Title
Development of Secretion System for Heterologous Recombinant Protein Expression in E.coli Bacteria and Pichia pastoris Yeast

Permalink
https://escholarship.org/uc/item/299047nt

Author
Qu, He

Publication Date
2013

Peer reviewed|Thesis/dissertation
Development of Secretion System for Heterologous Recombinant Protein Expression in

*E.coli* Bacteria and *Pichia pastoris* Yeast

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science
in
Bioengineering
by
He Qu

March 2014

Thesis Committee:
Dr. Jiayu Liao, Chairperson
Dr. Kaustabh Ghosh
Dr. William H. Grover
The Thesis of He Qu is approved:

__________________________________________

__________________________________________

__________________________________________

Committee Chairperson

University of California, Riverside
ACKNOWLEDGMENTS

I would like to express my sincere appreciation to my thesis chair, Dr. Jiayu Liao. This thesis would not have happened without his directing, advising and supporting. His continuous patience and encouragement helped me grow professional. I truly appreciate he gave me the opportunities to be part of his group.

I also would like to thank my other thesis committee members, Dr. Kaustabh Ghosh and Dr. William H. Grover for their advice, expertise and time. And a special appreciation to my parents for financially and spiritually supports.

Last, I would like to thank my colleagues, Dr. Ling Jiang; Jose Alanis; Hilda Wiryawan; Nicole Duong; Harbani Kaur Malik; Raphael Kung; Amanda Saavedra; George Way. Specially, Dr. Ling Jiang for her wise counsel with regards to experimental techniques; Jose Alanis and Nichole Duong for assistance towards expedition of experimentation.

It has been an honor and pleasure to be present in such a friendly working environment like Dr. Liao’s lab. Thanks for all the remarkable memories everyone has been created for me.
ABSTRACT OF THE THESIS

Development of Secretion System for Heterologous Recombinant Protein Expression in *E.coli* Bacteria and *Pichia pastoris* Yeast

by

He Qu

Master of Science, Graduate Program in Bioengineering
University of California, Riverside, March 2014
Dr. Jiayu Liao, Chairperson

Small ubiquitin-like modifier (SUMO) plays a crucial role in many cellular activities, such as regulation of cell cycle, transcription, apoptosis and DNA damage response. SUMOylation pathway is an important mechanism in modifying protein function and regulates protein activities in eukaryotes. Deregulation of SUMOylation pathway is found in many diseases, for instance, neurodegenerative disease, human cancers, heart disease and diabetes. SUMOylation pathway is a three-step process involves a multi-enzyme catalyzed cascade which consists of the following enzymatic steps: activation (E1), conjugation (E2) and ligation (E3). Only some E3 ligation proteins are understood. One of the E3 ligase types have extensively studied is PIAS. It is a protein inhibitor of activated STAT (PIAS1). Intracellular expression of PIAS1 with the bacterial cells does not work well due to the toxicity PIAS1 can create within the cells.

We established two secretion systems to test SUMO-related recombinant proteins expression: Bacteria secretion system and yeast secretion system. We used type II
secretion system which includes Sec, Tat and SRP- pathways to secreted recombinant protein production in *E.coli* K-12 strain. Recombinant proteins also been tested on *Pichia pastoris* yeast cell line. Since yeast is a eukaryote, it is a suitable host for produce eukaryotic heterologous protein like SUMO-related proteins.
# TABLE OF CONTENTS

Acknowledgement ............................................................................................................iv

Abstract of Thesis ...........................................................................................................v

Introduction .......................................................................................................................8

Chapter 1

1.1 Introduction .............................................................................................................20

1.2 Material and Method ..............................................................................................23

1.3 Results ....................................................................................................................29

1.4 Discussion ...............................................................................................................35

CHAPTER 2

2.1 Introduction .............................................................................................................37

2.2 Material and Method ..............................................................................................39

2.3 Results ....................................................................................................................41

2.4 Discussion ...............................................................................................................44

References ......................................................................................................................45
LIST OF FIGURES

Figure 1: Three-dimensional structures of Ubiquitin and SUMO-1...............................3
Figure 2: Mechanism of reversible SUMOylation......................................................4
Figure 3: Förster Resonance Energy Transfer process..............................................8
Figure 4: Overview of the *E.coli* Sec, SRP and Tat translocations from cytoplasm to periplasm.................................................................12
Figure 5: List of secretion signal sequences used in *E.coli*...................................15
Figure 6: Engineered plasmid contains bacteria Sec-pathway secretion signal ..........26
Figure 7: Engineered plasmid contain *PelB* signal sequences, gene of interested Ubc9 tagged with YPet and Arabinose promoter...........................................26
Figure 8: Fluorescence activity difference between non-induced and induced protein secretion in the periplasmic spaces.........................................................30
Figure 9: YPet Fluorescence intensity protein secreted in periplasmic space by Sec, Tat, and SRP pathway..............................................................31
Figure 10: Fluorescence intensity difference between non-induced and induced intracellular protein expression............................................................32
Figure 11: YPet Fluorescence intensity protein in cytoplasmic space by Sec, Tat, and SRP pathways.................................................................33
Figure 12: YPet Fluorescence intensity protein secretion in culture medium by Sec, Tat, and SRP pathways.................................................................34
Figure 13: Fluorescence intensity yield of YPet-Ubc9 protein expression distribution...35
Figure 14: Yeast secretion vector pPIC9 cloned with gene of interest Ubc9..............40
Figure 15: Fluorescence intensity of YPet-Ubc9 protein secretion in *Pichia pastoris* at 4 different time points……………………………………………………………………...42

Figure 16: Fluorescence intensity of YPet-PIAS protein secretion in *Pichia pastoris* at 4 different time points……………………………………………………………………...43
INTRODUCTION

1.1 Post-translation protein Modification

Posttranslational modifications (PTMs) play a very crucial role for generating protein to be fully functional and help utilize protein for different cellular activities. PTMs on protein occur after translation process, during which the polypeptide chain is mortified into mature functional proteins. The common processes to modify polypeptide in order to make protein fully functional are glycosylation, phosphorylation, methylation and small polypeptide modification [1, 2]. Cell use PTMs as a critical tool to control the function of protein by determining its activity state, subcellular localization, turnover, as well as interactions with other proteins. This mechanism can either be reversible or irreversible, enzymatic catalysis is required in most modification cases [1-4].

1.2 Ubiquitin

Ubiquitin is one of the well-known examples of Post Translation Modifier. It is a small 74 amino acid long cells (molecular weight of ~8.5kDa ) regulatory protein that has been found in almost all eukaryotic cells[5]. Studies have showed ubiquitin plays a significant role in covalently linking to many proteins for proteasome dependent proteolytic degradation. Several other proteins have been discovered which belong to ubiquitin family and share similar sequences, including SUMO (small ubiquitin-related modifier), Nedd8 (also called Rub1), Apg8, parkin ect. Most ubiquitin-related protein modifiers require enzymes to help conjugate with substrate proteins [2, 6].
1.3.1 SUMO Protein and its Function

SUMO (Small Ubiquitin-related MOdifier) could be the most intriguing Ubiquitin-related protein modifiers (Ubls). It was identified as a reversible post-translational protein as well [7, 8]. It regulates diverse cellular processes directly or indirectly, including participate in transcription, DNA repair, protein activation, apoptosis, stress response and signal transduction in eukaryotes [9-11]. The SUMO gene(SMT3) was first discovered in 
*Saccharomyces cerevisiae* in the mid-1900s [12]. Research has found that this gene is conserved throughout most of the eukaryotic kingdom, such as yeast, plants and vertebrates. SUMO has also been discovered attached to more than 1000 proteins, such as androgen receptor, IKBα, c-jun, p53, etc. [7, 9].

Humans express 4 SUMO paralogue proteins, SUMO1-SUMO4. SUMO 1 is ~45% identical to SUMO2 and 3, while SUMO 2 and SUMO3 are share 97% similarity of the sequences identity for human proteins [13]. They appear to be conjugate with protein more preferentially if under stress conditions [2, 13, 14]. Mature SUMO2 is 92 amino acids long, and mature SUMO3 consist 93 amino acids [8]. These proteins are about 12 KDa in size. Research has being mainly focus on SUMO-1 since it is the most prominently conjugated isoform.

Compare with ubiquitin, all SUMO proteins contain a 10-25 amino-acid tail at their N-terminal domains which is not found in any other ubiquitin related proteins [7, 15]. Another prominent feature is the surface charge distributions of these two proteins are quiet different, which means they do not share the same enzymes and substrates [2].
Moreover, even though SUMO shares 18% sequence identity with ubiquitin, it folded structure can virtually superimpose to ubiquitin (Figure 1).

Figure 1: Three-dimensional structures of Ubiquitin and SUMO-1[2]. The ribbon diagrams compare the Ubiquitin and SUMO-1 similarity of their structure. Even though their sequences only 18% identity, their three-dimensional structures are very similar. Notably, SUMO has a long and flexible N-terminal extension which is not found in ubiquitin.

1.3.2 Mechanisms of SUMOylation Pathway

The reversible SUMOylation pathway occurs as a three-step enzymatic cascade. Before the reversible cascade begins, SUMO has to synthesize as a precursor protein catalyzed by SUMO-specific proteases (SENPs) to become mature. As a result, the nascent SUMO has to be proteolytically cleaved to expose the C-terminal glycine-glycine(GG) motif to form an isopeptide bond with ε-amino group of a Lys residue in its substrates [7, 8]. Posttranslational conjugation of mature SUMOs occurs through the cascade step 2-5 represent in diagram 2. First of all, the mature SUMO protein activates its C-terminal
glycine motif and Cys residue of heterodimer SUMO-specific E1 activating enzyme (Uba2/Aos1) to form a thioester bond. This reaction undergo ATP-dependent. Next activated SUMO is transferred from E1 enzyme to a SUMO- conjugating enzyme E2 (Ubc9), forming a thioester linkage between the catalytic Cys residue of Ubc9 and the C-terminal carboxy group of SUMO [16-18]. Then, SUMO transfers to the substrate to form an isopeptide bond between SUMO C-terminal glycine residue and a lysine side chain of the substrate, usually together with specific SUMO E3 ligase [7, 8]. Last, SENPs will cleave SUMO from conjugated target substrate which release SUMO to free for further cycles[19].

Figure 2: A schematic represent the mechanism of reversible sumoylation [7]. It summarized the steps of SUMO (small ubiquitin-related modifier) maturation,
conjugation and deconjugation. First SUMO is proteolytically processes by SENPs to remove its C-terminal tail form the diglycine motif, represent as the step1. Mature SUMO in ATP-dependent to form a thioester bond between di-glycine and the activating enzyme Uba2/Aos1 in step 2. SUMO is then transferred to the cysteine residue of E2 (Ubc9), step3. In step 4, E3 ligases enzyme help SUMO to conjugate to the substrate protein. Finally, SUMO can be removed from conjugated species by the action of SENP (Step5).

1.3.3 Enzymes involved in SUMOylation pathway

Most organisms contain a single SUMO-activating enzyme (E1) which required for conjugation of SUMO to substrates. The E1 enzyme usually is a heterodimer and composed of two subunits, one is known as Aos1 located at the N-terminal and the other subunit is Uba2 which corresponds at the C-terminal contains the active site cysteine [8, 9, 14, 20]. The second enzyme called conjugation enzyme (E2) involved in the step after SUMO being activated. Conjugation enzyme E2 plays an essential role in SUMOylation cascade. E2 (Ubc9) enzyme is the only SUMO-conjugating enzyme in yeast, invertebrates and most of vertebrates. It directly involved in the selection of SUMO targets. These targets carry a consensus site which Ubc9 can bind directly to become sumoylated [8, 9].

The third enzyme involved in SUMOylation cascade is SUMO ligases enzyme(E3) which catalyze the transfer Ubl proteins from an E2 enzyme to a substrate [21]. One of the functions of E3 ligase is mediates or stabilizes the interaction of the substrate with the E2. Three distinct types of SUMO ligases(E3s) have been discovered which are PIAS (protein inhibitor of activated STAT) family, nuclear pore protein RanBP2/Nup358, and polycomb group protein Pc2 [22-24].PIAS proteins is the most type which been
extensively studied and it were described as negative regulators of STAT transcription factor initially. Research as discovered at least four genes in human encode PIAS proteins, which are PIAS1, PIAS3 PIASxα, PIASxβ and PIASy [25-27]. This family shares a similarity of structure SP-RING domain which is crucial for directly binding to Ubc9. Different PIAS proteins may be sumoylating different substrates, however currently studies have not fully understand the combination of substrate and E3 ligase[28, 29]. Studies have found PIAS3 ha is a specific inhibitor of Stat3 signaling; PIAS1 inhibits Stat1-mediate gene activation[27].

1.3.4 Roles of SUMO in Human Health Diseases

Since more than a thousand proteins can be targeted by SUMO, disruption of SUMO pathway can have a humongous impact on many diseases. Studies have discovered that SUMO is involved in many neurodegenerative diseases, such as Huntington’s, Alzheimer’s, and Parkinson’s disease [15]. Many key proteins associate in neuronal disorder can be modified by SUMO. For example, SUMO’s E2 enzyme Ubc9 has been found in several types of human cancer. It associated with tumor growth in breast carcinoma as well as it been described to be involved in stress response mechanisms which are related to tumor genesis [15, 30, 31]. SUMO not only plays a critical role in caners and neuronal diseases, but also associates with heart disease and diabetes [32, 33]. Fully understanding this pathway will create a significant impact on pathogenesis investigation and advance research for a potential target for therapeutic application.
1.4 Förster Resonance Energy Transfer (FRET)

Förster Resonance Energy Transfer is a nonradioactive energy transfer between the excited chromophore (donor) to a proximal ground state chromophore (acceptor) through long-range dipole-dipole interaction (figure 3). It is an electrodynamic phenomenon which explained by using classical physics. When the donor and acceptor with favorable orientations are close to each other, 2-10nm, the energy evoke from the excited donor will transfer to the acceptor to induce emission. The range in 2-10nm distance between donor and acceptor is comparable to the size of biomolecules or the distance of protein sites [34]. Therefore, FRET-based techniques have been extensively used in lots biological researches, such as study protein folding, monitor protein-protein interactions, high-throughput screening for drug usage, and sensory of signaling pathway [35, 36].

Suitable fluorophore FRET pairs are the key to the success of a FRET application. Traditional Cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) is the commonly used FRET pair. An engineered FRET pair, CyPet and YPet, was developed from CFP and YFP by mutagenesis shows a significant improvement of FRET efficiency[37].
Figure 3[38]: Schematic of the FRET process: the excited state donor molecule transfers energy non radiatively to acceptor molecule located at distance $r$ from the donor. The acceptor releases the energy either through fluorescence or nonradiative channels.
2. 1 Recombinant Protein Secretion in Escherichia coli

*Escherichia coli* is the most robust biological expression system for the successful production of many recombinant proteins for pharmaceuticals and industries. As a production host organism, it has numerous desirable characteristics such as easy manipulation, fast cell growth and simple fermentation [39]. However, some proteins contain complex disulfide bonds which help in structure proper fold or require post-translational modification. These types of modifications cannot be achieved in *E.coli*[40]. To overcome the problems, variety strategies such as using different promoters to regulate the expression level, changing host strains, reducing culture medium condition, co-expressing chaperones, or secreting protein out of cell membrane have been developed [40].

Compare with intracellular recombinant proteins expression, periplasm and extracellular secretory recombinant proteins in *E.coli* have several advantages. Primarily include facilitate disulfide-bond formation folding correctly, for example proinsulin increased 10-fold expression when the protein was secreted to the periplasmic [41]. Secretory recombinant proteins purification can be much simpler and less challenge for contamination. It may also have higher product stability and solubility, for instance production of penicillin amidase as an secreted protein avoided severe degradation problem as compared to intracellular production [42]. Moreover, secreted recombinant proteins can prevent incorrect formation of disulfide bonds since periplasmic space provides more oxidative environment than the cytoplasm [43-46]. Further, secreted
protein can guarantee the reliability of express polypeptide on N-terminal since it often used for cleavage of signal sequence [45].

2.2 Type I-V Secretion System in Bacteria

In Gram-negative bacteria, secreted proteins have to across the two membranes of the cell envelope and five types, I through V, of secretion pathways are used for protein transport from cytosol into periplasm or extracellular. Type I secretory pathway has been widely used for secretion of high-molecular-weight toxins, proteases and lipases secretion. This pathway secreted proteins directly from the cytoplasm across the outer membrane without a periplasmic intermediate and requires ATP [47-49]. This pathway can secrete protein ranging from 50 to over 4000 amino acids [50]. Type II contains Sec-dependent pathway, the signal recognition particles (SRP), and the twin-arginine translocation (TAT) pathway. To transport proteins in this system requires two steps, periplasmic translocation and extracellular transport [39]. Type III system allow extracellular bacteria adhering to the surface of a host cell to inject specialized proteins across the plasma membrane. The injected proteins subvert the functioning of the cell and destroy its communication, favoring the survival of the invading bacteria. It’s an essential for the pathogenicity of many pathogenic bacteria[51]. Type IV secretion system had found in both Gram-negative and Gram-positive bacteria. In Gram-negative bacteria, they mediate the secretion of monomeric proteins, multisubunit protein toxins and nucleoprotein complexes [52]. The last system is Type V; the protein secretion mechanisms are the simplest one amount others. Type V system includes autotransporter pathway and two–
partner secretion pathway. Proteins secreted through this system have similarities in their primary structures as well as their modes of biogenesis [53].

2.3 Sec- and SRP-pathway of Type II Secretion System

Type II secretion system is the most widely used one for secreted recombinant protein production in *E.coli* K-12 strain [45]. As mention earlier, this system contains three secretion pathways which are Sec, SRP and Tat pathways. In *E.Coli*, more than 90% of secreted proteins translocated out of the cytosol are through SecYEG pore. Proteins are targeted to the SecYEG pore is processed either by Secretion (Sec) or Signal recognition particle (SRP) pathways. Protein exported by Sec pathway occurs after translational of polypeptide whereas SRP pathway is co-translational targeting [46, 54]. For both Sec and SRP pathway, single stranded mRNA first synthesized by ribosome to become a pre-protein. The pre-protein must be maintained in an unfolded conformation which accomplished by interacting with either cytosolic chaperone SecB or SRP. One the chaperone binds to the pre-protein containing a signal peptide at N-terminal and moves them to inner membrane-bound SecA[55].

When the pre-protein travels across the inner membrane through SecYEG complex channel, chaperon detach the pre-protein and the signal peptide is removed by a signal peptidase. The pre-protein being transfer to periplasmic will finish folding itself [54].
Figure 4 [55]: Schematic overview of the *Escherichia coli* Sec, SRP and Tat translocations from cytoplasm to periplasm. a: Proteins post-translational delivery to Section (Sec) or Twin-arginine translocation (TAT) translocase. In Sec pathway, chaperone secB maintain the proteins unfold. SecA uses ATP hydrolysis to drive the protein through SecYEG channel. Tat pathway is able to secreted folded protein. b: Protein transfer by SRP pathway is co-translational deliver. Signal Recognition Particle (SRP) used as a signal sequence to help protein secretion through SRP pathway.

### 2.4 Tat-pathway of Type II Secretion System

Twin-arginine translocation (TAT) pathway is a capable system for secret protein across the cytoplasmic membrane by using the signal peptide contains a twin-arginine sequence [40]. This system is present in some, but not all, bacteria and archaea organisms. It also been identified as an essential system in human pathogen *Mycobacterium tuberculosis*, and the predatory bacterium *Bdellovibrio bacteriovorus*. Protein translocated by Tat-pathway is by a post-translational mechanism and the substrates often synthesized as precursor proteins that contain similar signal sequences as Tat at N-terminal. The biggest
substrate numbers use this pathway are species of the Gram-positive genus *Streptomyces coelicolor* and *Streptomyces scabies* [56, 57].

The most crucial element about Tat-pathway is it can transport fully-folded proteins across the ironically tight membrane, whereas Sec pathway is restricted to translocating polypeptides in an unstructured state. Transporting fully folded proteins is such a difficult task that so far only one other protein transport system which named peroxisomal import pathway has been found to exhibit [57-59]. Tat leader peptides are on average 14 amino acids longer than Sec-dependent leader peptides and contain a distinctive amino acid sequence (S/T-R-R-X-F-L-K). Also the complex channel for transport substrates in Tat-pathway can be reached up to 70Å or more in diameter which is much larger than SecYEG translocon. However, proteins transport through Tat pore appears to be less efficient and slower [46, 60].

One of the interesting questions is why some bacteria proteins in a folded state using the Tat pathway when most proteins can be transported in an unfolded conformation by the Sec pathway. The reasons can fall into three major possibilities. One reason is metal ions could compete for binding sites in proteins. Use Tat-pathway can control the metal ions condition in the cytoplasm. Secondly, proteins can insert their complex cofactor before transport out of membrane. Lastly, this pathway allows hetero-oligomeric complexes to form in the cytoplasm and then be transported [57, 58].
2.5 Secretion Signal Sequences

Generally, the premature protein get synthesized in the cytoplasm before they found in the outer membrane or periplasmic spaces. These premature protein usually contain a 15-30 specific amino acid sequence which allows protein to export outside of cytoplasm. Signal sequence is necessary for the protein to get correctly targeted to the translocation pathway. The one used for efficient secretory productions of recombinant proteins in \textit{E.coli} are PelB, OmpA, PhoA, endoxylanase, and StII. Common features about these signal sequences are they contain a 10-20 amino acids hydrophobic H-domain, 2-10 amino acids positively charged N-domain and a hydrophilic carboxy-terminal region which has a cleavage site for the signal peptidase (Table) [40].

Studies show increase the positive charge in N-terminal region enhance translocation rates, this could be due to increase in the interaction of the pre-protein with SecA [61, 62]. And, translocation efficiency increase with the length and hydrophobicity of the H-domain [63, 64].

<table>
<thead>
<tr>
<th>Signal sequences</th>
<th>Amino acid sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>PelB (pectate lyase B) from \textit{Erwinia carotovora}</td>
<td>MKYLLPTAAAGLLLLAAQPAMA</td>
</tr>
<tr>
<td>OmpA (outer-membrane protein A)</td>
<td>MKKTAIAVALAGPATVAQA</td>
</tr>
<tr>
<td>StII (heat-stable enterotoxin 2)</td>
<td>MKKNIAFLASMFVFS1ATNAYA</td>
</tr>
<tr>
<td>Endoxylanase from \textit{Bacillus sp.}</td>
<td>MFKKKKKFLVGLTAAFSMISMSFASATASA</td>
</tr>
<tr>
<td>PhoA (alkaline phosphatase)</td>
<td>MKQSTIALALLPLLFTPVTKA</td>
</tr>
<tr>
<td>OmpF (outer-membrane protein F)</td>
<td>MMKRNILAVIPALLVAGTANA</td>
</tr>
<tr>
<td>PhoE (outer-membrane pore protein E)</td>
<td>MKKSTLVLVMGIVASHVQA</td>
</tr>
<tr>
<td>MalE (maltose-binding protein)</td>
<td>MKITGARILALSALS1TMFSASALA</td>
</tr>
<tr>
<td>OmpC (outer-membrane protein C)</td>
<td>MKVKLVLVPALLVAGAANA</td>
</tr>
<tr>
<td>Lpp (murein lipoprotein)</td>
<td>MKATKLVLGAVLGLSTLAG</td>
</tr>
<tr>
<td>LamB (\lambda receptor protein)</td>
<td>MMITLRKPLAVAAAGVMSAQAMA</td>
</tr>
<tr>
<td>OmpT (protease VII)</td>
<td>MRKLLGIVUTTPIAISSFA</td>
</tr>
<tr>
<td>LTB (heat-labile enterotoxin subunit B)</td>
<td>MNKVCYVLFTALLSSLYAHG</td>
</tr>
</tbody>
</table>
Figure 5 [45]: Table shows a list of secretion signal sequences used for produce recombinant protein in E.coli. The signal sequence composes N-, H- and C-domain. The N-domain represent in bold, C-domain are underlined.

3.1 Recombinant protein Secretion in Yeast system

Scientists have been using different microorganism systems to produce heterologous proteins. Bacterial expression is the most common, efficient, and widely used system. However some post-translational processing of eukaryotic proteins is not achievable by bacteria [65]. Unlike bacteria expression system, P.pastoris generates many of the posttranslational modifications typically associated with higher eukaryotes, which includes correct folding, disulfide bond formation, O- and N- linked glucosylation and processing of signal sequences [66, 67]. Many heterologous proteins have been successfully expressed in Pichia pastoris. These proteins can be expressed either intracellularly or extracellularly.

3.2 Pichia pastoris expression system background history

Koichi Ogata was the first scientist to describe methanol as having the ability to utilize a sole carbon source in certain yeast species approximately thirty years ago [68]. Soon after, it caught some attention as potential resource for generation of single-cell proteins to be used in animal feed since the methanol is inexpensive to synthesize. Phillips Petroleum Company was the first to develop media and protocols for growing Pichia pastoris on methanol in continuous culture at high cell densities (>130 g/L dry cell weight) [66]. Later on, Phillips Petroleum Company together with Salk Institute
Biotechnology/Industrial Associates, Inc. (SIBIA) developed *P. pastoris* as a heterologous gene expression system [66]. Researchers from SIBIA were able to isolate the gene and promoter for alcohol oxidase (AOX); develop the vectors, strains and methods for molecular genetic manipulation of *P. pastoris* [66, 68].

### 3.3 Common expression host Strains

Most *P. pastoris* host strains can be grouped as methanol utilized strains and Protease-deficient strains. They generally contain a mutation in the histidinol dehydrogenase gene (HIS4) for the purpose of the expression vectors containing HIS4 to be selected for transformation [66].

Methanol utilizing strains which are identified at the present belong to only four genera: Hansenula, Pichia, Candida, and Torukopsis and they share a specific pathway including several unique enzymes [67, 69]. GS115 (*his4*) is one of the most commonly used expression hosts. It grows on methanol at the wild-type rate (*Mut*⁺, methanol utilization plus phenotype) to utilize approximately 85% of methanol by the alcohol oxidase enzyme generated from alcohol oxidase 1 gene (AOX1) [68, 70]. KM71 (*his4, arg4 aox1:ARG4*) is another common used methanol utilized host strain. Since chromosomal AOX1 gene in KM71 strain has been deleted and replaced with the *S. cerevisiae ARG4* gene, an alternative weaker gene AOX2 for AOX is been used for methanol metabolism [66, 68]. Compare with AOX1, AOX2 consume carbon source methanol slower (*Mut*⁻, methanol utilization slow phenotype) and express at lower level rate [71, 72]. The third strain is MC100-3 (*his4 arg4 aox1Δ::SARG4 aox1Δ:: Phis4*) unable to grow on methanol (*Mut*⁻,
methanol utilization minus phenotype) [66, 68, 73]. Research has shown that majority strains have been commonly using are the Mut* phenotype [74-76]. Determine methanol utilization phenotype Mut*, Mut* or Mut is important in order to decide on the culture condition. Methanol utilization phenotype strains also been used to characterize expression level of heterologous protein in my experiments.

3.4 Choice of Promoter

Since the majority of heterologous protein produced in P.pastoris base require methanol present for the metabolism, AOX promoters have been the most widely utilized amongst others. AOX1 gene is a strong promoter which can drive expression of single integrated copy recombinant protein to high levels. One of the studies compared expression level of human granulocyte-macrophage colony-stimulating factor (hGM-CSF) by using AOX1 promoter with another promoter GAP (glyceraldehydes 3-phosphate dehydrogenase). The result showed the biomass level for protein expressed under AOX1 promoter is 180mg/l and 90mg/l for using GAP promoter[77, 78].

AOX2 gene also produces alcohol oxidase, but in general the expression levels of foreign proteins yields are 10-20times less than the AOX1 gene. One success reports using AOX2 promoter to increase the expression level is producing recombinant human serum albumin. The expression level was 2 times larger than the original one [67, 70].

Another key enzyme involved in methanol metabolic pathway is Formaldehyde dehydrogenase (FLD) from Pichia [79]. This gene is inducible by both methanol and
certain methylated amines such as methylamine as a sole nitrogen source. So the protein expressed by FLD1 promoter use glucose or glycerol as a carbon source [80].

3.5 Secretion signals of P. pastoris

P. pastoris can produce expressed foreign proteins extracellularly and similar like bacteria, secretion signals can be attached to the interested protein which target to the secretion pathway to be exported out of the cell. Foreign genes can be cloned into P. pastoris vector with either their native secretion signal S. cerevisiae α-factor prepro-peptide, the P. pastoris acid phospatase (PHO1) signal or the invertase signal sequence (SUC2) [67, 81, 82]. The S. cerevisiae α-factor is the most commonly used secretion signal, the sequence contains a 19 amino acid signal peptide and followed by a 60 amino acid pro-region. α-factor signal sequence has been successfully used for producing heterologous protein, in some cases, the protein expression level can be even better than using their own native leader sequence [83]. For instance, activin receptor has been successfully secreted extracellular by using α-factor. Study has been successfully use PHO1 signal peptide to express mouse 5-HT5A serotonin receptor and pepsinogen [84, 85]. The SUC2 signal peptide was able to secrete human interferon hybrid proteins. Another study also showed that secreted levels of α- amylase by SUC2 signal peptide was 20% more efficient compare with secrete by its native signal peptide [86, 87]. Proteins that are highly disulphide-bonded are normally produced by secretion system in P. pastoris, such as coagulation protease (Factor XII), a gragment of thrombomodulin. Glycosylaiton is one the most common post-translational modifications performed in P. pastoris. Correct
glycosylation patterns ensure necessary biological activities for many native proteins in mammalian cells [88, 89].
Chapter 1: Development of Secretion System for Recombinant Protein Expression in E.coli Bacteria

Abstract
This first goal of work in this chapter is to establish the bacteria secretion system as an alternative way to improve expression level of SUMO-related recombinant protein secretion in E.coli. The second goal is to identify the secretion pathway which has the best secretion performance as measured by fluorescence intensity. We used Type II bacteria secretion system to examine on the protein Ubc9 secretion levels. Three pathways involved in Type II bacteria secretion system: Sec, Tat and SRP pathways. PelB, DsbA and Tat signal peptide were chosen and used to test protein secretion through Sec, Tat and SRP pathway respectively. Protein secretion levels were detected by the fluorescence spectrum measurement of YPet fluorescence protein. Fluorescence intensities were measured from three samples: cell culture medium, periplasmic space and cell lysate.

Introduction
Posttranslational modification by ubiquitin (Ub) or ubiquitin-link proteins (Ubls), such as SUMO, plays a crucial role in many cellular activities, such as regulation of cell cycle, transcription, apoptosis, DNA damage response and signal transduction. SUMOylation pathway is an important mechanism in modifying protein function and regulates protein activities in eukaryotes. Deregulation of SUMOylation pathway is found in many diseases, for instance, neurodegenerative disease, human cancers, heart disease and
diabetes. Similar to the ubiquitination pathway, protein modification by SUMO is reversible and is a highly dynamic process, which requires enzymatic catalysis for efficient modification. Three enzymes involved in the sequential actions of SUMOylation cascade are E1 activating enzyme (AOS1/Uba2), E2 conjugating enzyme (Ubc9) and E3 ligase. E1 (AOS1/Uba2) is required for conjugating SUMO to substrate. E2 plays a critical role in SUMOylation cascade by directly involved in the selection of SUMO targets. These selected targets carry a consensus site which Ubc9 can bind directly [8, 9]. The third enzyme is SUMO ligases (E3) that helps to mediate and stabilize the interaction between the substrate and E2. Unlike E2 enzyme (Ubc9) which is the only enzyme identified in most species, there are three distinct types of E3 enzymes have been discovered. One of the E3 ligase types have extensively studied is PIAS. It is a protein inhibitor of activated STAT (PIAS1) functions by blocking the DNA-binding activity of nuclear factor kB (NF-kB) and STAT1 on gene promoters, targeting PIAS1 SUMOylation may give us a better understanding to treat inflammatory disorders. However, the expression of PIAS1 with the bacterial cells does not work well due to the toxicity PIAS1 can create within the cells. Currently, the Gram-negative bacteria E.coli is the most widely used host organism for secretory production of recombinant proteins in bacteria. Among various other hosts, this system has been well characterized and used for 20 years to promote protein folding, to manage stress-induced protein misfolding, and to support protein translocation. It has the ability to accumulate many recombinant protein products, up to 50% of the total cell protein [89]. In comparison to traditional expression of recombinant proteins from whole
cell lysates, secretory and extracellular expression of recombinant protein has some advantages, such as simplified downstream process and proper folding of disulfide bond containing proteins such as peptide hormones or antibody fragments [90, 91]. Amount five different types of secretion systems in Gram-negative bacteria, type II secretion system is the most common used one for secreted recombinant protein production in *E.coli* K-12 strain [45]. Type II secretion system includes Sec-dependent pathway, the signal recognition particles (SRP), and the twin-arginine translocation (TAT) pathway. There are two consecutive steps required to transport proteins in this system: periplasmic translocation and extracellular transport [39].

The Sec pathway use a cytosolic chaperone SecB to bind with preprotein to help it unfold [92]. Preprotein is equipped with a signal sequence at N-terminal that guides them to the cytoplasmic membrane which can recognized by where SecA. SecYEG is a protein-conducting channel located in the membrane. Protein guided through the channel either by post-translational Sec-B targeting pathway or the co-translational SRP-targeting pathway.

In the other case of post-translational Tat-pathway, the signal sequence leads the preprotein to the Tat-complex pore instead of SecYEG. Preprotein is then translocated out into periplasmic space and the signal sequence is cleaved off by leader peptidase [93, 94].

The purposes of this project is to test on the expression level of SUMO-related proteins (Ubc9) by using different type II bacteria secretion system and determine the best performance for production of soluble YPet-Ubc9 protein in *E.coli*. Fluorescence protein
was used to tag the proteins in these experiments to be assist in quantifying the expression level by measuring the fluorescence intensity. We used Ubc9 protein as a testing sample since this protein already been determined has high yield expression level. Use it can evaluate whether the pathways has feasibility for SUMO-rated proteins. The fluorescence protein tag used in the experiment is YPet. It is one of the engineered FRET pair by mutagenesis from yellow fluorescent protein (YFP). The fluorescent efficiency of YPet is higher and more stable than the transitional YFP [37]. The fluorescent intensity of secretory protein expressed from relative Sec, SRP and Tat pathway were measured and compared to determine the suitable secretion pathway for SUMO-related proteins. Establishing the bacteria secretion system in this study creates another opportunity to examine and possibly to improve the expression level of difficult proteins.

**Material and Methods**

Two experimental tests were examined: (1) expression of 10 colonies of each pET28(b)-Sec-YUbc9, pET28(b)-Tat-YUbc9 and peT28(b)-SRP-YUbc9 into BL21 (DE3) E.coli cells. 80µl of each culture medium, periplasmic extract and cell lysate samples were tested on their YPet fluorescence intensity. (2) pET28(b)-Sec-YUbc9, pET28(b)-Tat-YUbc9 and peT28(b)-SRP-YUbc9 cells cultured in 500ml medium individually. pET28(b), BL21(DE3) cells cultured as controls. YPet-Ubc9 protein intracellular expressed counted as a positive control. 80µl of each culture medium, periplasmic extract and cell lysate samples were tested on their YPet fluorescence intensity..
Molecular cloning of DNA constructs

The fragment encoding PelB leader sequence in pET22(b) vector was extracted by Bgl II and NcoI digestion sites. Then, it was inserted into a linearized pET28(b) plasmid with the same BglII/NcoI restriction sites. This pET28(b) plasmid already has YPet-Ubc9 fragment been cloned in and together with 6x His on N-terminus. DNA sequencing were used to confirm the insertion of pelB leader sequences into pET28(b) (Fig.6).

The engineered plasmid pET28(b) contains PelB leader sequence was then used as a template. Arabinose promoter was designed into the upstream primer to replace with the T7 promoter which originally was used in pET28(b) vector. The open reading frame of arabinose promoter together with PelB leader sequence was amplified by performing Polymerase Chain Reaction (PCR). This PCR product was then cloned into PCR2.1-TOPO vector (Invitrogen). Once the DNA sequences were confirmed, the cDNA encoding the fragment of arabinose promoter and PelB leader sequence was extracted by Bgl II/BamH I and cloned at the same restriction sites in the pET28(b) vector with an engineered 6x Histidine-tag region on N-terminus (Novagen). YPet-Ubc9 and YPet-PIAS1 inserts were digested and extracted by Bgl II/BamH I restriction sites from PCR2.1 products. These YPet-Ubc9 and YPet-PIAS1 were subsequently ligated to the engineered pET28(b)-Arabinose-pelB fragment by T4 DNA ligase. The ligation mixture were used to transformed into E.coli TOP10 cell line and plated on Lysogeny broth (LB) agar plates containing kanamycin antibiotic (Fig.7)
In order to have a critical comparison of the heterologous protein expression levels between Sec-pathway, SRP-pathway and TAT-pathway, a cold-inducible promoter and phoA (alkaline phosphatase gene) inducible promoter were also engineered separately into pET28(b) vector obtained by similar strategy along with their secretion signal.
Figure 6: Schematic diagram to engineered plasmid that contains bacteria Sec-pathway secretion. The upper diagram indicates a close form plasmid map contains interested gene Ubc9 tagged with fluorescence protein YPet which digested at NheI/NotI for insertion. PelB secretion signal peptides cloned at BglII/NcoI into vector pET28(b) as well. The diagram below show part of the pET28(b) vector DNA sequence used for various cloning. The original vector’s size is 5369bp from Novagen®, and the PelB-YPet-Ubc9 is 1264bp.

Figure 7: Schematic used as an example of engineered plasmid contain signal sequences, PelB, gene of interest Ubc9 tagged with YPet and Arabinose promoter.

pET-28b-Arabinose-pelB cloning/expression region
Original T7 promoter contained in pET28(b) from Novagen® were replaced by Arabinose promoter.

**Expression and Purification of Secretory production of Recombinant Protein**

The pET28(b), pET28(b)-Sec, pET28(b)-Tat and pET28(b)-SRP plasmids encode with the YPet-Ubc9 gene of interest was respectively transformed into BL21 (DE3) *Escherchia coli* (*E.coli*) cell operated by electroporation techniques. Empty vector pET28(b) was also transformed into BL21 cell lines as a negative control. The transformed bacteria were plated on Lysogeny broth (LB) agar plates containing Kanamycin. Single colony was picked from the plate and inoculated into 5ml of 50µg/ml kanamycin-containing LB medium and vigorously shake at 37°C overnight.

The next day, the seed culture (1%, v/v) was transferred into a 500ml shake-flask contains 2×YT medium at 37°C. The culture was shook vigorously at 37°C until the optical density (*OD*$_{600}$) reached 0.4-0.5. 0.1 mM isopropyl β-D-galactopyranoside (IPTG) was added to the initiate protein expression. The culture was grew overnight and shook at 25°C for 16 hours.

After fermentation, 500ml culture medium was collected and cells were separated from the supernatant by centrifugation at 8000rpm for 10 min of 4°C. Three different fractions from this sample were collected: supernatant, periplasmic proteins, and cell lysates. 1ml of supernatant was collected for fluorescence spectrum test. Protein pellet was resuspended in 15ml Milli-Q H$_2$O for osmotic shock to release the remaining proteins in the periplasmic region. The addition of water creates a hypotonic condition because the
concentration of salt in cytoplasm is higher than the outer membrane hence large amount of water diffuse into the cells. This hypotonic condition increases permeability of the cell membrane causes the cells to swell up. The remaining proteins will then be burst out to periplasmic space. The resuspended cells were incubated on ice for 40 min and spun down at 8000rpm for 15 min of 4°C. 1ml of the water was collected. The cells was resuspended in buffer containing 20mM Tris-HCl (pH 7.4), 0.5M NaCl and 5mM imidazole and sonicated with an ultrasonic liquid processor (Misonix). Cell lysate containing recombinant proteins was cleared by centrifugation at 25000 g for 30 min. 1ml sample was also saved.

Paralleled examination of the YPet-Ubc9 expression was determined to further quantify and compare of secretory protein expression level in different bacterial secretion pathway. Ten colonies of each secretion pathway were picked up and inoculated in a test tube with 2mL LB that contained kanamycin antibiotic for plasmid maintenance. The culture tubes were inoculated in a 37°C for 16 hours. 1mL of the culture consists of the recombinant proteins expressed was added into a fresh 3mL LB media and inoculated at the same temperature for 3 hours. 2µL of 1mM IPTG was added and vigorous shaking at 25°C for 16 hours before being harvested for analysis.

3 ml cells culture media were separated from the supernatant by centrifugation at 13.2rpm for 10 min. Protein pellet was saved and resuspended in 240µl Milli-Q H$_2$O for osmotic shock as described above. Eighty µl water was collected for fluorescence activity test. The cells pellets were then resuspended one more time with 240µl 1x Phosphate
buffered saline (PBS) buffer and 80µl of it was collected to test on intracellular expression level.

**Fluorescence Measurement**

The solutions containing recombinant YPet-Ubc9 proteins were collected at three different steps during purification which were used to measure fluorescence to check the identify presences of the protein. pET28(b)-YPet-Ubc9 intracellularly expression was used as a positive control, empty vector pET28(b) and BL21 *E.coli* cells were used as a negative controls. 80µl of each sample were added in each well in a 384-well plate. The fluorescence emission spectrum of each well was determined by fluorescence multi-well plate reader Flexstation 384 (Molecular Devices, Sunnyvale, CA). YPet fluorophore were detected at the excitation wavelength of 475nm and the maximum emission intensity at 530nm with a cutoff filter of 495nm. This experiment was repeated three times. The same strategy was used in detecting fluorescence activity of single colony, the same strategy was used.

**Results**

**Production of YPet-Ubc9 protein into periplasmic space in Sec, SRP and Tat pathways**

YPet-Ubc9 encoded with specific signal peptide are able to be secreted into peripalsmic region respectively to the Sec-pathway, Tat-pathway and SRP-pathway. Production of secretory protein by Sec-pathway has fluorescence intensity around $8 \times 10^4$ RFU. This fluorescence intensity is about 2.5 fold higher compared to the periplasmic extract protein
from non-induced cells (Figure10A). In contrast to the Sec-pathway, the expression level of secretory protein through Tat-pathway varies between $6 \times 10^5$ to $1.2 \times 10^6$ RFU. One of the possible explanations is protein expression level is high, protein naturally secreted out without using secretion pathway (Figure 8B). The average fluorescence intensity for protein secreted by SRP-pathway is approximately two fold higher than non-induced samples represented by the higher relative fluorescence intensity at the 530nm emission (Figure 8C).

Further analyses on secretory protein in periplasmic space in different secretion pathways were tested on cells cultured in 500ml 2xYT medium. Both periplasmic extract of BL21 and BL21-pET28(b) negative control display no fluorescence (Figure 9 purple and blue lines). Blue fluorescence spectrum peaks appeared for Sec, SRP and Tat pathway (red,
gray and yellow lines). In comparison to cytoplasmic protein expression (green line),
secretory protein intensity is about three times higher. It represents proteins that were
successfully secreted into periplasmic space through the secretion systems. Product of
protein secretion by Sec-pathway indicates the highest fluorescence intensity, two times
higher RFU than the other two pathways.

![Fluorescence Activity in Periplasmic Extract](image)

Figure 9: Fluorescence intensity of YPet-Ubc9 protein secreted into periplasmic space by
Sec, Tat, and SRP pathway in E.coli. Non-trnasformed BL21 bacteria and BL21 empty
vector pET28(b) transformed BL21 served as negative controls. The way protein
extracted from periplasmic space is by adding 15ml of water to the cells. Osmotic shock
burst protein out to water and 80µl of waster fraction was detected for the YPet
fluorescence intensity.

**Production of YPet-Ubc9 protein intracellular production use Sec, Tat and SRP pathways**

Fluorescence intensity of YPet-Ubc9 protein measured in inducted culture (gray bars) is
much higher than non-induced culture (orange bars) figure 10A-C. This result indicates
the remaining protein in intracellular. The expression level of colony #5 in Tat-pathway,
fluorescence spectrums are very low for both aqueous solution and cells solution. This could due to various reasons, such as the colony was self ligased which means the gene only contain an empty vector; other possibilities are culture medium was contaminated during the fermentation handling or IPTG was not added which cause the cells un-appropriate grow.

Figure 10: fluorescence intensity difference between non-induced and induced intracellular protein expression. Orange bars indicate non-induced culture and grey bars represent induced culture. (A) Ten colonies of recombinant protein Sec-YPet-Ubc9 cultured in medium individually. (B) Tat-YPet-Ubc9 colonies were cultured at the same condition. (C) SRP -YPet-Ubc9 colonies were cultured as well.

Quantity of the proteins inside cytosol was examined by measuring the fluorescent intensity from cell lysates from 500ml cell culture. The result showed a large proportion of protein remains inside cytosol for Tat-pathway (Figure 12-yellow line). The other two secretion pathways, Sec and SRP pathways, show no or very low fluorescence activity, which indicate that most of the YPe-Ubc9 proteins were no longer remain in the cytoplasm. They may be either secreted to periplasmic space or extracellular space (Figure 12-red and gray lines).
Figure 11: YPet fluorescence measurement at maximum excitation 475nm wavelength. Protein in cytoplasmic released out of cell by broke the cell wall.

**Secretion of YPet-Ubc9 protein into the culture medium in Sec, SRP and Tat pathways**

YPet fluorescence spectrums of each 500ml cell culture medium were determined. The culture medium 2xYT fluorescence intensity was measured and subtracted as a background. BL21 cell and empty pET28(b) vector play a role as negative controls (Figure12-purple and blue lines). Protein YPet-Ubc9 shows similar fluorescence intensity as the negative control (blue and green lines). This illustrates cytoplasmic proteins were not being secreted out to the medium. Expression levels of secretory protein into the culture medium are higher than intracellular expression level (yellow, red and gray lines). Protein secretion level of Tat-pathway (yellow line) is proximately 20% higher compare with the other two secretion pathways.
Figure 12: Fluorescence intensity of YPet-Ubc9 protein secreted into culture media after background subtraction. Culture medium was collected after protein expressed overnight. Protein and supernatant were separated by centrifuge. 80µl of supernatant was detected for the YPet fluorescence intensity.

**YPet-Ubc9 protein Fluorescence intensity yield distribution in Sec, SRP and Tat pathway**

Fluorescence yield distribution calculated by divide the fluorescence intensity measure from one place to the total fluorescence observed in total three places by the same pathway. In graph A, Secretory protein in the culture medium through Sec and SRP pathways has higher yield than Tat-pathway. Two of them has similar yield of 0.2.

Protein secreted in periplasmic space shows the same that Sec and SRP pathways contain high protein expression protein either secretion through Tat-pathway or they intracellular expressed. Data indicate Sec-pathway is more efficient than SRP pathway. Most 100 percent of protein remained inside cytoplasmic by using Tat-pathway and same phenomena happened to intracellular expression, indicate in graph C.
Figure 13: Fluorescence intensity yield of YPet-Ubc9 protein expression distribution. Proteins are expression through four different ways: intercellular, Sec-pathway, Tat-pathway and SRP-pathway (A) Yield distribution of protein expressed into culture medium. (B) Yield distribution of protein expressed into periplasmic space. (C) Yield distribution of protein remains in cytoplasmic space.

**Discussion**

In this study, express recombinant protein had been tested use Sec, Tat and SRP pathways secretion system in *E.coli*. Protein expression through Tat-pathway shows large proportion of protein remain in cytoplasmic space (Figure 10B and Figure 11 green line) for both 10 colonies expression tests and 500 culture expression level. Production of secretory protein by Sec- and SRP pathways has similar fluorescence intensity average in $8 \times 10^4$ RFU (figure 8 A and C). However, secretory protein through Sec-pathway is 1 time folded higher than secretory protein through SRP-pathway in 500ml culture (Figure 9 red and gray lines). Protein expression in larger volume is more accurate compare with small volume culture, since small volume culture has more variations. Furthermore, 60% proteins expressed through Sec-pathway have either
secreted into culture medium or periplasmic space, similar, 50% protein being secreted through SRP-pathway as well (Figure 13). Since Sec and SRP pathway secretion process are very similar, it is reasonable to see similar pattern of protein secretion. Overall, protein secreted through Sec-pathway shows as a better system suit for YPet-Ubc9 protein secretion.

There are two main reasons to produce heterologous protein in the periplasm rather than in the cytoplasm of E.coli. First, the isolation of protein from the periplasmic is usually easier than the isolation of proteins from total cell lysate. Secondly, the disulfide bonds for lot of eukaryotic protein can be formed appropriately in cytoplasmic space. Unfortunately the productions of recombinant secretory proteins in E.coli have some negative effect, such as low yields of production in the periplasmic space. Studies have showed secretion efficiency depends on the protein characteristic, incomplete processing of signal sequences. To improve periplasmic yields of recombinant proteins, some strategies can be used. For instance, change the promoter, selection of different signal polypeptides or manipulate culture condition [65].
Chapter 2: Recombinant Protein Secretion in Pichia Pastoris Yeast system

Abstract
As mentioned earlier, PIAS1 function as a protein inhibitor. It involved in many physiological processes directly or indirectly. The expression of PIAS1 in bacterial cells results in a non-functional protein which can not be used to identify SUMO proteases. In the present study, we use yeast secretion system to test on SUMO-related protein expression level. Two of the protein been examined are YPet-Ubc9 and YPet-PIAS. The control YPet-Ubc9 was successfully expressed, but the expression of PIAS1 left with inconclusive results. More experiments are required to determine whether PIAS1 with the yeast secretion pathway is viable.

Introduction
Even though E.coli has been extensively used as a protein expression, it still facing to some problems, such as proteins derived from eukaryotic genomes cannot be appropriately expressing in E.coli. Because it is a simple microorganism and may not be able to correctly fold the foreign protein and protein product typically obtained as insoluble, miss-folded inclusion bodies. Therefore E.coli generally is not a suitable express protein require post-translational modification [70]. Since yeast is a eukaryote, it is a suitable host for produce eukaryotic heterologous protein and SUMO-related proteins are eukaryote. Yeast host could be a expression possibility since host organism combines the molecular genetic manipulation and growth
characteristic of prokaryotic to perform post-translational protein modification of eukaryotes [72].

Research have been started using *Sachharmyces cerevisiae* yeast as an expression system as early as 1981 and improvements of using methylotrophic yeast *Pichia pastoris* for foreign gene expression was developed later on. *P. pastoris* is type of yeast grows on methanol as a sole carbon and energy source and it possesses a highly inducible methanol utilization pathway. It became one of the mainstream expression host used alongside *Escherichia coli, Sachharmyces cerevisiae and baculovirus* [95]. Methanol-regulated promoter from *P. pastoris* such as alcohol oxidase I gene (AOX1) have been shown to be highly expressed as well as tightly regulated [96, 97]. AOX1 was identified as a strong promoter which can drive expression of single copy recombinant protein to high levels, as high as 30% of the total soluble protein. This promoter has been used to construct expression vectors for variety heterologous protein production in *P. pastoris* [96].

Expression of foreign gene in *P. pastoris* normally requires three basic steps: (1) insertion of gene into an expression vector; (2) expression vector insert into *P. pastoris* genome (3) examination of potential expression strains contain foreign gene product [77]. Recombinant protein can be produced either intercellularly or extracellularly in *P. pastoris*. There are many commercially available vectors and host strains which can be used for foreign proteins expression in *P. pastoris*.

In this chapter, yeast secretion system is used to test on YPet-Ubc9 and YPet-PIAS1 proteins expression level. Plasmids were contracted for the yeast cell line and allows the protein to be secreted so they can avoid the toxicity of PIAS1 creates. prepro α-factor
signal sequence was utilized as the secretion signal. Protein expression level was observed by measure the fluorescence of YPet fluorophore at maximum excitation 530nm.

**Material and Methods**

**DNA constructs cloning**

To construct the plasmid which can express recombinant protein Ubc9 tagged with fluorescence tag-YPet, an open reading frame of YPet-Ubc9 were used as a template and amplified by PCR. *EcoR I* and *Not I* restriction site was designed into forward primer and reverse primer respectively. The PCR products were cloned into PCR2.1-TOPO vector (Invitrogen). After the DNA sequence were checked, this cDNA encoded YPet-Ubc9 was cloned into the pET28(b) vector at *EcoR I* / *Not I* restriction sites. YPet-Ubc9 insert in pET28(b) was digested and extracted by *EcoR I* / *Not I* sites. Common used secretion vector from P.pastoris (Invitrogen) pPIC9 was been digested and extracted by the same enzymes. YPet-Ubc9 insert was ligased with pPIC9 vector and transferred into E.coli Top10 cells and plated on LB agar plates contain Ampicillin resistant (Figure 2-a).

To transfer the recombinant plasmid into the yeast genome requires ~15µg of the DNA, pPIC9-YPet-Ubc9 was expressed in100ml of LB broth over night. It was purified and digested by *Bgl II* to make it linearized. pPIC9–Yubc9 was transformed into a GS115/*His*⁺/*Mut*⁺ cell line, plated on minimal dextrose medium plates (MD plates) and incubated at 30°C until colonies appear.
Figure 14: Schematic shows yeast secretion vector pPIC9 cloned with gene of interest Ubc9 tagged with at EcoRI/NotI sites. It is Ampicillin resistance vector and uses AOX1 promoter for protein secretion.

**Expression and Purification of YPet-Ubc9**

The protocol used for the expression of our protein through the yeast secretion pathway was obtained through Invitrogen®. 10 single colonies were picked from the plate and inoculated individually in 3mL Buffered Glycerol-complex Medium (BMGY) for fermentation. Culture tubes shook for 16 hours for $OD_{600}$ reaches 2-6 at 30°C incubator. The cells were harvested by centrifuging at 1320rpm for 5 minutes at room temperature. Supernatant were decanted and cell pellets were resuspended once in 3mL BMM medium for wash. The process of pellated down the cells repeated one more time and cells were resuspended in 3mL buffered Methanol-complex Medium (BMMY) to induce expression.

100% filtered methanol was added to final concentration of 0.5% methanol accordingly to the 3mL culture medium every 24 hours. 200μL of culture samples were collected of
each 12-hour inoculation period for analyze expression levels and to determine the optimal time post-induction to harvest. The time points of samples been collected after during post-induction: 24hours, 48hours, 72hours and 96hours. Supernatant and cells were separated by centrifuging and supernatant were saved.

**Measurement of YPet Fluorescence**

The supernatant containing recombinant YPet-Ubc9 proteins were collected at 4 different days were used to measure fluorescence to check the presences of secretory protein. 80µl of each sample were injected in each well in a 384-well plate. The fluorescence emission spectrum of each well was determined by fluorescence multi-well plate reader Flexstation \(^{384}\) (Molecular Devices, Sunnyvale, CA). Same method was used to detect YPet fluorophore intensity. By excite at 475nm wavelength and the maximum emission intensity at 530nm with a cutoff filter of 495nm. Relative fluorescence units were compared and shows in the result section.

**Result:**

**YPet-Ubc9 protein secretion in Pichia Postoris yeast system**

Average fluorescence intensity of cells in supernatant is \(6\times10^5\) after 24 hours culture (Figure 15 A). Some of them grew rapidly started from their second day but some have not much changes (Figure 15B). Compare with previous day, average intensity of secretory proteins appropriately grew increased 1 times higher. Protein continuously secreted out after 72 hours culturing, most production of secretory protein increased,
indicated in table C. Some of fluorescence intensity of the cells shows a decrease pattern after grew for 96 hours (Figure 15C). Too much protein may start accumulation cause cell death.

Figure 15: fluorescence spectrum of YPet-Ubc9 protein secretion in *Pichia pastoris* yeast. (A) fluorescence intensity of protein expressed for 24 hours. (B) fluorescence intensity of protein expressed for 48 hours. (C) fluorescence intensity of protein expressed for 72 hours. (D) fluorescence intensity of protein expressed for 96 hours.
**YPet-PIAS protein secretion in Pichia Postoris yeast system**

After four days of expression, fluorescent intensity of YPet-PIAS1 for 20 colonies expressions are determined. The relative fluorescence intensity for YPet is not very distinguishable over the four days expression. Nine out of twenty colonies RFU value reached to $4 \times 10^5$ in the after 24 hours. Ten out of twenty colonies reached to $4 \times 10^5$ RFU value after 48 hours growing. After 72 hours, seven out twenty colonies stay around $4 \times 10^5$ FRU and only four colonies remained after 96 hours.

![Fluorescence spectrum of YPet-PIAS1 protein secretion in Pichia pastoris yeast.](image)

Figure 16: fluorescence spectrum of YPet-PIAS1 protein secretion in *Pichia pastoris* yeast. (A) Fluorescence intensity of protein expressed for 24 hours. (B) Fluorescence
intensity of protein expressed for 48 hours. (C) Fluorescence intensity of protein expressed for 72 hours. (D) Fluorescence intensity of protein expressed for 96 hours.

**Discussion:**

The protein of interest Ubc9 tagged with YPet fluorescence was successfully expressed within the *Pichia pastoris* cell line. Most protein did not start secrete out until proteins have expressed for 24 hours. 13 out of 20 colonies grew continuously until 72 hours after induction. 5 out of the 13 colonies, fluorescence intensity decrease after expressed for 72 hours. This could be due to protein degradation. Average fluorescence intensity of protein expressed after 48 hours is 2 times higher than cultured for 24 hours.

Recombinant protein YPet-Ubc9 secreted in *Pichia pastoris* has the best expression level on 48 hours. Also, not all the colonies were able to express protein appropriately.

Scanning for the right colonies is necessary before fermenting in large volume.

YPet-PIAS1 protein secreting in Pichia pastors however is not as successful as YPet-Ubc9 protein. There is not much different between the cells cultured for four days. 7 out of 20 colonies grew after 48 hours expression and 3 colonies continuously grew after 72 hours. Compare with YPet-Ubc9, YPet-PIAS1 proteins degrade rapidly. Several reasons could cost protein degradation such as formaldehyde formation during the methanol metabolism which break some hydrogen bonding between amino acids, lead denature protein. Culture condition may be another reason, since protein excreted out the cells synthesized at relatively high temperature, higher temperature can lead to deterioration of bonds between amino acid.


