watzele, M. (2004). RNA stem–loop enhanced expression of previously non-expressible genes. **Nucleic Acids Research**. 32:(9), e78.

Rechsteiner, M. (1991). Natural substrates of the Ubiquitin proteolytic pathway. **Cell**. 66:(4), 615-618.

Rechsteiner, M., and Rogers, S. W. (1996). PEST sequences and regulation by proteolysis. **Trends in Biochemical Sciences**. 21:(7), 267-271.

Rogers, S., Wells, R., and Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. **Science**. 234:(4774), 364-368.

Rooyen, R. V., Hagerdal, B. H., Grange, D. C. L., and Zyl, W. H. V. (2005). Construction of cellobiose-growing and fermenting *Saccharomyces cereviciae* strains. **Journal of Biotechnology**. 120, 284-295.

Salama, S. R., Hendricks, K. B., and Thorner, J. (1994). G1 cyclin degradation: the PEST motif of yeast Cln2 is necessary, but not sufficient, for rapid protein turnover. **Molecular and Cellular Biology**. 14:(12), 7953-7966.

Sanchez-Venegas, J., Navarrete, A., Dinamarca, J., Bravo Ramirez, L., Moraga, A., and Gidekel, M. (2009). Cloning and constitutive expression of *Deschampsia antarctica* Cu/Zn superoxide dismutase in *Pichia pastoris*. **BMC Research Notes**. 2:(1), 207.

Seo, S. W., Yang, J., and Jung, G. Y. (2009). Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and

translational efficiency in *Escherichia coli*. **Biotechnology and Bioengineering**. 104:(3), 611-616.

Spirin., A. S. (1999). Translational control in Prokaryotes, in, **Ribosomes**, Volume. 1, page. 310, Kluwer academic/Plenum Publishers, New York.

Streit, S., Michalski, C. W., Erkan, M., Kleeff, J., and Friess, H. (2008). Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. **Nature Protocols**. 4:(1), 37-43.

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. **Science**. 329:(5991), 533-538.

Tarahomjoo, S., Katakura, Y., and Shioya, S. (2008). Expression of C-terminal Repeat Region of Peptidoglycan Hydrolase of *Lactococcus lactis* IL1403 in Methylotrophic Yeast *Pichia pastoris*. **Journal of Bioscience and Bioengineering**. 105:(2), 134-139.

Thomashow, M. F. (1998). Role of Cold-Responsive Genes in Plant Freezing Tolerance. **Plant Physiology**. 118:(1), 1-8.

Thomashow, M. F. (1999). PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. **Annual Review of Plant Physiology and Plant Molecular Biology**. 50:(1), 571-599.

Thorlby, G., Fourrier, N., and Warren, G. (2004). The *SENSITIVE TO FREEZING2* Gene, Required for Freezing Tolerance in *Arabidopsis thaliana*, Encodes a β-glucosidase. **Plant Cell**. 16:(8), 2192-2203.

Walsh, G. (2000). **Proteins Biochemistry and Biotechnology**, John Wiley & Son, Inc., New York.

Wang, X.-R., Zhou, Y.-B., Liu, F., Wang, K.-S., Shen, Y., Liu, J.-H., and Han, Z.-G. (2006). A novel recently evolved gene *C19orf24* encodes a non-classical secreted protein. **Cellular & Molecular Biology Letters.** 11:(2), 161-170.

Weickert, M. J., Doherty, D. H., Best, E. A., and Olins, P. O. (1996). Optimization of heterologous protein production in *Escherichia coli*. **Current Opinion in Biotechnology**. 7:(5), 494-499.

Wu, X., Jörnvall, H., Berndt, K. D., and Oppermann, U. (2004). Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. **Biochemical and Biophysical Research Communications**. 313:(1), 89-96.

Wurch, T., Lestienne, F., and Pauwels, P. J. (1998). A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. **Biotechnology Techniques**. 12:(9), 653-657.

Xu, Z., Escamilla-Treviño, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., Mohamed, A., Cheng, C.-L., Shih, M.-C., Poulton, J., and Esen, A. (2004). Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. **Plant Molecular Biology**. 55:(3), 343-367.

Yadav, S. K. (2010). Cold stress tolerance mechanisms in plants. A review. **Agronomy for Sustainable Development**. 30:(3), 515-527.

Zhang, W., Xiao, W., Wei, H., Zhang, J., and Tian, Z. (2006). mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in *Escherichia coli*. **Biochemical and Biophysical Research Communications**. 349:(1), 69-78.

Zhao, R., Zhang, Z., Zhang, R., Li, M., Lei, Z., Utsumi, M., and Sugiura, N. (2010). Methane production from rice straw pretreated by a mixture of acetic-propionic acid. **Bioresource Technology**. 101:(3), 990-994.

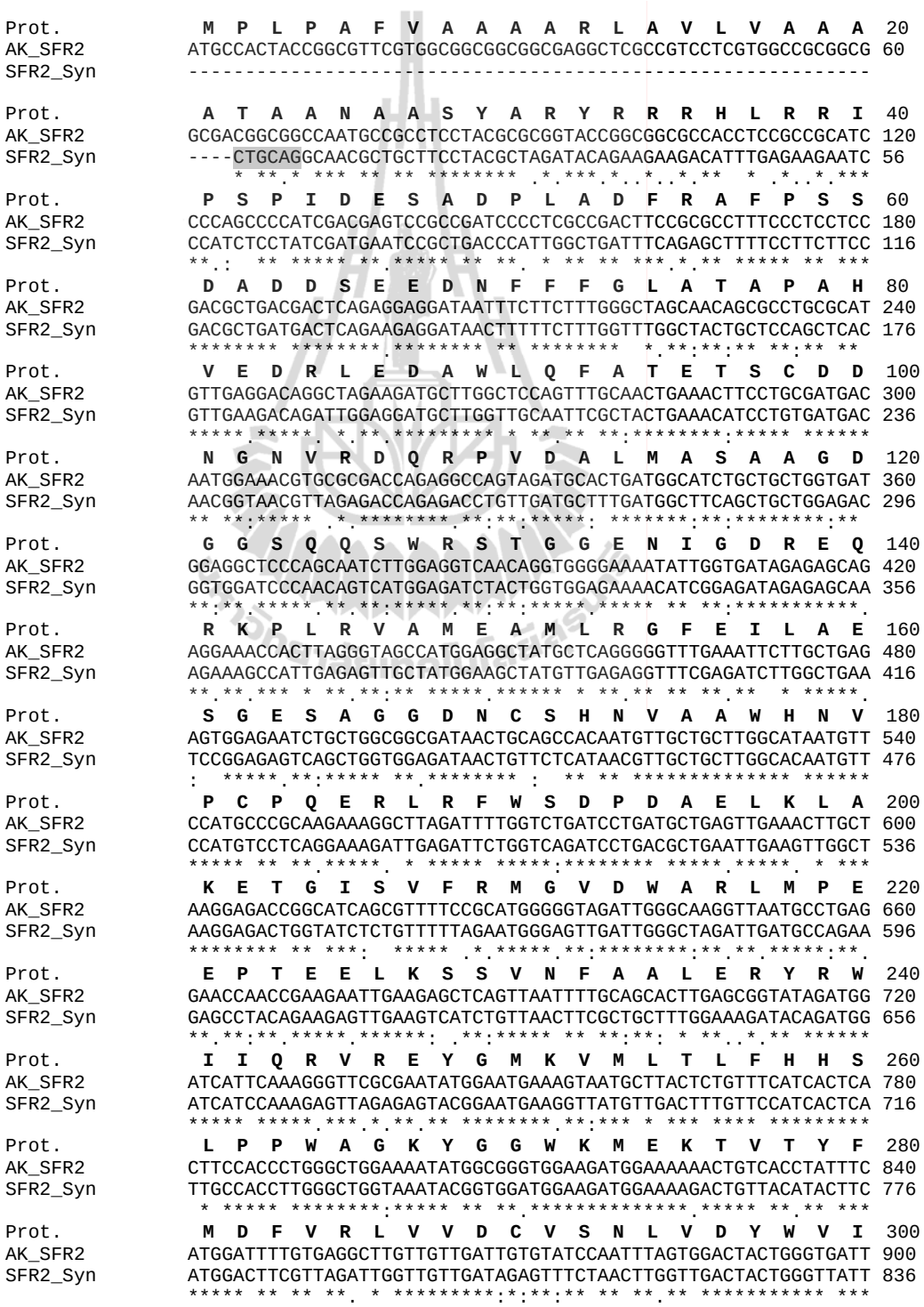
Zhao, W., Wang, J., He, X., Huang, X., Jiao, Y., Dai, M., Wei, S., Fu, J., Chen, Y., Ren, X., Zhang, Y., Ni, P., Zhang, J., Li, S., Wang, J., Wong, G. K.-S., Zhao, H., Yu, J., Yang, H., and Wang, J. (2004). BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. **Nucleic Acids Research**. 32:(suppl_1), D377-382.

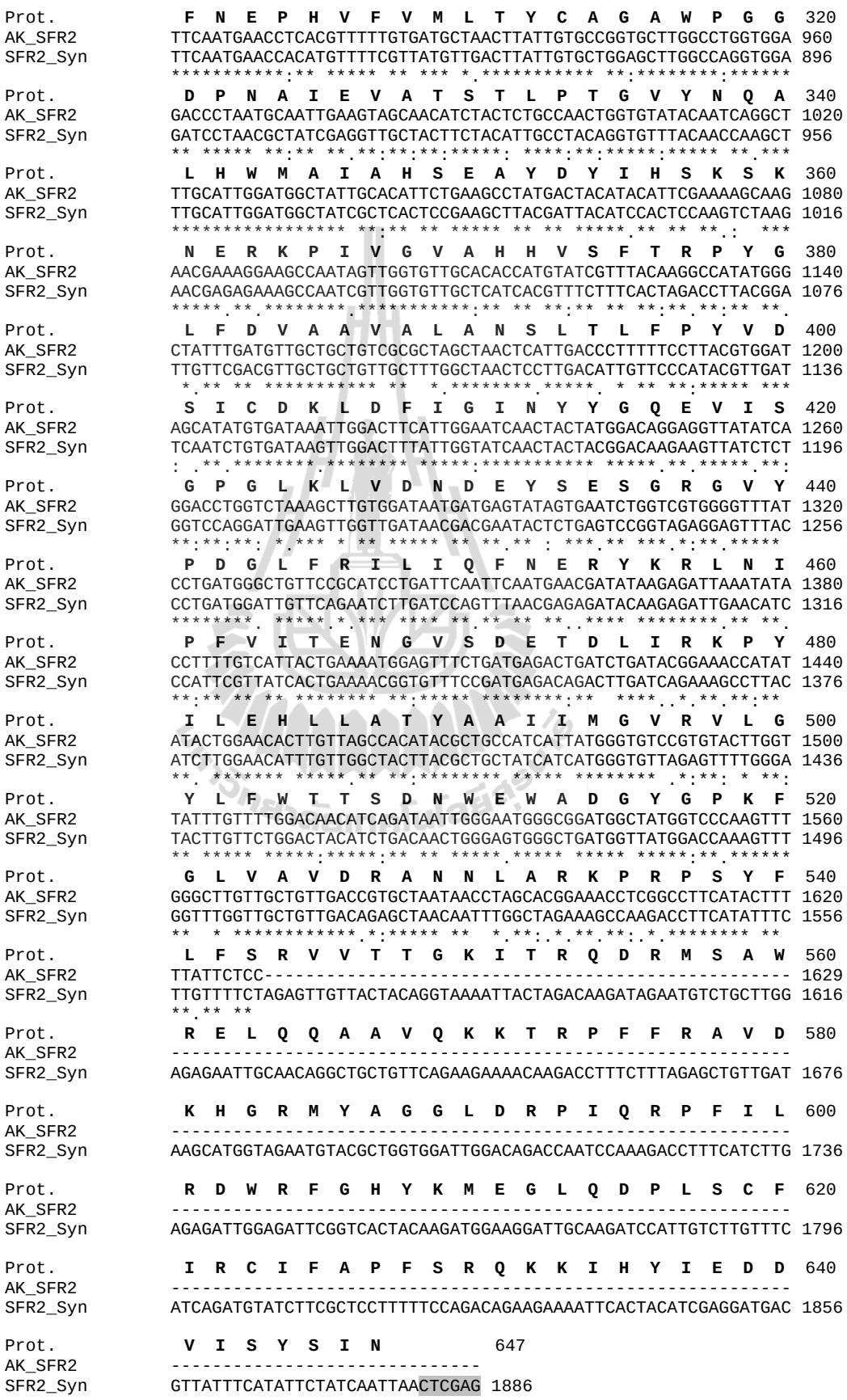
Zhou, Z., Schnake, P., Xiao, L., and Lal, A. A. (2004). Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization. **Protein Expression and Purification**. 34:(1), 87-94.

# APPENDICES

# Appendix I

## 1. Sequence of synthesized SFR2 gene




Prot. F N E P H V F V M L T Y C A G A W P G G 320
AK_SFR2 TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
SFR2_Syn TTCAATGAACCACATGTTTTCGTTATGTTGACTTATTGTGCTGGAGCTTGGCCAGGTGGA 896
Prot. D P N A I E V A T S T L P T G V Y N Q A 340
AK_SFR2 GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 1020
SFR2_Syn GATCCTAACGCTATCGAGGTTGCTACTTCTACATTGCCTACAGGTGTTTACAACCAAGCT 956
Prot. L H W M A I A H S E A Y D Y I H S K S K 360
AK_SFR2 TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 1080
SFR2_Syn TTGCATTGGATGGCTATCGCTCACTCCGAAGCTTACGATTACATCCACTCCAAGTCTAAG 1016
Prot. N E R K P I V G V A H H V S F T R P Y G 380
AK_SFR2 AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 1140
SFR2_Syn AACGAGAGAAAGCCAATCGTTGGTGTTGCTCATCACGTTTCTTTCACTAGACCTTACGGA 1076
Prot. L F D V A A V A L A N S L T L F P Y V D 400
AK_SFR2 CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 1200
SFR2_Syn TTGTTCGACGTTGCTGCTGTTGCTTTGGCTAACTCCTTGACATTGTTCCCATACGTTGAT 1136
Prot. S I C D K L D F I G I N Y Y G Q E V I S 420
AK_SFR2 AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 1260
SFR2_Syn TCAATCTGTGATAAGTTGGACTTTATTGGTATCAACTACTACGGACAAGAAGTTATCTCT 1196
Prot. G P G L K L V D N D E Y S E S G R G V Y 440
AK_SFR2 GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 1320
SFR2_Syn GGTCCAGGATTGAAGTTGGTTGATAACGACGAATACTCTGAGTCCGGTAGAGGAGTTTAC 1256
Prot. P D G L F R I L I Q F N E R Y K R L N I 460
AK_SFR2 CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 1380
SFR2_Syn CCTGATGGATTGTTCAGAATCTTGATCCAGTTTAACGAGAGATACAAGAGATTGAACATC 1316
Prot. P F V I T E N G V S D E T D L I R K P Y 480
AK_SFR2 CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 1440
SFR2_Syn CCATTCGTTATCACTGAAAACGGTGTTTCCGATGAGACAGACTTGATCAGAAAGCCTTAC 1376
Prot. I L E H L L A T Y A A I I M G V R V L G 500
AK_SFR2 ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 1500
SFR2_Syn ATCTTGGAACATTTGTTGGCTACTTACGCTGCTATCATCATGGGTGTTAGAGTTTTGGGA 1436
Prot. Y L F W T T S D N W E W A D G Y G P K F 520
AK_SFR2 TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1560
SFR2_Syn TACTTGTTCTGGACTACATCTGACAACTGGGAGTGGGCTGATGGTTATGGACCAAAGTTT 1496
Prot. G L V A V D R A N N L A R K P R P S Y F 540
AK_SFR2 GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1620
SFR2_Syn GGTTTGGTTGCTGTTGACAGAGCTAACAATTTGGCTAGAAAGCCAAGACCTTCATATTTC 1556
Prot. L F S R V V T T G K I T R Q D R M S A W 560
AK_SFR2 TTATTCTCC---------- 1629
SFR2_Syn TTGTTTTCTAGAGTTGTTACTACAGGTAAAATTACTAGACAAGATAGAATGTCTGCTTGG 1616
Prot. R E L Q Q A A V Q K K T R P F F R A V D 580
AK_SFR2 ----------
SFR2_Syn AGAGAATTGCAACAGGCTGCTGTTCAGAAGAAAACAAGACCTTTCTTTAGAGCTGTTGAT 1676
Prot. K H G R M Y A G G L D R P I Q R P F I L 600
AK_SFR2 ----------
SFR2_Syn AAGCATGGTAGAATGTACGCTGGTGGATTGGACAGACCAATCCAAAGACCTTTCATCTTG 1736
Prot. R D W R F G H Y K M E G L Q D P L S C F 620
AK_SFR2 ----------
SFR2_Syn AGAGATTGGAGATTCGGTCACTACAAGATGGAAGGATTGCAAGATCCATTGTCTTGTTTC 1796
Prot. I R C I F A P F S R Q K K I H Y I E D D 640
AK_SFR2 ----------
SFR2_Syn ATCAGATGTATCTTCGCTCCTTTTTCCAGACAGAAGAAAATTCACTACATCGAGGATGAC 1856
Prot. V I S Y S I N 647
AK_SFR2 ----------
SFR2_Syn GTTATTTCATATTCTATCAATTAACTCGAG 1886

Note: The restriction sites for *Pst*I and *Xho*I were added to SFR2_Syn at 5’ and 3’ end, respectively (grey highlight).

**2. Amino sequence alignment between the SFR2 as reference (first row) and the N-terminal gene deletion (second row) from DNA sequencing.**

- AK_SFR2 alignment

```
SFR2            MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
AK_SFR2_seq     MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
                ************************************************************

SFR2            DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
AK_SFR2_seq     DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
                ************************************************************

SFR2            GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
AK_SFR2_seq     GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
                ************************************************************

SFR2            PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
AK_SFR2_seq     PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
                ************************************************************

SFR2            IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
AK_SFR2_seq     IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
                ************************************************************

SFR2            FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 360
AK_SFR2_seq     FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 360
                ************************************************************

SFR2            NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 420
AK_SFR2_seq     NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 420
                ************************************************************

SFR2            GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 480
AK_SFR2_seq     GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 480
                ************************************************************

SFR2            ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 540
AK_SFR2_seq     ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 540
                ************************************************************

SFR2            LFSRVVTTGKITRQDRMSAWRELQQAAVQKKTRPFFRAVDKHGRMYAGGLDRPIQRPFIL 600
AK_SFR2_seq     LFS------------------------------------------------------------ 543
                ***

SFR2            RDWRFGHYKMEGLQDPLSCFIRCIFAPFSRQKKIHYIEDDVISYSIN- 647
AK_SFR2_seq     ------------------------------------------------
```

- 26_SFR2 alignment

```
SFR2            MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
26_SFR2_seq     -----------------------------------------------------DFRAFPSS 8
                                                                     ********

SFR2            DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
26_SFR2_seq     DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 68
                ************************************************************
```

```
SFR2            GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
26_SFR2_seq     GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 128
                ************************************************************

SFR2            PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
26_SFR2_seq     PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 188
                ************************************************************

SFR2            IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
26_SFR2_seq     IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 248
                ************************************************************

SFR2            FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 360
26_SFR2_seq     FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 308
                ************************************************************

SFR2            NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 420
26_SFR2_seq     NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 368
                ************************************************************

SFR2            GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 480
26_SFR2_seq     GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 428
                ************************************************************

SFR2            ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 540
26_SFR2_seq     ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 488
                ************************************************************

SFR2            LFSRVVTTGKITRQDRMSAWRELQQAAVQKKTRPFFRAVDKHGRMYAGGLDRPIQRPFIL 600
26_SFR2_seq     LFS------------------------------------------------------------ 491
                ***

SFR2            RDWRFGHYKMEGLQDPLSCFIRCIFAPFSRQKKIHYIEDDVISYSIN- 647
26_SFR2_seq     ------------------------------------------------
```

- 27_SFR2 alignment

```
SFR2            MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
27_SFR2_seq     ------------------------------------------------------------

SFR2            DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
27_SFR2_seq     ------------------------------------------------------------

SFR2            GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
27_SFR2_seq     ----------------------------------------------------HNVAAWHNV 9
                                                                     *********

SFR2            PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
27_SFR2_seq     PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 69
                ************************************************************

SFR2            IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
27_SFR2_seq     IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 129
                ************************************************************

SFR2            FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 360
27_SFR2_seq     FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 189
                ************************************************************

SFR2            NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 420
27_SFR2_seq     NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 249
                ************************************************************

SFR2            GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 480
27_SFR2_seq     GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 309
                ************************************************************
```

```
SFR2            ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 540
27_SFR2_seq     ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 369
                ************************************************************

SFR2            LFSRVVTTGKITRQDRMSAWRELQQAAVQKKTRPFFRAVDKHGRMYAGGLDRPIQRPFIL 600
27_SFR2_seq     LFS------------------------------------------------------------ 372
                ***

SFR2            RDWRFGHYKMEGLQDPLSCFIRCIFAPFSRQKKIHYIEDDVISYSIN- 647
27_SFR2_seq     ------------------------------------------------
```

- 28_SFR2 alignment

```
SFR2            MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
28_SFR2_seq     ------------------------------------------------------------

SFR2            DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
28_SFR2_seq     ------------------------------------------------------------

SFR2            GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
28_SFR2_seq     ------------------------------------------------------------

SFR2            PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
28_SFR2_seq     ------------------------------------------------------------

SFR2            IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
28_SFR2_seq     -------------------------------------------VRLVVDCVSNLVDYWVI 17
                                                           *****************

SFR2            FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 360
28_SFR2_seq     FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 77
                ************************************************************

SFR2            NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 420
28_SFR2_seq     NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 137
                ************************************************************

SFR2            GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 480
28_SFR2_seq     GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 197
                ************************************************************

SFR2            ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 540
28_SFR2_seq     ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 257
                ************************************************************

SFR2            LFSRVVTTGKITRQDRMSAWRELQQAAVQKKTRPFFRAVDKHGRMYAGGLDRPIQRPFIL 600
28_SFR2_seq     LFS------------------------------------------------------------ 260
                ***

SFR2            RDWRFGHYKMEGLQDPLSCFIRCIFAPFSRQKKIHYIEDDVISYSIN- 647
28_SFR2_seq     ------------------------------------------------
```

- 29_SFR2 alignment (the dot (●) on top indicate the position of 26, 27 and 28_SFR2)

```
                                                                ●
SFR2            MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
29_SFR2_seq     ------------------------------------------------------------

SFR2            DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
29_SFR2_seq     ------------------------------------------------------------
```

```
SFR2            GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
29_SFR2_seq     ------------------------------------------------------------

SFR2            PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
29_SFR2_seq     ------------------------------------------------------------

SFR2            IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
29_SFR2_seq     ------------------------------------------------------------

SFR2            FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 360
29_SFR2_seq     ------------------------------------------------------------

SFR2            NERKPIVGVAHHVSFTRPYGLFDVAAVALANSLTLFPYVDSICDKLDFIGINYYGQEVIS 420
29_SFR2_seq     ------------------------------------------------------------

SFR2            GPGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 480
29_SFR2_seq     -PGLKLVDNDEYSESGRGVYPDGLFRILIQFNERYKRLNIPFVITENGVSDETDLIRKPY 59
                 ************************************************************

SFR2            ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 540
29_SFR2_seq     ILEHLLATYAAIIMGVRVLGYLFWTTSDNWEWADGYGPKFGLVAVDRANNLARKPRPSYF 119
                ************************************************************

SFR2            LFSRVVTTGKITRQDRMSAWRELQQAAVQKKTRPFFRAVDKHGRMYAGGLDRPIQRPFIL 600
29_SFR2_seq     LFS------------------------------------------------------------ 122
                ***

SFR2            RDWRFGHYKMEGLQDPLSCFIRCIFAPFSRQKKIHYIEDDVISYSIN- 647
29_SFR2_seq     ------------------------------------------------
```

**3. Amino sequence alignment between the AK_SFR2 as reference (first row) and the C-terminal gene deletion (second row) from DNA sequencing.**

- 31_SFR2_delC(mP) alignment

```
AK_SFR2                 MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
31_SFR2_delC_mP__seq    ----------------------------------------------------DFRAFPSS 8
                                                                             ********

AK_SFR2                 DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
31_SFR2_delC_mP__seq    DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 68
                        ************************************************************

AK_SFR2                 GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
31_SFR2_delC_mP__seq    GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 128
                        ************************************************************

AK_SFR2                 PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
31_SFR2_delC_mP__seq    PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPQQPTQQLKSSVNF-------- 180
                        ************************************::**::*******

AK_SFR2                 IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
31_SFR2_delC_mP__seq    ------------------------------------------------------------
```

- 32_SFR2_delC(mP) alignment

```
AK_SFR2                 MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
32_SFR2_delC_mP__seq    ------------------------------------------------------------
```

```
AK_SFR2                 DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
32_SFR2_delC_mP__seq    ------------------------------------------------------------

AK_SFR2                 GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
32_SFR2_delC_mP__seq    ----------------------------------------------HNVAAWHNV 9
                                                                      *********

AK_SFR2                 PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
32_SFR2_delC_mP__seq    PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPQQPTQQLKSSVNF-------- 61
                        ************************************::**::*******

AK_SFR2                 IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
32_SFR2_delC_mP__seq    ------------------------------------------------------------
```

- 33_SFR2_delC alignment

```
AK_SFR2                 MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
33_SFR2_delC_seq        ---------------------------------------------------DFRAFPSS 8
                                                                           ********

AK_SFR2                 DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
33_SFR2_delC_seq        DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 68
                        ************************************************************

AK_SFR2                 GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
33_SFR2_delC_seq        GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 128
                        ************************************************************

AK_SFR2                 PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
33_SFR2_delC_seq        PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 188
                        ************************************************************

AK_SFR2                 IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
33_SFR2_delC_seq        IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDC---------- 238
                        ********************************************

AK_SFR2                 FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 360
33_SFR2_delC_seq        ------------------------------------------------------------
```

- 34_SFR2_delC alignment

```
AK_SFR2                 MPLPAFVAAAARLAVLVAAAATAANAASYARYRRRHLRRIPSPIDESADPLADFRAFPSS 60
34_SFR2_delC_seq        ------------------------------------------------------------

AK_SFR2                 DADDSEEDNFFFGLATAPAHVEDRLEDAWLQFATETSCDDNGNVRDQRPVDALMASAAGD 120
34_SFR2_delC_seq        ------------------------------------------------------------

AK_SFR2                 GGSQQSWRSTGGENIGDREQRKPLRVAMEAMLRGFEILAESGESAGGDNCSHNVAAWHNV 180
34_SFR2_delC_seq        ---------------------------------------------------HNVAAWHNV 9
                                                                           *********

AK_SFR2                 PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 240
34_SFR2_delC_seq        PCPQERLRFWSDPDAELKLAKETGISVFRMGVDWARLMPEEPTEELKSSVNFAALERYRW 69
                        ************************************************************

AK_SFR2                 IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDCVSNLVDYWVI 300
34_SFR2_delC_seq        IIQRVREYGMKVMLTLFHHSLPPWAGKYGGWKMEKTVTYFMDFVRLVVDC---------- 119
                        ********************************************

AK_SFR2                 FNEPHVFVMLTYCAGAWPGGDPNAIEVATSTLPTGVYNQALHWMAIAHSEAYDYIHSKSK 360
34_SFR2_delC_seq        ------------------------------------------------------------
```

**4. DNA sequence alignment between the SFR2 as reference (first row) and the N-terminal gene deletion (second row) from DNA sequencing.**

- AK_SFR2 alignment

```
AK_SFR2         ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
AK_SFR2_seq     ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
                ************************************************************

AK_SFR2         GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
AK_SFR2_seq     GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
                ************************************************************

AK_SFR2         CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
AK_SFR2_seq     CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
                ************************************************************

AK_SFR2         GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
AK_SFR2_seq     GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
                ************************************************************

AK_SFR2         GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
AK_SFR2_seq     GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
                ************************************************************

AK_SFR2         AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
AK_SFR2_seq     AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
                ************************************************************

AK_SFR2         GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
AK_SFR2_seq     GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
                ************************************************************

AK_SFR2         AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
AK_SFR2_seq     AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
                ************************************************************

AK_SFR2         AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
AK_SFR2_seq     AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
                ************************************************************

AK_SFR2         CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
AK_SFR2_seq     CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
                ************************************************************

AK_SFR2         AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
AK_SFR2_seq     AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
                ************************************************************

AK_SFR2         GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
AK_SFR2_seq     GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
                ************************************************************

AK_SFR2         ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
AK_SFR2_seq     ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
                ************************************************************

AK_SFR2         CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 840
AK_SFR2_seq     CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 840
                ************************************************************

AK_SFR2         ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 900
AK_SFR2_seq     ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 900
                ************************************************************

AK_SFR2         TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
AK_SFR2_seq     TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
                ************************************************************

AK_SFR2         GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 1020
AK_SFR2_seq     GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 1020
                ************************************************************
```

```
AK_SFR2          TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 1080
AK_SFR2_seq      TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 1080
                 ************************************************************

AK_SFR2          AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 1140
AK_SFR2_seq      AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 1140
                 ************************************************************

AK_SFR2          CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 1200
AK_SFR2_seq      CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 1200
                 ************************************************************

AK_SFR2          AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 1260
AK_SFR2_seq      AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 1260
                 ************************************************************

AK_SFR2          GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 1320
AK_SFR2_seq      GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 1320
                 ************************************************************

AK_SFR2          CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 1380
AK_SFR2_seq      CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 1380
                 ************************************************************

AK_SFR2          CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 1440
AK_SFR2_seq      CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 1440
                 ************************************************************

AK_SFR2          ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 1500
AK_SFR2_seq      ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 1500
                 ************************************************************

AK_SFR2          TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1560
AK_SFR2_seq      TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1560
                 ************************************************************

AK_SFR2          GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1620
AK_SFR2_seq      GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1620
                 ************************************************************

AK_SFR2          TTATTCTCC 1629
AK_SFR2_seq      TTATTCTCC 1629
                 *********
```

- 26_SFR2 alignment

```
AK_SFR2          ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
26_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
26_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
26_SFR2_seq      ----------------------------------GACTTCCGCGCCTTTCCCTCCTCC 24
                                                   ************************

AK_SFR2          GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
26_SFR2_seq      GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 84
                 ************************************************************

AK_SFR2          GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
26_SFR2_seq      GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 144
                 ************************************************************

AK_SFR2          AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
26_SFR2_seq      AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 204
                 ************************************************************
```

```
AK_SFR2          GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
26_SFR2_seq      GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 264
                 ************************************************************

AK_SFR2          AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
26_SFR2_seq      AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 324
                 ************************************************************

AK_SFR2          AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
26_SFR2_seq      AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 384
                 ************************************************************

AK_SFR2          CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
26_SFR2_seq      CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 444
                 ************************************************************

AK_SFR2          AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
26_SFR2_seq      AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 504
                 ************************************************************

AK_SFR2          GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
26_SFR2_seq      GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 564
                 ************************************************************

AK_SFR2          ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
26_SFR2_seq      ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 624
                 ************************************************************

AK_SFR2          CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 840
26_SFR2_seq      CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 684
                 ************************************************************

AK_SFR2          ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 900
26_SFR2_seq      ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 744
                 ************************************************************

AK_SFR2          TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
26_SFR2_seq      TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 804
                 ************************************************************

AK_SFR2          GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 1020
26_SFR2_seq      GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 864
                 ************************************************************

AK_SFR2          TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 1080
26_SFR2_seq      TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 924
                 ************************************************************

AK_SFR2          AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 1140
26_SFR2_seq      AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 984
                 ************************************************************

AK_SFR2          CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 1200
26_SFR2_seq      CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 1044
                 ************************************************************

AK_SFR2          AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 1260
26_SFR2_seq      AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 1104
                 ************************************************************

AK_SFR2          GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 1320
26_SFR2_seq      GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 1164
                 ************************************************************

AK_SFR2          CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 1380
26_SFR2_seq      CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 1224
                 ************************************************************

AK_SFR2          CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 1440
26_SFR2_seq      CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 1284
                 ************************************************************
```

```
AK_SFR2          ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 1500
26_SFR2_seq      ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 1344
                 ************************************************************

AK_SFR2          TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1560
26_SFR2_seq      TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1404
                 ************************************************************

AK_SFR2          GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1620
26_SFR2_seq      GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1464
                 ************************************************************

AK_SFR2          TTATTCTCC 1629
26_SFR2_seq      TTATTCTCC 1473
                 *********
```

- 27_SFR2 alignment

```
AK_SFR2          ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
27_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
27_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
27_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
27_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
27_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
27_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
27_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
27_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
27_SFR2_seq      --------------------------------CACAATGTTGCTGCTTGGCATAATGTT 27
                                                 ***************************

AK_SFR2          CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
27_SFR2_seq      CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 87
                 ************************************************************

AK_SFR2          AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
27_SFR2_seq      AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 147
                 ************************************************************

AK_SFR2          GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
27_SFR2_seq      GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 207
                 ************************************************************

AK_SFR2          ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
27_SFR2_seq      ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 267
                 ************************************************************
```

```
AK_SFR2          CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 840
27_SFR2_seq      CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 327
                 ************************************************************

AK_SFR2          ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 900
27_SFR2_seq      ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 387
                 ************************************************************

AK_SFR2          TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
27_SFR2_seq      TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 447
                 ************************************************************

AK_SFR2          GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 1020
27_SFR2_seq      GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 507
                 ************************************************************

AK_SFR2          TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 1080
27_SFR2_seq      TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 567
                 ************************************************************

AK_SFR2          AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 1140
27_SFR2_seq      AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 627
                 ************************************************************

AK_SFR2          CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 1200
27_SFR2_seq      CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 687
                 ************************************************************

AK_SFR2          AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 1260
27_SFR2_seq      AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 747
                 ************************************************************

AK_SFR2          GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 1320
27_SFR2_seq      GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 807
                 ************************************************************

AK_SFR2          CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 1380
27_SFR2_seq      CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 867
                 ************************************************************

AK_SFR2          CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 1440
27_SFR2_seq      CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 927
                 ************************************************************

AK_SFR2          ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 1500
27_SFR2_seq      ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 987
                 ************************************************************

AK_SFR2          TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1560
27_SFR2_seq      TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1047
                 ************************************************************

AK_SFR2          GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1620
27_SFR2_seq      GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1107
                 ************************************************************

AK_SFR2          TTATTCTCC 1629
27_SFR2_seq      TTATTCTCC 1116
                 *********
```

- 28_SFR2 alignment

```
AK_SFR2          ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
28_SFR2_seq      ------------------------------------------------------------
```

```
AK_SFR2          GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 840
28_SFR2_seq      ------------------------------------------------------------

AK_SFR2          ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 900
28_SFR2_seq      ---------GTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 51
                          **************************************************

AK_SFR2          TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
28_SFR2_seq      TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 111
                 ************************************************************

AK_SFR2          GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 1020
28_SFR2_seq      GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 171
                 ************************************************************

AK_SFR2          TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 1080
28_SFR2_seq      TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 231
                 ************************************************************

AK_SFR2          AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 1140
28_SFR2_seq      AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 291
                 ************************************************************

AK_SFR2          CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 1200
28_SFR2_seq      CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 351
                 ************************************************************

AK_SFR2          AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 1260
28_SFR2_seq      AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 411
                 ************************************************************
```

```
AK_SFR2          GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 1320
28_SFR2_seq      GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 471
                 ************************************************************

AK_SFR2          CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 1380
28_SFR2_seq      CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 531
                 ************************************************************

AK_SFR2          CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 1440
28_SFR2_seq      CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 591
                 ************************************************************

AK_SFR2          ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 1500
28_SFR2_seq      ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 651
                 ************************************************************

AK_SFR2          TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1560
28_SFR2_seq      TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 711
                 ************************************************************

AK_SFR2          GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1620
28_SFR2_seq      GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 771
                 ************************************************************

AK_SFR2          TTATTCTCC 1629
28_SFR2_seq      TTATTCTCC 780
                 *********
```

- 29_SFR2 alignment

```
AK_SFR2          ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
29_SFR2_seq      ------------------------------------------------------------
```

```
AK_SFR2          AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 840
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 900
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GACCCTAATGCAATTGAAGTAGCAACATCTACTCTGCCAACTGGTGTATACAATCAGGCT 1020
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          TTGCATTGGATGGCTATTGCACATTCTGAAGCCTATGACTACATACATTCGAAAAGCAAG 1080
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AACGAAAGGAAGCCAATAGTTGGTGTTGCACACCATGTATCGTTTACAAGGCCATATGGG 1140
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          CTATTTGATGTTGCTGCTGTCGCGCTAGCTAACTCATTGACCCTTTTTCCTTACGTGGAT 1200
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          AGCATATGTGATAAATTGGACTTCATTGGAATCAACTACTATGGACAGGAGGTTATATCA 1260
29_SFR2_seq      ------------------------------------------------------------

AK_SFR2          GGACCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 1320
29_SFR2_seq      ---CCTGGTCTAAAGCTTGTGGATAATGATGAGTATAGTGAATCTGGTCGTGGGGTTTAT 57
                    ********************************************************

AK_SFR2          CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 1380
29_SFR2_seq      CCTGATGGGCTGTTCCGCATCCTGATTCAATTCAATGAACGATATAAGAGATTAAATATA 117
                 ************************************************************

AK_SFR2          CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 1440
29_SFR2_seq      CCTTTTGTCATTACTGAAAATGGAGTTTCTGATGAGACTGATCTGATACGGAAACCATAT 177
                 ************************************************************

AK_SFR2          ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 1500
29_SFR2_seq      ATACTGGAACACTTGTTAGCCACATACGCTGCCATCATTATGGGTGTCCGTGTACTTGGT 237
                 ************************************************************

AK_SFR2          TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 1560
29_SFR2_seq      TATTTGTTTTGGACAACATCAGATAATTGGGAATGGGCGGATGGCTATGGTCCCAAGTTT 297
                 ************************************************************

AK_SFR2          GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 1620
29_SFR2_seq      GGGCTTGTTGCTGTTGACCGTGCTAATAACCTAGCACGGAAACCTCGGCCTTCATACTTT 357
                 ************************************************************

AK_SFR2          TTATTCTCC 1629
29_SFR2_seq      TTATTCTCC 366
                 *********
```

**5. DNA sequence alignment between the AK_SFR2 as reference (first row) and the C-terminal gene deletion (second row) from DNA sequencing.**

- 31_SFR2_delC(mP) alignment

```
AK_SFR2                 ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
31_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
31_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
31_SFR2_delC_mP_seq     -----------------------------------GACTTCCGCGCCTTTCCCTCCTCC 24
                                                           ************************

AK_SFR2                 GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
31_SFR2_delC_mP_seq     GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 84
                        ************************************************************

AK_SFR2                 GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
31_SFR2_delC_mP_seq     GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 144
                        ************************************************************

AK_SFR2                 AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
31_SFR2_delC_mP_seq     AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 204
                        ************************************************************

AK_SFR2                 GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
31_SFR2_delC_mP_seq     GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 264
                        ************************************************************

AK_SFR2                 AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
31_SFR2_delC_mP_seq     AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 324
                        ************************************************************

AK_SFR2                 AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
31_SFR2_delC_mP_seq     AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 384
                        ************************************************************

AK_SFR2                 CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
31_SFR2_delC_mP_seq     CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 444
                        ************************************************************

AK_SFR2                 AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
31_SFR2_delC_mP_seq     AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 504
                        ************************************************************

AK_SFR2                 GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
31_SFR2_delC_mP_seq     GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGC---------------------- 542
                        ************************************

AK_SFR2                 ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
31_SFR2_delC_mP_seq     ------------------------------------------------------------
```

- 32_SFR2_delC(mP) alignment

```
AK_SFR2                 ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
32_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
32_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
32_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
32_SFR2_delC_mP_seq     ------------------------------------------------------------
```

```
AK_SFR2                 GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
32_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
32_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
32_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
32_SFR2_delC_mP_seq     ------------------------------------------------------------

AK_SFR2                 AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
32_SFR2_delC_mP_seq     --------------------------------CACAATGTTGCTGCTTGGCATAATGTT 27
                                                         ***************************

AK_SFR2                 CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
32_SFR2_delC_mP_seq     CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 87
                        ************************************************************

AK_SFR2                 AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
32_SFR2_delC_mP_seq     AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 147
                        ************************************************************

AK_SFR2                 GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
32_SFR2_delC_mP_seq     GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGC---------------------- 185
                        ***********************************

AK_SFR2                 ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
32_SFR2_delC_mP_seq     ------------------------------------------------------------
```

- 33_SFR2_delC alignment

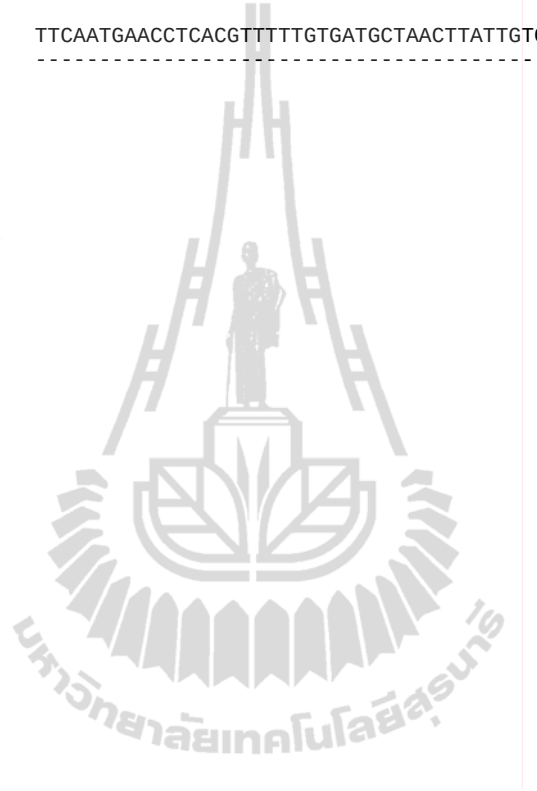
```
AK_SFR2                 ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
33_SFR2_delC_seq        ------------------------------------------------------------

AK_SFR2                 GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
33_SFR2_delC_seq        ------------------------------------------------------------

AK_SFR2                 CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
33_SFR2_delC_seq        -----------------------------------GACTTCCGCGCCTTTCCCTCCTCC 24
                                                           ************************

AK_SFR2                 GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
33_SFR2_delC_seq        GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 84
                        ************************************************************

AK_SFR2                 GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
33_SFR2_delC_seq        GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 144
                        ************************************************************

AK_SFR2                 AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
33_SFR2_delC_seq        AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 204
                        ************************************************************

AK_SFR2                 GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
33_SFR2_delC_seq        GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 264
                        ************************************************************

AK_SFR2                 AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
33_SFR2_delC_seq        AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 324
                        ************************************************************

AK_SFR2                 AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
33_SFR2_delC_seq        AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 384
                        ************************************************************

AK_SFR2                 CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
33_SFR2_delC_seq        CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 444
                        ************************************************************
```

```
AK_SFR2             AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
33_SFR2_delC_seq    AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 504
                    ************************************************************

AK_SFR2             GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
33_SFR2_delC_seq    GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 564
                    ************************************************************

AK_SFR2             ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
33_SFR2_delC_seq    ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 624
                    ************************************************************

AK_SFR2             CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 840
33_SFR2_delC_seq    CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 684
                    ************************************************************

AK_SFR2             ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 900
33_SFR2_delC_seq    ATGGATTTTGTGAGGCTTGTTGTTGATTGT---------------------------- 714
                    ******************************

AK_SFR2             TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
33_SFR2_delC_seq    ------------------------------------------------------------
```

- 34_SFR2_delC alignment

```
AK_SFR2             ATGCCACTACCGGCGTTCGTGGCGGCGGCGGCGAGGCTCGCCGTCCTCGTGGCCGCGGCG 60
33_SFR2_delC_seq    ------------------------------------------------------------

AK_SFR2             GCGACGGCGGCCAATGCCGCCTCCTACGCGCGGTACCGGCGGCGCCACCTCCGCCGCATC 120
33_SFR2_delC_seq    ------------------------------------------------------------

AK_SFR2             CCCAGCCCCATCGACGAGTCCGCCGATCCCCTCGCCGACTTCCGCGCCTTTCCCTCCTCC 180
33_SFR2_delC_seq    ------------------------------------------------------------

AK_SFR2             GACGCTGACGACTCAGAGGAGGATAATTTCTTCTTTGGGCTAGCAACAGCGCCTGCGCAT 240
33_SFR2_delC_seq    ------------------------------------------------------------

AK_SFR2             GTTGAGGACAGGCTAGAAGATGCTTGGCTCCAGTTTGCAACTGAAACTTCCTGCGATGAC 300
33_SFR2_delC_seq    ------------------------------------------------------------

AK_SFR2             AATGGAAACGTGCGCGACCAGAGGCCAGTAGATGCACTGATGGCATCTGCTGCTGGTGAT 360
33_SFR2_delC_seq    ------------------------------------------------------------

AK_SFR2             GGAGGCTCCCAGCAATCTTGGAGGTCAACAGGTGGGGAAAATATTGGTGATAGAGAGCAG 420
33_SFR2_delC_seq    ------------------------------------------------------------

AK_SFR2             AGGAAACCACTTAGGGTAGCCATGGAGGCTATGCTCAGGGGGTTTGAAATTCTTGCTGAG 480
33_SFR2_delC_seq    ------------------------------------------------------------

AK_SFR2             AGTGGAGAATCTGCTGGCGGCGATAACTGCAGCCACAATGTTGCTGCTTGGCATAATGTT 540
33_SFR2_delC_seq    --------------------------------CACAATGTTGCTGCTTGGCATAATGTT 27
                                                    ***************************

AK_SFR2             CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 600
33_SFR2_delC_seq    CCATGCCCGCAAGAAAGGCTTAGATTTTGGTCTGATCCTGATGCTGAGTTGAAACTTGCT 87
                    ************************************************************

AK_SFR2             AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 660
33_SFR2_delC_seq    AAGGAGACCGGCATCAGCGTTTTCCGCATGGGGGTAGATTGGGCAAGGTTAATGCCTGAG 147
                    ************************************************************

AK_SFR2             GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 720
33_SFR2_delC_seq    GAACCAACCGAAGAATTGAAGAGCTCAGTTAATTTTGCAGCACTTGAGCGGTATAGATGG 207
                    ************************************************************
```

```
AK_SFR2             ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 780
33_SFR2_delC_seq    ATCATTCAAAGGGTTCGCGAATATGGAATGAAAGTAATGCTTACTCTGTTTCATCACTCA 267
                    ************************************************************

AK_SFR2             CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 840
33_SFR2_delC_seq    CTTCCACCCTGGGCTGGAAAATATGGCGGGTGGAAGATGGAAAAAACTGTCACCTATTTC 327
                    ************************************************************

AK_SFR2             ATGGATTTTGTGAGGCTTGTTGTTGATTGTGTATCCAATTTAGTGGACTACTGGGTGATT 900
33_SFR2_delC_seq    ATGGATTTTGTGAGGCTTGTTGTTGATTGT------------------------------ 357
                    ******************************

AK_SFR2             TTCAATGAACCTCACGTTTTTGTGATGCTAACTTATTGTGCCGGTGCTTGGCCTGGTGGA 960
33_SFR2_delC_seq    ------------------------------------------------------------
```



# APPENDIX II

## 1. Media

### 1.1 LB-ampicillin/kanamycin/tetracycline broth

| | | |
|---|---|---|
| Yeast extracts | 5 | g/L |
| Tryptone | 10 | g/L |
| Sodium chloride | 10 | g/L |

Autoclave at 121 °C, for 15 min. Kanamycin, Tetracycline, and Ampicillin are added in the warm medium at 100 μg/ml, 12.5 μg/ml and 50 μg/ml final concentrations, respectively after autoclaving. If the agar medium is desired, 15 g/L of agar is added into the medium before autoclaving.

### 1.2 LB-ampicillin/chloramphenicol broth

| | | |
|---|---|---|
| Yeast extracts | 5 | g/L |
| Tryptone | 10 | g/L |
| Sodium chloride | 10 | g/L |

Autoclave at 121 °C for 15 min. Chloramphenicol and Ampicillin are added in the warm medium at 34 μg/ml and 50 μg/ml final concentrations, respectively after autoclaving. If the agar medium is desired, 15 g/L of agar is added into the medium before autoclaving.

**1.3 Yeast extract peptone dextrose (YPD) Zeocin broth**

| | | |
|---|---|---|
| Yeast extracts | 10 | g/L |
| Peptone | 20 | g/L |
| Dextrose | 20 | g/L |

Autoclave at 121°C for 20 minutes then cooling down to 55°C and add zeocin at 100 μg/ml final concentrations. If the agar medium is desired, 15 g/L of agar is added into the medium before autoclaving.

**1.4 Buffered glycerol complex (BMGY) medium**

| | | |
|---|---|---|
| Yeast extracts | 10 | g/L |
| Peptone | 20 | g/L |
| Glycerol | 10 | g/L |

Dissolved in 100 mM potassium phosphate buffer pH 6 and the final volume is adjusted to 1 L. The media is sterilized for 15 min at 121°C. Zeocin is added into the sterilized medium at 100 μg/ml final concentration.

**1.5 Reagent for SDS-PAGE**

**1.5.1 SDS-gel loading buffer 5X stock, 8 ml**

Dissolve 0.30 g Tris Base, 1 g SDS, 0.05 g Bromophenol blue, 5 ml glycerol and adjust pH to 6.8 with HCL and the volume to 8 ml with distilled water. Before used add 20 μl of 2-mercapthoethanol to 80 μl of solution mixture. Store at room temperature.

### 1.5.2 1.5 M Tris-Cl pH 8.8 100 ml

Dissolve 18.17 g Tris Base in 80 ml distilled water. Adjust pH to 8.0 with HCl and adjust the volume to 100 ml with distilled water. Store 4 °C.

### 1.5.3 0.5 M Tris-Cl pH 6.8 100 ml

Dissolve 6.06 g Tris Base in 80 ml distilled water. Adjust pH to 6.8 with HCl and adjust the volume to 100 ml with distilled water. Store 4 °C.

### 1.5.4 30 % Acrylamide solution 100 ml

Dissolve 29 g acrylamide and 1 g N, N.-methyene-bis-acrylamide in distilled water and adjust the volume to 100 ml. Mix the solution by stirring in dark bottle for 1 h or till homogenious and filter through Whatman membrane No. 1. Store in the dark bottle at 4 °C.

### 1.5.5 Tris-Glycine electrode buffer 5x stock, 1 L

Dissolve 15.1 g Tris base, 94 g glycine and 5 g SDS in distilled water. Adjust pH to 8.3 with HCl and adjust the volume to 1 L with distilled water.

### 1.5.6 Staining solution with Coomassie brilliant blue for protein 1 L

Mix 1 g Coomassie brilliant blue R-250, 400 ml methanol, 500 distilled water and 100 ml glacial acetic acid.

### 1.5.7 Destaining solution for Coomassie Stain 1 L

Mix 400 ml methanol, 100 ml glacial acetic acid and add distilled water to a final volume of 1 L.

### 1.5.8 10% w/v Ammonium persulfate 1 ml

Dissolve 100 mg ammonium persulfate in 1 ml distilled water. Store at -20 °C.

## 1.6 Reagent for western blot

### 1.6.1 Transfer buffer 500 ml

Mix 50 ml methanol, 100 ml 5X Tris-Glycine and add distilled water to a final volume 1 L.

### 1.6.2 10X PBS stock

Dissolve 80 g NaCl, 2 g KCl, 14.4 g Na2HPO4 and 2.4 g KH2PO4 then add distilled water to final volume 1 L. Adjust pH to 7.4 with HCl. Sterile by autoclaving at 121 °C for 15 min.

### 1.6.3 PBST

Mix 100 ml of 10X PBS and add 500 µl Tween-20 then add distilled water to final volume 1 L.

## 1.7 Reagent for northern blot

### 1.7.1 10X FA gel buffer

Dissolve 41.8 g MOPS, 6.8 g sodium acetate and 3.72 g EDTA in 500 ml RNase free water then adjust pH with NAOH to 7.0. Add RNase free water to final volume 1 L. Sterile by autoclaving at 121 °C for 15 min.

### 1.7.2 1X FA running buffer

Mix 100 ml of 10X FA gel buffer, 20 ml 37% (12.3 M) formaldehyde and 880 ml RNase free water.

### 1.7.3 1% FA gel

Mix 1 g agarose, 10 ml of 10X FA gel buffer and add RNase-free water to final volume 100 ml. Heat the mixture to melt agarose. Cool to 65 °C and add 1.8 ml of 37% formaldehyde. Mix thoroughly and pour onto gel support.

### 1.7.4 5X RNA loading buffer

Dissolve 16 µl of saturated aqueous bromophenol blue solution, 80 µl of 500 mM EDTA, 720 ml 37% formaldehyde, 2 ml of 100% glycerol, 3.084 ml formamide and 4 ml of 10X FA gel buffer. Add RNase-free water to 10 ml. Store at 4 °C.

# BIOGRAPHY

Sasiprapa Kanjanawattana was born on April 26th, 1984 in Nakhon Ratchasima. She obtained her Bachelor degree of Science in Food Technology from the Institution of Agricultural Technology, Suranaree University of Technology, in 2006. After graduation, she had opportunity to study in Master Degree in the School of Biotechnology, Institution of Agricultural Technology at Suranaree University of Technology. Her thesis title is "The expression and detection of rice SFR2 in *Escherichia coli* and *Pichia pastoris*". Some parts of this study have been presented as a poster presentation at The Annual Meeting and International Conference of the Thai Society for Biotechnology (TSB 2008) at Maha Sarakham, The 15th National Graduated Research Conference 2009, Rajaphat Nakhon Ratchasima and the 3rd SUT Graduate Conference 2010, SUT.