Identification of Factors Limiting Heterologous Lipase Expression in the Cytoplasm and the Periplasm as well as Display on Cell Surface of *Escherichia coli*

by

Yali Xu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 2008

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Lipase B from *Pseudozyma antarctica* (PalB), had been expressed in several recombinant protein hosts and showed very good transesterification activity for biodiesel production. However, the functional expression could not be demonstrated until recently in the most popular recombinant protein expression system, e.g. *Escherichia coli*, and the expression performance stands improvement. The probable reason is that PalB is a lipase with more hydrophobic surface and three disulfide bonds and thus may not be easily expressed in *E. coli* functionally. This thesis focuses on the identification of factors limiting heterologous expression of PalB in *E. coli* through a systematic study by using several strategies, including the different expression compartments, fusion tags, folding factors, and host strains.

Functional expression of PalB in the cytoplasm of *E. coli* was explored using BL21(DE3) and its mutant derivative Origami B(DE3) as the host. Bioactive PalB was obtained in the reduced cytoplasm of BL21(DE3), implying that the formation of disulfide bond was not strictly required for functional expression. However, the expression was ineffective and was primarily limited by formation of PalB inclusion bodies and growth arrest, both of which were associated with PalB misfolding and deteriorated physiology. The culture performance in terms of cell growth and PalB expression level could be significantly improved by simultaneous coexpression of multiple chaperones of trigger factor and GroEL/ES, but not individual coexpression of either one of them. It was proposed that the two chaperones mediate the early stage and late stage of cytoplasmic PalB folding and would be required simultaneously for boosting both the overall PalB synthesis rate and the cytoplasmic folding efficiency. On the other hand, a much higher bioactive PalB was produced in Origami B(DE3) harboring the same PalB expression vector. Furthermore, the significant high bioactive PalB was produced by coexpression of periplasmic folding factor without a signal peptide (e.g., coexpression of DsbA, and DsbC). Coexpression of DsbA was found to be effective in enhancing PalB expression and such an improvement was more pronounced in Origami B(DE3), suggesting that both folding and disulfide bond formation could be the major factors limiting PalB expression. The fusion tag technique was also explored by constructing several PalB fusions for the evaluation of their expression performance. While the solubility was enhanced for most PalB fusions, only the DsbA tag was effective in boosting PalB activity possibly via both enhanced solubility and correct disulfide bond formation. Our results suggest that solubilization of PalB fusions did not necessarily result in the development of PalB activity which could be closely associated with correct disulfide bond formation.

While PalB was stably expressed in the cytoplasm, most of the expressed gene product aggregated in cells as inactive inclusion bodies. In contrast, PalB was extremely unstable when expressed in the periplasm, also leading to poor expression performance. Such unstable PalB can be rescued by coexpression of several periplasmic folding factors, such as DegP, FkpA, DsbA, and DsbC, but not cytoplasmic chaperones. As a result, the performance for functional PalB expression in the periplasm was significantly improved. This is the first report demonstrating the use of folding factors to rescue the extremely unstable gene product that is otherwise completely degradable. On the other hand, functional expression of PalB in the periplasm was explored using four fusion tags, e.g., DsbC, DsbA, maltose binding protein (MBP), and FLAG in the sequence of increasing expression efficacy. Amongst these fusion tags for functional expression of PalB, FLAG and MBP appear to be the most effective ones in terms of boosting enzyme activity and enhancing solubility of gene products, respectively. Overexpression of these PalB fusions often resulted in concomitant formation of insoluble inclusion bodies. Coexpression of a selection of periplasmic folding factors, including DegP (and its mutant variant of $DegP_{S210A}$), FkpA, DsbA, DsbC, and a cocktail of SurA, FkpA, DsbA, and DsbC, could improve the expression performance. Coexpression of DsbA appeared to be the most effective in reducing the formation of inclusion bodies for the four PalB fusions, implying that functional expression of PalB could be limited by initial bridging of disulfide bonds. Culture performance for functional expression of PalB was optimized by overexpressing FLAG-PalB with DsbA coexpression, resulting in a high volumetric PalB activity of 360 U/liter.

Without extracting protein from cells the whole cell can be directly used as a platform for the immobilized enzyme. Proof-of-concept experimentation was conducted by PalB display on the *E. coli* cell surface. By fusing the *palB* gene in between the signal peptide *phoA* and an autotransporter Protein EstA's gene under the *lac* promoter, PalB was successfully displayed on the *E. coli* cell surface. However, cells encountered a severe physiological stress. Coexpression of various periplasmic folding factors, e.g., DegP, SurA, DsbA and DsbC could erease the physiological stress, but only DsbA was demonstrated to be effective to restore cell physiology and increase PalB expression level.

Key words: enzyme, *Escherichia coli*, chaperone, folding factor, fusion tag, gene expression, *Pseudozyma antarctica*, lipase B, recombinant protein production

Acknowledgements

First and foremost I would like to express my deepest gratitude and appreciation to my supervisors, Professor C. Perry Chou, Professor Murray Moo-Young, and Professor Jeno M. Scharer for their unlimited support, critical guidance, and heartfelt encouragement. Without their help, I could not go through the difficulties in this innovative research at the University of Waterloo. I consider myself fortunate to have had the chance to learn from such excellent mentors. Their scientific vision, keen attitude, and charisma toward academic work will inspire me in my life now and in the future.

I would also like to thank Professor William A. Anderson, Professor Raymond Legge, and Brendan J. McConKey for serving as my committee members and providing me with helpful suggestions for my research and constructive comments for this thesis.

I would like to present my appreciations to external examiner Professor Yen-Han Lin for his valuable comments.

I would like to thank Europe Molecular Biology Laboratory (EMBL), Didier Busso, Arne Skerra, Deb K. Chatterjee, David S. Waugh and T. Yura for kindly providing numerous plasmids used in this study.

I would like to extend my special thanks to Ms. Jana Otruba, technician of the biotechnology laboratory, Mrs. Patricia Anderson, graduate administrative assistant in the Department of Chemical Engineering for clerical helps, as well as, Mr. Bert Habischer, Mr. Ralph Dickhout, Mr. Dennis Herman, Mr. Ron Neill, Mr. Ravindra Singh, and Mr. Rick Hecktus for their various technical support throughout the duration of my research.

I would like to extend my sincere gratitude to our group members, Niju Narayanan, Reza Gheshlaghi, Keyvan Nowruzi, Victoria Zhang, and co-op students: Amrita Yasin, Raymond Tang, Thomas Wucherpfenning, Darrell Lewis, and Stefan Rosenkranz, for their friendship, and a invaluable assistance in the research experiments.

I also would like to thank Dr. Tzannwei Wang for his advice in solving the problems which I had met in molecular biological techniques and Mr. D. Moyle for his assistance in establishing the TEM protocol.

I appreciate the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chair (CRC) program for their financial supports.

Finally, I would like to thank my husband, my parents, and my son, who have supported me all the time both emotionally and academically. Thank you so much for your constant love and support.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Overview

The depletion and recent skyrocketing prices of fossil fuels make the search for alternative and renewable energy sources imperative [1]. Bioethanol has been rapidly advancing worldwide in the last decade although the process is still open for improvement[*]. Biodiesel is the other biofuel and is on the road to commercial use. Biodiesel is a mixture of fatty esters made by transesterification of animal fats or plant oils with methanol or ethanol. As a renewable fuel, biodiesel is considered to be environment-friendly since it reduces exhaust emissions upon burning, particularly carbon monoxide, black smoke, smog-causing particulate matter, odor, harmful carcinogens, and even sulphur dioxide [2, 3]. Due to the characteristics of the biodiesel, it can be directly used in diesel engines as a 6~20% mixture with conventional diesel fuel[†].

Currently, most biodiesel production via transesterification reaction are catalyzed by alkaline or acidic catalysts or with supercritical methanol at high temperatures and pressures [2, 4, 5]. However, chemical processes generally are high energy-consuming, require harsh reaction conditions and suffer various separation issues caused by side reactions [2, 4, 6-9]. With several attractive features, such as mild reaction conditions and high reaction specificity, biocatalysts are gaining more and more attention nowadays and have the potential to outperform chemical catalysts in the future [10, 11]. There are three major advantages for biocatalyst: (1) mild reaction conditions. Usually the reaction is catalyzed at room temperature; (2) high specificity. For enzymatic reaction, the enzyme works with substrate in a lock and key function, so high quality can be reached; (3) no unwanted byproduct formation. It makes the biodiesel production more environmental friendly. On the other hand, there are also some disadvantages, such as low biocatalyst activities, and frequent inactivation and denaturation by alcohol [10, 12]. To make the bioprocess more competitive, it is imperative to develop an effective biocatalyst production system.

Innovative design of genetically-engineered strains can significantly increase the recombinant protein yield with minimum investment in capital and operating costs. For developing recombinant enzyme production using microorganisms, generally three steps should be included (1) generating a

* Net Energy Balance for Bioethanol Production and Use. U.S. Department of Energy. Available Online: http://www1.eere.energy.gov/biomass/net_energy_balance.html [2007, Dec. 08]

† Available Online: http://www.biodiesel.org/ [2007, Dec. 08]

robust microbial strain, (2) developing a cultivation process with high-level gene expression and high-cell-density, and (3) optimizing a purification process. First and foremost importance is engineering a microbial host. Once a microorganism is chosen, the cultivation conditions are predetermined, and fine tuning can be done in a very rational manner. In addition, the purification process is also depending on the characteristics of the microbe and the enzyme.

Lipases (triacylglycerol acylhydrolase, E.C.3.1.1.3) are enzymes that catalyze the synthesis or hydrolysis of glycerol esters (e.g., triglycerides), such as fats and oils [13, 14]. Several microorganisms, such as bacteria, fungi, and yeasts, have been found to produce lipases, including *Pseudomonas fluorescens* [15, 16], *Pseudomonas cepacia* [17], *Rhizopus oryzae* [18], *Rhizomucor miehei* (e.g., the immobilized form commercially available from Lipozyme RM IM, Bagsvaerd, Demark) [18], *Thermomyces lanuginosus* (e.g., the immobilized form commercially available from Lipozyme TL IM, Bagsvaerd, Demark) [19], *Candida rugosa* (ex. *Candida cylindracea*), and *Pseudozyma antarctica* (ex. *Candida antarctica*). These lipases have numerous applications in organic chemistry and biotechnology [20, 21]. Apparently, *P. antarctica* lipase B (PalB, ex. CalB [22]) has been found to be the most active microbial lipase with good transesterification activity. The enzyme and its corresponding immobilized form (e.g., Novozyme 435) are commercially available from Novozyme A/S, Bagsvaerd, Demark. This enzyme has been the most common biocatalyst for biodiesel production [17, 23, 24]. While the biocatalyst is gaining more attention and has a potential to outperform chemical catalysts, extensive application remains impractical due to its high cost and several technical issues (e.g., inactivation by alcohol). To make the bioprocess economically feasible, there is a motivation to overproduce PalB.

Previously, recombinant PalB was expressed heterologously in *Aspergillus oryzae* [25], *Pichia pastoris* [26], and *Saccharomyces cerevisiae* [27]. However, the productivities of PalB from these microorganisms could not meet the quantity requirement for biodiesel production. Given the popularity of *Escherichia coli* as a workhorse for recombinant protein production, functional expression of PalB in this host was not demonstrated until recently [28, 29] and the expression performance stands improvement. A major technical issue limiting the PalB expression level in *E. coli* is associated with its intracellular misfolding. PalB is a lipase destined to be secreted extracellularly in the original psychrophilic and eukaryotic source of *P. antarctica*. Consequently, the intracellular overexpression might prevent it from being structurally adaptable to the mesophilic and prokaryotic expression system of *E. coli*, resulting in protein misfolding. In addition, wild-type PalB has three intermolecular disulfide bonds potentially associated with its folding process although these disulfide bonds might not be strictly required for bioactivity. Finally, cell growth is often arrested due to deterioration in cell physiology associated with

high-level expression and misfolding of PalB. Considering PalB's potential application and all the pros and cons of different expression systems, this thesis focused on identification of factors limiting recombinant PalB expression in *E. coli*.

By using molecular biological techniques, twenty-five expression plasmids, e.g., eighteen of them for cytoplasmic expressions, five for periplasmic expressions, and two for cell surface display, were made during the course of this study for identifying the limiting factors for recombinant PalB expression in *E. coli* (Table 1.1).

Table 1.1 Twenty-five expression plasmids for recombinant PalB expression in *E. coli*

No.	Name	Vector	Promoter	Fusion tag
Cytoplasmic expression plasmid				
1	pTrcG	pTrc99A	*trc*	–
2	pETKnL	pETKn 20b(+)	*T7*	–
3	pETL	pET 20b(+)	*T7*	–
4	pGEXC4S-M	pGEXC4S	*trc*	GST
5	pSkp-G	pDest556	*T7*-lacO	Strep-II-SKP
6	pT7PK-G	pDest555	*T7*-lacO	Strep-II-T7PK
7	pRset-G	pRSET	*T7*-lacO	6×His
8	pMBP-G	pMGWA	*T7*-lacO	MBP
9	pHMBP-G	pHMGWA	*T7*-lacO	6×His-MBP
10	pGST-G	pGGWA	*T7*-lacO	GST
11	pHGST-G	pHGGWA	*T7*-lacO	6×His-GST
12	pNusA-G	pNGWA	*T7*-lacO	NusA
13	pHNusA-G	pHNGWA	*T7*-lacO	6×His-NusA
14	pTRX-G	pXGWA	*T7*-lacO	TRX
15	pHTRX-G	pHXGWA	*T7*-lacO	6×His-TRX
16	pHisMBP-G	pDEST-HisMBP	*tac*	6×His-MBP
17	pETllDsbA-G	pET52A	*T7*-lacO	DsbA-6×His

18	pETllDsbC-G	pETM82	*T7*	DsbC-6×His
Periplasmic expression plasmid				
19	pETG	pET 20b(+)	*T7*	–
20	pFlag-P	pFLAG	*trc*	FLAG
21	pETDsbA-G	pET50A	*T7*-lacO	DsbA-6×His
22	pETDsbC-O	pETM80	*T7*	DsbC-6×His
23	pHisperiMBP-G	pDEST-periHisMBP	*trc*	6×His-MBP
Cell surface display plasmid				
24	pEstCmN	pEST100	*lac*	EstA
25	pEstKnN	pEST100	*lac*	EstA

In order to systematically study the functional PalB expression in *E. coli*, the following strategies were applied in designing the expression plasmids. First, targeting PalB in different compartments by constructing appropriate expression plasmids, including cytoplasmic expression plasmids (#1~3), and a periplasmic expression plasmid (#19). In addition, cell surface display of PalB was conducted to determine the feasibility of whole cell biocatalyst. Second, adding a fusion tag for improving the solubility of PalB. The hydrophilic tags were included His (#7), GST (#4, 10, and 11), MBP (#8, 9and 16), NusA (#12 and 13), TRX (#14, and 15), Skp (#5), T7PK (#6), llDsbA (#17), and llDsbC (#18) for the cytoplasmic expression, and Flag (#20), DsbA (#21), DsbC (#22), and MBP (#23) for the periplasmic expression. Third, coexpression of folding factors, e.g., chaperones, proteases, and catalysts, for mediating misfolded PalB. In principle, the constructs listed in Table 1.1 can be coexpressed with cytoplasmic chaperones or periplasmic folding factors. The cytoplasmic chaperones included the trigger factor (TF), DnaK/J-GrpE, GroEL/ES, leaderless DsbA (llDsbA), and llDsbC, and the combinations of TF and GroEL/ES, and GroEL/ES and DnaK/J-GrpE. The periplasmic folding factors were DegP, $DegP_{S210A}$, FkpA, SkP, SurA DsbA, DsbC, and a cocktail (SurA, FkpA, DsbA, and DsbC).

1.2 Research Objectives

The overall objectives of this thesis are to

1. Explore PalB expression in *E. coli* system, including the construction of the PalB expression plasmids in the cytoplasm, periplasm, and for cell surface display.
2. Systematically study the expression systems.
3. Study the expression systems mediated by fusion tags
4. Study the effect of cytoplasmic chaperones and periplasmic folding factors on PalB expression.
5. Based on the above studies, identify the factors limiting the PalB expression in *E. coli* and provide comments for the further improvement of PalB expression.

1.3 Outline of the Thesis

This thesis consists of eight chapters. The scope of each chapter is as follows:

Chapter 1 gives an introduction to this thesis, including an overview of biodiesel production, advantages and disadvantages of biocatalyst for biodiesel production. The hypothesis, objectives and the scope of the thesis are also given in this chapter.

Chapter 2 reviews the recombinant protein production in *E. coli*. Why is it suitable for PalB expression and what are the potential challenges for this study?

Chapter 3 presents the construction of PalB expression plasmid and expression results for identification of factors limiting heterologous production of PalB in the cytoplasm.

Chapter 4 presents the improvement of PalB cytoplasmic expression by using a mutant host strain and/or by adding a fusion tag.

Chapter 5 presents the construction of PalB expression plasmid and expression results for showing the effect of folding factors in rescuing unstable PalB to enhance its overexpression in the periplasm.

Chapter 6 presents the construction of PalB fusion expression plasmids and expression results for improvement of solubility and bioactivity of PalB by addition of fusion tag in the periplasm.

Chapter 7 presents PalB displaying on *E. coli* cell surface and the potential application.

Chapter 8 presents the conclusions of this study, contributions of this research and recommendations for future work.

Chapter 2
Literature Review

2.1 Choice of Host Strain for Recombinant Protein Production

Among various protein expression systems, such as bacteria (both gram-positive or gram-negative), fungi, insect cells, mammalian cells, plant cells and even whole (transgenic) animals/plants, the gram-negative bacterium of *Escherichia coli* is the most widely used as host for biotechnological and pharmaceutical applications [30]. Generally, *E. coli* is used for cost-efficient production of large amounts of proteins that are limited in size and have a relatively simple structure. However, even for therapeutic protein production, which typically has several options in selecting the expression system [31, 32], nine out of the 31 EPA approved biopharmaceuticals during 2003 and 2006 were produced in *E. coli*, which is second only to mammalian cell systems [33]. In addition, approximately 80% of the proteins used to solve 3-dimensional structures submitted to the protein data bank (PDB) in 2003 were prepared in *E. coli* expression systems [34]. Therefore, *E. coli* is always one of the first choices as a host for recombinant protein research and production.

Three major factors result in the popularity of *E. coli* used as a host for recombinant protein research and production. First, its genetics are far better characterized than any other microorganism and the mature technologies (in genetic engineering, metabolic engineering, and protein engineering) for manipulation and construction of the host/vector system make it possible for a variety of attempts to improve the expression performance. Second, the fast growth rate with simple nutritional requirements and easy cultivation technology for *E. coli* make it suitable for many biomanufacturing applications [35, 36]. Third, the successful applications of recombinant protein production in *E. coli* inspire and motivate more in-depth exploration to improve the *E. coli* expression system and investigate potential applications. Theoretically, the maximum cell density that can be reached is approximately 200 g-dry-cell-weight (DCW)/liter in fed-batch cultivation [37] and accumulation of intracellular recombinant proteins can be up to 50% of total cellular protein by using a strong promoter [38]. Therefore, *E. coli* already is a well established protein overproducer but potential improvements are possible.

Several drawbacks need to be paid attention to when choosing *E. coli* as a host for recombinant protein production, such as the pathogenicity of this microorganism, lack of posttranslational processing abilities (particularly glycosylation) which required for many eukaryotic proteins [32], lack of a secretion

mechanism for the efficient release of protein into the culture medium [39], and limited ability to facilitate extensive disulfide bond formation [40].

Considering the global demand to produce large quantities of PalB with high activity at low cost for competing with chemical catalysts for biodiesel production, this work focuses on heterologous expression of PalB in the *E. coli* systems.

2.2 Genetic Strategies for Recombinant Protein Expression in *E. coli*

As mentioned before, innovative design of genetically engineered strains can greatly increase the recombinant protein yield with minimum investment in capital and operating costs [41]. Based on the central dogma of molecular biology that DNA makes RNA makes protein (although RNA can also "make" DNA), the strategies for high level gene expression in *E. coli* are: (i) gene replication, (ii) transcriptional regulation, (iii) translational regulation, (iv) protein targeting, (v) folding factor (e.g., molecular chaperone, protease, and foldase), (vi) fusion tag, and (vii) adjustment of cultivation conditions (e.g., temperature, oxygen level, pH, induction time, inducer concentration, and medium) [38, 42]. The idea is that in addition to improving the efficiency of each gene expression step, a 'balanced' protein synthesis flux throughout these steps should be properly maintained to avoid the accumulation of any protein species along the formation pathway of targeted protein. Steps (v) to (vii) are commonly used for mediating high level protein expression and will be reviewed later. In this section the review focuses only on the essential elements in design of recombinant expression systems, including steps (i) to (iv) techniques (Table 2.1) [42-44].

Table 2.1 Techniques for genetically optimizing recombinant protein overexpression in *E. coli*

Techniques	Comments
Plasmid copy number	Gene dosage, as manipulated through plasmid copy number, affects expression
Selection antibiotic	Choice of antibiotic resistance on the expression plasmid can influence heterologous protein expression
Promoter	Strong/weak, inducible/constitutive, promoter and regulation is a major influence on protein expression, which is also affected by relative orientation and strength of promoters on the plasmid
Transcription termination	Effectiveness and spacing of transcription terminators affect expression
mRNA stability	The stability of the mRNA impacts yield. Secondary structure, especially at the 5´ end of the message, often plays a critical role
Translation signal	The ribosome-binding site affects the level of ribosome loading and clearance, and hence, expression. Secondary structure at the 5´ end of the message can affect the accessibility of the ribosome binding site

For replication, the gene dosage can be increased with the use of high-copy-number plasmids [45, 46] or "runaway" plasmids [47, 48]. Using the plasmids with high-copy-number always mean high efficiency, e.g., pUC (copy number 500~700) in the cloning application. However, the primary carbon and energy metabolism in the host cell are significantly affected by gene expression with high-copy-number plasmids [49-52] and the metabolic overload not only exhausts the cell itself but also causes the loss of plasmid and decreases protein yield dramatically [53]. For protein expression, therefore, it is necessary to not use the high-copy-number plasmids [54, 55], e.g., the plasmids derived from pBR322 (copy number 15~20) [56, 57] or pACYC (copy number 10~12) [50, 58]. When the design includes two plasmids coexisting in a host for coexpression of genes (e.g., one for target gene, and another for chaperone), they should be compatible for replication (e.g., one from pBR322 and another from pACYC) and also should have different antibiotic resistance markers. However, it was recently reported that

incompatible plasmids carrying the same replicon coexisted stably and coexpressed two proteins efficiently with different selection marker [59, 60].

Antibiotic resistance marker (or selection marker) is designed for screening the plasmid in the cell. Generally used antibiotics are ampicillin, tetracycline, chloramphenicol, and kanamycin. Ampicillin binds to and inhibits a number of enzymes in the bacterial membrane that are involved in the synthesis of the cell wall. Tetracycline binds to a protein of the 30S subunit of the ribosome and inhibits ribosomal translocation. Chloramophenicol binds to the ribosomal 50S subunit and inhibits protein synthesis. Kanamycin binds to ribosomal components and inhibits protein synthesis [61]. Only the cells containing the plasmids, in which the resistance marker confers resistance to the drug, can survive in the culture supplemented with the corresponding drug. When the design involves two plasmids coexisting in a host for coexpression of genes (e.g., one for target gene, and another for chaperone), they should be compatible in replication origin (e.g., one from pBR322 and another from pACYC), and also should have different antibiotic resistance markers.

2.2.1 Transcription is the Major Step Limiting the Overall Gene Expression.

In the ensuing section the *lac* operon is used as an example to depict how transcription works (Figure 2.1). It consists of the control sequences *lac*P and *lac*O followed by the tandem arranged genes for *β*-galactosidase (*lac*Z), galactoside permease (*lac*Y), and thiogalactoside transacetlylase (*lac*A). In the absence of an inducer, for example the physiological allolactose such as IPTG, the *lac* repressor, the product of the *lac*I gene, binds to operator (*lac*O) so as to prevent the transcription of the *lac* operon by RNA polymerase. The binding of the inducer causes the repressor to release the operator that allows the *lac* structural genes to be transcribed to a single polycistronic mRNA. The mRNAs transiently associate with the ribosome so as to direct them to synthesize the encoded polypeptides. The *E. coli* RNA polymerase has the σ subunit, which initiates transcription on the sense strand of a gene at a position designated by its promoter. The most conserved region of the promoter is -10 and -35 region. After the initiation of RNA synthesis, the *σ* subunit dissociates from the RNA polymerase, which then autonomously catalyzes chain elongation in the 5'→3' direction. RNA synthesis is terminated by a segment of the transcript that forms a G+C rich hairpin. Therefore, a strong promoter is the key for increasing the efficiency of transcription.

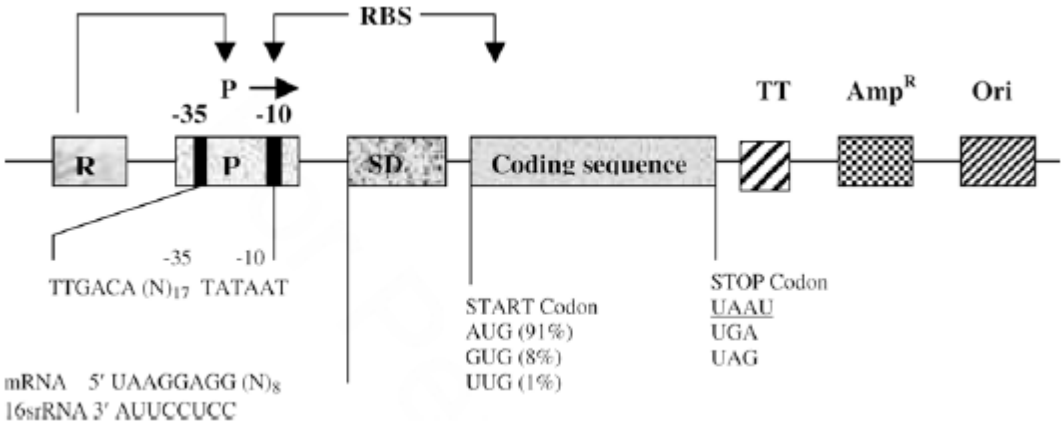


Figure 2.1 Schematic presentation of the salient features and sequence elements of *E. coli* expression vector. (Modified after [42]) Shown as an example is the hybrid *tac* promoter (P) consisting of the -35 and -10 sequences, which are separated by a 17-base spacer. The arrow indicates the direction of transcription. The *lac*O is a short region lies partially within the promoter and interacts with repressor that controls the transcription of the operon. The ribosome-binding site (*RBS*) consists of the Shine-Dalgarno (SD) sequence, followed by an A+T rich translational spacer, which has an optimal length of approximately 8 bases. The SD sequence interacts with the 3' end of the 16S rRNA during translational initiation, as shown. The three start codons are shown, along with the frequency of their usage in *E. coli*. The repressor is encoded by a regulatory gene (R), *lac*I, or *lac*I^q, modulates the activity of the promoter. The transcription terminator (*TT*) serves to stabilize the mRNA and the vector. An antibiotic resistance gene, e.g., for ampicillin (Ap^R), facilitates phenotypic selection of the vector. The origin of replication (*Ori*) determines the vector copy number. The coding sequence, e.g., *lac*Z, *lac*Y, and *lac*A, is the gene(s) for protein(s) of interest.

A useful promoter should exhibit several desirable features: (i) be strong, resulting in the accumulation of protein making up 10-30% or more of the total cellular protein; (ii) exhibit a minimal level of basal transcriptional activity; (iii) possess inducibility in a simple and cost-effective manner. To render it suitable for high-level protein synthesis, generally speaking, recombinant expression plasmid requires a strong transcriptional promoter. A number of strong promoter systems are available for gene expression in *E. coli* with various induction factors, such as chemicals (e.g., IPTG, arabinose or tetracycline), heat, and pH, to increase the transcriptional efficiency (perhaps translational efficiency as

well) and, in turn, to increase the expression level (Table 2.2) [38, 62]. The first three promoters in Table 2.2 are the most prominent and have frequently been used for the *E. coli* system, *lac* is a natural promoter [63], and *tac/trc* are synthetic promoters, which only differ by 1-bp in the length of spacer domain [64-66]. They are regulated by *lac*I, *lac*I^q by induction of isopropyl-β-D-thiogalactopyranoside (IPTG), and also regulated by catabolite repression and the metabolic state, which is represented by the cyclic AMP level. Therefore significant variations of the level of expression from a given vector may be observed depending on the host strain used. The *tac/trc* promoters are at least 5-fold more efficient than *lacUV5* promoter and allow the accumulation of up to 15~30% of total cell protein. There are many recombinant proteins successfully expressed under these promoters, e.g., penicillin acylase [67-69], PalB [28], human CD83 [70].

Table 2.2 Commonly used promoters for recombinant protein expression

Promoter	Induction	Characteristics
lac (*lacUV5*)	IPTG, thermal (0.05-2.0 mM)	Low-level expression; leaky expression
tac/trc	IPTG, thermal (0.05-2.0 mM)	High-level, but lower than T7; leaky expression
T7	IPTG, thermal (0.05-2.0 mM)	High-level expression, leaky expression; difficult to achieve high cell densities
*T7/lac*O	IPTG, thermal (0.05-2.0 mM)	High-level expression, tighly controlled expression; difficult to achieve high cell densities
λp_L	Thermal From 30-42 °C	Induction cannot be performed at low temperatures; partial induction cannot be achieved
cspA	Thermal From 15-29 °C	High-level expression, tightly controlled expression;

tetA	Anhydrotetracycline 200 μg/l	Tightly controlled expression
araBAD	L-Arabinose (0.001-5%)	Expression level can be controlled in a dose-dependent manner; catabolite repressed by glucose; few vectors available

The *T7* gene expression system is the most used in recombinant protein production. It is based on bacteriophage T7 RNA polymerase, which elongates chains about 5 times faster than the native *E. coli* RNA polymerase. In order to produce T7 polymerase, the gene of the T7 RNA polymerase, *T7* gene 1, is embedded in the *E. coli* chromosome under the control of a *lac* promoter derivate *lacUV5*, which genetic modification is called DE3. This *lacUV5* promoter, which has three point mutations to the wild-type *lac* promoter, increases the promoter strength and decreases its dependence on cyclic AMP, and less sensitive to glucose. T7 RNA polymerase is highly selective for its own promoters, which do not occur naturally in *E. coli*. A relatively small amount of a T7 RNA polymerase is sufficient to direct high-level transcription from a *T7* promoter in a multicopy plasmid. Such transcription can proceed several times around the plasmid without terminating. Due to T7 polymerase activity the transcription by *E. coli* RNA polymerase is greatly decreased. A specific mRNA produced by T7 RNA polymerase can rapidly saturate the translational machinery of *E. coli*, such that the rate of protein synthesis from such an mRNA will depend primarily on the efficiency of its translation. When the mRNA is efficiently translated, a target protein can accumulate to greater than 50% of the total cell protein in three hours or less. The T7 RNA polymerase seems to be capable of transcribing almost any DNA linked to *T7* promoter, so that T7 expression system should be capable of transcribing almost any gene or its complement in *E. coli*. There are thousands of recombinant proteins successfully expressed in BL21(DE3) under the control of T7 promoter, e.g., penicillin acylase [71], CalB [29], [72].

The λp_L is a tightly regulated phage promoter by the cI repressor. By up-shift the growth temperature, from 30 to 42 °C, the repressor is inactivated, and the gene expression is induced. The more interesting thing is that when the cultivation is at low temperature, λp_L becomes constitutive and the gene can be expressed. In contrast, *csp*A promoter is a cold-shock promoter directing recombinant protein production at low temperature [73]. pCold™ is cold-shock expression vector under *csp*A promoter [74]. PalB was constructed in pColdIII and functionally expressed in cytoplasm of *E. coli* at 15 °C [29].

Unlike the above on/off promoters, *araBAD* promoter is an inducer-control promoter, which means that the expression level is a function of L-arabinose concentration, and is tightly shut off in media containing glucose but lacking arabinose. For full induction of *araBAD*, the L-arabinose concentration varied from 0.001~5% [75, 76]. This promoter is used in many applications to control gene expression, e.g., penicillin acylase [77], chaperone coexpression [67, 68, 71]. In this study, it is used for coexpressing the folding factor for mediating the expression of PalB.

The *tetA* promoter is useful for the tightly regulated high-level synthesis of a foreign gene product in variety of *E. coli* strains [78]. Its strength was comparable to that of *lacUV5*, but it overcomes the weakness of *lac* promoter. Anhydrotetercycline binds the promoter about 35-fold higher than tetracycline and its antibiotic activity is 100-fold lower. In contrast to other systems, the basal level is very low and independent on the *E. coli* strain and the metabolic state [79]. All chaperones and proteases in this study were constructed under the control of *araBAD* promoter or *tetA* promoter. There are many other promoters used in *E. coli* system for recombinant protein productions. For example, the pH promoter is very strong promoter and recombinant proteins are produced at levels of up to 40~50% of the total cellular protein [80]. It also overcomes the concern of toxicity of the inducer, such as IPTG in the applications of final protein product. However, none of them is ideal for recombinant protein expression in *E. coli*. For a particular application, one still needs to do extensive screening in order to find the best promoter [42].

2.2.2 Translation Regulation is also Important for the Overall Gene Expression.

Translation can be limited by several factors, such as the initiation efficiency, the number of available ribosomes and/or tRNAs, the secondary structure and/or stability of mRNAs and the presence of rare codons. Genetic modification of the regulatory elements has been proved to be effective for improving the translational efficiency [42, 81, 82]. Due to the expression vectors used in this study that were either purchased or supplied as gifts, it is assumed that the most parameters, except rare codons, have been optimized already. In addition, the effect of rare codons in *palB* gene expression is beyond the scope of this study.

2.2.3 Expression Vector

An expression vector, usually a man-made bacterial plasmid, is a double-stranded closed circular DNA molecule and is used for expressing a gene encoding the protein of interest [61, 83]. It contains all of the

features shown in Figure 2.1, and replication origin (Ori) [for controlling the copy number], and multiple cloning sites [for inserting the target gene in frame] (Figure 3.1).

For most ready-for-use plasmid expression vectors, the systems have been optimized for different purposes and choosing the appropriate systems is needed. For example, the plasmids with pUC (copy number 500~700) origin is suitable for cloning work, while the plasmids derived from pBR322 (copy number 15~20) or pACYC (copy number 10~12) are good for gene expression.

There are also some other considerations for choosing expression vectors. First, depending on the protein targeting, the expression vector should contain a signal peptide for leading the recombinant protein translocate into the periplasm or secrete into the medium. A signal peptide plays an important role in protein export [84]. It is highly hydrophobic but with some positively charged residues and it generally contains 3~60 amino acid residues, most often between 15~30. After directing the nascent polypeptide chain translocation, it is removed and rapidly degraded. Many eukaryotic proteins have signal peptides, but most of those signal peptides do not efficiently translocate nascent proteins when they are expressed in *E. coli* [28]. For the heterologous expression in the periplasm of *E. coli*, the most well-characterized signal peptides are *pelB* (the signal peptide region of PelB of *Erwinia carotovora*), *ompA* (the signal peptide of a major outer membrane protein, OmpA, of *E. coli*), *phoA* (the signal peptide of alkaline phosphatase of *E. coli*). In this study, five signal peptides are used, including *pelB* for pETG, *ompA* for pFlag-P, *mbp* for pHisperiMBP-G, *dsbA* for pETDsbA-G, and *dsbC* for pETDsbC-O.

Second, leakiness frequently results in the overgrowth of plasmid-free segregants and/or segregants that have a reduced capacity for high-level expression, an impediment for high-level protein production. For overexpression of recombinant protein in *E. coli*, the multicopy plasmids are generally desired for maximum gene expressions: (i) they are small size, usually less than 5 kbp, and easily manipulated for controlled gene expression; (ii) high copy number generally results a good expression. However, the metabolic burden, resulted from the overexpression of gene, is detrimental for maximum productivity. Furthermore, the presence of a high copy plasmid may also cause a significant shift in the normal metabolism of the host cell, which can increase the risk of plasmid instability. A novel type of expression vectors with a dual regulation of both the plasmid copy number and gene expression are described [85]. When they are not induced, they are maintained as a single-copy plasmid.

Third, downstream processing is another major issue that should be taken into the account when designing the expression vector. Several affinity tags, such as 6×His, GST, FLAG, MBP. have been developed to facilitate downstream purification of recombinant proteins [86, 87]. The expression vector is specifically designed so that the target protein would be expressed as a protein-tag fusion, which can be

easily recovered using the proper affinity chromatography. The target protein moiety can be released using a specific protease to cleave the junction between the target protein and the tag.

Upon designing the expression plasmid, the above genetic strategies can be integrated to simultaneously enhance gene expression and protein purification. In conclusion, most of the technical strategies developed for recombinant protein production in *E. coli* focus on innovative design of the expression plasmids. For the most ready-for-use plasmid expression vectors, the systems have been optimized for different purposes.

2.3 Protein Targeting in *E. coli* Expression System

E. coli consists of a cytoplasm surrounded by a lipid bilayer called the cytoplasmic membrane (or inner membrane) and a cell envelope, which consists of a comparatively thin peptidoglycan layer, the periplasm membrane (or outer membrane). Therefore, the two membranes provide *E. coli* three possible compartments for the expression of recombinant protein, the cytoplasm, periplasm, and cell surface display [88-91]. The major biological steps involved for recombinant protein expression in *E. coli* are replication of the gene encoding the protein of interest, transcription of DNA to mRNA, translation of mRNA to polypeptide, and posttranslational processing (Figure 2.2) [38, 92]. Each of the compartments offers advantages and disadvantages for gene expression, depending on the experimental objectives and the characteristics of the targeted protein.

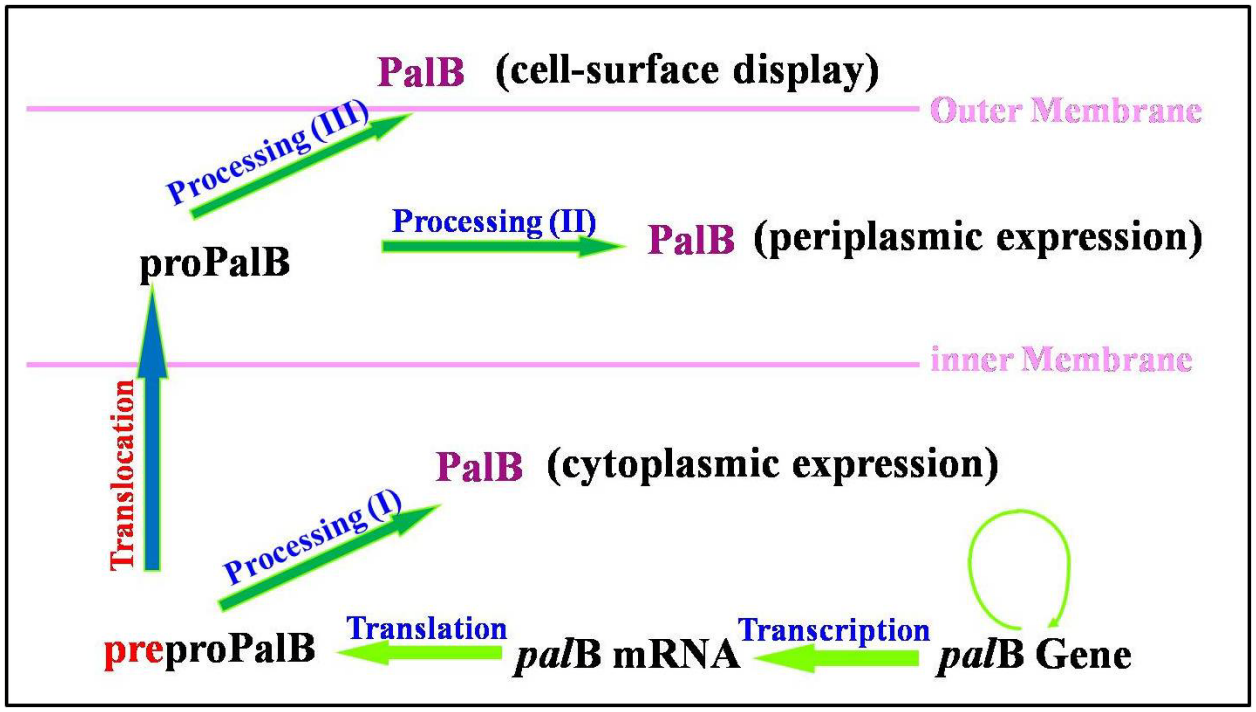


Figure 2.2 The major biological steps involved for recombinant protein synthesis in *E. coli*. The expression plasmid, e.g., *palB* gene, is duplicated with the cell replication. After induction, the RNA polymerase binds to the operator on the sense strand of plasmid to transcript *palB* gene into mRNA, e.g., *palB* mRNA. The translation starts when the ribosome attaching the mRNA, which is decoded to produce a polypeptide, e.g., preproPalB (with a signal peptide), or proPalB (w/o a signal peptide). If the expression is designed in the cytoplasm, the proPalB will be folded into PalB. If the expression is in the periplasm, the preproPalB will translocate into the periplasm with the leading of the signal peptide, and be folded into PalB. If the destination is surface display, the translocated PalB will be carried out of the cell wall with a specific carrier.

2.3.1 Cytoplasmic Compartment

Cytoplasm is the compartment where all recombinant proteins are first synthesized and are possibly located as its final expression destination [93]. Related to periplasmic expression, it is simpler for plasmid construction (e.g., no signal peptide is required), and generally higher protein expression is reached [94]. However, the proteins at concentrations of 200~300 mg/ml in the cytoplasm suggests a highly crowded

and unfavorable protein folding environment, especially during the recombinant protein overexpression, in which accumulation of intracellular recombinant proteins at a level up to 50% of total cellular protein [38]. Therefore it is not surprising that the formation of inclusion bodies remains a significant barrier for recombinant protein expression in the cytoplasm.

Several experimental approaches have been developed to reduce inclusion body formation. For example, (1) Reduction of the rate of protein synthesis by lowering cultivation temperature or using a moderately strong promoter [38, 92, 95]; (2) Coexpression of chaperones [96]; (3) Addition of fusion tag [87]; (4) Substitution of amino acids in the polypeptide chain [97-99]; (5) Reduction of acetate's effect by changing chemicals in the growth medium[100, 101]; (6) Alteration of the pH [102, 103]; or (7) Replacement of the bacterial strain [104].

Due to disulfide bond formation in *E. coli* requiring the disulfide bond formation catalyst, which could not be found in the reduced environment, the disulfide bond formation is strongly disfavored in the cytoplasm [105, 106]. But Robinson and King found the transient disulfide bonds were presented in a folding intermediate of a non-disulfide bonded protein, P22 tailspike endorhamnosidase [107]. Jonathan Beckwith' group in Harvard Medical School identified two thioredoxins, TrxA and TrxC, and three glutaredoxins in *E. coli* cytoplasm. The oxidized form of these proteins can catalyze the formation of disulfide bonds in peptides. However, in the cytosol, both the thioredoxins and the glutaredoxins are maintained in a reduced state by the action of thioredoxin reductase (TrxB) and glutathione, respectively. In a *trxB* mutant, the two thioredoxins are oxidized and serve as catalysts for the formation of disulfide bonds [104]. They found that a protein with multiple disulfide bonds can be fully expressed actively in the host strain with *gshA, or gor* and *trxB* mutated [108]. In *E. coli*, glutathione is synthesized by *gshA* and *gshB* gene products and the enzyme glutathione oxidoreductases, the product of the *gor* gene, is required to reduce oxidized glutathione and complete the catalytic cycle of the glutathione-glutaredoxin system. However, the growth was very poor for the double mutants with *gshA* and *trxB*. The strains FA112 (*trxB gor supp*) and FA113 (*trxB gshA supp*) almost reached the same growth rate as the wild-type strain bearing a secretion plasmid, e.g., human tissue plasminogen activator [109] and the single chain variable fragment antibiotics [110]. Novagen took over the technology and developed a series of commercial strains called Origami^{TM}, which are used widely for recombinant protein expression in the cytoplasm and was used in this study. Coexpression of chaperones in mutant strain were studied and it was found that the effect was limited, e.g., Fab antibody fragment [111], PalB [29].

2.3.2 Periplasmic Compartment

Periplasmic expression of recombinant proteins provides several advantages compared to cytoplasmic expression. First, an authentic N-terminal amino acid protein can be obtained after cleavage of the signal sequence by a specific signal peptidase. Second, there is much less protease activity in the periplasmic space than in the cytoplasm. Third, recombinant protein purification is simpler due to less contaminating proteins in the periplasm. Fourth, oxidative environment and existing catalysts facilitate the disulfide bonds formation.

There are several advantages for expressing eukaryotic proteins in the periplasm because its oxidative environment is suitable for disulfide bond formation and it is feasible to obtaining proteins with authentic N-termini. This compartment also contains fewer proteins such that downstream purification will be facilitated. Microbial cell surface display has been extensively explored due to its significant impact on various biotechnological and industrial applications, such as vaccine development, biosensor development, high-throughput screening of macromolecular libraries and preparation of whole-cell biocatalysts [91]. Extracellular release of gene products offers an alternative for recombinant protein production in *E. coli* [90] and this approach is particularly valid for periplasmic proteins. The proteins released are less subject to intracellular proteolysis and selective secretion of the target protein would facilitate downstream purification. Also, reducing the local protein concentration via protein release can alleviate intracellular protein misfolding.

Generally proteins found in the periplasm are synthesized in the cytoplasm as premature proteins, which contain a signal peptide in front of the mature proteins. With the leading of the signal peptide, the protein is translocated through the inner membrane into the periplasm for folding and maturation [89]. The efficiency of protein secretion varies depending on the host, the signal peptide, and the type of protein to be secreted. To date, there is no general rule in selecting a proper signal sequence for a given recombinant protein to guarantee its successful secretion.

The main problems in secretarial recombinant protein expression are: (i) incomplete translocation across the inner membrane [92], (ii) insufficient capacity of the export machinery [112], (iii) low or undetectable amounts of recombinant protein secretion, (iv) formation of inclusion bodies in the cytoplasm and periplasm when using strong promoters, and (v) incorrect formation of disulfide bonds, (vi) proteolytic degradation [113]. However, without trial-and-error, it is somewhat difficult to select a proper host/vector system and a signal sequence for the secretion of a desired protein [88].

2.3.3 Cell Surface Display

Recently microbial surface display has been extensively studied due to its significant impact on various biotechnological and industrial applications, including vaccine, biosensor, screening of macromolecular library, and whole-cell biocatalyst [91, 114-116]. Except a signal peptide, it needs a carrier (or anchoring motif), which usually is an outer membrane protein possessing unique membrane-spanning structure, e.g., OmpA, OmpF, OmpS, and LamB. Depending on the properties of the carrier and the passenger (or displayed protein), the protein can be fused at an N-terminus, a C-terminus, or sandwiches fused at both N- and C-termini to expose the passenger protein to extracellular environment with its bioactive formation.

The passenger protein should be maintained in an unfolded or translocation-competent form until it is displayed [117]. The misfolded passenger-carrier-related proteins could induce periplasmic stress. In addition, overexpression of cell surface display could generate much higher cell stress than the cytoplasmic or periplasmic overexpression, because the outer membrane becomes fragile or even disintegrated upon frequent insertion of passenger-carrier fusions [118]. Therefore the cell defense system could sense the stresses, most likely react to the periplasmic stress, through σ^E and Cpx, which are partially overlapping [119-122]. The response to the extracytoplasmic stress is driven by synthesizing the proteins, e.g., DegP, FkpA, and DsbA, Skp, and SurA, who are in charge of releasing the stresses through degradation of misfolded proteins and/or mediation of misfolded proteins folding.

Cell surface display was interested for whole cell biocatalyst, PalB. In this way, the cell surface displayed PalB could be directly used in enzyme immobilization, and don’t need to go through the purification process.

2.4 Technical Limitations for Production of Recombinant Protein

Some often encountered problems of recombinant protein expression in E. coli and possible solutions for those problems are summarized in Table 2.3. Theoretically, all the biological steps, depicted in Figure 2.2, have to be effective to result in a high recombinant protein yield. The step limiting the overall gene overexpression should be targeted for improving expression performance. However, enhancing the efficiency of the original limiting step could imply that another step becomes limiting. A typical example is that protein folding becomes limiting, resulting in the formation of misfolded protein, when strong promoters are used for boosting transcription. This raises an important issue that, for the overproduction of recombinant proteins, a 'balanced' protein synthesis flux throughout all the gene expression steps (e.g.,

transcription, translation and post-translational steps) should be properly maintained to avoid the accumulation of polypeptide intermediates in addition to boosting the individual limiting step.

Misfolding of the target gene products can result in the formation of insoluble protein aggregates as non-bioactive inclusion bodies. This represents a major technical hurdle for recombinant protein production in *E. coli* [123]. One of the reasons for inclusion body formation is that the overexpressed gene products cannot be suitably processed by folding modulators to develop the proper protein structure [43]. Inclusion bodies can be observed by optical microscopy as refractile particles of up to about 2 μm^3 [124]. Most inclusion bodies are found in the cytoplasm but the secreted proteins can also form aggregates in the periplasm. Inclusion bodies do offer several advantages, such like facile isolation of protein in high purity and concentration, target protein protected from proteases, desirable for production of proteins that if active are lethal to host cell. In principle, these insoluble aggregates can be easily purified and refolding can be explored to regain the protein's biological activity [125]. However, the potential of this approach is often limited since renaturation of these misfolded proteins is often ineffective [38, 123]. Therefore, *in vitro* protein refolding is a choice only if all the efforts to solubilize the aggregates and gain the native conformation *in vivo* were unsuccessful. In other words, the formation of inclusion bodies keeps recombinant protein away from proteolytic degradation, and refolding can recover active proteins from inactive inclusion bodies, the initial expression of soluble, correctly folded recombinant protein is ideal [126].

High-level gene expression can also induce a stress response [127]. The major physiological stresses include heat shock, starvation and stationary-phase stress, pH stress, and oxidative stress. The overexpressed heterologous gene products can challenge cells with different levels of toxicity and metabolic burden, and cell growth is hindered. In general the specific growth rate of cells expressing a product correlates inversely with the rate of recombinant protein synthesis [127, 128]. On the other hand, due to the limitation in nutrient and/or oxygen availability, cells in dense cultures often have lower cellular activities and metabolic energy that are required for effective biosynthesis. A compromise in balancing the levels of gene expression and cell growth needs to be reached at a certain point to maximize the volumetric recombinant protein productivity. Studies have been conducted to characterize cellular responses to recombinant protein overproduction [129, 130]. Stress can be reduced by slow adaptation of cells to a specific production task, by gradually increasing the level of inducer, or by slowly increasing the plasmid copy number during cultivation.

Table 2.3 Some problems of recombinant protein production in *E. coli* and possible solutions (modified after [79])

Symptom	Possible problems	Solutions
Cell death or no colonies	Toxic protein, high basal expression	More stringent control over basal expression Tightly controlled promoter system Weaker promoter Lowering temperature Lowering inducer concentration
Insoluble disulfide protein (inclusion bodies)	Reduction of disulfide bonds	Minimize reduction in cytoplasm Accumulation in the periplasm
Insoluble protein (inclusion bodies)	Too much expression	Attenuate expression by weaker promoter, lowering temperature, or inducer concentration, decrease plasmid copy number, adding fusion tag
No activity	Misfolded protein, fusion tag can decrease activity	Minimize reduction in cytoplasm Accumulation in the periplasm Attenuate expression Change fusion tag
No protein, truncated protein	*E. coli* codon usage (codon bias)	Supply rare tRNAs Stronger promoter Increase plasmid copy number Lower temperature Tightly controlled promoter system

In a conclusion, most of the technical strategies developed for recombinant protein production in *E. coli* focus on innovative design of the expression vectors. However, the physiological deterioration associated with gene overexpression can result in a number of negative cellular responses, such as growth inhibition, cell lysis or even death, which are detrimental to protein productivity. To maximize the protein productivity, emphasis should also be given for high-cell-density cultivation by improving cell physiology upon gene overexpression. The integrative approaches in the aspects of both strain and bioprocess development for effective biomanufacturing form the major scope of a new recombinant protein production in *E. coli*.

2.5 Resolutions for the Recombinant Protein Overexpression in *E. coli*

The goal for recombinant protein production in *E. coli* is to simultaneously achieve a high-level gene expression and high-cell density. However, this goal is really difficult to reach because high-level gene expression and high-cell-density always conflict with each other. The resolution is making compromise that to optimize each gene expression step based on a 'balanced' protein synthesis flux being properly maintained to avoid the accumulation of any protein species along the protein formation pathway. Therefore the inclusion body formation and the deteriorated cell caused by the overexpression of the recombinant protein could be prevented or mediated by coexpression of the periplasmic folding factor, adding fusion tag, and/or the techniques mentioned in the section of 2.3.

2.5.1 Fusion Tag is a Tool to Assist Heterologous Protein Expression

Recombinant proteins with low or no structural content can often be produced in *E. coli* whereas proteins containing both structured and large unstructured regions always encounter problems, especially for the eukaryotic proteins. It is reported that only 22.7% of heterologous proteins of they have expressed in *E. coli* are soluble (1452 in total of 6386 proteins) [126]. Among the strategies for improving expression of recombinant protein in *E. coli*, including choosing strong promoter, coexpressing chaperone, using fusion tag, and adjusting cultivation conditions, no other technology has been as effective at improving the solubility of recombinant proteins as fusion tag does, especially for difficult-to-express protein [126].

Fusion tag technology is able to improve heterologous expression by overcoming many challenges, including the low translation rate, the proteolytic degradation of target proteins, protein misfolding, poor solubility, and the necessity for good purification methodologies [87, 126]. The ribosomes can efficiently initiate translation at the N-terminus methionine residue of the fusion tag. Attachment of a highly evolved translational frame at the N-terminus of an inefficiently translated protein improves the latter's efficiency

of translation [131], which inefficiency is often caused by a secondary structure in the mRNA interfering with the binding of ribosomes [87, 132, 133].

It is not uncommon that the overexpressed heterologous proteins are attacked by the proteases in the cell defense system [134, 135]. Proteolysis is highly regulated and plays critical roles in maintaining cellular homeostasis, including removing unwanted or incorrectly folded proteins from the cell [136]. Fusion protein could protect the target protein from being degraded by solubilizing the fused protein and improving folding to avoid it off-productive pathway and aggregation [38, 86, 137].

How the fusion tags improve solubility, promote protein folding and enhance recombinant protein expression is unknown, maybe because the fusion tag efficiently and rapidly reaches a native conformation as it emerges from the ribosome (or soon after its release), and this the highly conserved structure of the fusion tag acts as chaperone to stabilize and promote the acquisition of correct structure in downstream folding units by favoring on-pathway isomerization reactions [92, 137].

More excited about fusion tag is that some bifunctional fusion tags (e.g., MBP, FLAG, and GST) not only are the solubility enhancer to improve the target protein production, but also offer the opportunities to facilitate purification and detection of recombinant proteins. However, fusion tag technology also has disadvantages.

The major disadvantage of protein fusion technology is proteolytic cleavage. Generally it is low yield for generating free protein from fusion protein, the protein of interest could be precipitated after cleaving from the fusion tag, optimization of cleavage conditions is labor-intensive, expense of protease is high, and it may fail for recovering the active, structurally intact protein. Another problem is the generation of non-native N-terminal amino acids. Finally, the large fusion partners impose a high metabolic cost on the cells, resulting in cell deterioration or even lysis. The improved fusion tag technology by using the Ubiquitin and SUMO fusion tag systems can overcome the disadvantages resulting efficient, accurate cleavage, and authentic N-terminus of target protein [126].

2.5.2 N-Terminus Fusion Tag is More Powerful

Originally it was believed that virtually any highly soluble protein could act as a solubility enhancer, but this seems not true [138-140]. Both N-terminus and C-terminus fusion tags are reported successfully improving the fused protein's solubility and enhancing the recombinant protein expression in *E. coli*. N-terminus fusion tags are even more effective [141] because translational efficiency is known depending on the mRNA sequence around the ribosome binding site including the 5' end of the coding sequence.

Therefore, sequence encoding the N-terminus protein fusion provides good translation initiation properties to generate a high expression level [137, 138].

An N-terminus fusion tag can help to alleviate two obstacles at once by enhancing translation initiation and improving solubility to reach high activity and productivity for recombinant protein expression in *E. coli* [87]. However, the abilities for solubility enhancement are different for the N-terminus fusion tags [142]. For the same fusion tag, the effect depends on the fused protein's properties, generally on the size of the target protein. The larger target protein (e.g., above 25 kDa), the less amount of soluble protein even though the total expression levels are unchanged. Many published papers show the comparisons about the solubility capacities between the fusion tags for same target protein and/or for different proteins [72, 138, 142, 143]. For example, Hammarstrom *et al.*, overexpressed 43 genes for eukaryotic proteins with four fusion tags (e.g., His, TRX+His, GST+His, GBI+His) in *E. coli* resulted that 14 of them were predominantly insoluble, six predominantly soluble with all four fusion tags, indicating the complexity of heterologous proteins expressed in *E. coli* [138].

The properties of the N-terminus fusion tags used in this study for expression of PalB are summarized in Table 2.4. His-tag and GST tag are commonly used as purification purpose, not expecting the solubility enhancement [72, 144, 145]. Tags of MBP, NusA, and TRX earned very high reputations for improving the solubility and enhancing the stability and bioactivity of the fused proteins [93, 94, 137, 139, 140, 146]. Thioredoxin (TRX), [94], DsbA and DsbC [72], are used specifically for improving disulfide bonds formation.

The larger fusion tags can be more efficient for improving solubility, but generate more metabolic burden. In contrast, smaller fusion tags could be directly used without being removed if the fused protein is an enzyme, but may not offer much for the solubility enhancement [86].

Table 2.4 The properties of the N-terminal fusion tags used in this study

Name	Advantage*	Disadvantage*	Molecular weight (kDa)	Purification method	Reference
6×His	A-, C+	F-, D-	0.84	Ni-NTA	[87]
Strep-II	A-, C+, F+	D-, E+, G+,	1	Strep-Tactin	[87]
DsbA	B+, D+	A+, C-	21		[72]
DsbC		A+, C-	23.4		[72]
FLAG	A-, B+; C+, D+; F+	E+, G+	1	Anti-Flag antibodies	[87]
GST	B+, C+, E-,G-	A+, D-	27.3	Glutathione-Sepharose	[87]
MBP	B+, C+, E-, D+, G-	A+	45	Amylose	[87]
NusA	B+, D+	A+, C-	55		[87]
Skp	A-	C-	21		[147]
T7PK	B+	A+, C-	30		[147]
TRX	B+, D+	C-	14.3	ThioBond	[87]

*Symbols in the table represent for

+: Positive effect; -: Negative effect;

A: Metabolic burden; B: Efficient translation initiation; C: affinity tag;

D: Enhances solubility; E: Expensive affinity resin; F: High specificity;

G: Mild elution conditions;

2.5.3 Coexpression of Cytoplasmic Chaperone to Mediate the Misfolded Protein in the Cytoplasm

As mentioned before, the cytoplasm is a really crowded compartment. About 10~20% of newly synthesized polypeptides require interaction with the molecular chaperones to prevent their misfolding and aggregation, and to promote their folding [135]. The well-characterized cytoplasmic chaperones are TF, DnaK/J-GrpE, and GroEL/ES (Figure 2.3) [43, 148, 149].

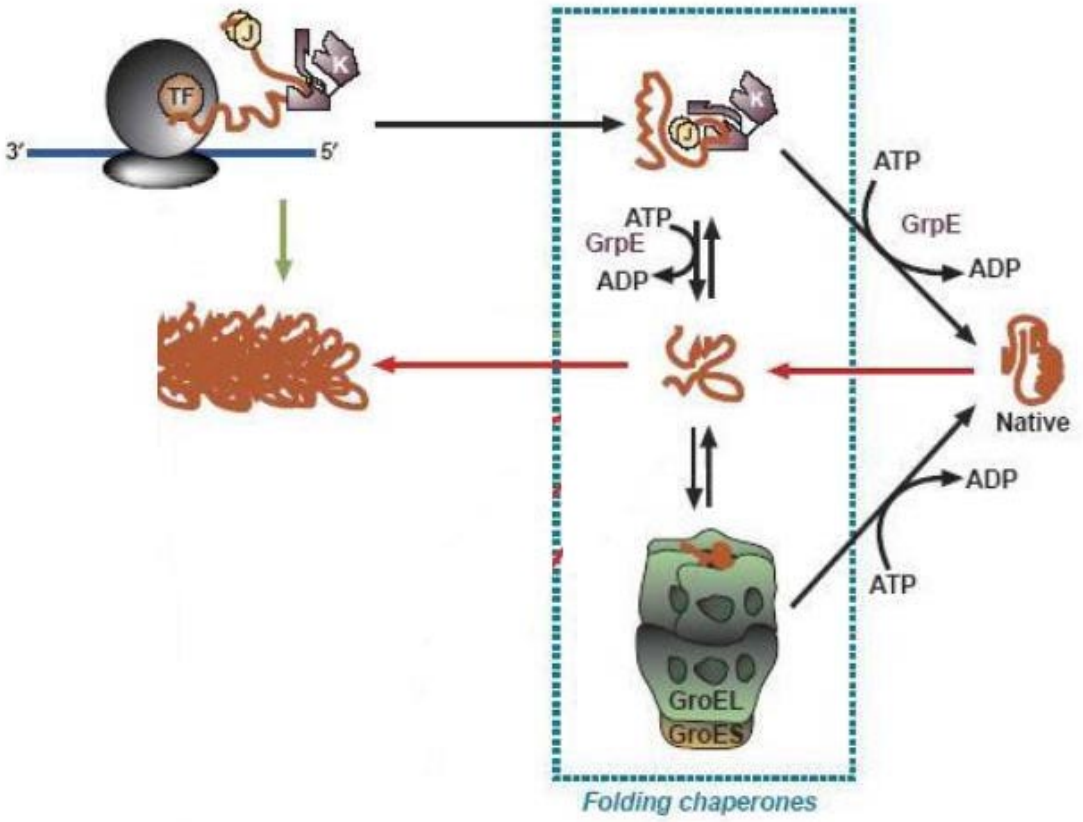


Figure 2.3 Cytoplasmic chaperone-assisted protein folding in the cytoplasm of *E. coli* (modified after[43]). The newly synthesized polypeptides require the assistance of TF or DnaK/J-GrpE to avoid the solvent and each other *in vitro*. A native conformation could be reached after releasing from TF and DnaK. The misfolded proteins are transferred to GroEL for unfolding. Some of them will be back to the productive pathway, others will be degraded.

TF is composed of three domains: an N-terminal domain for association with the large ribosomal subunit, a central substrate binding and peptidyl-prolyl-*cis-trans*-isomerase (PPIase) domain, and an unknown function C-terminal domain [150-152]. A recent paper reported that all of three domain of TF interact with nascent chains during translation [153]. TF is the first chaperone that interacts moderately and stabilizes elongating nascent peptide chain protruding from 50S ribosome in a non-aggregated state [152, 154]. In contrast to DnaK/J-GrpE and GroEL/ES, TF is not a heat-shock induced chaperone. Deletion of the *tig* gene encoding TF does not impair cell growth [155].

DnaK (Hsp70), DnaJ (Hsp40) (involved in substrate binding and presentation) and GrpE, (a nucleotide-exchange factor) always collaborate together as DnaK/J-GrpE because overexpression of DnaJ does not affect protein folding, but overexpression of DnaK in the absence of DnaJ may result in cell death [156]. DnaK/J-GrpE cooperated in folding of newly synthesized proteins and possessed overlapping substrate pools and binding specificities with TF [157, 158].

The polypeptide chains could be folded after released from TF or DnaK/J-GrpE, or are transferred to GroEL/ES [43, 154]. Together with GroES, GroEL (Hsp60) provides a physically defined compartment in which interacts hydrophobically with the captured an aggregation-prone intermediates or misfolded proteins and facilitates their unfolding or refolding in an ATP-dependent manner while they are sequestered from the cytoplasm [149, 159].

Coexpression of TF and GroEL/ES or DnaK/J-GrpE and GroEL/ES form a cooperative team and apply a synergetic effect on folding intermediates or at later folding stages and maximizing recovery of the target protein in a soluble form [160, 161]. It was reported that coexpression of TF and GroEL/ES is more effective [76] because on the basis of further association with GroEL, TF can strengthen its function for facilitating protein folding through efficient binding to GroEL substrates [157, 162, 163].

It was observed in penicillin acylase (PAC) studies [68] that all three cytoplasmic chaperones enhanced the proPAC solubilization process. TF not only had the most prominent solubilization effect but also significantly increased PAC activity, suggesting that TF might be helpful for preserving the specific folding state for PAC maturation [164]. It appeared difficult for the misfolded proPAC to be reshifted back into the productive pathway for PAC maturation, even though it could be solubilized by DnaK/J-GrpE or GroEL/ES possibly due to the failure in recovering the critical folding state (mediated by TF) of newly synthesized proPAC. In the long-term run experiment (50 h after induction in a batch fermenter) of the same study, the effect of proPAC solubilization was rather prominent upon coexpression of DnaK/J-GrpE, particularly in the exponential phase. Both soluble proPAC and mature PAC were protected from

intracellular proteolysis by DnaK/J-GrpE, and GroEL/ES, but not TF, particular during the stationary phase. The chaperone team of TF and GroEL/ES acted collaboratively and applied a significant effect in facilitating the proper folding of proPAC, protecting proPAC from proteolysis, and improving PAC maturation in the cytoplasm [67].

Another function, at least for most, if not all, of chaperones is for responding to cell stress. The amount of a native protein reflects an equilibrium status of protein synthesis, folding and stability. Stress, from overexpression of protein, or elevated temperature, causes an imbalance in this equilibrium, resulting cell control system initiating stress response. The heat shock response controls levels of chaperones and proteases to ensure a proper cellular environment for protein folding. In the cytoplasm of *E. coli*, this response is mediated by the classic heat-shock regulon σ^{32} (or σ^{H} or RpoH), a heat-shock transcription factor regulating the expression of many cytoplasmic heat-shock gene and its production is only induced through the σ^{32} pathway [165]. DnaK chaperone regulates both the amount and activity of σ^{32} [166-170]. DnaK cooperated with ClpB to resolubilize the protein aggregation, and mediate them back to the productive pathway [171]. GroEL/ES is an additional regulator of σ^{32} [172]. The degradation of aggregates also requires chaperones[136], e.g., DnaK, probably function in substrate recognition [113], or GroEL/ES, for maintaining substrates in a soluble easily digestible state [135, 162, 163].

2.5.4 Coexpression of Periplasmic Folding Factor to Assist Recombinant Protein Expression in the Periplasm of *E. coli*

Periplasmic folding factors function in an ATP-independent manner, in contrast to their cytoplasmic counterparts, because ATP is not present in the periplasm. It has been reported that several folding factors interact with secreted proteins for functional expression [88, 96, 173, 174]. There are more than 10 different folding factors characterized in the periplasm, but here only the folding factors used in this study will be reviewed, including DegP, FkpA, SurA, Skp, DsbA, and DsbC (Table 2.5 and Figure 2.4).

Table 2.5 Properties of the periplasmic folding factors and the cytoplasmic chaperones used in this study

Protein	No. of amino acids of mature protein	Molecular weight of monomer (kDa)	Oligomeric state	pI	PDB code
Periplasmic folding factor					
DegP	443	46	Hexamer	7.9	1ky9
Skp	180	21	Trimer	11.9	1sg2, 1u2m
SurA	408	45	Monomer	6.1	1m5y
FkpA	245	26	Dimer	6.7	1q6u, 1qq6h, 1q6i
DsbA	181	21	Monomer	5.4	2b3s, 1a2i
DsbC	223	23	Dimer	5.9	1jzo
Cytoplasmic chaperone					
TF		48	monomer		1i1p, 1oms
DnaK-J/GrpE		72			1q5i, 1bpr, 1bqz
GroEL/GroES		60			1kpo, 1fy9

PDB: protein data bank

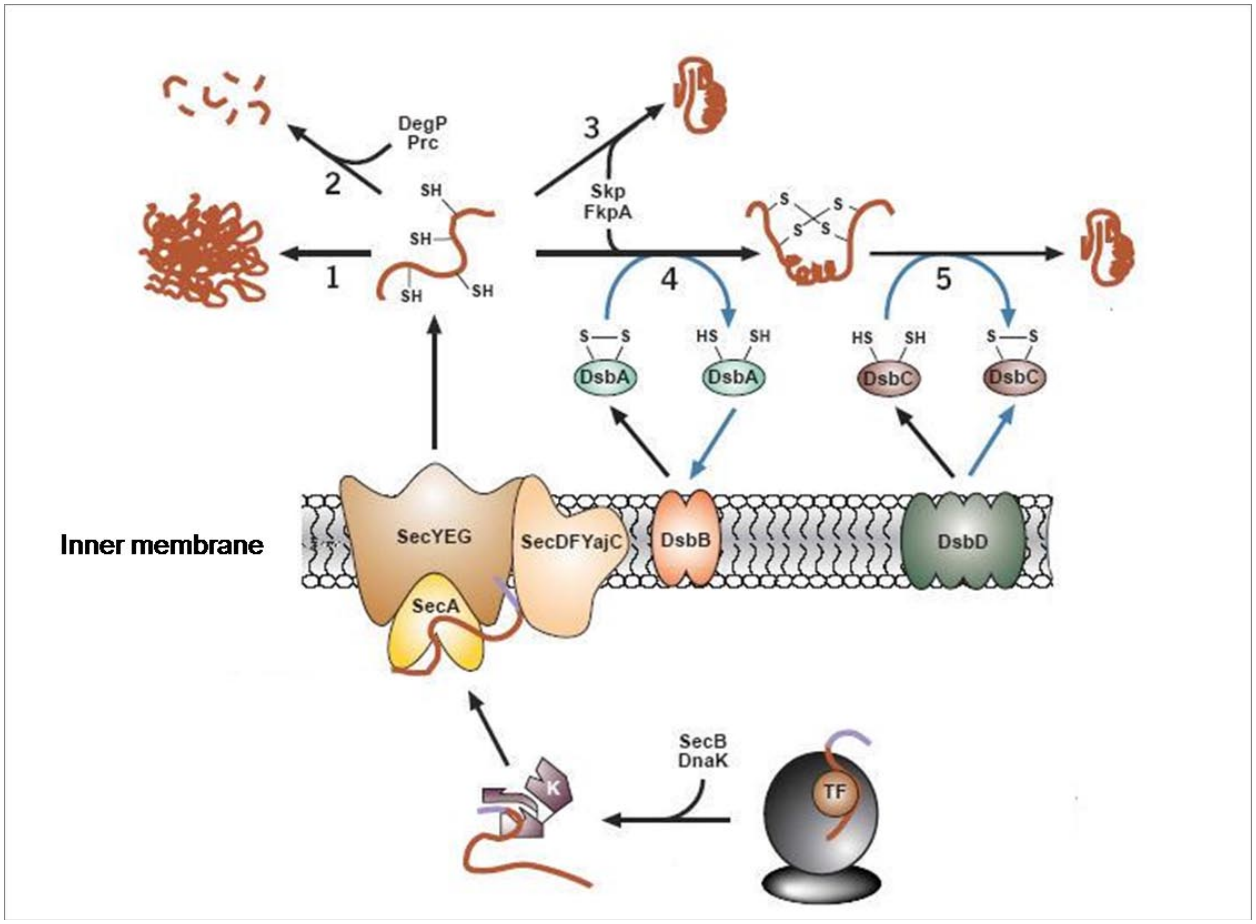


Figure 2.4 Export and periplasmic folding pathways in *E. coli* (modified after[43]). Preproteins with less hydrophobic signal sequence (lavender) undergo Sec-dependent export. After cleavage the signal sequence, partially folded periplasmic proteins may aggregate (1), undergo proteolysis (2) or reach a native conformation, possibly with the aid of folding factors (3), (4) and (5).

There are two different stress response pathways (σ^E and Cpx) in the envelope space of *E. coli*, but overlapped in both regulations and functions [174-177]. For the folding factors used in this study, *degP*, *skp* are under the control of both the σ^E and Cpx [174, 178]; *fkpA* is under the control of σ^E [120, 179]; *surA* is not a member of σ^E or Cpx regulons, but lack of SurA activates σ^E [173]; *dsbA* is under the control of Cpx [174], and *dsbC* is under control of σ^E [178]. Among these folding factors, Skp is the only chaperone. DegP possesses protease or chaperone activity. FkpA and SurA are PPIases and chaperones, but they belong to different PPIase. DsbA and DsbC are catalysts for disulfide bond formation and

isomerization, and may have chaperone activities. In addition, the functions of DsbA and DsbC are overlapped in certain situations.

2.5.4.1 DegP

DegP is essential for *E. coli* survival in high temperature [180, 181]. DegP also degrades transiently denatured or unfolded proteins which accumulate in the periplasm following heat shock or other stress conditions, and/or newly secreted proteins prior to folding and disulfide bond formation [182]. With temperature shift a downward, DegP changes its role from protease to chaperone [182, 183]. DegP function in the periplasm could be similar to that of TF in the cytoplasm, namely, having an interaction with newly synthesized (or newly translocated) protein, since DegP is often located on the side of the cytoplasmic membrane facing the periplasm [184]. DegP deficient strain is useful for the protease-sensitive recombinant proteins production [185, 186]. In contrast, DegP is required and even need to be coexpressed with the target protein. For example, for efficient export of overexpressed alkaline phosphatase (PhoA), DegP functioned as a 'translocation channel cleaner' which degrades misfolded proteins accumulated near the periplasmic side of the channel upon extracytoplasmic stresses (e.g., gene overexpression resulting overloading on protein export machinery) [184].

In this lab, the results from penicillin acylase (PAC) studies showed DegP protease activity was primarily accountable for the enhancement of the periplasmic processing for PAC maturation. While the culture performance for production of PAC, particularly the volumetric PAC activity, was improved by $degP_{S210A}$ coexpression, the mutant DegP (only having chaperone activity without protease activity) could not suppress several negative cellular responses caused by *pac* overexpression (e.g., inclusion body formation, growth arrest, and cell lysis) as effectively as the wild-type DegP could. The results suggest that both protease and chaperone activities are required for DegP to effectively improve the production of PAC [187, 188].

2.5.4.2 FkpA

FkpA is another periplasmic heat-shock protein with both PPIase and chaperone activities [189-192]. It is reported that FkpA assisted MalE31 formation [189], and increased antibiotic fragment expression dramatically by interacting with early folding intermediate for preventing their aggregation, or by reacting with aggregated protein, and unfolding and releasing them on a productive pathway [193, 194]. FkpA was used in penicillin acylase (PAC) study in this lab, and found that it neither reduced the amount of PAC inclusion bodies nor increased the *pac* expression level. However, coexpression of FkpA released cell

stress and improved cell physiology upon *pac* overexpression, so the overall productivity was increased [195].

2.5.4.3 Skp

Skp is a characterized periplasmic chaperone [196, 197] that interacts with unfolded proteins translocated into the periplasm via Sec secretion system [179] for generating and maintaining the solubility of early folding intermediates [198]. Skp plays a key role in proper biogenesis of the envelope by mediating the folding and targeting of several outer membrane proteins to the membrane [199-202]. On the other hand, Skp has been shown to improve the folding of recombinant proteins or antibody fragments in the periplasm of *E. coli* [111, 203-205].

2.5.4.4 SurA

As a more important component in the biogenesis of OMPs, SurA possesses PPIase and chaperone activity. Recent study showed its main function is as the primary chaperone responsible for the periplasmic transit of the bulk mass of OMPs to the YasT complex [206] or even to later OMPs maturation [207]. Its substrate specificity relates to the certain secreted OMPs [208] and to the chaperone domain, not PPIase [209].

2.5.4.5 DsbA and DsbC

While all above are general folding factors for assisting recombinant protein expression in the periplasm of *E. coli*, DsbA and DsbB, as oxidoreductases, specifically mediate the recombinant protein with multiple disulfide bonds [40, 210]. DsbA and DsbC are well-characterized members of Dsb-family [40, 211]. DsbA is essential for direct bridging of disulfide bonds in newly translocated proteins and DsbC assists subsequent rearrangement of aberrant disulfide bonds because the newly formed disulfide bonds can be formed among incorrectly paired cysteines, trapping substrate proteins in a misfolded conformation [40]. Recent studies revealed that the overexpression of Dsb proteins increased secretion efficiency, folding, and the solubility of recombinant proteins in the periplasmic space [212-214]. Although they have defined roles in the formation of disulfide bonds, DsbA and DsbC might share certain biological functions in some particular conditions [104, 215-217]. In addition, DsbC may also have chaperone activity [218].

The redundant periplasmic chaperones actually do function in parallel or work as a team. For example, the role of DegP/Skp was amplified in the absence of SurA, but in the normal situation, DegP and Skp worked for rescuing OMPs that fall off the SurA pathway [206]. SurA and FkpA are helpful for

disulfide bond formation [219]. DsbA and DegP have overlapping effect on the MalS folding process [183].

2.6 The Recombinant PalB Expression System

Lipase (E.C.3.1.1.3) belongs to the α/β-hydrolase-fold superfamily (E.C.3.X.X.X) which contains enzymes to catalyze reactions such as hydrolysis ester, thioesters, peptides, epoxides, and alkyl halides or cleavage of carbon bonds in hydroxynitriles. PalB is built up of 317 amino acids and the molecular weight is 33 kDa without glycosylation [220] (Table 2.6). It contains the catalytic triad Ser^{105}-His^{224}-Asp^{187}, common to all serine hydrolases (Figure 2.5) [220]. There are six cysteine residues in the sequence of PalB and all involved in disulfide bonds, Cys^{22}-Cys^{64}, Cys^{216}-Cys^{258}, and Cys^{293}-Cys^{311} for stabilizing the N-terminal, the surface, and the C-terminal respectively. The central β–sheet is composed of seven strands of which the last six are parallel, most connections between strands are formed by the right handed β-α-β structural motif (Figure 2.5) [221, 222]. PalB has a large hydrophobic surface surrounding the entrance of the active site channel, so that it has no interfacial activity, which means PalB activity in the solvent is similar to that in water.

Table 2. 6 Amino acid sequence of the mature PalB [226]

1	LPSGSDPAFS	QPKSVLDAGL	TCQGASPSSV	SKPILLVPGT	GTTGPQSFDS
51	NWIPLSAQSG	YTPCWISPPP	FMLNDTQVNT	EYMVNAITAL	YAGSGNNKLP
101	VLTWSQGGLV	AQWGLTFFPS	IRSKVDRLMA	FAPDYKGTVL	AGPLDALAVS
151	APSVWQQTTG	SALTTALRNA	GGLTQIVPTT	NLYSATDEIV	QPQVSNSPLD
201	SSYLFNGKNV	QAQAVCGPLF	VIDHAGSLTS	QFSYVVGRSA	LRSTTGQARS
251	ADYGITDCNP	LPANDLTPEQ	KVAAAALLAP	AAAAIVAGPK	QNCEPDLMPY
301	ARPFAVGKRT	CSGIVTP			

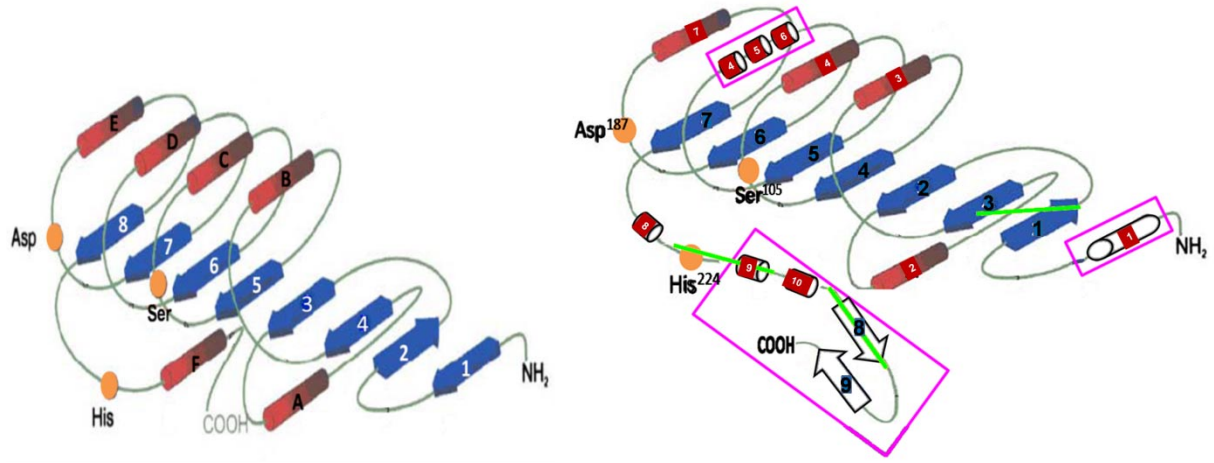
Table 2.7 Properties of PalB on an Amino Acid Base (given by Vector NTI advance™, Invitrogen)

Amino Acid(s)	Number count	% by weight	% by frequency
Charged (RKHYCDE)	51	19.85	16.09
Acidic (DE)	18	6.34	5.68
Basic (KR)	17	7.01	5.36
Polar (NCQSTY)	105	34.40	33.12
Hydrophobic (AILFWV)	116	36.32	36.59
A Ala	37	8.53	11.67
C Cys	6	1.88	1.89
D Asp	14	4.82	4.42
E Glu	4	1.52	1.26
F Phe	10	4.27	3.15
G Gly	26	5.05	8.20
H His	1	0.40	0.32
I Ile	11	3.73	3.47
K Lys	9	3.40	2.84
L Leu	30	10.18	9.46
M Met	4	1.54	1.26
N Asn	14	4.78	4.42
P Pro	30	8.93	9.46
Q Gln	18	6.81	5.68
R Arg	8	3.60	2.52
S Ser	32	8.70	10.09
T Thr	26	8.01	8.20
V Val	23	6.97	7.26
W Trp	5	2.64	1.58
Y Tyr	9	4.22	2.84
B Asx	28	9.60	8.83
Z Glx	22	8.33	6.94
X Xxx	0	0.00	0.00

Genetic code used: Standard

Amino Acid(s)	Codon(s) Used
A Ala	GCA(12), GCC(10), GCG(10), GCT(5)
C Cys	TGC(5), TGT(1)
D Asp	GAC(10), GAT(4)
E Glu	GAG(4)
F Phe	TTC(7), TTT(3)
G Gly	GGA(2), GGC(9), GGG(1), GGT(14)
H His	CAT(1)
I Ile	ATC(10), ATT(1)
K Lys	AAA(2), AAG(7)
L Leu	CTA(1), CTC(16), CTG(8), CTT(5)
M Met	ATG(4)
N Asn	AAC(13), AAT(1)
P Pro	CCA(4), CCC(17), CCG(4), CCT(5)
Q Gln	CAA(2), CAG(16)
R Arg	AGG(2), CGA(3), CGC(2), CGT(1)
S Ser	AGT(3), TCA(2), TCC(12), TCG(14), TCT(1)
T Thr	ACA(2), ACC(20), ACG(3), ACT(1)
V Val	GTA(2), GTC(12), GTG(6), GTT(3)
W Trp	TGG(5)
Y Tyr	TAC(8), TAT(1)

A1 **A2**



B

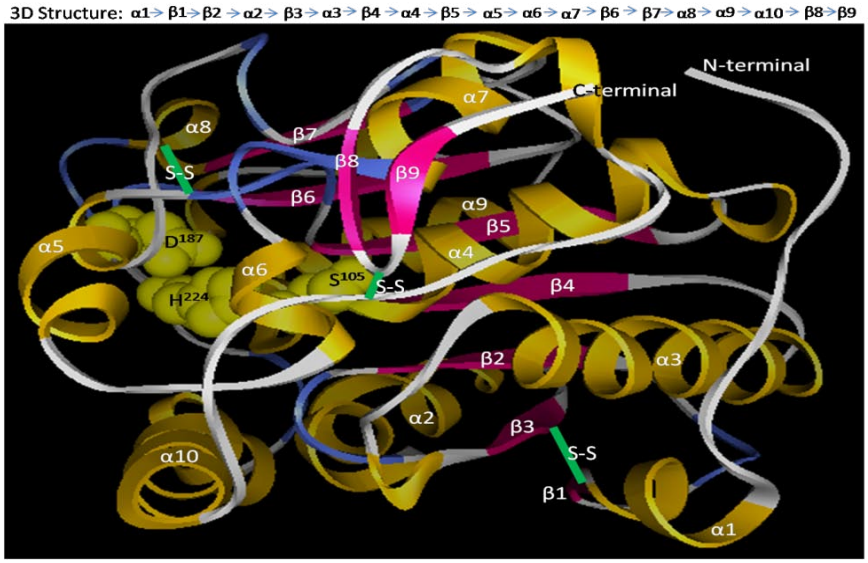


Figure 2.5 The structure diagram of α/β-hydrolase fold of lipase. Panel A is the secondary structure of a general lipase (Panel A1) and PalB (Panel A2). α-helix and β-sheet are drawn as red cylinder and blue arrows respectively. The central β-sheet is composed of six are parallel, and most connections between strands are formed by the right handed β-α-β structural motif. The amino acids (Ser, Asp, His) belonging to the catalytic triad are shown in orange solid oval. The big differences of PalB with general lipase are in the pink boxes, which might be the cause for the specific PalB characteristics. Three disulfide bonds (green bar) in PalB are for stabilizing the N-terminal, the surface, and the C-terminal respectively. Panel B is the three dimensional structure of PalB.

Due to the unique properties of PalB in catalyzing hydrolysis of triacylglycerols, and transesterification and other organic synthetic reactions to produce new kinds of lipids [223, 224], it is widely used in organic chemistry (e.g., cosmetics, food) [225, 226]. However, the PalB product from the original microbe is really low. Previously, recombinant PalB was expressed heterologously in *Aspergillus oryzae* [25], *Pichia pastoris* [26], and *Saccharomyces cerevisiae* [27, 227]. Intensive studies of PalB have been focused on characterizing the structures [221, 222, 225, 228], investigating thermal stability, enantioselectivity and activity [223, 225, 226, 229]. and improving PalB activity by applying fusion tag (e.g., cellulose-bind domain) [26] and mutagenesis [227, 230, 231].

PalB is an attractive enzyme for biodiesel production for catalyzing the transesterification reaction of vegetable oil or animal fat with methanol. It is inspired and motivated to investigate the feasibility of *E. coli* as host because of its capability of massive production of recombinant protein. First, the concern of pathogenicity, or endotoxin of this host is not a problem for PalB quality used as biocatalyst in biodiesel production. Second, although the expression performance stands improvement, the functional expression of PalB in *E. coli* [28, 29] sparks a light. The glycosylation is not essential for PalB bioactivity [28, 29] and disulfide bond formation can be mediated by biochemical and genetic means.

A major technical issue limiting the PalB expression level in *E. coli* is associated with its intracellular misfolding. PalB is a cold active lipase destined to be secreted extracellularly in the original psychrophilic and eukaryotic source of *P. antarctica*. Consequently, the intracellular overexpression might prevent it from being structurally adaptable to the mesophilic and prokaryotic expression system of *E. coli*, resulting in protein misfolding. In addition, wild-type PalB has three intermolecular disulfide bonds (Figure2.5) potentially associated with its folding process though they might not be strictly required for developing the bioactivity. Finally, cell growth is often arrested due to deteriorated cell physiology associated with high-level expression and misfolding of PalB.

Considering the PalB's application purpose and the benefit of different expression systems, a systematic study of recombinant PalB expression in *E. coli* system was focused on, including promoter, expression vector in different compartments, fusion tag, chaperone, folding factor, and cultivation condition.

Chapter 3*

Identification of Factors Limiting Heterologous production of Lipase in the Cytoplasm of *Escherichia coli*

3.1 Introduction

The bacterium *Escherichia coli* remains popular as a workhorse for recombinant protein production. Formation of active proteins involves a series of intracellular steps for gene expression (e.g., transcription and translation) and posttranslational processing (e.g., folding, translocation and others). Strategies based on boosting gene expression steps are often adopted to enhance recombinant protein production [31, 38, 92]. However, it is not uncommon that the overexpressed gene products cannot be suitably processed during posttranslational processing, resulting in the formation of insoluble protein aggregates known as inclusion bodies [43, 232]. Theoretically, the precursors, intermediates, or final gene products can possibly form inclusion bodies in the cytoplasm and/or periplasm particularly during the course of gene overexpression. This brings up a technical issue that, in addition to improving the efficiency of each gene expression step, a 'balanced' protein synthesis flux throughout these steps should be properly maintained to avoid the accumulation of any protein species along the protein formation pathway.

Lipases (triacylglycerol acylhydrolase, E.C.3.1.1.3) are enzymes that catalyze the synthesis or hydrolysis of glycerol esters (e.g., triglycerides), such as fats and oils [14]. Microbial cells, including bacteria, fungi, and yeasts, are the major source for lipases with various enzymatic characteristics, such as synthetic and hydrolytic activity, stability, substrate specificity, for numerous applications in organic chemistry and biotechnology [20, 21]. Our interest in lipase arises from its application on biotransformation as an alternative to chemical conversion for biodiesel production. Lipase B from *Pseudozyma antarctica* (PalB) is a cold active microbial lipase with the transesterification activity and its corresponding immobilized form (e.g., Novozyme 435 commercially available from Novozyme A/S, Bagsvaerd, Demark) has been recognized as the most common biocatalyst for biodiesel production. While the biocatalyst is gaining more attention and has a potential to outperform chemical catalysts, extensive application remains impractical due to its high cost and several technical issues (e.g., inactivation by alcohol). To make the bioprocess economically feasible, there is a motivation to overproduce PalB.

* This chapter is based on a paper "Y. Xu, M. Scharer, M. Moo-Young, C. P. Chou, Identification of Factors Limiting Heterologous Production of Lipase in the Cytoplasm of *Escherichia coli*. 2008." (under review).

Previously, recombinant PalB was expressed heterologously in *Aspergillus oryzae* [25], *Pichia pastoris* [26], and *Saccharomyces cerevisiae* [27]. Given the popularity of *E. coli* as a workhorse for recombinant protein production, functional expression of PalB in this host was not demonstrated until recently [28, 29] and the expression performance stands improvement. A major technical issue limiting the PalB expression level in *E. coli* is associated with its intracellular misfolding. PalB is a cold active lipase destined to be secreted extracellularly in the original psychrophilic and eukaryotic source of *P. antarctica*. Consequently, the intracellular overexpression might prevent it from being structurally adaptable to the mesophilic and prokaryotic expression system of *E. coli*, resulting in protein misfolding. In addition, wild-type PalB has three intermolecular disulfide bonds potentially associated with its folding process though they might not be strictly required for developing the bioactivity. Finally, cell growth is often arrested due to deteriorated cell physiology associated with high-level expression and misfolding of PalB.

In this study, we explored functional PalB expression in the cytoplasm of *E. coli* to gain more insight on relevant technical issues limiting the production of PalB. A selection of cytoplasmic chaperones, including TF, GroEL/ES, and DnaK/J-GrpE, were coexpressed with PalB for investigation of their individual and synergistic effect on the production of active PalB.

3.2 Materials and Methods

3.2.1 Strains and Plasmids

The strains, plasmids and oligonucleotides used in this study are summarized in Table 3.1 and briefly described here. BL21(DE3) was used as the host strain for PalB expression. Molecular cloning was performed according to standard protocols [61] using XL10-Gold® (Stratagene, Dedar Creek, TX, USA) as the cloning host. *P. antarctica* strain ATCC 32657 containing the *palB* source gene was purchased from ATCC (Manassas, VA, USA). *P. antarctica* chromosomal DNA was extracted and purified by using DNeasy® tissue Kit (Qiagen, Valencia, CA, USA). PCR was performed in an automated thermal cycler (GeneAmp® PCR System 9700; Applied Biosystems, Foster City, CA, USA). The PCR product was purified by StrataPrep® PCR purification kit (Stratagene). Restriction enzymes were purchased from New England Biolabs (Ipswich, MA, USA). Plasmid DNAs were purified using a spin column kit (Clontech, Mountain View, CA, USA). Plasmid transformation was carried out using an electroporator (Bio-Rad, Hercules, CA, USA). All the chemicals used in this study were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise specified.

Table 3.1 Strains, plasmids, and oligonucleotides.

Strains, plasmid, and oligonucleotides	Relevant genotype or phenotype^	Source and reference
E. coli		
BL21(DE3)	*E. coli* B, F^- *ompT [dcm][lon] hsdS*$_B$(r_{B^-} m_{B^-}) *gal* (λ cIts857 *ind*1 Sam7 *nin*5 *lacUV5-T7* gene 1)	Novagen
XL10-Gold®	Tet^R Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* Hte [F' *pro*AB *lacI*^{q}ZΔ*M*15 Tn10 (Tet^R) Amy Cm^R]	Stratagene
P. antarctica		
ATCC 32657	The source of *palB* gene	ATCC
Plasmid		
pAR3KJ	P_{araB}::*dnaK/J*::*grpE*, Ori (pACYC184), Cm^R	[233]
pAR3GRO	P_{araB}::*groEL/ES*, Ori (pACYC184), Cm^R	[233]
pETKn-20b(+)	A P_{T7}-expression vector derived from pET-20b(+), Ori (pBR322), Kn^R	This lab [71]
pETKnL	P_{T7}::*palB*, Ori (pBR322), Kn^R	This study
pG-KJE8	P_{araB}::*dnaK/J*::*grpE*, $P_{zt\text{-}lp}$::*groEL/ES*, Ori (pACYC184), Cm^R	T. Yura [76]
pG-Tf2	$P_{zt\text{-}lp}$:: *grpE*::*groEL/ES*::*tig*, Ori (pACYC184), Cm^R	T. Yura [76]

pG-Tf3	P_{araB}::*tig*, $P_{zt\text{-}lp}$::*groEL/ES*, Ori (pACYC184), Cm^R	T. Yura [76]
pTf16	P_{araB}::*tig*, Ori (pACYC184), Cm^R	T. Yura [76]
Oligonucleotide		
P17 and P11	5'-TGCA*TATG*CTACCTTCCGGTTCGG-3' and 5'-*CTGAATTC*TCAGGGGGTGACGATGCCGGAGCAGG-3' primer pair for amplification of *palB*	This study

^ Designed restriction sites are underlined and the introduced mutations are in italic.

The leaderless *palB* gene was PCR-amplified using *Pfu* DNA polymerase (Stratagene), the primer pair of P17 and P11 (Table 3.1), and *P. antarctica* chromosomal DNA as the template. The 970-bp PCR product flanked with *Nde*I and *Eco*RI was purified and cloned into the corresponding restriction sites of pETKn-20b(+) [71], respectively, resulting in a transcriptional fusion of pETKnL (Figure 3.1). DNA sequencing was performed to ensure no mutations occurred during PCR. The coexpression plasmids containing a selection of cytoplasmic chaperone gene(s) were derivatives of pAR3 [233]. pAR3GRO and pAR3KJ [233], and pTf16 [76] contain the gene of *groEL/ES*, *dnaK/J-grpE*, and *tig* (encoding TF), respectively, fused with the *araB* promoter. pG-Tf2 [76] contains the *groES/EL::tig* fused with the *zt-lp* promoter. pG-Tf3 [76] contains *tig* and *groES/EL*, fused with the *araB* and *zt-lp* promoters, respectively. pG-KJE8 [76] contains the *dnaK/J-grpE* and *groES/EL*, fused with the *araB* and *zt-lp* promoters, respectively. All the chaperone-gene-containing plasmids have a pACYC184 replication origin and a chloramphenicol-resistant (Cm^R) marker and are, therefore, compatible with pETKnL which have a pBR322 replication origin and a kanamycin-resistant (Kn^R) marker.

3.2.2 Cultivation

Cells were revived by streaking the stock culture stored at -80 °C on a Luria-Bertani (LB) agar plate (5 g/liter NaCl, 5 g/liter Bacto yeast extract, 10 g/liter Bacto tryptone, 15 g/liter Bacto agar). The plate was incubated at 37 °C for approximately 15 h. An isolated single colony was picked and inoculated with 25 ml of LB medium in an Erlenmeyer flask, which was then incubated on a rotary shaker at 37 °C and 200 rpm for approximately 15 h to form a seed culture. The medium was supplemented with 50 μg/ml kanamycin and/or 34 μg/ml chloramphenicol when necessary. Erlenmeyer flasks containing 25 ml of LB medium were inoculated with 1 ml of seed culture and were incubated in a rotary shaker at 37 °C and 200

rpm. When the cell density reached approximately 0.5 OD_{600}, the culture was chilled on ice and supplemented with isopropyl *β*-D-thiogalactopyranoside (IPTG), arabinose, and/or tetracycline for induction of the gene regulated by the T7, *araB*, and/or *zt-lp* promoters, respectively. After induction, the Erlenmeyer flasks were further shaken for another 4 h at 28 °C or 24 h at 15 °C. All cultivations were conducted at least in duplicate.

3.2.3 Analytical Methods

The culture sample was appropriately diluted with saline solution for measuring cell density at OD_{600} with a spectrophotometer (DU 520; Beckman Coulter, Fullerton, CA, USA). For the preparation of the cell extract, cells at 20 OD_{600} units (defined as $OD_{600} \times$ ml) were centrifuged at 4 °C and 12,000 × g for one min. The cell pellet was resuspended in 0.75 ml of sodium phosphate buffer (PB buffer, 0.05 M, pH 7.5). The cell suspension was sonicated for 4 min (0.5 s/0.5 s pulse on/off) using an ultrasonic processor (Misonix, Farmingdale, NY, USA) and then centrifuged at 4 °C and 12,000 × g for 15 min. The supernatant containing soluble proteins was assayed for PalB activity. The pellet containing insoluble proteins and cell debris was washed with PB buffer, resuspended in TE-SDS buffer (10 mM Tris HCl, pH 8.0, 1 mM EDTA, 1% sodium dodecyl sulfate), and heated to 100 °C for 5 min. The supernatant and the protein content of the pellet were analyzed as the soluble and insoluble fractions, respectively.

PalB activity was qualitatively evaluated by the size and the transparency of the halo formed on a tributyrin agar plate (LB agar plate containing 1% emulsified tributyrin). The soluble fraction of cell lysate at 25 µl was loaded on the plate and incubated at 37 °C for overnight. In addition, PalB was assayed at 37 °C using *p*-nitrophenyl palmitate (pNPP) as a substrate [234]. The 0.5 mM pNPP solution was prepared by emulsifying 5 mg pNPP into 1 ml of *β*-propanol by sonication, and adding it dropwise into 25 ml of 50 mM Tris HCl (pH 8.0) and 0.1% Triton X-100 solution along with the emulsification. The enzymatic reaction was started by adding lysate sample (20~100 µl) to the pNPP solution. The amount of enzymatic reaction product of *p*-nitrophenol (pNP) was quantified colorimetrically with the absorbance at 410 nm (A_{410}). One unit was defined as the amount of enzyme that hydrolyzed pNPP to produce 1 µmole of pNP per min at 37 °C. The reaction was carried out in a shaker at 37 °C and 200 rpm. All assays were conducted in duplicate.

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed in a Mini-PROTEIN®III electrophoresis cell (Bio-Rad) using a 12.5% polyacrylamide separating gel stacked with a 4% polyacrylamide stacking gel. Protein samples of the cell extract from cells of 0.008 OD_{600} unit for the soluble fraction and 0.005 OD_{600} unit for insoluble fractions, respectively, were loaded for SDS-PAGE

and Western blotting analysis. Electrophoresis was conducted under a constant voltage of 200 V for 45 min. The gel was stained with Coomassie blue and dried in a hood. The dried gel was then scanned.

To conduct Western blotting, proteins on the polyacrylamide gel were electroblotted to a polyvinylidene difluoride membrane using a Mini Trans-Blot cell (Bio-Rad) according to a standard protocol [235]. The electrophoretic transfer was conducted at a constant voltage of 100 V for 1 h. Primary anti-PalB antibodies were raised in a rabbit intermittently immunized with the antigen of recombinant PalB expressed from *A. oryzae* (Sigma-Aldrich). It was further purified by SDS-PAGE and PalB used for immunization was obtained by polypeptide elution of the corresponding band in the polyacrylamide gel slice using an Electro-Eluter (model 422; Bio-Rad). The secondary antibody was goat anti-rabbit immunoglobulin G conjugated with horseradish peroxidase (HRP). PalB-related polypeptides were probed by a colorimetric method using 3,3´-diaminobenzidine tetrahydrochloride (DAB) as the substrate. The processed membrane was scanned.

3.3 Results

3.3.1 Functional Expression of PalB in BL21(DE3) Harboring pETKnL

Figure 3.1 shows the construction of pETKnL both containing the leaderless *palB* gene from *P. antarctica* ATCC 32657 and the expression of which was under the regulation of the T7 promoter. With this design, the gene product is expected to contain an extra methionine in the N terminus and, presumably, this won't affect PalB activity. Various recombinant BL21(DE3) strains, either single-plasmid or binary-plasmid, were cultivated under designated culture conditions for PalB expression. Qualitative visualization of PalB activity was conducted using tributyrin plates and the results are summarized in Figure 3.2. Halos were developed for all the lysate samples corresponding to various BL21(DE3) strains containing pETKnL except the two control strains [e.g., BL21(DE3) and BL21(DE3) harboring pETKn-20b(+)], indicating PalB was functionally expressed in the cytoplasm of *E. coli*. In general, the area and transparency of the halo corresponding to the lysate sample potentially correlate with the expressed PalB activity. However, the expressed PalB cannot be secreted since halos were not observed for the whole cells growing on tributyrin plates (data not shown).

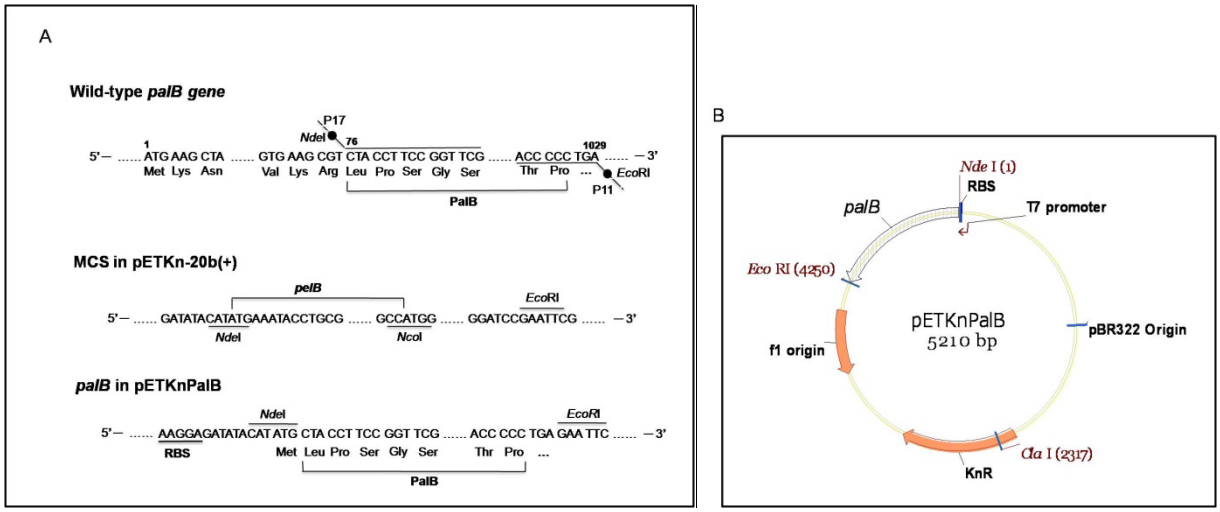
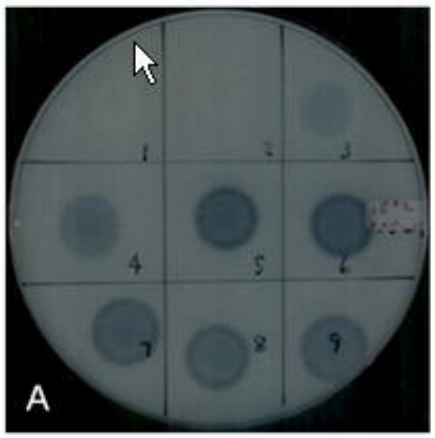
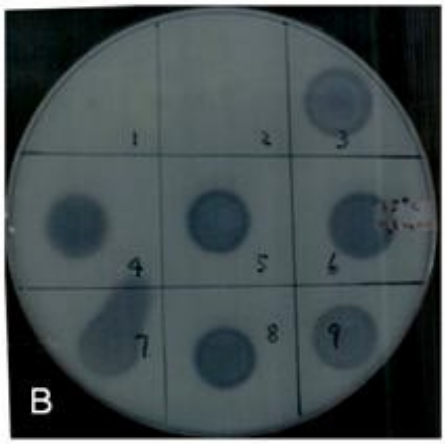


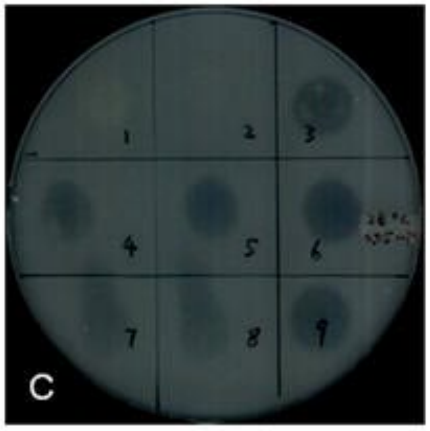
Figure 3.1 Construction of *palB* expression plasmid of pETKnL. Panel A: The DNA region from nucleotides 76 to 1029 of the wild-type *palB* gene on the chromosome of *P. antarctica* strain ATCC 32657 was PCR-amplified with primer pair P17 and P11. The PCR product was cloned into pETKn-20b(+) using *Nde*I and *Eco*RI to form a transcriptional fusion of pETKnL. Due to the lack of a signal peptide, pETKnL is used for cytoplasmic expression of PalB and the heterologously expressed PalB has an extra methionine at the N-terminus. MCS, multiple cloning sites. RBS, ribosome binding site. Panel B: pETKnL contains the leaderless *palB* gene in-frame fused with the T7 promoter. It has a pBR322 replication origin and a kanamycin-resistant (Kn^R) marker and is compatible with various chaperone-coexpression plasmids with a pACYC184 replication origin and a chloramphenicol-resistant (Cm^R) marker.



0.05 mM IPTG, 15 °C



0.1 mM IPTG, 15 °C



0.05 mM IPTG, 28 °C

No	Expression system
1	BL21(DE3)
2	BL21(DE3)(pETKn-20b(+))
3	BL21(DE3)(pETKnPalB)
4	BL21(DE3)(pETKnPalB, pTf16)
5	BL21(DE3)(pETKnPalB, pG-Tf2)
6	BL21(DE3)(pETKnPalB, pG-Tf3)
7	BL21(DE3)(pETKnPalB, pG-KJE8)
8	BL21(DE3)(pETKnPalB, pAR3KJ)
9	BL21(DE3)(pETKnPalB, pAR3GRO)

Figure 3.2 Qualitative visualization of PalB activity using the tributyrin plate. Preparation of the plate for halo development is described in Materials and Methods. The area and intensity of the halo corresponding to the lysate sample correlates with the PalB expression level. Cultivation conditions are: 0.05 mM IPTG, 15 °C (Panel A), 0.1 mM IPTG, 15 °C (Panel B), and 0.05 mM IPTG, 28 °C (Panel C). Additional 5 g/l of arabinose and/or 0.02 mg/l of tetracycline were supplemented for induction of the *araB* and/or *zt-lp* promoters that were used for regulation of chaperone coexpression (see Table 1). Lysate samples from the following expression systems were used for the test: (1) BL21(DE3); (2) BL21(DE3) (pETKn-20b(+)); (3) BL21(DE3) (pETKnL); (4) BL21(DE3) (pETKnL, pTf16), (5) BL21(DE3) (pETKnL, pG-Tf2); (6) BL21(DE3) (pETKnL, pG-Tf3); (7) BL21(DE3) (pETKnL, pG-KJE8); (8) BL21(DE3) (pETKnL, pAR3KJ); (9) BL21(DE3) (pETKnL, pAR3GRO).

Using BL21(DE3) harboring pETKnL, the results of cultivation performance for PalB expression under various culture conditions and their corresponding Western blotting analysis are summarized in Figure 3.3 and Figure 3.4, respectively. Apparently, heterologous PalB expression favored the non-productive pathway for inclusion body formation. Cell growth was arrested upon the induction for PalB expression and such deterioration in cell physiology could be caused by the accumulation of inclusion bodies, particularly in the case of 28 °C. Comparing the two expression runs with 0.05 mM IPTG at 15 °C and 28 °C, while the PalB solubility was somewhat improved at a lower temperature (Figure 3.4), the specific PalB expression level decreased, resulting in a lower total PalB activity (Figure 3.3). Though increasing the IPTG concentration to 0.1 mM could boost the PalB expression level at 15 °C, cells appeared to suffer more growth arrest due to an increase amount of PalB inclusion bodies. Nevertheless, the results demonstrated the feasibility of using a relatively high cultivation temperature at 28 °C for functional expression of the cold active PalB in *E. coli*.

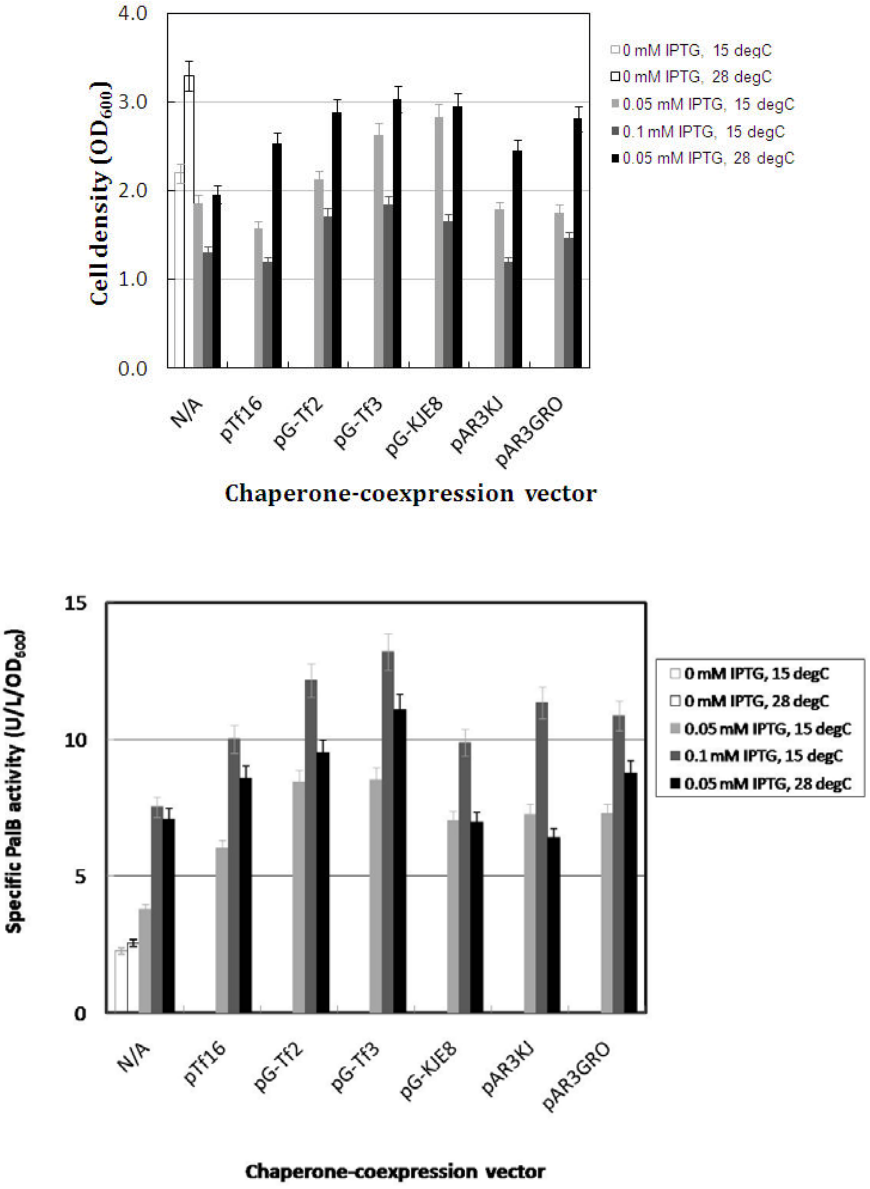


Figure 3.3 Cultivation performance for PalB expression for various host/vector systems under different culture conditions. Cell density (Panel A) and specific PalB activity (Panel B) are presented. BL21(DE3) and pETKnL were used as PalB expression host and vector, respectively (e.g., group N/A). Various chaperones were also coexpressed in different groups: "TF" for group pTf16; "TF and GroEL/ES" for groups pG-Tf2 and pG-Tf3; "DnaK/J-GrpE and GroEL/ES" for group pG-KJE8; "DnaK/J-GrpE" for group pAR3KJ; "GroEL/ES" for group pAR3GRO. The error bar represents a range of the data variance.

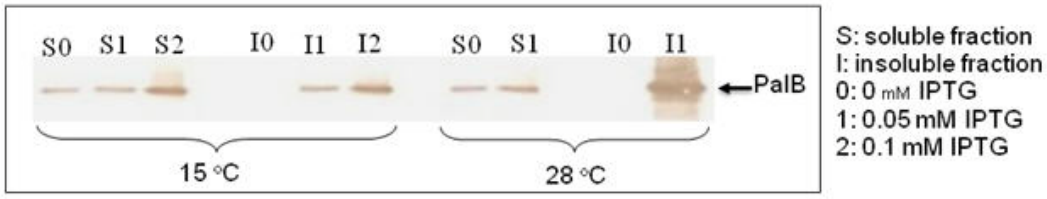


Figure 3.4 Western blotting analysis of soluble (S) and insoluble (I) fractions of BL21(DE3) harboring pETKnL under various culture conditions for PalB expression: (0) 0 mM IPTG, (1) 0.05 mM IPTG, and (2) 0.1 mM IPTG.

3.3.2 Effect of Chaperon Coexpression on PalB Expression

Since protein folding appeared to be the major factor limiting PalB expression, a selection of cytoplasmic chaperones, including TF, GroEL/ES, and DnaK/J-GrpE, were coexpressed and the results are summarized in Figure 3.3 and Figure 3.5. Coexpression of a single chaperone (e.g., TF from pTf16, DnaK/J-GrpE from pAR3KJ or GroEL/ES from pAR3GRO) had almost no effect on alleviating growth inhibition at 15 °C, but a slight improvement at 28 °C. On the other hand, the specific PalB activity was significantly increased at 15 °C, but the expression level remained more or less the same at 28 °C. Note that, though the total PalB expression (e.g., the sum of soluble and insoluble fractions) was significantly increased upon coexpression of TF, most of the expressed PalB went to the non-productive pathway to form inclusion bodies. This resulted in an even higher amount of inclusion bodies compared to the control experiment, particularly for the case of 0.05 mM IPTG and 15 °C. On the other hand, the amount of inclusion bodies was significantly reduced upon coexpression of DnaK/J-GrpE or GroEL/ES.

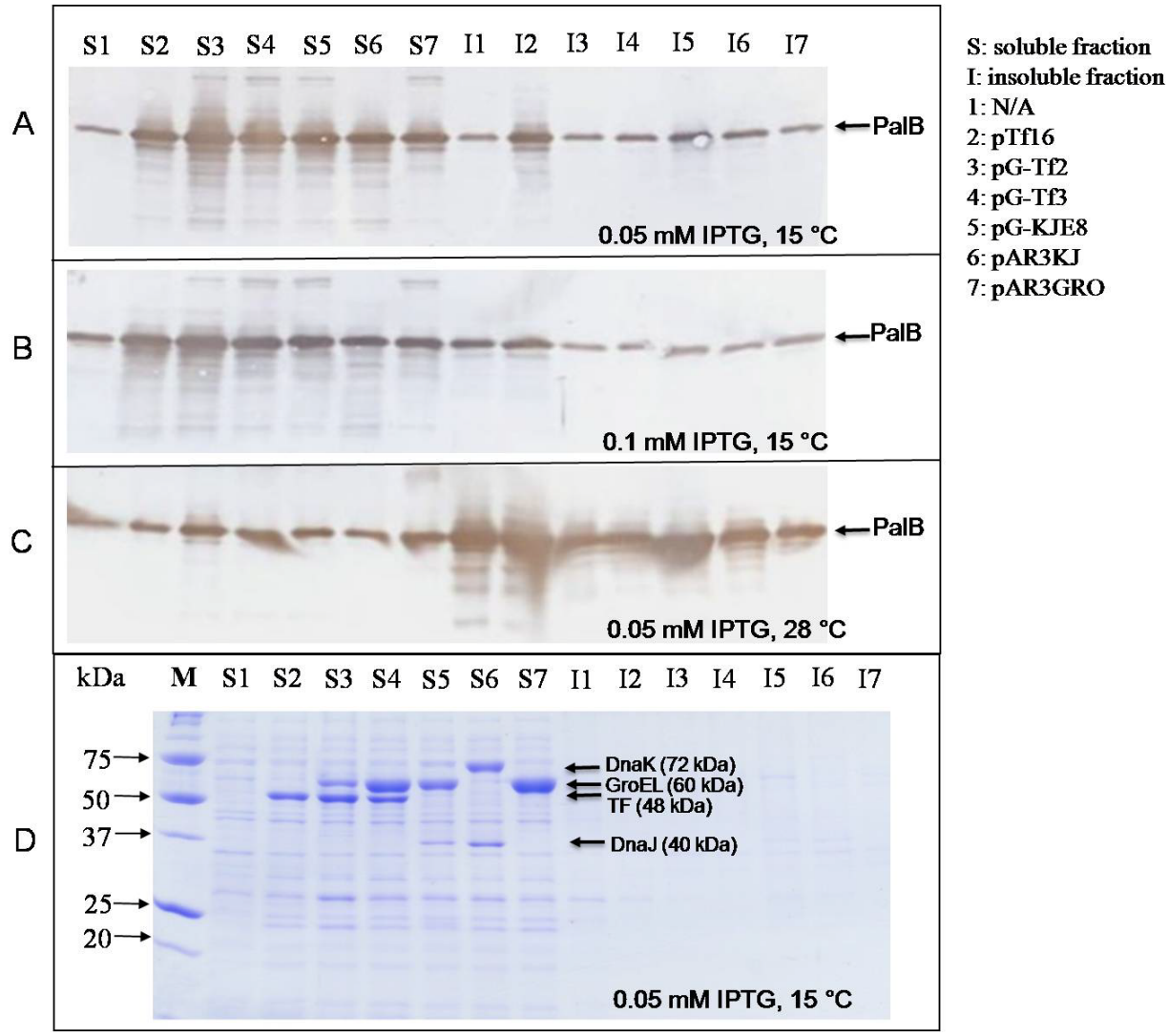


Figure 3.5 Western blotting (Panel A~C) and SDS-PAGE (Panel D) analysis of soluble (S) and insoluble (I) fractions of BL21(DE3) harboring pETKnL and chaperone-coexpression plasmid under various culture conditions for PalB expression. Panel A and Panel D: 0.05 mM IPTG, 15 °C; Panel B: 0.1 mM IPTG, 15 °C; Panel C: 0.05 mM IPTG, 28 °C. Designation of chaperone groups is the same as that of Figure 3.3. Lane M, protein marker.

The culture performance for PalB expression had a significant improvement upon simultaneous coexpression of TF and GroEL/ES (based on the use of either pG-Tf2 or pG-Tf3); reflected by enhanced cell growth, less amounts of inclusion bodies, and higher specific PalB expression levels under almost all

the culture conditions. Such synergistic effect was also significant upon simultaneous coexpression of DnaK/J-GrpE and GroEL/ES based on the use of pG-KJE8 at 15 °C, but less observable at 28 °C.

3.4 Discussion

PalB is a cold active lipase destined to be secreted extracellularly in the original psychrophilic and eukaryotic source of *P. antarctica*. Consequently, the intracellular overexpression might prevent it from being structurally adaptable to the mesophilic and prokaryotic expression system of *E. coli*, resulting in protein misfolding. This could be the reason why cultivation was conducted at low temperatures via the use of a cold-inducible expression system for functional expression of PalB [29]. Practical application with this host/vector system could be limited due to arrested cell growth and low PalB synthesis rate under low temperatures. In this study, functional expression of PalB was demonstrated in the cytoplasm of BL21(DE3), a regular host with a reduced cytoplasm, implying that disulfide bond formation might not be strictly required to develop PalB bioactivity. This result was not previously reported. In addition, cultivation temperature for such functional expression can be as high as 28 °C, a common temperature for recombinant protein production. Nevertheless, culture performance stands improvement due to the two technical issues, e.g., growth arrest and intracellular PalB misfolding, which are interrelated.

PalB misfolding could result from the overwhelming of the cytoplasmic folding machinery due to a fast protein synthesis rate associated with the use of the strong T7 promoter system at a relatively high temperature of 28 °C. Hence, lowering the cultivation temperature to 15 °C to substantially reduce protein synthesis rate, but not protein folding rate, was explored to test this hypothesis. In general, the amount of insoluble inclusion bodies was significantly reduced upon temperature lowering (Figure 3.4), implying a more balanced flux into the productive pathway for protein formation. However, the solubility of PalB was enhanced at an expense of a decrease in the total amount of PalB synthesis, suggesting that this operational strategy might not be practically feasible.

The overexpression of PalB resulted in growth arrest which could be alleviated to some extent by chaperone coexpression. In general, it appears that coexpression of single chaperone had a minimum effect on improving PalB expression performance. On the other hand, simultaneous coexpression of TF and GroEL/ES was effective in restoration of cell growth by maintaining or even boosting the specific lipase activity, resulting in significant improvement in total lipase activity. The positive effect based on simultaneous coexpression of DnaK/J-GrpE and GroEL/ES was also observable at 15 °C possibly because TF and DnaK could share certain improving functions by possessing overlapping substrate pools and binding specificities [158]. The results suggest that an appropriate chaperone cocktail is required for

not only alleviation of physiological deterioration but also more soluble PalB overexpression. The enhancement in recombinant protein production based on the synergistic effect of multiple chaperones was previously reported [68, 76].

With the peptidyl-prolyl cis/trans isomerase (PPIase) activity, TF plays a role *in vivo* as a ribosome-associated chaperone through the interaction with newly synthesized proteins and 50S ribosome [152]. In this study, coexpression of TF resulted in more accumulation of insoluble PalB without restoring cell growth or increasing the specific lipase activity, particularly for the expression at 15 °C. The results suggest that coexpression of TF could have actually boosted the overall PalB synthesis rate via the enhanced interaction with the nascent PalB. Nevertheless, these PalB folding intermediates could overwhelm the subsequent folding machinery, such as GroEL/ES, resulting in more accumulation of insoluble PalB. On the other hand, coexpression of another single chaperone, either GroEL/ES or DnaK/J-GrpE which mediates the subsequent folding event in the cytoplasm, improved the PalB solubility without increasing the overall PalB synthesis rate. It was proposed that protein synthesis on ribosome should be coordinated with the activities of various chaperone systems for not only stabilizing nascent polypeptides but also promoting subsequent folding [149]. In addition, by further association with GroEL, TF can strengthen its function for facilitating protein folding through efficient binding to GroEL-substrates [157]. The results of this study lead to a conclusion that, in order to effectively enhance PalB expression performance, both TF and GroEL/ES respectively mediating the early stage and late stage of cytoplasmic PalB folding would be simultaneously required for boosting both the overall PalB synthesis rate and the cytoplasmic folding efficiency.

3.5 Abstract of the submitted paper

Lipase B from *Pseudozyma antarctica* (PalB) was actively expressed in the reduced cytoplasm of a regular *Escherichia coli* strain of BL21(DE3), implying that disulfide bond formation was not strictly required for functional expression. However, the expression was ineffective and was primarily limited by the formation of PalB inclusion bodies and growth arrest, both of which were associated with PalB misfolding and deteriorated physiology. The culture performance in terms of cell growth and PalB expression level could be significantly improved by simultaneous coexpression of multiple chaperones of trigger factor and GroEL/ES, but not individual coexpression of either one of them. It was proposed that the two chaperones respectively mediating the early stage and late stage of cytoplasmic PalB folding would be simultaneously required for boosting both the overall PalB synthesis rate and the cytoplasmic folding efficiency.

Keywords: chaperone, *Escherichia coli*, gene expression, inclusion body, lipase B, *Pseudozyma antarctica*, recombinant protein production

Chapter 4*
Improvement of PalB Cytoplasmic Expression by Using a Mutant Host and/or by Adding a Fusion Tag

4.1 Introduction

As mentioned before, PalB expression performance was limited by several factors. PalB is a cold active lipase destined to be secreted extracellularly in the original psychrophilic and eukaryotic source of *P. antarctica*. Consequently, the intracellular overexpression might prevent it from being structurally adaptable to the mesophilic and prokaryotic expression system of *E. coli*, resulting in protein misfolding. In addition, wild-type PalB has three intramolecular disulfide bonds potentially associated with its folding process though they might not be strictly required for developing the bioactivity. Finally, cell growth is often arrested due to deteriorated cell physiology associated with high-level expression and misfolding of PalB. It was the first report for a functional PalB expression in the cytoplasm of a regular *E. coli*, as showed in Chapter 3. However, a huge amount of inclusion bodies associated with its intracellular misfolding were produced that not only limited the PalB expression but also severely deteriorated the cell physiology. Although the cell performance in terms of cell growth and PalB expression was significantly improved by coexpression of multiple chaperones of TF and GroEL/ES, the overall activity was still too low to reach the requirement for PalB production.

The above results are not surprising. In wild-type *E. coli*, disulfide bond formation only occurs in the oxidative periplasm, but not reductive cytoplasm. The periplasm contains several Dsb-family members whose biological functions involved in disulfide bond formation are well defined [40]. Even with the pathway available for disulfide bond formation, proteins with complex disulfide patterns are generally difficult to express in *E. coli*. In fact, for many eukaryotic proteins heterologously expressed in *E. coli*, folding and disulfide bond formation, either individually or together, can seriously limit the expression performance since correct formation of disulfide bonds can facilitate protein folding, stabilize protein structure, and develop protein bioactivity. These limitations will result in intracellular accumulation of inactive or insoluble protein species (e.g. inclusion bodies) and/or induce physiological stress [43, 174, 236, 237].

* This chapter is based on a paper "Y. Xu, A. Yasin, R. Tang, M. Moo-Young, C. P. Chou, Heterologous Expression of Lipase in *Escherichia coli* is Limited by Folding and Disulfide Formation. *Applied Microbiology and Biotechnology*. 2008" (accepted).

Among the strategies for improving expression of recombinant protein in *E. coli*, no other technology has been as effective at improving the solubility of recombinant proteins as fusion tags [86, 126]. Fusion tag technology is able to improve heterologous expression by overcoming many challenges, e.g., the low translation rate, the proteolytic degradation of target proteins, protein misfolding, poor solubility, and the necessity for good purification methodologies [87, 126]. While C-terminus fusion tags are reported successfully improving the fused protein's solubility, an N-terminus fusion tag can be more effective by enhancing translation initiation and improving solubility at once to reach high activity and productivity for recombinant protein expression in *E. coli* [87]. Several fusion tags have been demonstrated to be effective in promoting protein folding so that the solubility of fusion proteins is enhanced though the mechanism leading to the enhanced solubility remains unclear [87].

While solubility is prerequisite for gene products further being processed on the productive pathway, the folding process in a bioactive protein production is the determining factor, which dependes on the specificity of the final protein. In other words, without an efficient folding process, a final bioactive protein could be only a small portion or even none of the soluble protein. It was discovered that disulfide bond formation can occur in strains with the *trxB* and *gor* mutations [104, 108, 109, 238-240] and these mutant strains, e.g., Origami series commercially available from Novagen, have been widely adopted for cytoplasmic expression of proteins with disulfide bonds. Nevertheless, the efficiency of disulfide bond formation can still be limited.

In this chapter, the functional expression of PalB in the cytoplasm with chaperone coexpression or fusion tag techniques was explored. Also, the expression performance based on the use of BL21(DE3) and Origami B(DE3) as the host was compared. Based on the results, disulfide bond formation was identified as an important factor potentially limiting functional PalB expression in *E. coli*.

4.2 Materials and Methods

4.2.1 Strains and Plasmids

The strains, plasmids and oligonucleotides used in this study are summarized in Table 4.1 and briefly described here. BL21(DE3) and Origami B(DE3) were used as the hosts for the expression of recombinant PalB and DH5α was the host for cloning. All of the protocols, kits, and chemicals for molecular cloning were described previously in Chapter 3.

DNA fragments containing the leaderless *palB* gene were PCR-amplified using *pfu* DNA polymerase (Stratagene), the primer pair of P17/P11, P15/P19, and P10/P11 (in Table 4.1), respectively,

and *P. antarctica* chromosomal DNA as the template. The PCR products were cloned into pPCR Script® (Stratagene). DNA sequencing was performed to ensure no mutations occurred during PCR. The *palB*-containing DNA fragments were obtained by digestion of the above pPCRScript® derivatives with *Nde*I/*Eco*RI, *Eco*RI/*Xma*I, and *Nco*I/*Eco*RI, respectively, subcloned into the corresponding restriction sites of pET-20b(+), pGEXC4S-6P-1 [241] and pENTR4 (Invitrogen) resulting in pETL, pGEXC4S-M, and pENTRG. The *palB* gene from the *palB*-containing entry vector of pENTRG was subcloned into various destination vectors of pHMGWA [242], pHNWA [242], pHXWA [242], pDest-555 [147], pDest-556 [147], pET-52A (EMBL, Heidelberg, Germany) using the Gateway technology (Invitrogen) to form various PalB fusion plasmids of pHMBP-G, pHNusA-G, pHTRX-G, pT7PK-G, pSkp-G, and pETllDsbA-G, respectively.

Table 4.1 Strains, plasmids, and oligonucleotides.

Strain, plasmid, and oligonucleotide	Relevant genotype or phenotype^	Source and reference
E. coli		
BL21(DE3)	F⁻ *ompT dcm lon hsdS*$_B$ (r_B^- m_B^-) *gal λ*(DE3[*lacI ind1 sam*7 *nin*5 *lacUV5-T7* gene 1])	Novagen
DH5α	F'/*(ϕ80 lacZ ΔM15) Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (r_K^-, m_K^+) phoA supE44 λ-thi-1 gyrA96 relA1*	Lab stock
Origami B(DE3)	F⁻ *ompT hsdS*$_B$(r_B- m_B-) *gal dcm lacY1 ahpC* (DE3) *gor*522::Tn10 (Tc^R) *trxB* (Kn^R)	Novagen
P. antarctica		
ATCC 32657	*palB* source gene	ATCC
Plasmid		
pAR3	Expression vector, P_{araB}, Ori (pACYC184), Cm^R	Lab stock [233]
pAR3KJ	P_{araB}::*dnaK/J*::*grpE*, Ori (pACYC184), Cm^R	Lab stock [233]
pAR3GRO	P_{araB}::*groEL/ES*, Ori (pACYC184), Cm^R	Lab stock [233]
pARllDsbA	P_{araB}::*lldsbA*, Ori (pACYC184), Cm^R	This study
pARllDsbC	P_{araB}::*lldsbC*, Ori (pACYC184), Cm^R	This study
pDest-555	Destination vector, P_{T7}::*lac*Op::*strep-II*::Δ*T7PK*, Ori (pBR322), Ap^R, Cm^R	[147]

pDest-556	Destination vector, P_{T7}::*lac*Op::*strep-II*::*skp*, Ori (pBR322), Ap^R, Cm^R	[147]
pENTR4	Entry vector, *att*L1-*rrn*B-*att*L2, Ori (pUC), Kn^R	Invitrogen
pENTRG	*palB*-containing entry vector, Ori (pUC), Kn^R	This study
pET-20b(+)	Expression vector, P_{T7}, Ori (pBR322), Ap^R	Lab stock
pETG-50A	Destination vector, P_{T7}::*lac*Op::*dsbA*::6×*his*, Ori (pBR322), Ap^R, Cm^R	EMBL
pETG-52A	Destination vector, P_{T7}::*lac*Op::*lldsbA*::6×*his*, Ori (pBR322), Ap^R, Cm^R	EMBL
pETllDsbA-G	P_{T7}::*lac*Op::*lldsbA*::6×*his*::*palB*, Ori (pBR322), Ap^R	This study
pETL	P_{T7}::*palB*, Ori (pBR322), Ap^R	This study
pETM-80	Expression vector, P_{T7}::*dsbC*::6×*his*, Ori (pBR322), Kn^R	EMBL [72]
pGEXC4S-6P-1	Expression vector, P_{tac}::*gst*, Ori (pBR322), Ap^R	[241]
pGEXC4S-M	P_{tac}::*gst*::*palB*, Ori (pBR322), Ap^R	This study
pG-KJE8	P_{araB}::*dnaK/J*::*grpE*, $P_{zt\text{-}lp}$::*groEL/ES*, Ori (pACYC184), Cm^R	T. Yura [76]
pG-Tf3	P_{araB}::*tig*, $P_{zt\text{-}lp}$::*groEL/ES*, Ori (pACYC184), Cm^R	T. Yura [76]
pHMBP-G	P_{T7}::*lac*Op::6×*his*::*mbp*::*palB*, Ori (pBR322), Ap^R	This study
pHMGWA	Destination vector, P_{T7}::*lac*Op::6×*his*::*mbp*, Ori (pBR322), Ap^R, Cm^R	[242]

pHNGWA	Destination vector, P_{T7}::*lac*Op::6×*his*::*nusA*, Ori (pBR322), Ap^R, Cm^R	[242]
pHNusA-G	P_{T7}::*lac*Op::6×*his*::*nusA*::*palB*, Ori (pBR322), Ap^R	This study
pHTRX-G	P_{T7}::*lac*Op::6×*his*::*trx*::*palB*, Ori (pBR322), Ap^R	This study
pHXGWA	Destination vector, P_{T7}::*lac*Op::6×*his*::*trx*, Ori (pBR322), Ap^R, Cm^R	[242]
pSkp-G	P_{T7}::*lac*Op::*strep-II*::*skp*::*palB*, Ori (pBR322), Ap^R	This study
pT7PK-G	P_{T7}::*lac*Op::*strep-II*::Δ*T7PK*::*palB*, Ori (pBR322), Ap^R	This study
pTf16	P_{araB}::*tig*, Ori (pACYC184), Cm^R	T. Yura [76]
Oligonucleotide*		
P10	5' – GG*CCATGG*GTCTACCTTCCGGTTCGG – 3'	This study
P11	5' – *CT*GAATTCTCAGGGGGTGACGATGCCGGAGCAGG – 3'	This study
P15	5' - *CG*GAATTCCTACCTTCCGGTTCGG - 3'	
P17	5' – TGCA*TATG*CTACCTTCCGGTTCGG – 3'	This study
P19	5' – *GGCCCGGG*T*T*AGGGGGTGACGATGCCGGAG – 3'	This study
Pf_{llDsbA}/Pr_{DsbA}	5' – *CCATGGA*ACAGTACACTACCCTGG – 3' and 5' – *AAGCTTTCA*TTTCTCGCTTAAGTATTTC – 3' primer pair for amplification of *lldsbA*	This study
Pf_{llDsbC}/Pr_{DsbC}	5' – *CCATGGTT*GATGACGCGGCAATTC – 3' and 5' – *AAGCTTTCA*ACCAGAACCACTAGT – 3' primer pair for amplification of *lldsbC*	This study

^ Designed restriction sites are underlined and the introduced mutations are in italic.

The coexpression plasmids containing a selection of cytoplasmic chaperone gene(s) were derivatives of pAR3 [233]. pAR3GRO [233] and pAR3KJ [233], and pTf16 [76] contain the gene of *groEL/ES*, *dnaK/J-grpE*, and *tig* (encoding TF), respectively fused with the *araB* promoter. pG-Tf3 [76] contains *tig* and *groES/EL*, respectively fused with the *araB* and *zt-lp* promoters. pG-KJE8 [76] contains the *dnaK/J-grpE* and *groES/EL*, respectively fused with the *araB* and *zt-lp* promoters. The coexpression plasmid containing the leaderless *dsbA* gene (*ll-dsbA*), and *ll-dsbC* (e.g., pARllDsbA, and pARllDsbC) was also derived from pAR3. To construct them, *ll-dsbA* and *ll-dsbC* were PCR-amplified using *Pfu* DNA polymerase, the primer pair Pf_{llDsbA}/Pr_{DsbA}, and Pf_{llDsbC}/P_{rDsbC}, and pETG-50A and pETM-80 as the template, respectively. The PCR product was cloned into pPCRScript® and DNA sequencing was performed to ensure no mutations occurred during PCR. The *ll-dsbA* and *ll-dsbC*-containing DNA fragments flanked with *Nco*I/*Hin*dIII were subcloned into the corresponding restriction sites of pAR3 to form pARllDsbA, and pARllDsbC, respectively. All the chaperone-gene(s)-containing plasmids have a pACYC184 replication origin and a chloramphenicol-resistant (Cm^R) marker and, therefore, are compatible with the PalB expression plasmid (e.g., pETL) containing a pBR322 replication origin and an ampicillin-resistant (Ap^R) marker.

4.2.2 Cultivation and Analytical Methods

The cultivation and analytical procedures were identical with the procedures described in Section 3.2.2 and 3.2.3.

4.3 Results

4.3.1 Effect of host on functional expression of PalB in the cytoplasm

Since *trxB* and *gor* mutation used kanamycin and tetracycline as the selection marker, respectively, in Origami B(DE3), the expression plasmid pETKnL, which has a kanamycin marker and had been characterized in Chapter 3, could not be used in Origami B(DE3). For consistent comparison, pETL, which has a same backbone with pETKnL but an ampicillin marker, was constructed (Figure 4.1) and used in this Chapter. We assume pETL and pETKnL have the same behavior for expression PalB in the cytoplasm of *E. coli*.

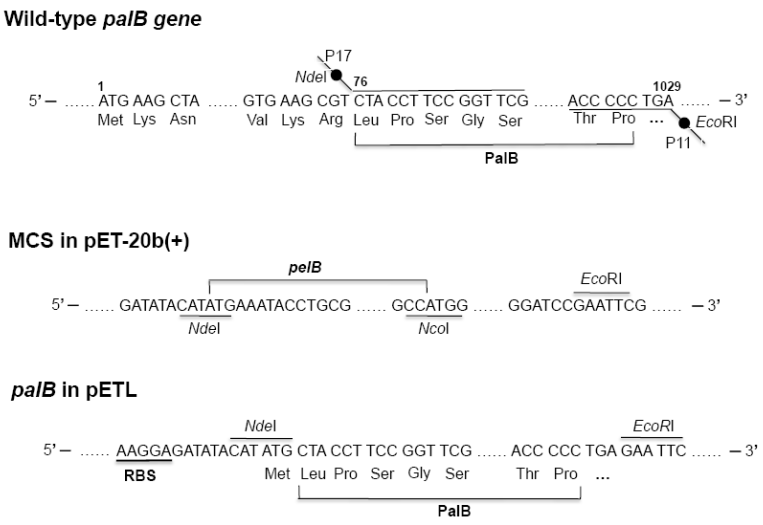
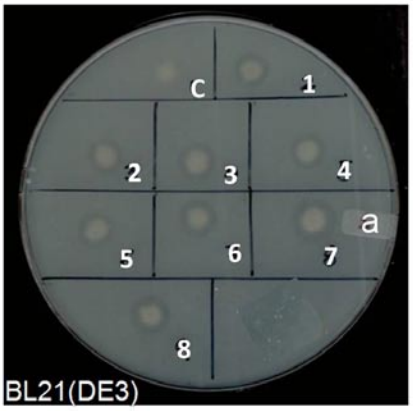
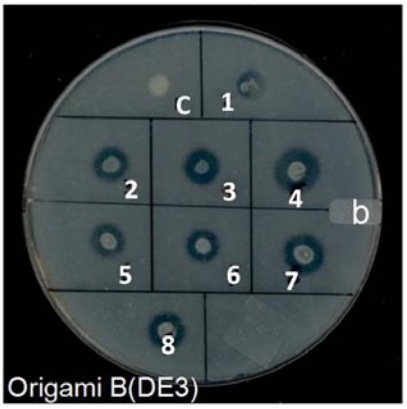


Figure 4.1 Construction of *palB* expression plasmid of pETL. The DNA region from nucleotides 76 to 1029 of the wild-type *palB* gene on the chromosome of *P. antarctica* ATCC 32657 was PCR-amplified with primer pairs of P17/P11, respectively. The PCR products were cloned into pET-20b(+) using *Nde*I/*Eco*RI to form a transcriptional fusion of pETL. MCS, multiple cloning sites. RBS, ribosome binding site.

To evaluate the effect of the host strain on cytoplasmic PalB expression, pETL expression performance using the host of BL21(DE3) and its mutant derivative, Origami B(DE3), is summarized in Figure 4.2. Halos were developed on the tributyrin plate for both expression systems (e.g., #2 on both a and b plates in Figure 4.2A), indicating that PalB was functionally expressed in the cytoplasm. Since the area and transparency of the halo potentially correlate with the expressed PalB activity, the results suggest that Origami B(DE3) outperformed BL21(DE3) as the host for cytoplasmic PalB expression. The specific PalB activity of Origami B(DE3) harboring pETL was more than 3-fold that of BL21(DE3) harboring pETL at 28 °C (Figure 4.2C). In general, Origami B(DE3) grew slower than BL21(DE3) under the same culture condition with a lower biomass yield (Figure 4.2B). A slight growth arrest was observed upon PalB expression. Though PalB can be functionally expressed, inclusion bodies at a significant amount were observed, implying that functional PalB expression could be limited by proper folding. Lowering the cultivation temperature to 15 °C slightly increased the specific PalB activity (#1 in Figure 4.2C) for Origami B(DE3) harboring pETL though PalB solubility appeared to be significantly improved (#1 in Figure 4.2D).

A

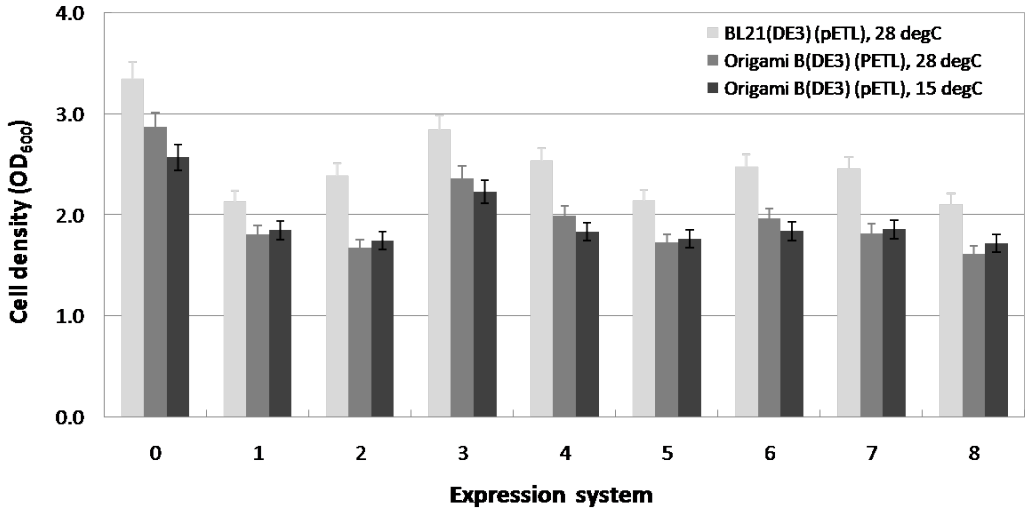




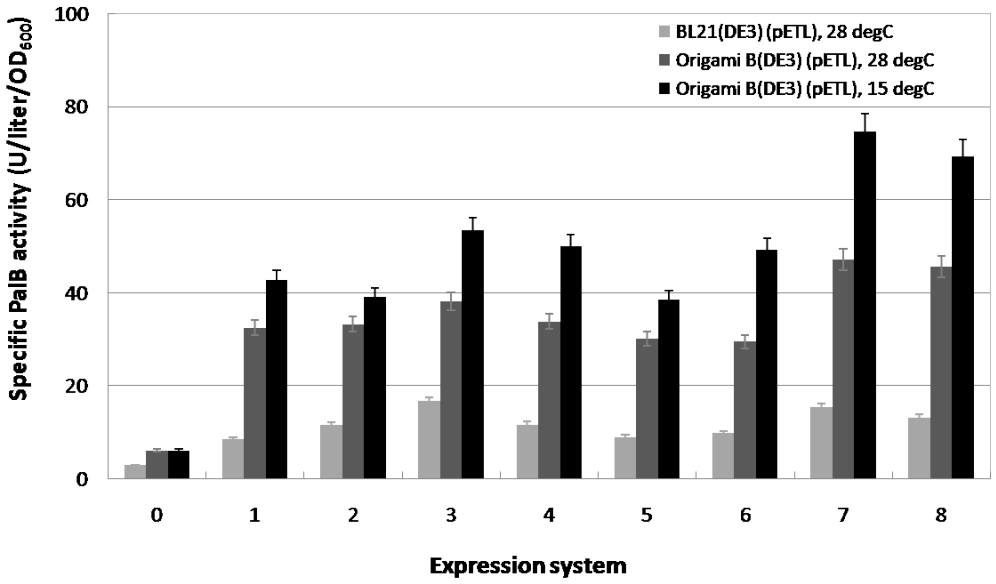
No	Expression vector
C*	-
1	pETL
2	pETL, pTF16
3	pETL, pG-TF3
4	pETL, pG-KJE8
5	pETL, pAR3KJ
6	pETL, pAR3GRO
7	pETL, pARIIDsbA
8	pETL, pARIIDsbC

* Host strain BL21(DE3) only

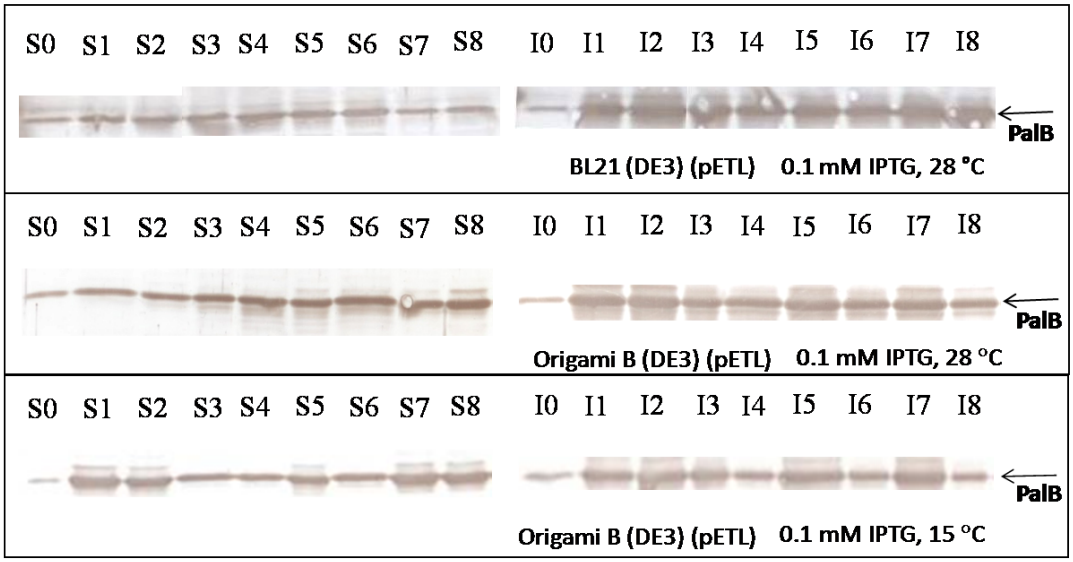
B



C


Specific PalB activity (U/liter/OD600)
100
80
60
40
20
0
0
1
2
3
4
5
6
7
8
Expression system
BL21(DE3) (pETL), 28 degC
Origami B(DE3) (pETL), 28 degC
Origami B(DE3) (pETL), 15 degC

D


S0 S1 S2 S3 S4 S5 S6 S7 S8
I0 I1 I2 I3 I4 I5 I6 I7 I8
PalB
BL21 (DE3) (pETL) 0.1 mM IPTG, 28 °C
S0 S1 S2 S3 S4 S5 S6 S7 S8
I0 I1 I2 I3 I4 I5 I6 I7 I8
PalB
Origami B (DE3) (pETL) 0.1 mM IPTG, 28 °C
S0 S1 S2 S3 S4 S5 S6 S7 S8
I0 I1 I2 I3 I4 I5 I6 I7 I8
PalB
Origami B (DE3) (pETL) 0.1 mM IPTG, 15 °C

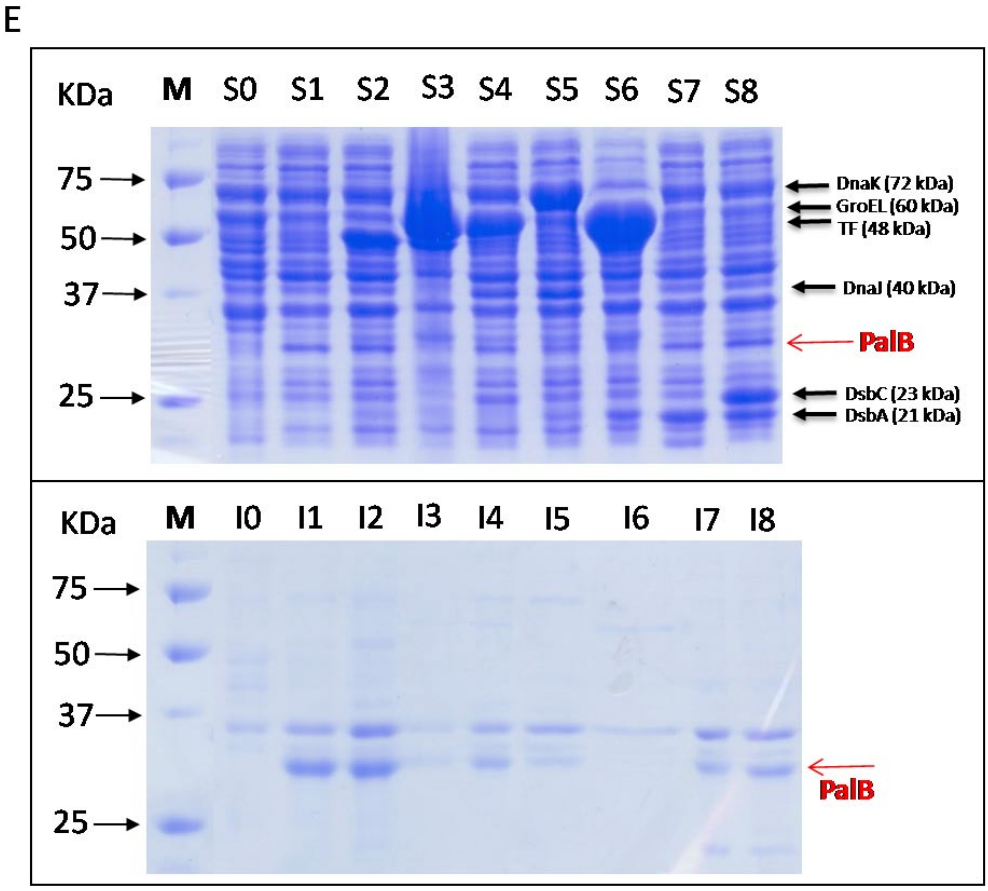


Figure 4.2 The expression performance for non-fused PalB and the effect of folding factors on the expression. Panel A: Qualitative visualization of PalB activity using the tributyrin plate for recombinant BL21(DE3) (plate a) and Origami B(DE3) (plate b), respectively. Expression plasmids are labeled in separate legend. Panel B: Cell density for various expression systems. Panel C: Specific PalB activity for various expression systems. Panel D: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various expression systems. Panel E: SDS-PAGE analysis of the soluble (S) and insoluble (I) fractions for PalB expression system hosted in Origami B(DE3) at 28 °C. M: protein molecular weight marker. Coexpression plasmids for Panels B, C, D, and E: 0&1/ none; 2/ pTf16; 3/ pG-Tf3; 4/ pG-KJE8; 5/ pAR3KJ; 6/ pAR3GRO; 7/ pARllDsbA. Inducer supplementation: 1/ IPTG; 2~7/ IPTG and arabinose. The error bar in Panels B and C represents a range of the data variance.

4.3.2 Effect of folding factors

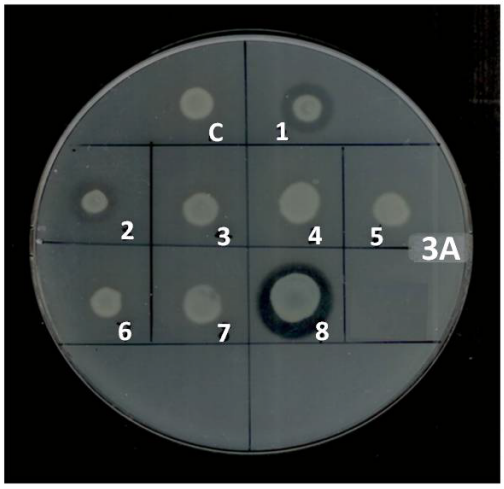
Since functional expression of PalB appeared to be limited by folding even in the oxidative cytoplasm of Origami B(DE3) which could mediate effective disulfide bond formation, coexpression of various folding factors (including TF, GroEL/ES, DnaK/J-GrpE, and DsbA) was performed to identify potential factors limiting the expression performance and the results are summarized in Figure 4.2. In general, functional PalB expression for all recombinant Origami B(DE3) strains outperformed the corresponding recombinant BL21(DE3) strains in terms of specific PalB activity (Figure 4.2A and C) though the biomass yield was slightly reduced for recombinant Origami B(DE3) strains. The results imply that the oxidative environment could mediate proper formation of disulfide bonds in the cytoplasm of Origami B(DE3). Among the folding factors investigated, the effect of coexpression of DsbA and DsbC (e.g., pARllDsbA, and pARllDsbC) were pronounced in both BL21(DE3) and Origami B(DE3) backgrounds, though the effect of coexpression of TF and GroEL/ES also appeared to be observable in BL21(DE3) background. Lowering the cultivation temperature to 15 °C also assisted the function expression by boosting the specific PalB activity. The synergetic effect mediated by the oxidative cytoplasm of Origami B(DE3) and the folding factor of DsbA and DsbC on improving PalB bioactivity support the above argument that functional expression of PalB could possibly be limited by disulfide bond formation.

4.3.3 Effect of fusion tags

Since functional expression of PalB in the cytoplasm appeared to be limited by protein solubility even with the coexpression of folding factors, fusion tags were explored to enhance PalB folding. Since the expression performance based on the use of fusion tag technique is unpredictable [87], seven N-terminal tags, including GST, MBP, NusA, TRX, T7PK, Skp, and DsbA, were selected to construct various translational fusions with PalB for evaluation, resulting in seven expression plasmids, including pGEXC4S-M, pHMBP-G, pHNusA-G, pHTRX-G, pT7PK-G, pSkp-G, and pETDsbA-G. Except GST-PalB whose expression was regulated by the *tac* promoter, the expression of all the other PalB fusions was regulated by the *T7* promoter. Using BL21(DE3) harboring these PalB fusion expression plasmids for cultivation, PalB expression performance was summarized in Figure 4.3. Qualitative visualization of the PalB activity on the tributyrin plate was conducted (Figure 4.3A). Visible halos were developed only for three PalB fusions, GST-PalB (#3), and DsbA-PalB (#9). Note that growth arrest was observed upon the expression of non-fused PalB and T7PK-PalB, but was not observed for the other PalB fusions (Figure 4.3B), implying that PalB toxicity could be reduced by most of the fusion tags. However, the specific

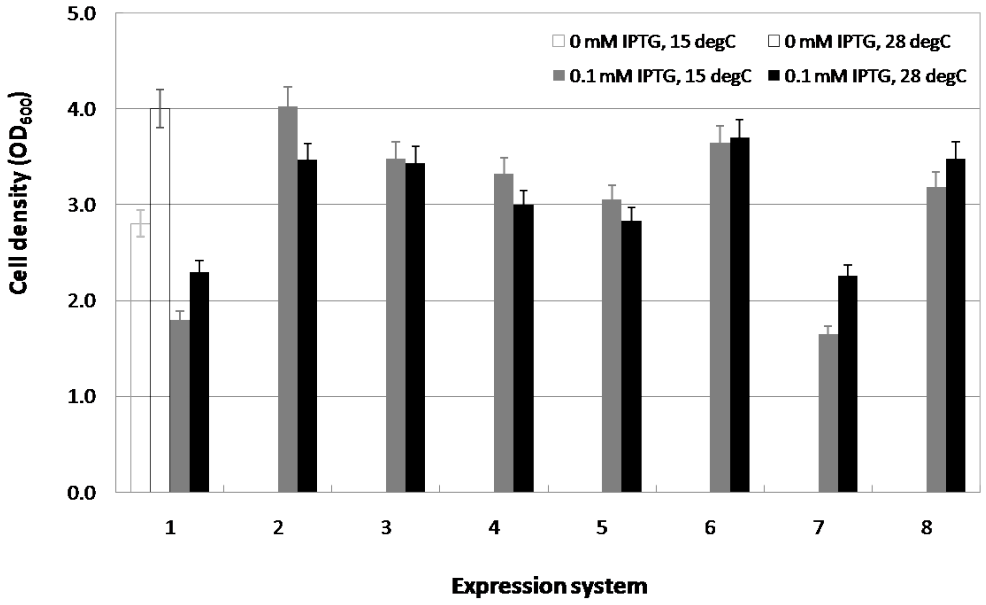
PalB activities for most of the PalB fusions were either approximately the same as or lower than the non-fused PalB except DsbA-PalB (Figure 4.3C). Note that, among these tags investigated here, MBP and T7PK significantly improved PalB solubility but not PalB activity, whereas DsbA significantly improved both PalB solubility and activity (Figure 4.3D) particularly under the low cultivation temperature at 15 °C.

A

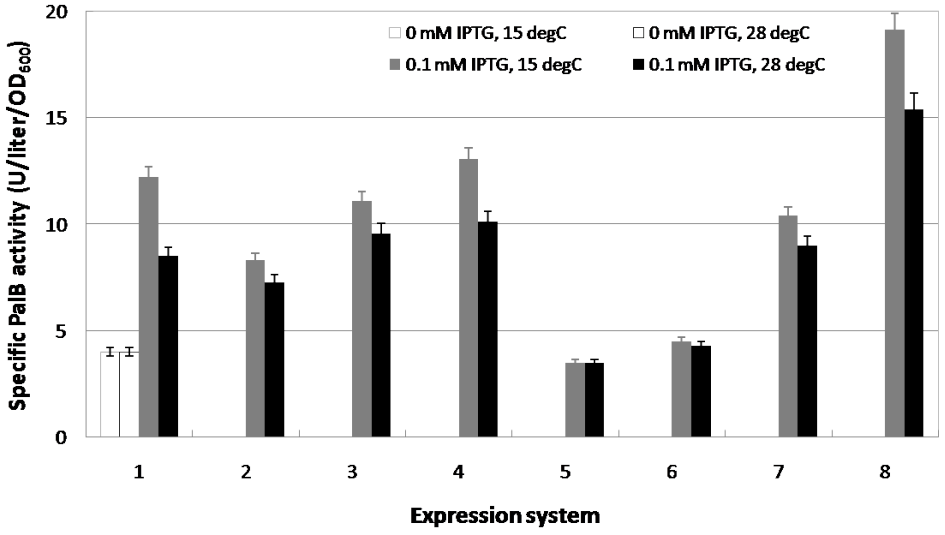


No	Expression system
C	BL21(DE3)
1	BL21(DE3) (pETL)
2	BL21(DE3) (pGEXC4S-M)
3	BL21(DE3) (pHMBP-G)
4	BL21(DE3) (pHNusA-G)
5	BL21(DE3) (pHTRX-G)
6	BL21(DE3)(pT7PK-G)
7	BL21(DE3) (pSkp-G)
8	BL21(DE3) (pETIIDsbA-G)

B


0 mM IPTG, 15 degC
0 mM IPTG, 28 degC
0.1 mM IPTG, 15 degC
0.1 mM IPTG, 28 degC
Cell density (OD600)
5.0
4.0
3.0
2.0
1.0
0.0
1
2
3
4
5
6
7
8
Expression system

C


0 mM IPTG, 15 degC
0 mM IPTG, 28 degC
0.1 mM IPTG, 15 degC
0.1 mM IPTG, 28 degC
Specific PalB activity (U/liter/OD600)
20
15
10
5
0
1
2
3
4
5
6
7
8
Expression system

D

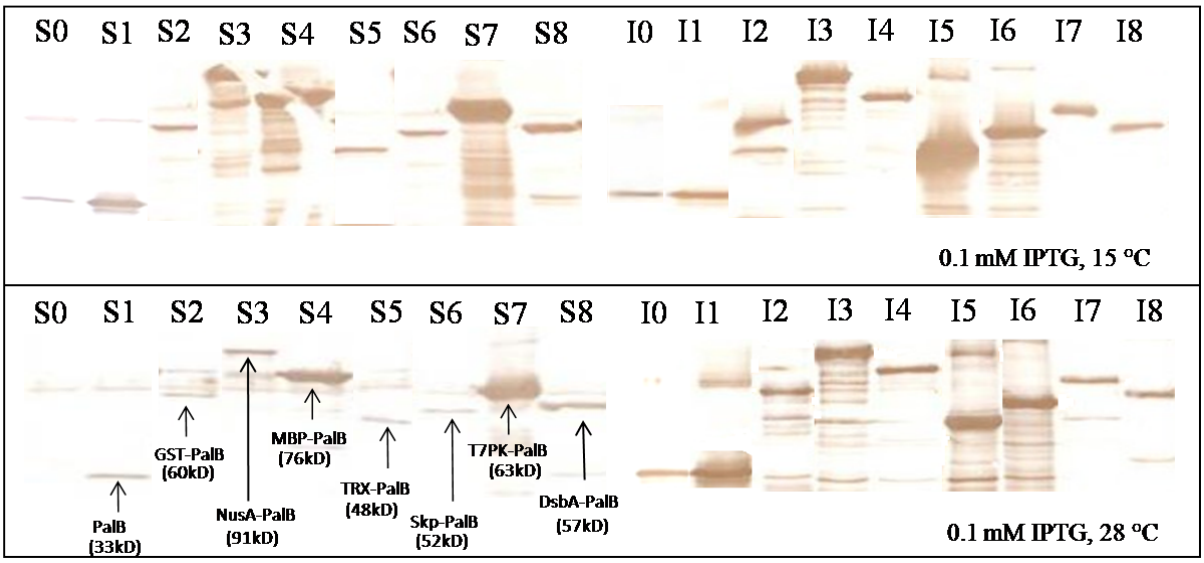
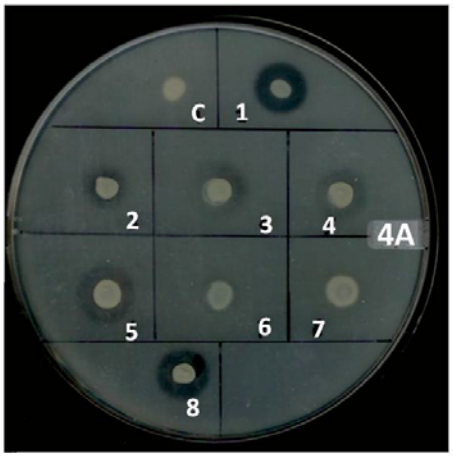


Figure 4.3 The expression performance for various PalB fusions using BL21(DE3) as the host. Panel A: Qualitative visualization of PalB activity using the tributyrin plate. Expression plasmids are labeled in separate legend. Panel B: Cell density for various PalB-fusion expression systems. Panel C: Specific PalB activity for various PalB-fusion expression systems. Panel D: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various PalB-fusion expression systems. Lipase fusions: 0&1/ PalB; 2/ GST-PalB; 3/ NusA-PalB; 4/ MBP-PalB; 5/ TRX-PalB; 6/ Skp-PalB; 7/ T7PK-PalB; 8/ DsbA-PalB. Inducer supplementation: 1~8/ IPTG. The error bar in Panel B and C represents a range of the data variance.

Compared to BL21(DE3), functional expression of PalB fusions in the cytoplasm of Origami B(DE3) (summarized in Figure 4.4) was more effective since visible halos were developed on the tributyrin plate for all the PalB fusions (Figure 4.4A). The results again suggest the importance of disulfide bond formation for functional expression of PalB. The growth arrest associated with the expression of non-fused PalB was alleviated upon the expression of all PalB fusions (except T7PK-PalB) in Origami B(DE3) (Figure 4.4B). Note that, compared to the use of BL21(DE3) as the host, both the solubility and specific PalB activity of GST-PalB were significantly improved when expressed in Origami B(DE3). On the other hand, while T7PK tag appeared to be effective in solubilizing PalB in both

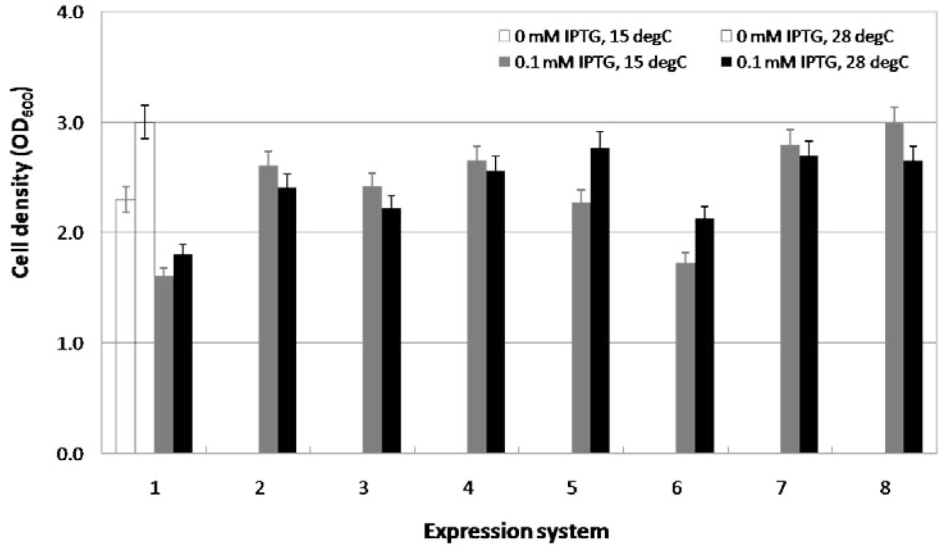
BL21(DE3) and Origami B(DE3) backgrounds, PalB activity was minimally induced out of this fusion, implying that solubilization of PalB does not necessarily develop PalB activity. Compared to the non-fused PalB as the control, only DsbA tag could enhance functional expression in terms of boosting the specific PalB activity; whereas the other tags could at most enhance PalB solubility (Figures 4.3 and 4.4). The specific PalB activity for DsbA-PalB could reach as high as 50 U/liter/OD_{600}, resulting in the highest volumetric PalB activity at 130 U/liter for functional expression of PalB in the cytoplasm of Origami B(DE3).

A

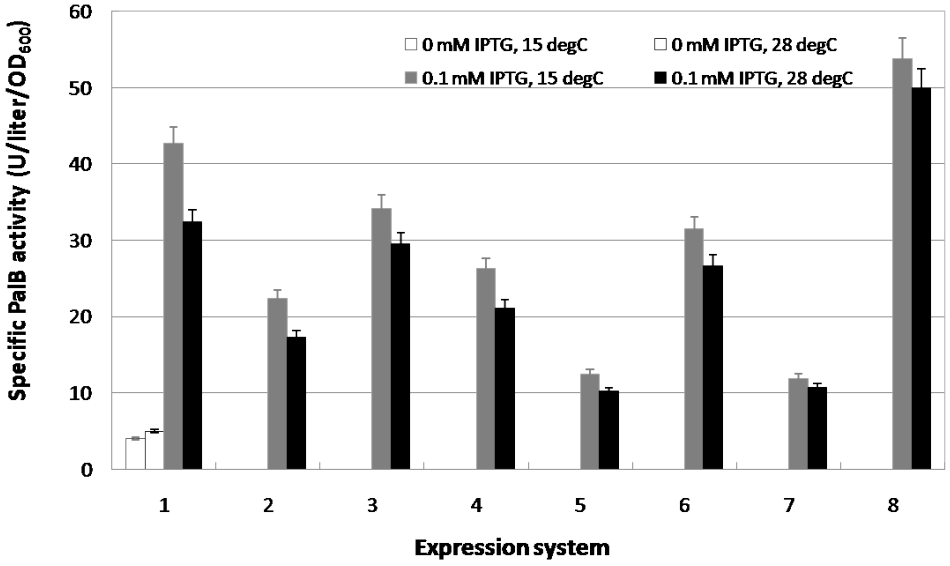


No	Expression system
C	Origami B(DE3)
1	Origami B(DE3) (pETL)
2	Origami B(DE3) (pGEXC4S-M)
3	Origami B(DE3) (pHMBP-G)
4	Origami B(DE3) (pHNusA-G)
5	Origami B(DE3) (pT7PK-G)
6	Origami B(DE3) (pHTRX-G)
7	Origami B(DE3) (pSkp-G)
8	Origami B(DE3) (pETllDsbA-G)

B


0 mM IPTG, 15 degC
0 mM IPTG, 28 degC
0.1 mM IPTG, 15 degC
0.1 mM IPTG, 28 degC
Cell density (OD600)
4.0
3.0
2.0
1.0
0.0
1
2
3
4
5
6
7
8
Expression system

C


0 mM IPTG, 15 degC
0 mM IPTG, 28 degC
0.1 mM IPTG, 15 degC
0.1 mM IPTG, 28 degC
Specific PalB activity (U/liter/OD600)
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
Expression system

D

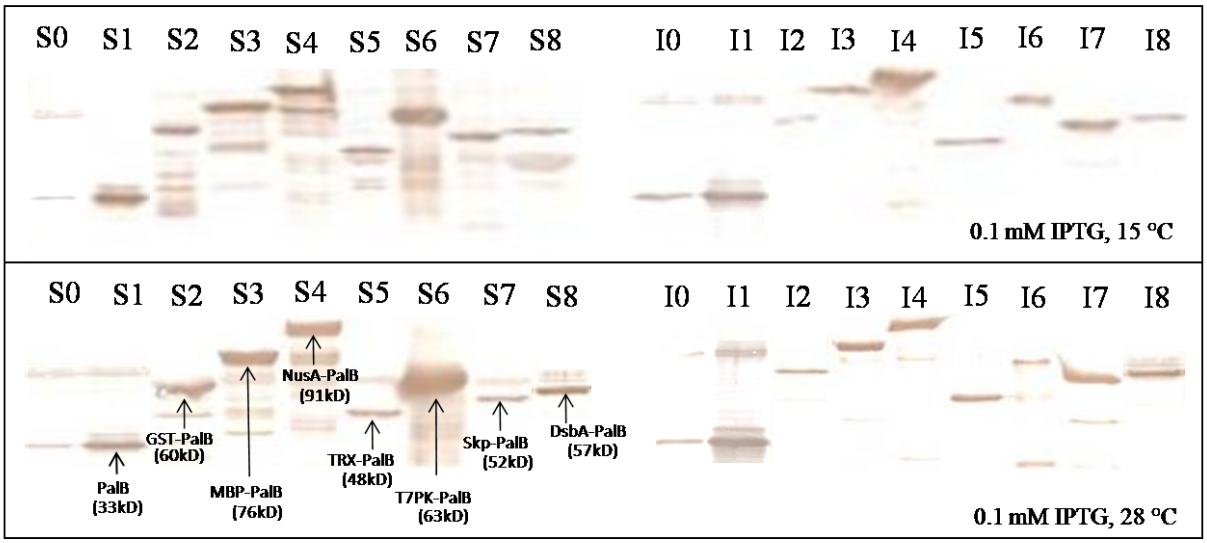


Figure 4.4 The expression performance for various PalB fusions using Origami B(DE3) as the host. Panel A: Qualitative visualization of PalB activity using the tributyrin plate. Expression plasmids are labeled in separate legend. Panel B: Cell density for various PalB-fusion expression systems. Panel C: Specific PalB activity for various PalB-fusion expression systems. Panel D: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various PalB-fusion expression systems. Lipase fusions: 0&1/ PalB; 2/ GST-PalB; 3/ MBP-PalB; 4/ NusA-PalB; 5/ TRX-PalB; 6/ T7PK-PalB ; 7/ Skp-PalB; 8/ DsbA-PalB. Inducer supplementation: 1~8/ IPTG. The error bar in Panel B and C represents a range of the data variance.

4.3.4 Bioactive PalB was obtained by purification of PalB from GST-PalB cytoplasmic expression system

Purification of PalB fusion by using Origami B(DE3) harboring pGEXC4S-M was demonstrated as an example. Basically, the GST-PalB binds to an affinity chromatographic column (e.g., GSTrap FF, GE Healthcare, Piscataway, NJ, USA) to be separated from the other *E. coli* proteins and then the GST tag was on-column digested using a protease (e.g., Prescission™ protease, GE Healthcare). The purification flowchart was shown in Figure 4. 5.

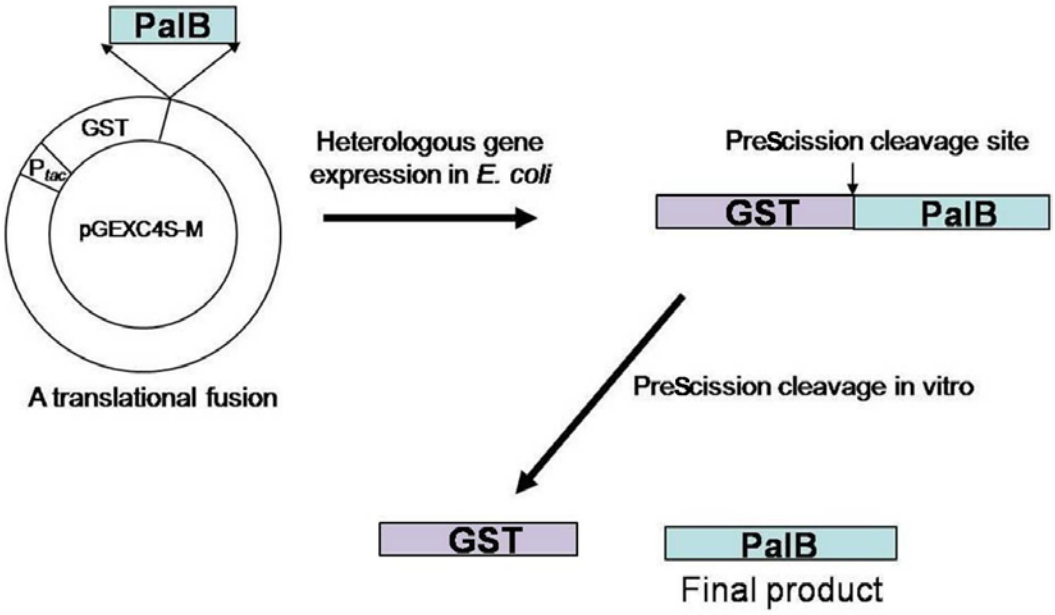


Figure 4.5 Flowchart for production of recombinant PalB in *E. coli*. The *palB* was fused with the *gst* gene to form a translational fusion in pGEXC4S-M. PalB is overexpressed as a GST fusion protein (e.g., GST-PalB) under the regulation of the *tac* promoter. The GST-PalB was captured using an affinity chromatographic column and subsequently processed with Prescission™ cleavage to release the moiety of PalB as a target protein.

The cultivation was performed with 100 ml of culture in a 500 ml of flask which was incubated at 15 °C, 200 rpm for 24 h after induction by supplemented 0.1 mM IPTG. The paste was collected, resuspended into 50 mM Tris-HCl (pH 7.0) buffer, and sonicated. The lysate was separated by high-speed centrifugation. All the procedures were the same as for preparing the sample for analysis purposes described before. The GST-PalB was captured in a 1 ml of GSTrap affinity chromatography from the lysate. Instead of being eluted, GST-PalB was in-situ cleaved by injecting 80U of Prescission™ protease in 1 ml of cleavage buffer (50 mM Tris-HCl, pH7.0, 150 mM NaCl, 1 mM ETDA, 1mM dithiothreitol) following the instructs given by GE Healthcare. After 3 h of incubation at room temperature, PalB was released into the bulk liquid phase, and collected by injecting 1ml of 50 mM Tris-HCl buffer. GST tag was removed from the GSTrap column by using glutathione solution (50mM Tris-HCl, 10 mM glutathione, pH8.0). The GSTrap column was cleaned with 6 M guanidine HCl.

The samples taken along the purification processing were analyzed by SDS-PAGE and results are shown in Figure 4.6B. Apparently, the enzymatic cleavage was not optimized and the purified PalB may

need to go through more separation steps to clean out the contaminants, e.g., uncut GST-PalB and GST tag. The released PalB moiety showed a 3-fold bioactivity compared with lysate GST-PalB.

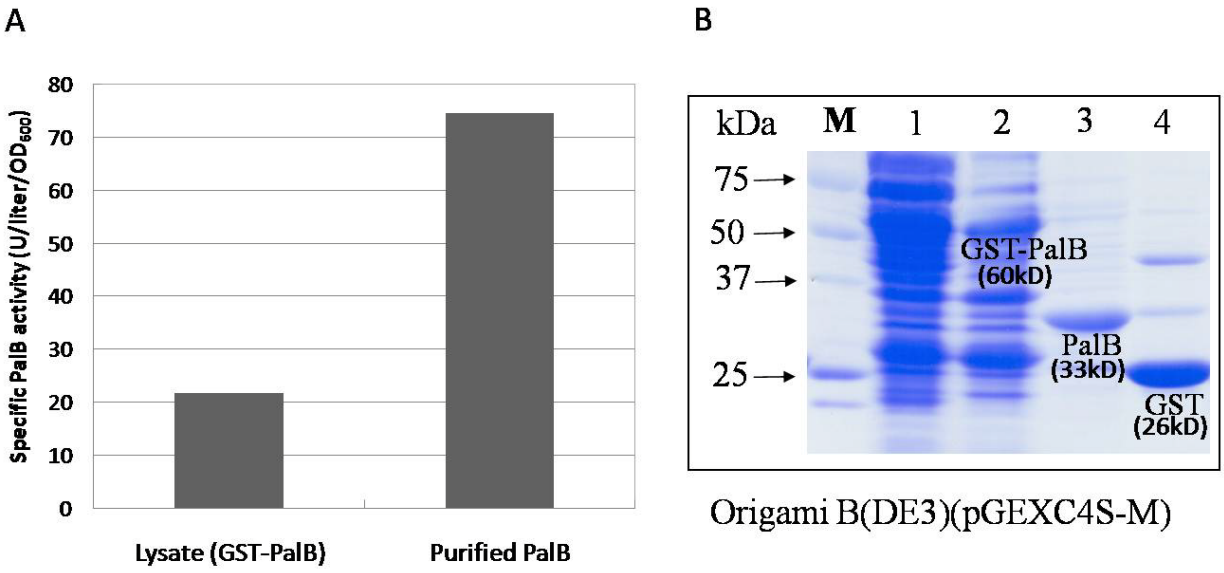


Figure 4.6 Panel A, comparison of PalB activity between GST-PalB and purified PalB. Panel B, results of SDS-PAGE stained by Coomassie blue for the sample along the purification process of PalB from GST-PalB. M, protein molecular weight marker. The samples loaded on the gel are labeled as: 1/ lysate; 2/ flow-through GSTrap column; 3/ elute sample after thrombin cutting; and 4/ elute with injection of glutathione.

4.4 Discussion

Though *E. coli* contains an oxidative compartment in the periplasm in which disulfide bonds can be formed, the expression performance can be limited by other steps such as translocation or periplasmic folding. The Origami strains with the mutations on thioredoxin reductase and glutathione reductase offer an alternative for overexpression of proteins with disulfide bonds in the oxidative cytoplasm [104, 238]. The Origami series had been widely used as a host in various recombinant protein expression in the cytoplasm [29, 110, 111, 137]. In this study, PalB expressed in the cytoplasm of BL21(DE3) harboring pETL could still develop PalB activity though at a low level, implying that disulfide bond formation was not strictly required for functional expression. The specific PalB activity was significantly enhanced by

using Origami B(DE3) harboring pETL, suggesting that disulfide bond formation could still be a critical step limiting the expression performance. Nevertheless, it appeared that PalB solubility (or PalB folding) was not necessarily improved with the enhanced PalB activity or disulfide bond formation. Similar to the expression of non-fused PalB, the specific PalB activity was significantly enhanced by using Origami B(DE3) as the host cell for all the PalB fusions, but the solubilization effect was not necessarily reflected by improved PalB activity. Given such unparallel but intertwined effects on PalB solubilization and activity, Origami B(DE3) was demonstrated to be a suitable host for functional expression of either non-fused PalB or PalB fusions, presumably due to effective disulfide bond formation.

In fact, the oxidizing environment might not be sufficient for disulfide bond formation. It is has been shown that a system consisting of at least two cellular envelope proteins, DsbA and DsbB, is necessary for the formation of disulfide bonds in the periplasm [40]. DsbA is a periplasmic protein (21 kDa) with oxidoreductase activity catalyzing direct bridging of disulfide bonds within newly translocated proteins and DsbC (homodimer, 23 kDa per monomer) is an isomerase for correcting the intra- and intermolecularly misfolded disulfide bonds. Though with distinctively major activities, the two proteins might share biological functions under certain conditions [104, 215-217]. Coexpression of DsbA and/or DsbC enhanced functional expression of several recombinant proteins, e.g., ribonuclease I [243], human leptin [244], insulin-like growth factor-I [215], human tissue plasminogen activator [104, 213], horseradish peroxidase [245] and human nerve growth factor β [212]. In addition, it has reported that DsbA and DsbC may have chaperone activity [90, 218]. Among the folding factors being coexpressed with PalB, DsbA and DsbC presented almost the similar positive effects on developing PalB activity and such effects were more pronounced in Origami B(DE3) than BL21(DE3), suggesting that DsbA and DsbC functioned properly in an oxidative environment. Such mediation was presumably more effective in the oxidative cytoplasm of Origami B(DE3). The results also suggest that both folding and disulfide bond formation could be critical for developing PalB activity.

Cells expressing non-fused PalB suffered growth arrest presumably due to physiological stress mediated by the toxicity of the gene product. Such growth arrest could be alleviated to a certain degree by expressing most PalB fusions except T7PK-PalB. The solubility of most PalB fusions, such as MBP-PalB, T7PK-PalB, and DsbA-PalB, was significantly higher than non-fused PalB, particularly in Origami B(DE3). However, only DsbA tag successfully induced PalB activity upon the fusion with PalB even though the solubility of most PalB fusions was enhanced.

It is interesting to observe that functional expression of PalB was enhanced using DsbA either as an independent folding factor or as a fusion tag. DsbA was previously explored as a fusion tag for enhancing

the functional expression of several proteins, such as bovine enterokinase [246], proinsulin [247], CD8+ cytotoxic T lymphocyties (CTL) [248], and human perforin [249], with an assumption that DsbA tag can potentially solubilize and stabilize its fusion partner. Nevertheless, DsbA has not been as common as other fusion tags due to the lack of proper affinity chromatography for purification of fusion proteins. The role of DsbA tag for improving the solubility of its fusion partner has been demonstrated using the DsbA mutant lacking its oxidoreductase activity [250]. However, it is still unclear whether DsbA tag could assist disulfide bond formation for its fusion partner via the oxidoreductase activity.

The current result that DsbA tag could enhance both the solubility and PalB activity (possibly via correct disulfide bond formation) upon the fusion with PalB strongly suggests that both protein folding and disulfide bond formation play critical roles for developing PalB activity. The DsbA-PalB expression performance observed in BL21(DE3) should purely reflect the effect of DsbA tag on solubilization, whereas the expression performance observed in Origami B(DE3) should reflect the synergistic effect on both solubilization and disulfide bond formation. Note that solubilization of PalB fusions did not necessarily result in the development of PalB activity which could be associated with correct disulfide bond formation. In other words, the use of DsbA tag can enhance both the solubility and disulfide bond formation for functional expression of PalB in Origami B(DE3) though the mechanism by which DsbA tag can assist correct formation of disulfide bonds for its fusion partner of PalB remains unclear.

The purification for PalB is not difficult as showed in the purification of PalB from GST-PalB. Before and after the cleavage, GST-PalB and PalB all showed the PalB activity, although the PalB activity with the tag could be underestimated. If the fusion tag was not interfered with PalB activity too much, it could be more benefit to directly use PalB fusion as an enzyme in the biodiesel production.

4.5 Abstract of the submitted paper

Functional expression of lipase B from *Pseudozyma antarctica* (PalB) in the cytoplasm of *Escherichia coli* was explored using BL21(DE3) and its mutant derivative Origami B(DE3) as the host. Coexpression of DsbA was found to be effective in enhancing PalB expression and such improvement was more pronounced in Origami B(DE3), suggesting that both folding and disulfide bond formation could be the major factors limiting PalB expression. Fusion tag technique was also explored by constructing several PalB fusions for evaluation of their expression performance. While the solubility was enhanced for most PalB fusions, only DsbA tag was effective in boosting PalB activity possibly via both enhanced solubility and correct disulfide bond formation. Our results suggest that solubilization of PalB fusions did not necessarily result in the development of PalB activity which could be closely associated with correct disulfide bond formation.

Keywords: disulfide bond formation, fusion tag, *Escherichia coli*, lipase, recombinant protein production

Chapter 5*

Effect of Folding Factors in Rescuing Unstable Heterologous PalB to Enhance its Overexpression in the Periplasm of *Escherichia coli*

5.1 Introduction

The bacterium *Escherichia coli* remains popular as a workhorse for recombinant protein production. Formation of active proteins involves a series of intracellular steps for gene expression (e.g., transcription and translation) and posttranslational processing (e.g., translocation, folding). While strategies based on boosting gene expression steps are often adopted to enhance recombinant protein production [31, 38, 92], it is not uncommon that any of the overexpressed polypeptide species, including precursors, intermediates, and mature gene products, cannot be suitably processed during posttranslational processing, resulting in protein misfolding and cellular stress. The cell defense system is activated in response to the stress, either by degradation or refolding of the misfolded gene products [136]. Misfolded proteins are often degraded by proteases in quality control processes [134, 135]. However, such protein degradation is still possible in protease-deficient strains [251-253], implying the existence of protease-independent pathways in *E. coli* [254]. When the amount of misfolded protein exceeds the cell's degradation capacity, insoluble protein aggregates known as inclusion bodies form in the cytoplasm and/or periplasm [134, 136, 255]. This brings up a technical issue that, in addition to improving the efficiency of each gene expression step, a 'balanced' protein synthesis flux throughout these steps should be properly maintained to avoid the degradation and accumulation of any protein species along the protein formation pathway.

Lipases (triacylglycerol acylhydrolase, E.C.3.1.1.3) are enzymes that catalyze the synthesis or hydrolysis of glycerol esters (e.g., triglycerides), such as fats and oils [14]. Microbial cells, including bacteria, fungi, and yeasts, are the major source for lipases with various enzymatic characteristics, such as synthetic and hydrolytic activity, stability, substrate specificity, for numerous applications in organic chemistry and biotechnology [20, 21]. Our interest in lipase arises from its application on biotransformation as an alternative to chemical conversion for biodiesel production. Lipase B from *Pseudozyma antarctica* (PalB) is a cold active microbial lipase with the transesterification activity and its

* This chapter is based on a paper "Y. Xu, D. Lewis, C. P. Chou, Effect of Folding Factors in Rescuing Unstable Heterologous Lipase B to Enhance its Overexpression in the Periplasm of *Escherichia coli*. *Applied Microbiology and Biotechnology*. 2008" (accepted).

corresponding immobilized form (e.g., Novozyme 435 commercially available from Novozyme A/S, Bagsvaerd, Demark) has been recognized as the most common biocatalyst for biodiesel production. While the biocatalyst is gaining more attention and has a potential to outperform chemical catalysts, extensive application remains impractical due to its high cost and several technical issues (e.g., inactivation by alcohol). To make the bioprocess economically feasible, there is a motivation to overexpress PalB.

Previously, recombinant PalB was expressed heterologously in *Aspergillus oryzae* [25], *Pichia pastoris* [224], and *Saccharomyces cerevisiae* [27]. Given the popularity of *E. coli* as a workhorse for recombinant protein production, functional expression of PalB in this host was not demonstrated until recently [28, 29] and the expression performance stands improvement. A major technical issue limiting PalB expression in *E. coli* is associated with its intracellular misfolding. PalB is a cold active lipase destined to be secreted extracellularly in the original psychrophilic and eukaryotic source of *P. antarctica*. Consequently, the intracellular overexpression might prevent it from being structurally adaptable to the mesophilic and prokaryotic expression system of *E. coli*, resulting in protein misfolding. In addition, wild-type PalB has three intermolecular disulfide bonds potentially associated with its folding process though they might not be strictly required for developing the bioactivity. Finally, cell growth is often arrested due to deteriorated cell physiology associated with high-level expression and PalB misfolding. Nevertheless, from an application standpoint, developing effective PalB expression strategies in *E. coli* is still worthy of exploration.

In contrast to the cytoplasm where all the proteins are first synthesized, the periplasm offers several advantages as a destination for heterologous protein expression, such as its oxidizing environment suitable for disulfide bond formation and a less amount of contaminating proteins in this compartment so that downstream purification would be facilitated [88, 90]. Due to the possible impact of disulfide bond formation on PalB bioactivity, it would be advantageous to express PalB in the periplasm. In this study, a key factor limiting functional expression of PalB in the periplasm of *E. coli* was identified. The use of various periplasmic folding factors to eliminate the limiting factor was explored to improve PalB expression performance.

5.2 Material and Methods

5.2.1 Strains and Plasmids

The strains, plasmids and oligonucleotides used in this study are summarized in Table 5.1 and briefly described here. BL21(DE3) was used as the host strain for PalB expression. Molecular cloning was performed according to standard protocols [61] using DH5α as the cloning host. *P. antarctica* strain

ATCC 32657 containing the *palB* source gene was purchased from ATCC (Manassas, VA, USA). *P. antarctica* chromosomal DNA was extracted and purified by using DNeasy® tissue Kit (Qiagen, Valencia, CA, USA). PCR was performed in an automated thermal cycler (GeneAmp® PCR System 9700; Applied Biosystems, Foster City, CA, USA). The PCR product was purified by StrataPrep® PCR purification kit (Stratagene). Restriction enzymes were purchased from New England Biolabs (Ipswich, MA, USA). Plasmid DNAs were purified using a spin column kit (Clontech, Mountain View, CA, USA). Plasmid transformation was carried out using an electroporator (Bio-Rad, Hercules, CA, USA). All the chemicals used in this study were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise specified.

Table 5.1 Strains, plasmids, and oligonucleotides.

Strain, plasmid, and oligonucleotide	Relevant genotype or phenotype[a]	Source and reference
E. coli		
BL21(DE3)	*E. coli* B, F^- *ompT [dcm][lon] hsdS*$_B$(r_B^- m_B^-) *gal* (λ cIts857 *ind*1 Sam7 *nin*5 *lacUV5-T7* gene 1)	Novagen [256]
DH5α	F'/*(ϕ80 lacZ ΔM15) Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17* (r_K^- m_K^+) *phoA supE44 λ-thi-1 gyrA96 relA1*	Lab stock [257]
P. antarctica		
ATCC 32657	The source of *palB* gene	ATCC
Plasmid		
pAR3	Expression vector, P_{araB}, Ori (pACYC184), Cm^R	Lab stock [233]

pARDegP	P_{araB}::*degP*, Ori (pACYC184), Cm^R	This lab [187]
pARDegP$_{S210A}$	P_{araB}::$degP_{S210A}$, Ori (pACYC184), Cm^R	This lab [187]
pARFkpA	P_{araB}::*fkpA*, Ori (pACYC184), Cm^R	This lab [195]
pARDsbA	P_{araB}::*dsbA*, Ori (pACYC184), Cm^R	This study
pARDsbC	P_{araB}::*dsbC*, Ori (pACYC184), Cm^R	This study
pET-20b(+)	Expression vector, P_{T7}::$pelB_{signal}$, Ori (pBR322), Ap^R	Lab stock Y. P. Chao
pETG	P_{T7}::$pelB_{signal}$::*palB*, Ori (pBR322), Ap^R	This study
pETG-50A	Destination vector, P_{T7}::*lac*Op::*dsbA*::6×*his*, Ori (pBR322), Ap^R, Cm^R	EMBL
pETM-80	Expression vector, P_{T7}::*dsbC*::6×*his*, Ori (pBR322), Kn^R	EMBL [72]
pETL	P_{T7}::*palB*, Ori (pBR322), Ap^R	This study
pTUM4	P_{fkpA}::*fkpA*::*surA*, P_{dsbA}::*dsbA*::*dsbC*, Ori (pACYC184), Cm^R	A. Skerra [258]
pG-KJE8	P_{araB}::*dnaK/J*::*grpE*, $P_{zt\text{-}lp}$::*groEL/ES*, Ori (pACYC184), Cm^R	T. Yura [76]
pG-Tf3	P_{araB}::*tig*, $P_{zt\text{-}lp}$::*groEL/ES*, Ori (pACYC184), Cm^R	T. Yura [76]
Oligonucleotide		
P10	5'-GG*CCATGG*GTCTACCTTCCGGTTCGG-3'	This study
P11	5'-*CTGAATTC*TCAGGGGGTGACGATGCCGGAGCAGG-3'	This study
P17	5'-TG*CATATG*CTACCTTCCGGTTCGG-3'	This study

Pf_{DsbA}/Pr_{DsbA}	5'-*CC*ATGG*A*AAAAGATTTGGCTGGCGC-3' and 5'-*AAGCTTTCA*TTTCTCGCTTAAGTATTTC-3'; primer pair for amplification of *dsbA*	This study
Pf_{DsbC}/Pr_{DsbC}	5'-*CC*ATGG*A*GAAAGGTTTTATGTTG-3' and 5'-*AAGCTTTCA*ACCAGAACCACTAGT-3'; primer pair for amplification of *dsbC*	This study

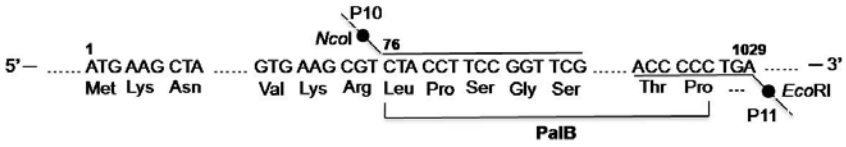
[a] Designed restriction sites are underlined and the introduced mutations are in italics.

The two DNA fragments containing the leaderless *palB* gene were PCR-amplified using *Pfu* DNA polymerase (Stratagene), the primer pairs of P10/P11 and P17/P11, respectively (Table 5.1), and *P. antarctica* chromosomal DNA as the template. The PCR products were cloned into PCR-Script® (Stratagene), respectively. DNA sequencing was performed to ensure no mutations occurred during PCR. The *palB*-containing DNA fragments were obtained by digestion of the above PCR-Script® derivatives with *Nco*I/*Eco*RI and *Nde*I/*Eco*RI, respectively, gel-purified and subcloned into the corresponding restriction sites of pET-20b(+), resulting in a translational fusion of pETG (Figure 5.1A) for periplasmic expression and a transcriptional fusion of pETL (Figure 5.1B) for cytoplasmic expression. Due to the design of the restriction site in the sense primers, the N-terminus of the heterologously expressed PalB is expected to have two extra amino acids (e.g., Met and Gly) for pETG and one extra amino acid (e.g., Met) for pETL. The coexpression plasmids containing a selection of cytoplasmic chaperone gene(s) were derivatives of pAR3 [233]. pG-Tf3 [76] contains *tig* and *groES/EL*, fused with the *araB* and *zt-lp* promoters, respectively. pG-KJE8 [76] contains the *dnaK/J-grpE* and *groES/EL*, fused with the *araB* and *zt-lp* promoters, respectively. Except pTUM4 [258], the coexpression plasmids containing a selection of periplasmic folding factor gene(s) were also derivatives of pAR3. Two of them, e.g., pARDsbA and pARDsbC, were constructed in this study. The *dsbA* gene was PCR-amplified using *Pfu* DNA polymerase, the primer pair Pf_{DsbA}/Pr_{DsbA}, and pETG-50A (EMBL) as the template. The *dsbC* gene was PCR-amplified using *Pfu* DNA polymerase, the primer pair Pf_{DsbC}/Pr_{DsbC}, and pETM-80 [72] as the template. The PCR products were cloned into PCR-Script®. DNA sequencing was performed to ensure no mutations occurred during PCR. The *dsbA*-containing and *dsbC*-containing DNA fragments, both flanked with *Nco*I/*Hin*dIII, were subcloned into the corresponding sites of pAR3 to form pARDsbA and pARDsbC, respectively. The design of the *Nco*I site in the sense primers resulted in a change in the second amino acid of the signal peptide (e.g., Lys→Glu) for both DsbA and DsbC, and the mutations did not appear to affect translocation. pARDegP [187], pARDegP$_{S210A}$ [187], and pARFkpA [195] contain the

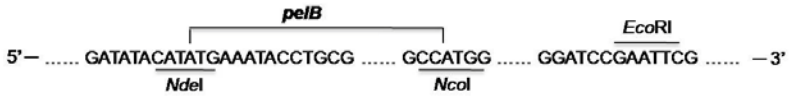
gene of *degP*, *degP*$_{S210A}$, and *fkpA*, respectively fused with the *araB* promoter. pTUM4 harbors two artificial dicistronic operons: one with the structural genes of *dsbA* and *dsbC* under the control of the constitutive *dsbA* promoter and the other one with the structural genes of *fkpA* and *surA* under the control of the constitutive *fkpA* promoter [258]. All the folding-factor-gene-containing plasmids have a pACYC184 replication origin and a chloramphenicol-resistant (Cm^R) marker. Therefore, they are compatible with pETG, which has a pBR322 replication origin and an ampicillin-resistant (Ap^R) marker.

A

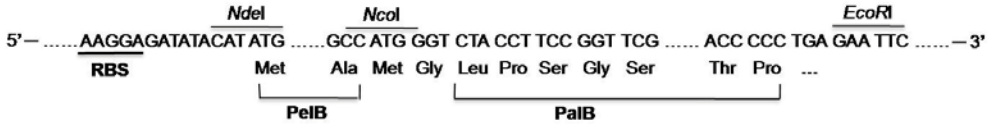
Wild-type *palB gene*



MCS in pET-20b(+)

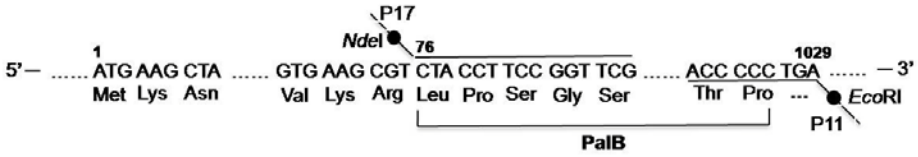


***palB* in pETG**

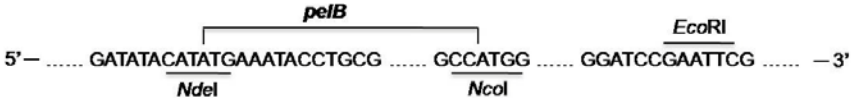


B

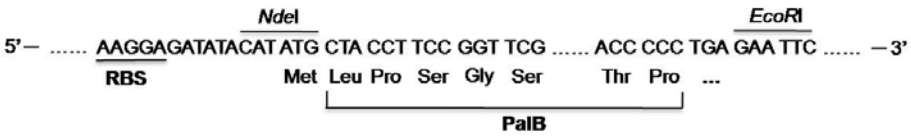
Wild-type *palB gene*



MCS in pET-20b(+)



***palB* in pETL**



C

the DNA sequence's differences between vector pETG and pETL from T7 promoter to T7 terminator portion

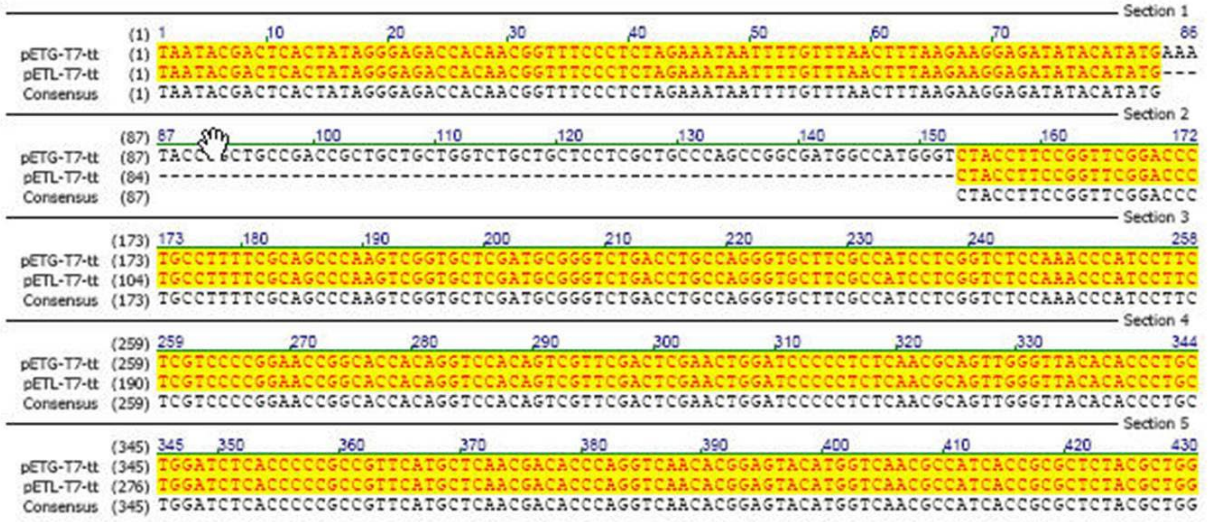


Figure 5.1 Construction of two PalB expression plasmids of pETG (Panel A: periplasmic expression) and pETL (Panel B: cytoplasmic expression), and comparison of their DNA sequence's differences from T7 promoter to T7 terminator portion (Panel C).The DNA region from nucleotides 76 to 1029 of the wild-type *palB* gene on the chromosome of *P. antarctica* ATCC 32657 was PCR-amplified with primer pairs of P10/P11 and P17/P11, respectively. The PCR products were cloned into pET-20b(+) using *Nco*I/*Eco*RI and *Nde*I/*Eco*RI, respectively, to form a translational fusion of pETG and a transcriptional fusion of pETL. MCS, multiple cloning sites. RBS, ribosome binding site.

5.2.2 Cultivation and Analytic Methods

The cultivation and analytical procedures were identical with the procedures described in Section 3.2.2 and 3.2.3.

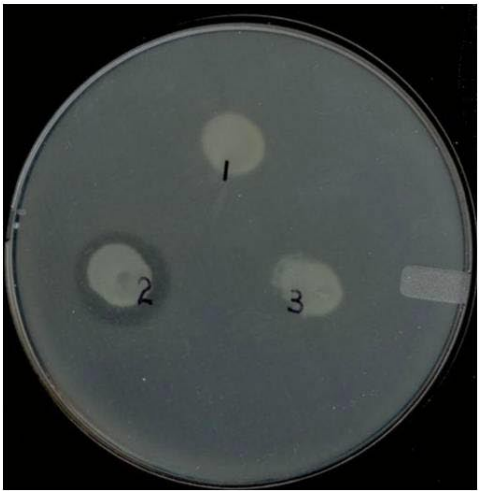
5.3 Results

5.3.1 Comparison of Cytoplasmic and Periplasmic Expression of PalB

To compare cytoplasmic and periplasmic expression of PalB in parallel, two similar expression plasmids of pETL and pETG were constructed (Figure 5.1). They both contain the *palB* gene from *P. antarctica* ATCC 32657 whose expression is under the regulation of the T7 promoter. However, pETG contains the *pelB* signal peptide ahead of the *palB* gene for secretion of PalB into the periplasm, whereas pETL does not. With this design, the cytoplasmic and periplasmic PalB will contain an extra residue (e.g., Met) and two extra residues (e.g., Met and Gly), respectively, in the N-terminus of the expressed gene product.

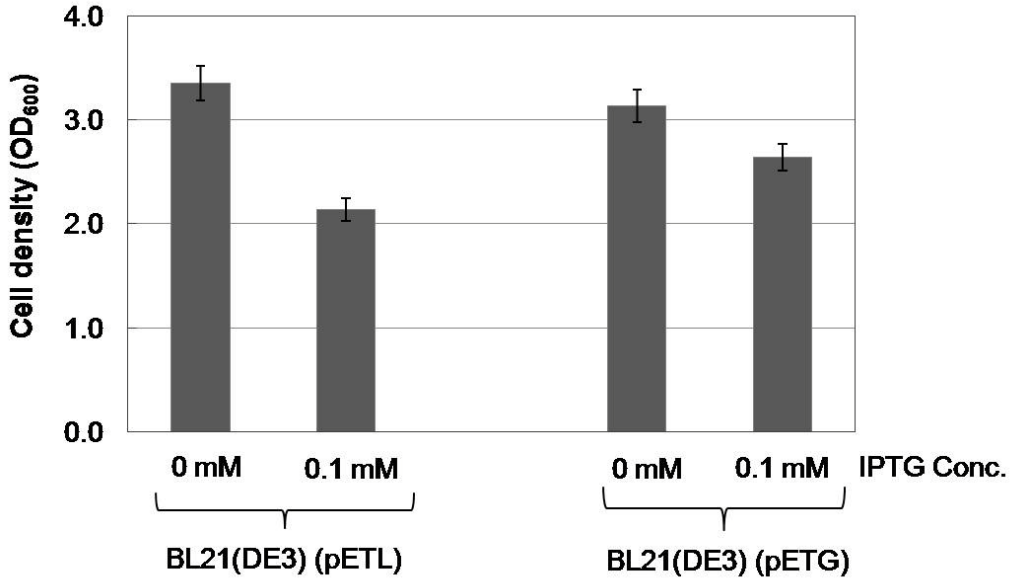
Using BL21(DE3) as the host to harbor these two expression plasmids, qualitative visualization of PalB activity of the recombinant cells was conducted on the tributyrin plate and the results are summarized in Figure 5.2A. In general, the area and transparency of the halo potentially correlate with the expressed PalB activity corresponding to the applied sample. A visible halo was observed for BL21(DE3) harboring pETL, indicating PalB was functionally expressed in the cytoplasm of *E. coli*. Surprisingly, BL21(DE3) harboring pETG did not show a visible halo though functional PalB was expected to be formed in the periplasm.

A

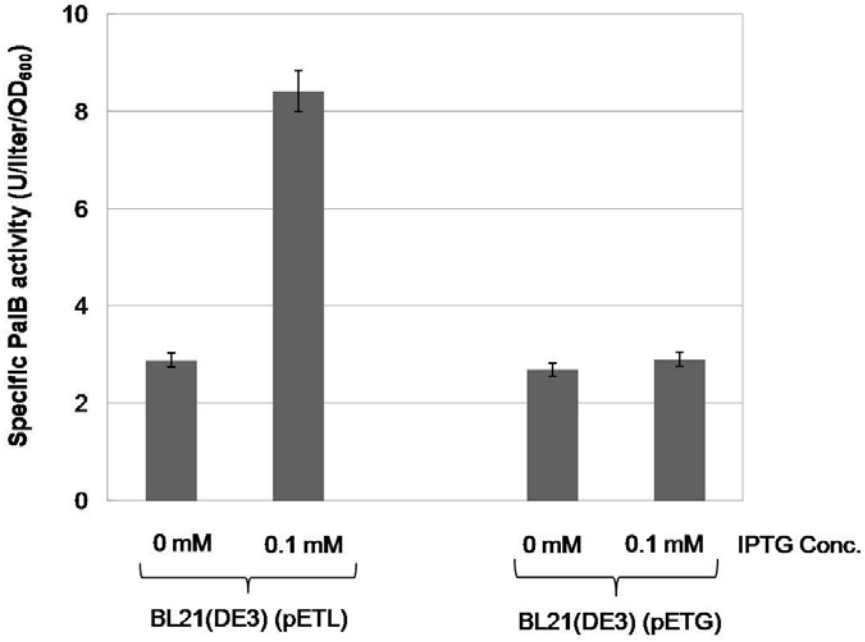


1	BL21(DE3)
2	BL21(DE3) (pETL)
3	BL21(DE3) (pETG)

B


4.0
3.0
2.0
1.0
0.0
Cell density (OD600)
0 mM
0.1 mM
0 mM
0.1 mM
IPTG Conc.
BL21(DE3) (pETL)
BL21(DE3) (pETG)

C


10
8
6
4
2
0
Specific PalB activity (U/liter/OD600)
0 mM
0.1 mM
0 mM
0.1 mM
IPTG Conc.
BL21(DE3) (pETL)
BL21(DE3) (pETG)

D

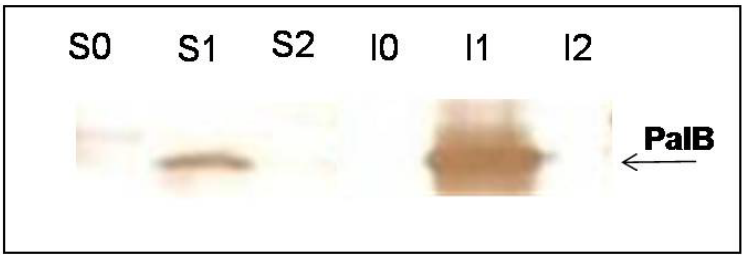


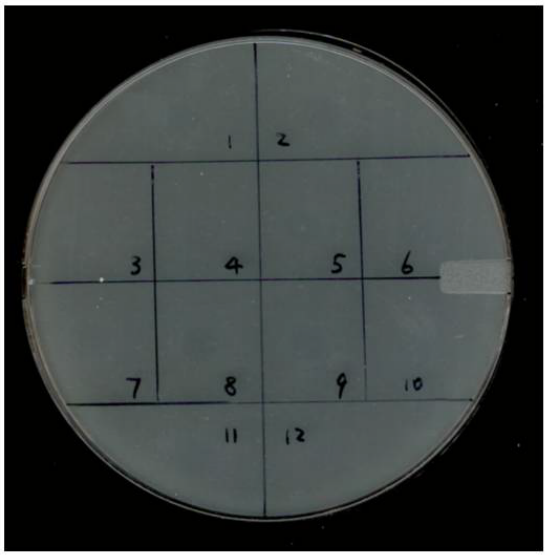
Figure 5.2 Comparison of PalB expression performance. Panel A: Qualitative visualization of PalB activity using the tributyrin plate. 1/ BL21 (DE3); 2/ BL21(DE3) (pETL); 3/ BL21(DE3) (pETG). Panel B: cell density and Panel C: specific PalB activity for cultivation of BL21(DE3) (pETL) and BL21(DE3) (pETG) at 28 °C. Panel D: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various culture samples. 0/ BL21(DE3); 1/ BL21(DE3) (pETL); 2/ BL21(DE3) (pETG). The error bar in Panels B and C represents a range of the data variance.

The PalB expression performance for BL21(DE3) harboring pETL and BL21(DE3) harboring pETG is summarized in Figure 5.2B~D. Though PalB was functionally expressed in BL21(DE3) harboring pETL, PalB activity was relatively low and most of the expressed gene product accumulated in cells as inclusion bodies. The non-productive pathway was favored for PalB expression possibly due to the reductive environment of the cytoplasm unsuitable for disulfide bond formation. In addition, cell growth was severely arrested due to the physiological deterioration upon induction. In principle, expression of PalB in the oxidative periplasm should not be worse, if not better, than that in the reductive cytoplasm. However, PalB activity was minimally detected for BL21(DE3) harboring pETG even though cell growth was not arrested upon induction. In addition, no PalB-related gene product was detected in soluble or insoluble fraction when Western blotting was conducted. The result suggests two possible expression issues, e.g., (i) an ineffective translation or (ii) the instability of the PalB gene product in the cytoplasm and/or the periplasm. In order to identify the factor(s) limiting PalB expression in the periplasm, coexpression of cytoplasmic and periplasmic folding factors to rescue the potentially unstable gene products was conducted.

5.3.2 Effect of Folding Factors on Rescuing PalB Gene Product

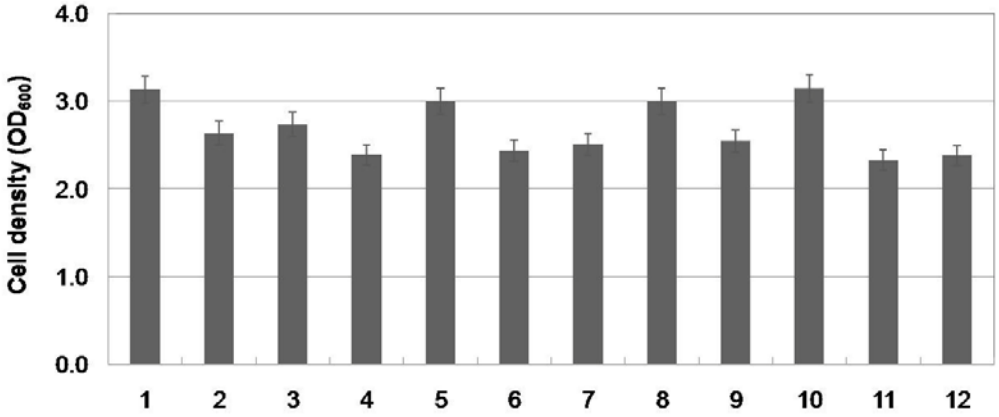
A selection of cytoplasmic chaperones (e.g., the TF, GroEL/ES, and DnaK/J-GrpE) and their combinations were coexpressed with the periplasmic expression of PalB with pETG as the expression plasmid and the results are summarized in Figure 5.3. Apparently, coexpression of any of these cytoplasmic chaperones or their combinations hardly improved PalB expression performance. PalB activity was not qualitatively visualized on the tributyrin plate. In addition, similar to the control culture of BL21(DE3) harboring pETG, neither PalB activity was increase (Figure 5.3C) nor PalB-related gene products were detected (Figure 5.3D), although the chaperones or their combinations were expressed (Figure 5.3E). The results suggest that PalB expression using pETG as the expression plasmid might not be limited by the instability of the PalB gene products in the cytoplasm.

A



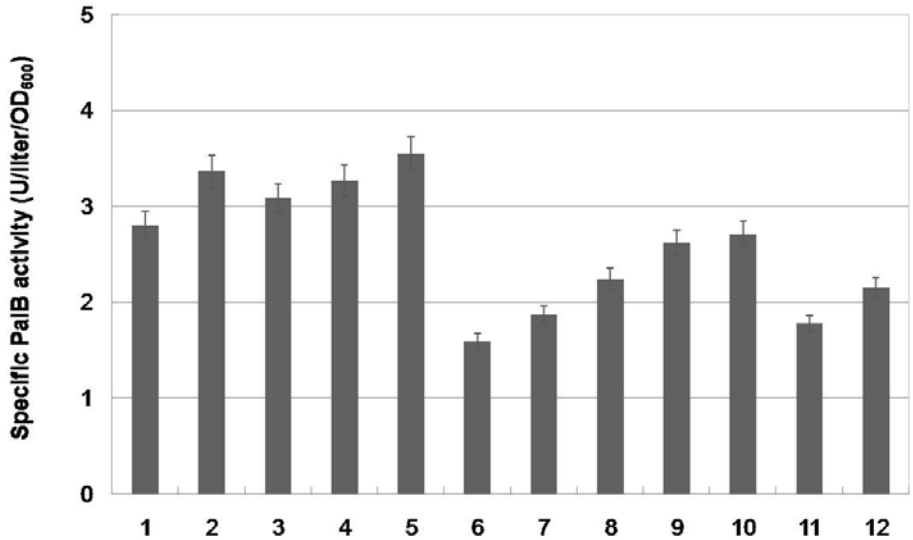
	1	2	3	4	5	6	7	8	9	10	11	12
PalB		+		+	+	+	+		+	+	+	+
Trigger factor					+		+					
DnaK/J										+		+
GroEL/ES						+	+				+	+

B



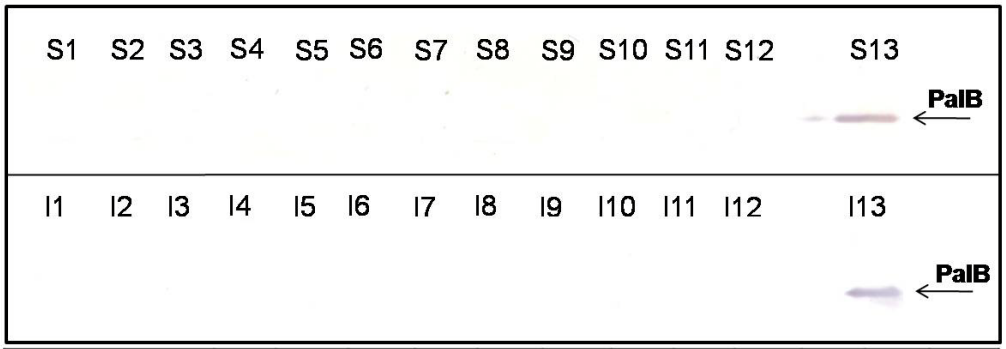
PalB		+		+	+	+	+		+	+	+	+
Trigger factor					+		+					
DnaK/J										+		+
GroEL/ES						+	+				+	+

C



PalB		+		+	+	+	+		+	+	+	+
Trigger factor					+		+					
DnaK/J										+		+
GroEL/ES						+	+				+	+

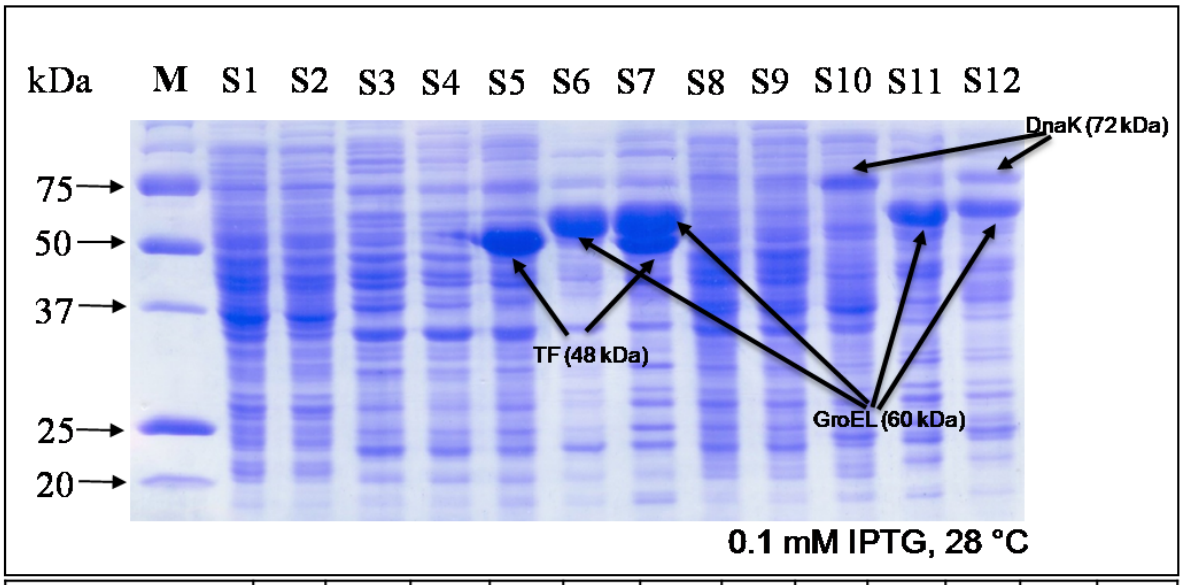
D



	1	2	3	4	5	6	7	8	9	10	11	12
PalB		+		+	+	+	+		+	+	+	+
Trigger factor					+		+					
DnaK/J										+		+
GroEL/ES						+	+				+	+

#13 is positive control sample from BL21(DE3) (pETL), 28 °C, 0.1 mM IPTG (Panel D in Figure 4.2 #1, p.61).

E

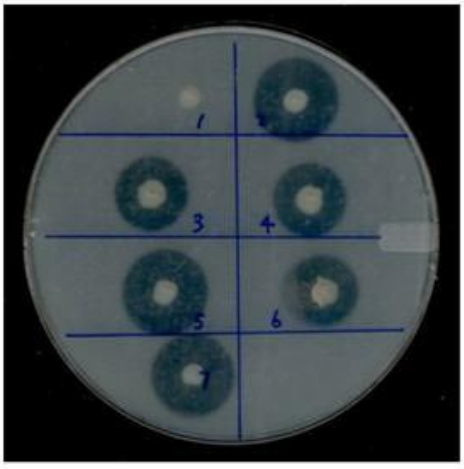


	1	2	3	4	5	6	7	8	9	10	11	12
PalB		+		+	+	+	+		+	+	+	+
Trigger factor					+		+					
DnaK/J										+		+
GroEL/ES						+	+				+	+

Figure 5.3 Effect of cytoplasmic chaperones on PalB coexpression performance using BL21(DE3) (pETG). Panel A: Qualitative visualization of PalB activity using the tributyrin plate. Panel B: cell density for cultivation of various expression systems at 28 °C. Panel C: specific PalB activity for cultivation of various expression systems at 28 °C. Panel D: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various culture samples. Panel E: SDS-PAGE results of the soluble fractions for various culture samples. Expression systems: 1&2/ BL21(DE3) (pETG); 3~7/ BL21(DE3) (pETG, pG-Tf3); 8~12/ BL21(DE3) (pETG, pG-KJE8). Inducer supplementation: 1, 3, 8/ no inducer; 2, 4, 9/ IPTG; 5&10/ IPTG and arabinose; 6&11/ IPTG and tetracycline; 7&12/ IPTG, arabinose, and tetracycline. M, protein molecular weight marker. The error bar in Panels B and C represents a range of the data variance.

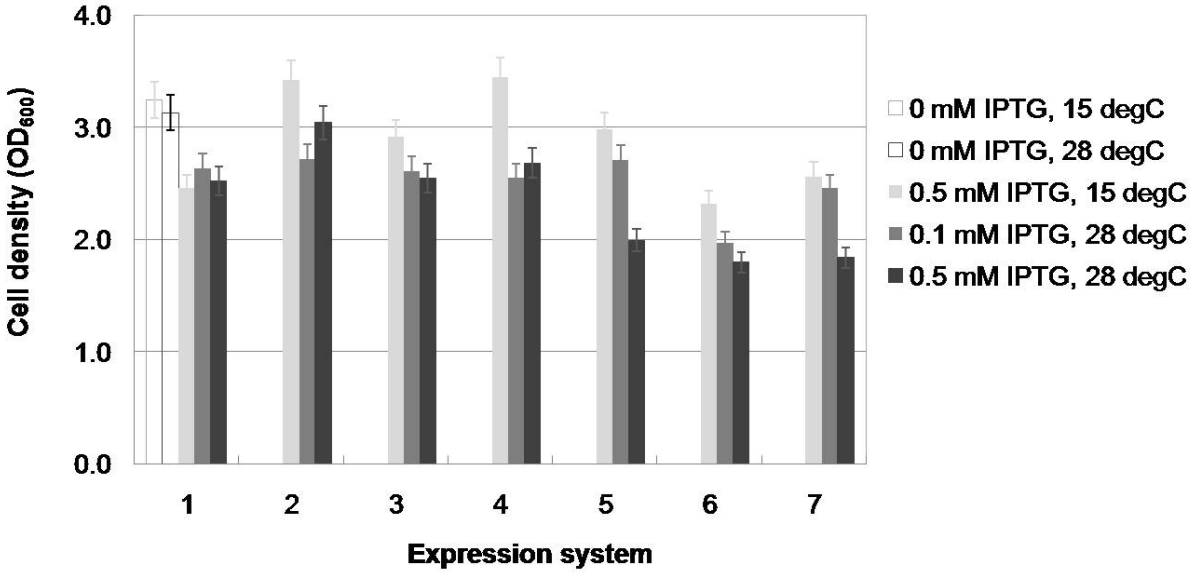
In contrast, PalB expression performance was significantly improved upon coexpression of a number of periplasmic folding factors, including DegP (and its mutant variant of $DegP_{S210A}$), FkpA, DsbA, DsbC, and a cocktail of SurA, FkpA, DsbA, and DsbC, and the results are summarized in Figure 5.4. Halos developed on the tributyrin plate for all the cell colonies of BL21(DE3) harboring pETG and folding-factor-coexpression plasmids, indicating that PalB was functionally expressed in the periplasm of *E. coli* (Figure 5.4A). Note that, for BL21(DE3) harboring pETG, cell growth was arrested to some extent upon IPTG-induction (Figure 5.4B). This physiological burden could be alleviated by coexpression of DegP and $DegP_{S210A}$ and such effect was particularly observable at 15 °C. More importantly, the specific PalB activity was significantly enhanced by approximately 6-fold (30 U/liter/OD_{600} for DegP coexpression and 33 U/liter/OD_{600} for $DegP_{S210A}$ coexpression vs. 5 U/liter/OD_{600} for the control culture grown at 15 °C and induced with 0.5 mM IPTG), resulting in a significant improvement in the volumetric PalB activity. A similar improvement was observed when cultivations were performed at 28 °C, though the improving effect was less significant. On the other hand, coexpression of DsbA or a folding factor cocktail (SurA, FkpA, DsbA, and DsbC) had approximately the same or even better improving effect compared to the culture with DegP coexpression, particularly in terms of the specific PalB expression level. In contrast, the enhancement on PalB expression performance was limited upon coexpression of FkpA or DsbC. Compared to BL21(DE3) harboring pETG, the growth arrest became even more severe upon coexpression of DsbC and the PalB band became invisible for one culture condition (e.g., 0.1 mM IPTG at 28 °C in Figure 5.3D), suggesting that the rescuing ability for DsbC might be limited.

A

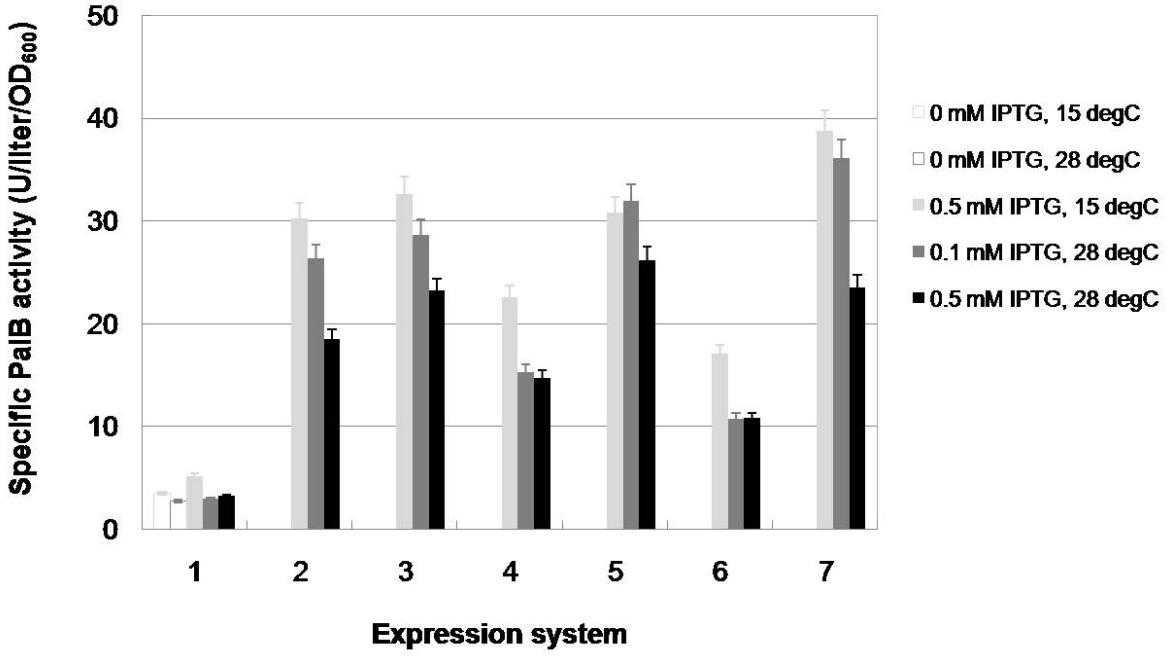


No	Expression system
1	BL21(DE3)(pETG)
2	BL21(DE3)(pETG, pARDegP)
3	BL21(DE3)(pETG, pARDegP$_{S210A}$)
4	BL21(DE3)(pETG, pARFkpA)
5	BL21(DE3)(pETG, pARDsbA)
6	BL21(DE3)(pETG, pARDsbC)
7	BL21(DE3)(pETG, pTUM4)

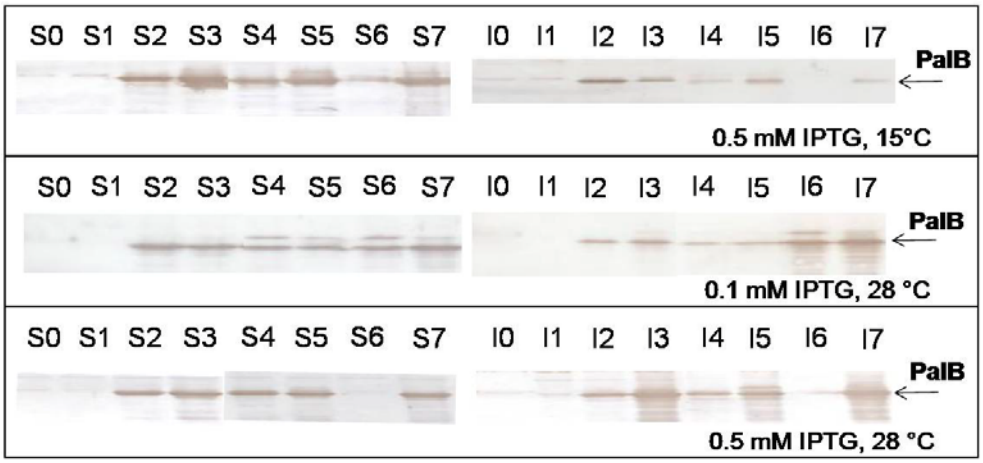
B


4.0
3.0
2.0
1.0
0.0
Cell density (OD600)
1
2
3
4
5
6
7
Expression system
0 mM IPTG, 15 degC
0 mM IPTG, 28 degC
0.5 mM IPTG, 15 degC
0.1 mM IPTG, 28 degC
0.5 mM IPTG, 28 degC

C


50
40
30
20
10
0
Specific PalB activity (U/liter/OD600)
1
2
3
4
5
6
7
Expression system
0 mM IPTG, 15 degC
0 mM IPTG, 28 degC
0.5 mM IPTG, 15 degC
0.1 mM IPTG, 28 degC
0.5 mM IPTG, 28 degC

D



E

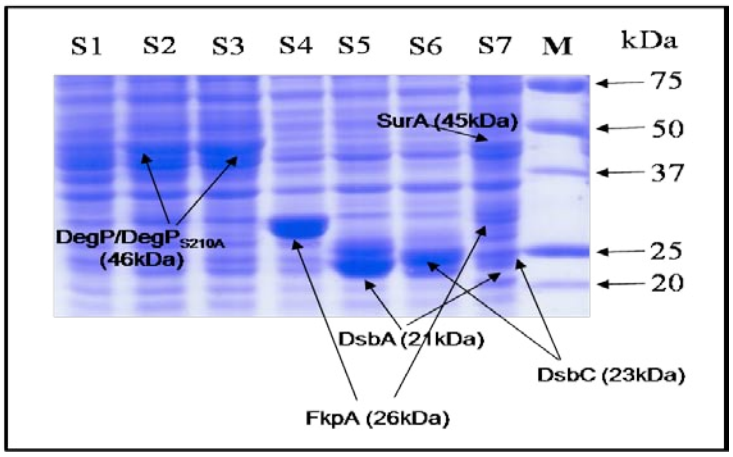


Figure 5.4 Effect of periplasmic folding factors on PalB coexpression performance using BL21(DE3) (pETG). Panel A: Qualitative visualization of PalB activity using the tributyrin plate. Panel B: cell density for cultivation of various expression systems at 15 and 28 °C. Panel C: specific PalB activity for cultivation of various expression systems at 15 and 28 °C. Panel D: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various culture samples. Panel E: SDS-PAGE results for the soluble (S) fractions for various culture samples. Expression systems: 0, 1/ BL21(DE3) (pETG); 2/ BL21(DE3) (pETG, pARDegP); 3/ BL21(DE3) (pETG, pARDegP$_{S210A}$); 4/ BL21(DE3) (pETG, pARFkpA); 5/ BL21(DE3) (pETG, pARDsbA); 6/ BL21(DE3) (pETG, pARDsbC); 7/ BL21(DE3) (pETG, pTUM4). Inducer supplementation: 1, 7/ IPTG; 2~6/ IPTG and arabinose. M, protein molecular weight marker. The error bar in Panels B and C represents a range of the data variance.

The enhancement in PalB expression performance using periplasmic folding factors was clearly demonstrated using Western blotting (Figure 5.4D). The PalB band was slightly visible at 15 °C and was hardly visible at 28 °C for BL21(DE3) harboring pETG, implying that PalB might be structurally unstable in the periplasm. The PalB band signal was significantly amplified and became clearly visible when various periplasmic folding factors were coexpressed, suggesting that unstable PalB was rescued by periplasmic chaperones. Note that PalB was detected in both soluble and insoluble fractions, suggesting that misfolding of PalB still occurred in the periplasm even in the presence of chaperones. In some coexpression cases (e.g., Figure 5.4D, lanes 4, 5, 6 with 0.1 mM IPTG at 28 °C), PalB precursor (presumably in the cytoplasm) was visible in both soluble and insoluble fractions.

5.4 Discussion

Biochemical and genetic strategies have been developed to enhance various gene expression steps for heterologous protein production in *E. coli*, including transcription, translation, posttranslational processing (e.g., translocation, periplasmic processing, and folding). To effectively enhance expression performance, the limiting expression step(s) will need to be identified first. This leads to a generic concept that not only the efficiency of each gene expression step should be maximized but also the expression flux leading to the mature protein should be balanced to avoid any potential diversion into the non-productive pathway.

Up to now, it is well recognized that overexpression of bioactive PalB in *E. coli* remains challenging primarily due to intracellular misfolding which demolishes its bioactivity. In this study, PalB was functionally expressed in the cytoplasm of *E. coli* using pETL as the expression plasmid although PalB inclusion bodies in a large amount were also observed. The result suggests that PalB was stable in the cytoplasm and disulfide bond formation might not be strictly required to develop the bioactivity. However, PalB expression performance was poor using pETG as the expression plasmid though the periplasm was presumed to be a more suitable compartment for functional expression of PalB due to its oxidative environment facilitating disulfide bond formation. The minor difference in DNA construction of the two similar PalB expression plasmids (e.g., pETL and pETG) and the major difference in PalB expression performance using these two plasmids leads to a hypothesis that either (or even both) of cytoplasmic PalB precursor (e.g., prePalB) and periplasmic PalB might be unstable. Therefore, cytoplasmic or periplasmic folding factors were used to identify and even rescue the potentially unstable PalB species.

Several cytoplasmic chaperones, including TF, DnaK/J-GrpE, and GroEL/ES, were coexpressed, either separately or in combination, with PalB expression using pETG. No major improvement was observed, implying that the factor limiting PalB expression might not be located in the cytoplasm. In contrast, coexpression of several periplasmic folding factors, in particular DegP, DsbA, and a cocktail of SurA, FkpA, DsbA, and DsbC, could not only rescue the unstable PalB species in the periplasm but also significantly improve PalB expression performance. It has been widely reported that several periplasmic folding factors can interact with secreted proteins to enhance functional expression via improved translocation and/or polypeptide stabilization [88, 96, 173, 174, 245].

DegP primarily acts as a serine protease for breakdown of aberrant periplasmic proteins arising upon extracytoplasmic stress [181, 182] and its expression can be activated via both σ^E and Cpx pathways [121]. Wild-type DegP is a periplasmic heat-shock protein with both protease and chaperone activities, whereas $DegP_{S210A}$ is a mutant variant with its protease activity being inactivated but chaperone activity being retained [183]. Coexpression of *degP* was applied to relieve the extracytoplasmic stress upon overexpression of several periplasmic gene products, such as alkaline phosphatase (PhoA) [184], DsbA'-PhoA [259], MalS [183], OmpF mutants [260], maltose-binding protein (MBP or MalE) variants [179, 261], and penicillin acylase [187]. In this study, DegP appears to be effective in stabilizing newly secreted and presumably unfolded PalB in the periplasm. Such improvement in PalB expression performance remained approximately the same when $DegP_{S210A}$ was coexpressed, suggesting that this improving effect was likely mediated by DegP chaperone activity. The stabilization effect associated with chaperone activity was further demonstrated by coexpression of other periplasmic folding factors.

FkpA is another periplasmic heat-shock protein with both *cis/trans* peptidyl-prolyl isomerase (PPIase) and chaperone activities [189, 190, 192]. Expression of *fkpA* is activated via the σ^E pathway [127, 199] and is capable of reducing the σ^E-dependent response induced by misfolded proteins in the periplasm [179]. Coexpression of *fkpA* was applied to improve the functional production of recombinant antibody fragments [193, 194] and penicillin acylase [195]. In this study, the above improving effect on functional PalB expression was also observable when FkpA was coexpressed, supporting the argument regarding the role that chaperones could play.

Many eukaryotic proteins, including PalB, contain disulfide bonds which often mediate the stabilization of folding conformation. It would be advantageous for these proteins to be expressed in the oxidative periplasm of *E. coli* to facilitate disulfide bond formation. DsbA and DsbC are two Dsb-family members in the periplasm and play important roles in disulfide bond formation [40]. DsbA catalyzes direct bridging of disulfide bonds in newly translocated proteins, whereas DsbC assists subsequent

rearrangement of aberrant disulfide bonds [40]. Though with distinctively major activities, the two proteins might share biological functions under certain conditions [104, 215-217]. Coexpression of DsbA and/or DsbC enhanced functional expression of several recombinant proteins, such as ribonuclease I [243], human tissue plasminogen activator [104, 213], horseradish peroxidase [245] and human nerve growth factor β [212]. In this study, the improvement on PalB expression associated with DsbA coexpression was more significant than DsbC, suggesting that initiation of a correct disulfide pattern could be critical for functional expression of PalB in the periplasm. Since the three disulfide bonds in PalB are quite distant from each other [228], the chance for disulfide bond rearrangement might be slim [211, 219, 245] and this could be the reason why the improvement associated with DsbC coexpression was not as significant as DsbA. The similar improving effect on PalB expression mediated by DegP coexpression and DsbA coexpression could be associated with shared function for the two chaperones previously reported [183]. In addition, a synergistic improvement on PalB expression level was observed by coexpression of a folding factor cocktail of DsbA, DsbC, SurA, and FkpA though cell growth was arrested to some extent. While this synergistic effect is not surprising, the individual contribution from SurA coexpression remains to be evaluated. SurA is another periplasmic protein with both chaperone and PPIase activities [209] and coexpression of SurA could stabilize and facilitate the folding of several periplasmic proteins, including MalE31 [237], OmpA, OmpF, and LamB [208].

It is somewhat surprising to observe that PalB was stable in the cytoplasm (e.g., for pETL), but was unstable in the periplasm (e.g., for pETG). Such periplasmic instability associated with PalB critically limited PalB expression performance. This is the first report demonstrating the rescue of very unstable polypeptide using folding factor in the periplasm for improving expression performance. Note that the specific PalB expression level for pETG with coexpression of periplasmic folding factor was significantly higher than that for pETL (Figure 5.2C vs. Figure 5.4C), suggesting that the periplasm is a more appropriate compartment for functional expression of PalB than the cytoplasm. In fact, we have a few expression plasmids with other signal peptides, leading to the production of stable PalB variants with modified N-terminus in the periplasm (data not shown). In addition, we had the expression results similar to those in this study (e.g., unstable PalB variant that can be rescued by coexpression of periplasmic chaperones) by using another expression plasmid (e.g., pGPSH, data not shown) containing the same signal peptide as pETG (e.g., *pelB*) but a different promoter (e.g., P_{tet}). These observations suggested that the choice of the signal peptide could possibly affect PalB stability in the periplasm.

5.5 Abstract of the submitted paper

Functional expression of recombinant *Pseudozyma antarctica* lipase B (PalB) in *Escherichia coli* was explored. While PalB was stably expressed in the cytoplasm, most of the expressed gene product aggregated in cells as inactive inclusion bodies. In contrast, PalB was extremely unstable when expressed in the periplasm, also leading to poor expression performance. Such unstable PalB can be rescued by coexpression of several periplasmic folding factors, such as DegP, FkpA, DsbA, and DsbC, but not cytoplasmic ones. As a result, the performance for functional PalB expression in the periplasm was significantly improved. To our knowledge, this is the first report demonstrating the use of folding factors to rescue the extremely unstable gene product that is otherwise completely degradable.

Keywords: chaperone, *Escherichia coli*, folding factor, gene expression, lipase B, *Pseudozyma antarctica*, recombinant protein production

Chapter 6*

Using Fusion Tags to Enhance Functional Expression of Heterologous PalB in the Periplasm of *E. coli*

6.1 Introduction

The periplasm, in contrast to the cytoplasm where all proteins are first synthesized, offers several advantages as a destination for heterologous protein expression, such as its isolated and oxidative environment facilitating protein purification and disulfide bond formation, respectively [88, 90]. Just like the cytoplasm, protein misfolding can still occur in the periplasm [174, 262]. Several periplasmic proteins, such as DegP, FkpA, and Dsb-family members, can potentially act as folding factors to increase the solubility, mediate the stability, prevent the aggregation of secreted proteins [43, 173]. The previous results (Chapter 5) indicated that PalB was extremely unstable in the periplasm, leading to poor expression performance. Though the unstable PalB species could be rescued for functional expression by several periplasmic folding factors, the expression performance was far from optimized.

Fusion tags, originally developed for protein purification and immobilization, offer another powerful tool to enhance the folding and stability of the target protein moiety because fusion tags can rapidly attain a native conformation [86]. Several fusion tags, such as maltose-binding protein (MBP) [137, 139], N-utilization substance protein A (NusA) [146], FLAG [28], Thioredoxin (TRX) [94], and DsbA/DsbC [72], have been demonstrated to be effective in practical applications. While the tags theoretically can be fused with the target protein at either the N- or C-terminus, the N-terminus has been adopted for most applications due to potentially more effective initiation for translation and a higher solubility of gene products [38, 86, 87, 126, 131, 134-138]. The solubility can also be increased by using a larger tag though such strategy could possibly limit the expression yield of the fusion protein [126]. On the other hand, fusion proteins can be used directly as enzymes without removing the tag provided the tag moiety is small and hardly affects the bioactivity [86, 87].

In this study, functional expression of PalB in the periplasm of *E. coli* was explored using four N-terminal fusion tags. In addition, coexpression of various periplasmic folding factors was used to enhance the PalB expression performance.

* This chapter is based on a paper "Y. Xu, A. Yasin, T. Wucherpfennig, C. P. Chou, Using Fusion Tags to Enhance Functional Expression of Heterologous Lipase in the Periplasm of *Escherichia coli*. *World journal of Microbiology and Biotechnology*. 2008" (accepted).

6.2 Materials and Methods

6.2.1 Strains and Plasmids

The strains, plasmids and oligonucleotides used in this study are summarized in Table 6.1 and briefly described here. BL21(DE3) was used as the host strain for PalB expression. Molecular cloning was performed according to standard protocols [61] using DH5α as the cloning host. All the cloning techniques and chemicals were the same as described in Chapter 3.

Table 6.1 Strains, plasmids, and oligonucleotides

Strain, plasmid, and oligonucleotide	Relevant genotype or phenotype[a]	Source and reference
E. coli		
BL21(DE3)	*E. coli* B, F^- *ompT [dcm][lon] hsdS*$_B$(r_B^- m_B^-) *gal* (λ cIts857 *ind*1 Sam7 *nin*5 *lacUV5-T7* gene 1)	Novagen [256]
DH5α	F'/*(ϕ80 lacZ ΔM15) Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (r_K^-, m_K^+) phoA supE44 λ-thi-1 gyrA96 relA1*	Lab stock [257]
P. antarctica		
ATCC 32657	The source of *palB* gene	ATCC
Plasmid		
pAR3	Expression vector, P_{araB}, Ori (pACYC184), Cm^R	Lab stock [233]
pARDegP	P_{araB}::*degP*, Ori (pACYC184), Cm^R	This lab [187]

pARDegP$_{S210A}$	P$_{araB}$::*degP$_{S210A}$*, Ori (pACYC184), Cm^R	This lab [187]
pARDsbA	P$_{araB}$::*dsbA*, Ori (pACYC184), Cm^R	This study
pARDsbC	P$_{araB}$::*dsbC*, Ori (pACYC184), Cm^R	This study
pARFkpA	P$_{araB}$::*fkpA*, Ori (pACYC184), Cm^R	This lab [195]
pENTR4	Entry vector,	Invitrogen
pENTRG	*palB*-containing entry vector	This study
pETDsbA-G	P$_{T7}$::*lac*Op::*dsbA$_{signal}$*::*dsbA*::6×*his*::*palB*, Ori (pBR322), Ap^R	This study
pETDsbC-O	P$_{T7}$::*dsbC$_{signal}$*::*dsbC*::6×*his*::*palB*, Ori (pBR322), Kn^R	This study
pETG-50A	Destination vector, P$_{T7}$::*lac*Op::*dsbA$_{signal}$*::*dsbA*::6×*his*, Ori (pBR322), Ap^R, Cm^R	EMBL
pETM-80	Expression vector, P$_{T7}$::*dsbA$_{signal}$*::*dsbC*::6×*his*, Ori (pBR322), Kn^R	EMBL [72]
pFLAG	Expression vector, P$_{tac}$::*OmpA$_{signal}$*::*flag*, Ori (pBR322), Ap^R	Sigma-Aldrich
pFlag-P	P$_{tac}$::*OmpA$_{signal}$*::*flag*::*palB*, Ori (pBR322), Ap^R	This study
pHisperiMBP	Destination vector, P$_{tac}$::*mbp$_{signal}$*::6×*his*::*mbp*, Ori (pBR322), Ap^R, Cm^R	D. Waugh [137]
pHisperiMBP-G	P$_{tac}$::*mbp$_{signal}$*::6×*his*::*mbp*::*palB*, Ori (pBR322), Ap^R	This study
pTUM4	P$_{fkpA}$::*fkpA*::*SurA*, P$_{dsbA}$::*dsbA*::*dsbC*, Ori (pBR322), Cm^R	A. Skerra [258]

Oligonucleotide		
P10	5' – GG*CCATGG*GTCTACCTTCCGGTTCGG – 3'	This study
P11	5' – *CTGAATTC*TCAGGGGGTGACGATGCCGGAGCAGG – 3'	This study
P22	5' – *CTGAGCTC*TCAGGGGGTGACGATGCCGGAG – 3'	This study
P23	5' – *GTCTCGAG*CTACCTTCCGGTTCGGAC– 3'	This study
Pf_{DsbA} /Pr_{DsbA}	5' – *CC*ATG*G*AAAAGATTTGGCTGGCGC – 3' and 5' – *AAGCTTTC*ATTTCTCGCTTAAGTATTTC – 3' primer pair for amplification of *dsbA*	This study
Pf_{DsbC}/Pr_{DsbC}	5' – *CC*ATG*G*AGAAAGGTTTTATGTTG – 3' and 5' – *AAGCTTTC*AACCAGAACCACTAGT – 3' primer pair for amplification of *dsbC*	This study

[a]Designed restriction sites are underlined and the introduced mutations are in italic.

DNA fragments containing the leaderless *palB* gene were PCR-amplified using pfu DNA polymerase (Stratagene), the primer pair (P10/P11, P10/22 or P23/P11 in Table 1), and P. antarctica chromosomal DNA as the template. The PCR products were cloned into pPCR Script® (Stratagene). DNA sequencing was performed to ensure no mutations occurred during PCR. The *palB*-containing DNA fragments were obtained by digestion of the above pPCRScript® derivatives with *Nco*I/*Eco*RI, *Nco*I/*Sac*I, and *Xho*I/*Eco*RI, and subcloned into the corresponding restriction sites of pENTR4 (Invitrogen), pETM-80 [72], and pFLAG (Sigma-Aldrich), resulting in pENTRG, pETDsbC-O, and pFlag-P, respectively. pENTRG was mixed with the destination vector pET-50A [72], and pHisperiMBP [137], respectively, in a LR reaction with LR Clonase™ using Gateway technology (Invitrogen), resulting pETDsbA-G and pHisperiMBP-G.

Except pTUM4 [258], the coexpression plasmids containing a selection of periplasmic folding factor gene(s) were derivatives of pAR3 [233]. Two of them, e.g., pARDsbA and pARDsbC, were constructed in this study. The dsbA gene was PCR-amplified using pfu DNA polymerase, the primer pair Pf_{DsbA}/Pr_{DsbA}, and pETG-50A as the template. The dsbC gene was PCR-amplified using pfu DNA polymerase, the primer pair Pf_{DsbC}/Pr_{DsbC}, and pETM-80 as the template. The PCR products were cloned into pPCRScript®. DNA sequencing was performed to ensure no mutations occurred during PCR. The

dsbA-containing and dsbC-containing DNA fragments, both flanked with NcoI/HindIII, were subcloned into the corresponding restriction sites of pAR3 to form pARDsbA and pARDsbC, respectively. The design of the NcoI site in the sense primers resulted in changes in the second amino acid of the signal peptides (e.g., Lys → Glu) for both DsbA and DsbC, and the mutations did not appear to affect translocation. pARDegP [187], pARDegP$_{S210A}$, [187], and pARFkpA [195] contain the gene of degP, degP$_{S210A}$ and fkpA, respectively, fused with the araB promoter. pTUM4 harbors two artificial dicistronic operons: one with the structural genes of DsbA and DsbC under the control of the constitutive dsbA promoter, and the other with the structural genes of fkpA and surA under the control of the constitutive fkpA promoter [258]. All the folding factor-gene(s)-containing plasmids have a pACYC184 replication origin and a chloramphenicol-resistant (Cm^R) marker and are therefore, compatible with all the PalB-fusion expression plasmids containing a pBR322 replication origin and an ampicillin-resistant (Ap^R) or kanamycin-resistant (Kn^R) marker.

6.2.2 Cultivation and Analytic Methods

The cultivation and analytical procedures were identical with the procedures described in Section 3.2.2 and 3.2.3.

6.3 Results

6.3.1 Effect of Fusion Tags on Functional Expression of PalB

Results in Chapter 5 suggested that PalB was extremely unstable in the periplasm and functional expression of PalB could be possibly limited by certain types of N-terminal sequence. In this study, four periplasmic N-terminal tags, e.g., FLAG, MBP, DsbA, and DsbC, were used to modify the N-terminus of PalB for more effective expression in the periplasm. To do this, four translational fusion of tags with PalB were constructed, resulting in four expression plasmids, e.g., pFlag-P, pHisperiMBP-G, pETDsbA-G, and pETDsbC-O. The expression of FLAG-PalB and MBP-PalB was under the regulation of the *tac* promoter whereas the expression of DsbA-PalB and DsbC-PalB was under the regulation of T7 promoter. All the fusions used the signal of their corresponding tags for translocation, except FLAG-PalB which used the OmpA signal.

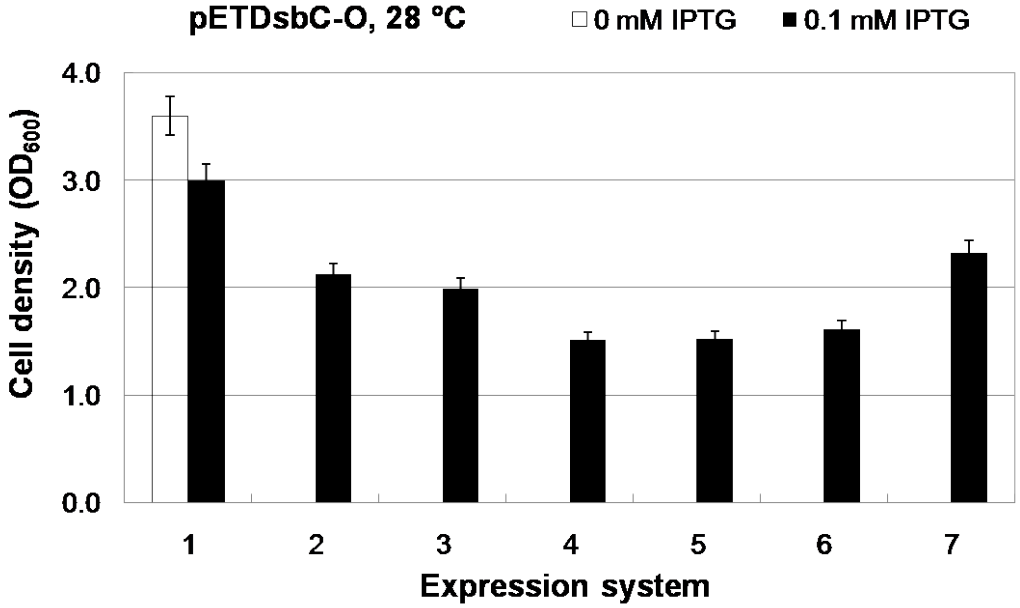
Using BL21(DE3) as the host to harbor these expression plasmids, qualitative visualization of PalB activity was conducted on the tributyrin plate and the result is summarized in Figure 6.1. Halos were developed for all the recombinant strains, but not for the control strain of BL21(DE3), indicating that all PalB fusions were functionally expressed in the periplasm with PalB activity. Since the area and

transparency of the halo potentially correlate with the expressed PalB activity of the applied sample, FLAG-PalB appeared to be the most active one among the four PalB fusions. Culture performance for the expression of DsbC-PalB, DsbA-PalB, MBP-PalB, and FLAG-PalB using BL21(DE3) harboring the corresponding expression plasmid is summarized in Figures 6.2~6.5, respectively (e.g., expression system 1 in each Figure). In general, a slight growth arrest was observed for all the recombinant BL21(DE3) with significant levels of PalB induction (e.g., 5 to 11-fold increase in specific PalB activity) upon IPTG supplementation. The specific PalB activity reached 2, 36, 47, and 82 U/liter/OD_{600} for DsbC-PalB, DsbA-PalB, MBP-PalB, and FLAG-PalB, respectively. Except MBP-PalB, most of these PalB fusions were overexpressed with concomitant accumulation of insoluble gene products as inclusion bodies.

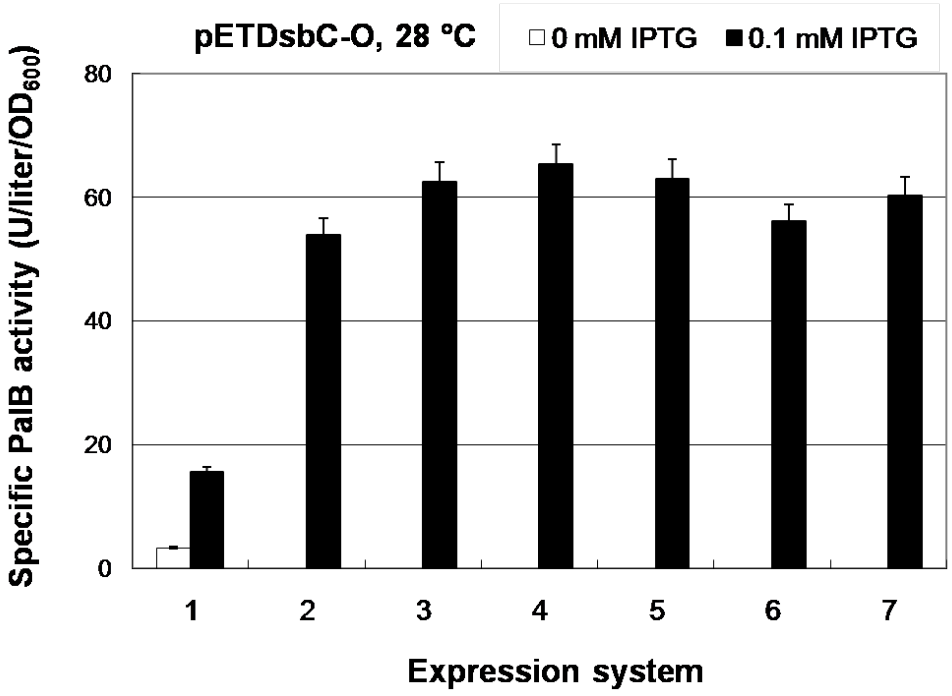
1	BL21 (DE3)
2	BL21 (DE3) (pFlag-P)
3	BL21 (DE3) (pETDsbA-G)
4	BL21 (DE3) (pETDsbC-O)
5	BL21 (DE3) (pHisperiMBP-G)

Figure 6.1 Qualitative visualization of PalB activity using the tributyrin plate for various PalB fusion expression systems.

A


pETDsbC-O, 28 °C
0 mM IPTG
0.1 mM IPTG
Cell density (OD600)
4.0
3.0
2.0
1.0
0.0
1
2
3
4
5
6
7
Expression system

B


pETDsbC-O, 28 °C
0 mM IPTG
0.1 mM IPTG
Specific PalB activity (U/liter/OD600)
80
60
40
20
0
1
2
3
4
5
6
7
Expression system

C

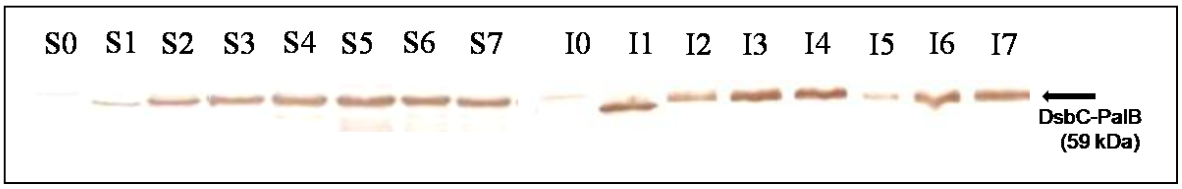
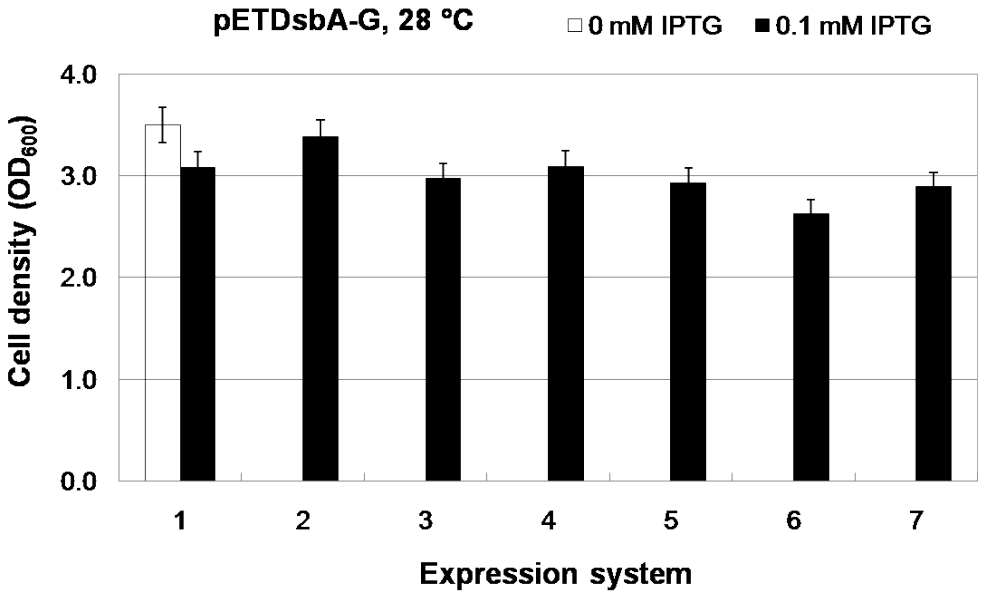
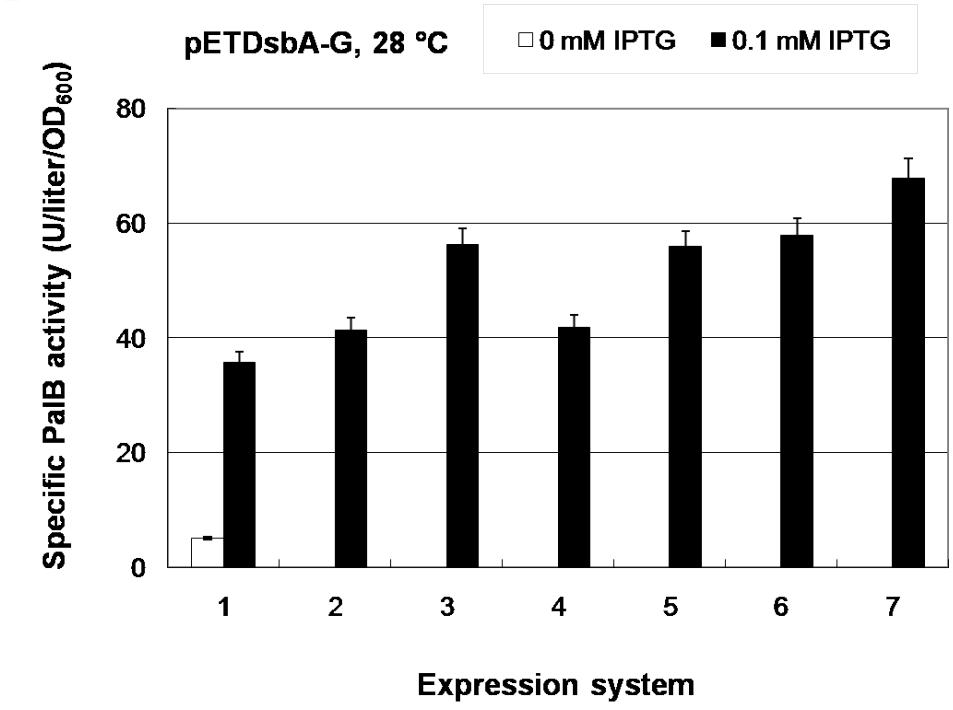


Figure 6.2 PalB expression performance for BL21(DE3) (pETDsbC-O) and the effect of coexpressing various periplasmic folding factors on the expression performance. Panel A: cell density and Panel B: specific PalB activity for various expression systems. Panel C: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various expression systems. 0, 1/ BL21(DE3) (pETDsbC-O); 2/ BL21(DE3) (pETDsbC-O, pARDegP); 3/ BL21(DE3) (pETDsbC-O, pARDegP$_{S210A}$); 4/ BL21(DE3) (pETDsbC-O, pARFkpA); 5/ BL21(DE3) (pETDsbC-O, pARDsbA); 6/ BL21(DE3) (pETDsbC-O, pARDsbC); 7/ BL21(DE3) (pETDsbC-O, pTUM4). Inducer supplementation: 1, 7/ IPTG; 2~6/ IPTG and arabinose. The error bar in Panels A and B represents a range of the data variance.

A


pETDsbA-G, 28 °C
□ 0 mM IPTG ■ 0.1 mM IPTG
Cell density (OD600)
4.0
3.0
2.0
1.0
0.0
1 2 3 4 5 6 7
Expression system

B


pETDsbA-G, 28 °C
□ 0 mM IPTG ■ 0.1 mM IPTG
Specific PalB activity (U/liter/OD600)
80
60
40
20
0
1 2 3 4 5 6 7
Expression system

C

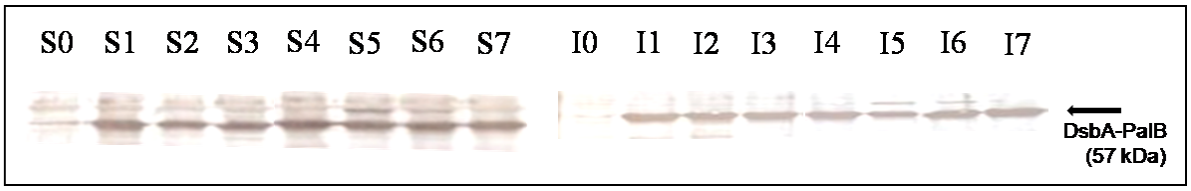
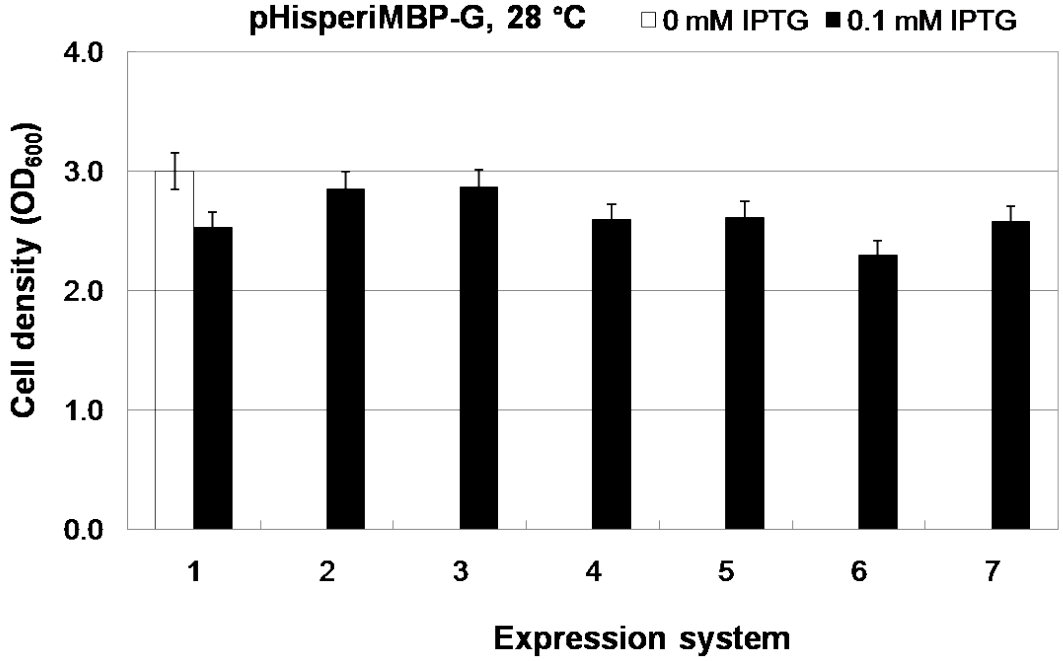
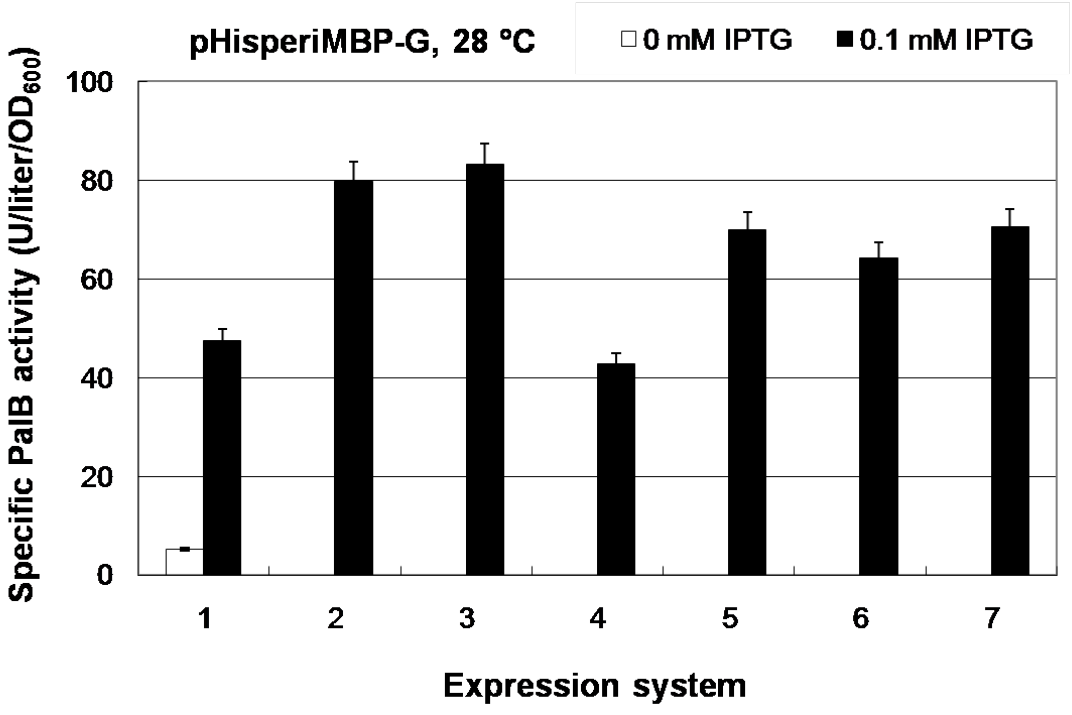


Figure 6.3. PalB expression performance for BL21(DE3) (pETDsbA-G) and the effect of coexpressing various periplasmic folding factors on the expression performance. Panel A: cell density and Panel B: specific PalB activity for various expression systems. Panel C: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various expression systems. 0, 1/ BL21(DE3) (pETDsbA-G); 2/ BL21(DE3) (pETDsbA-G, pARDegP); 3/ BL21(DE3) (pETDsbA-G, pARDegP$_{S210A}$); 4/ BL21(DE3) (pETDsbA-G, pARFkpA); 5/ BL21(DE3) (pETDsbA-G, pARDsbA); 6/ BL21(DE3) (pETDsbA-G, pARDsbC); 7/ BL21(DE3) (pETDsbA-G, pTUM4). Inducer supplementation: 1, 7/ IPTG; 2~6/ IPTG and arabinose. The error bar in Panels A and B represents a range of the data variance.

A


pHisperiMBP-G, 28 °C
0 mM IPTG
0.1 mM IPTG
Cell density (OD600)
4.0
3.0
2.0
1.0
0.0
1
2
3
4
5
6
7
Expression system

B


pHisperiMBP-G, 28 °C
0 mM IPTG
0.1 mM IPTG
Specific PalB activity (U/liter/OD600)
100
80
60
40
20
0
1
2
3
4
5
6
7
Expression system

C

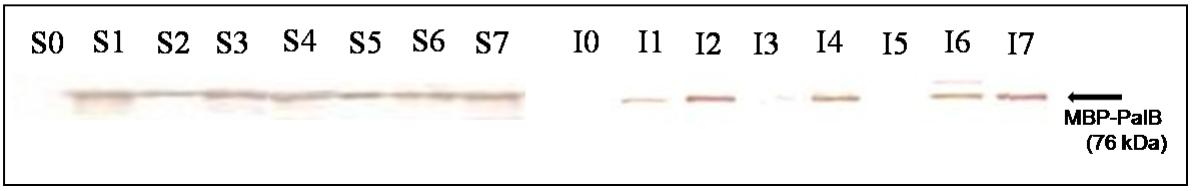
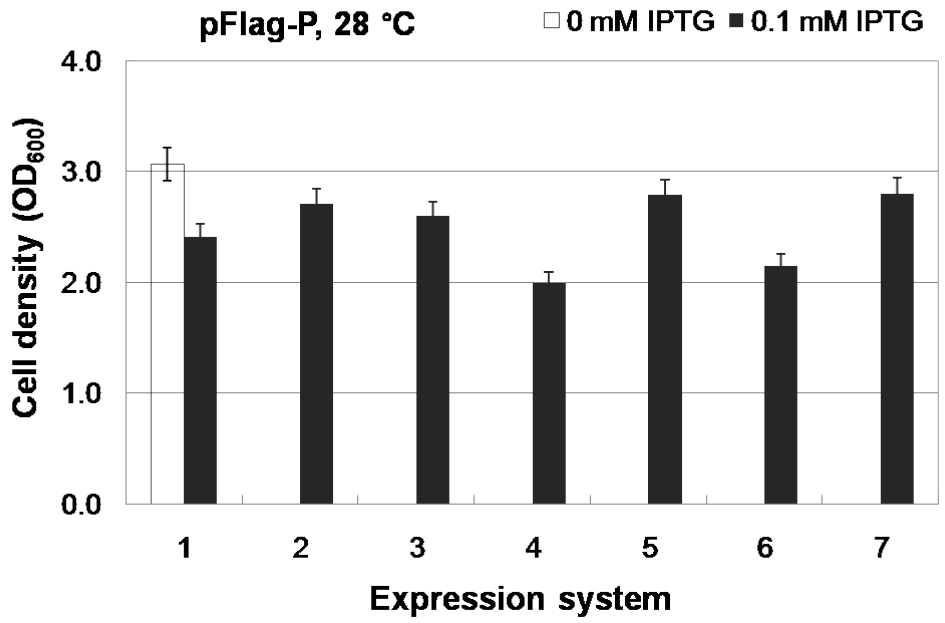
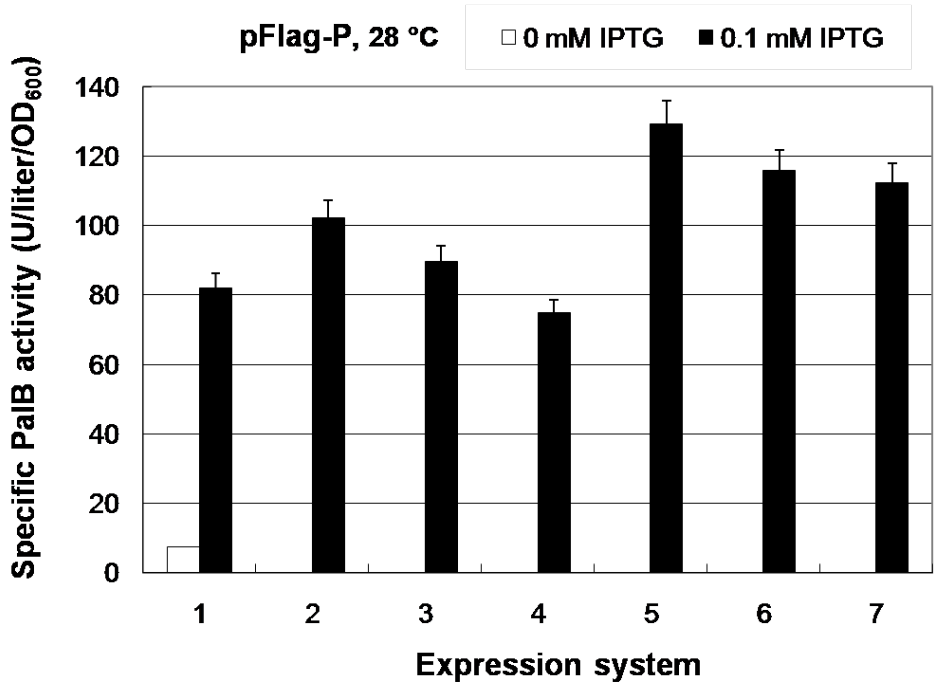


Figure 6.4. PalB expression performance for BL21(DE3) (pHisperiMBP-G) and the effect of coexpressing various periplasmic folding factors on the expression performance. Panel A: cell density and Panel B: specific PalB activity for various expression systems. Panel C: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various expression systems. 0, 1/ BL21(DE3) (pHisperiMBP-G); 2/ BL21(DE3) (pHisperiMBP-G, pARDegP); 3/ BL21(DE3) (pHisperiMBP-G, pARDegP$_{S210A}$); 4/ BL21(DE3) (pHisperiMBP-G, pARFkpA); 5/ BL21(DE3) (pHisperiMBP-G, pARDsbA); 6/ BL21(DE3) (pHisperiMBP-G, pARDsbC); 7/ BL21(DE3) (pHisperiMBP-G, pTUM4). Inducer supplementation: 1, 7/ IPTG; 2~6/ IPTG and arabinose. The error bar in Panels A and B represents a range of the data variance.

A


pFlag-P, 28 °C
0 mM IPTG
0.1 mM IPTG
Cell density (OD600)
4.0
3.0
2.0
1.0
0.0
1
2
3
4
5
6
7
Expression system

B


pFlag-P, 28 °C
0 mM IPTG
0.1 mM IPTG
Specific PalB activity (U/liter/OD600)
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
Expression system

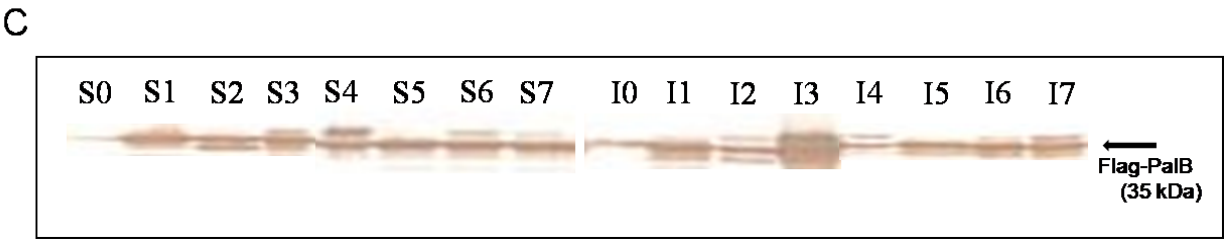


Figure 6.5. PalB expression performance for BL21(DE3) (pFlag-P) and the effect of coexpressing various periplasmic folding factors on the expression performance. Panel A: cell density and Panel B: specific PalB activity for various expression systems. Panel C: Western blotting analysis of the soluble (S) and insoluble (I) fractions for various expression systems. 0, 1/ BL21(DE3) (pFlag-P); 2/ BL21(DE3) (pFlag-P, pARDegP); 3/ BL21(DE3) (pFlag-P, pARDegP$_{S210A}$); 4/ BL21(DE3) (pFlag-P, pARFkpA); 5/ BL21(DE3) (pFlag-P, pARDsbA); 6/ BL21(DE3) (pFlag-P, pARDsbC); 7/ BL21(DE3) (pFlag-P, pTUM4). Inducer supplementation: 1, 7/ IPTG; 2~6/ IPTG and arabinose. The error bar in Panels A and B represents a range of the data variance.

6.3.2 Effect of Folding Factors on Expression PalB-Fusions

While these tags appeared to improve the solubility and bioactivity of the expressed PalB, functional expression could still be limited by ineffective posttranslational processing in the periplasm, such as folding and disulfide bond formation. Coexpression of several periplasmic folding factors, including DegP (and its mutant variant of $DegP_{S210A}$), FkpA, DsbA, DsbC, and a cocktail of SurA, FkpA, DsbA, and DsbC, was explored.

Among the four tags investigated here, DsbC appeared to be the least effective one for assisting functional expression of PalB, possibly due to a limited solubility of DsbC-PalB. Nevertheless, the culture performance was significantly improved upon coexpression of several periplasmic folding factors and the results are summarized in Figure 6.2. The specific PalB activity for the binary-plasmid systems was more than 3-fold that of the single plasmid system and the results were consistent with those of Western blotting. However, culture performance was limited by physiological stress since cell growth was seriously arrested upon coexpression of the periplasmic folding factors, in particular FkpA, DsbA, and DsbC. Coexpression of DegP and DsbA appeared to be effective in reducing the formation of inclusion bodies.

The effect of coexpressing periplasmic folding factors on enhancing DsbA-PalB expression performance was also noticeable (Figure 6.3). Compared to the single-plasmid system, the specific PalB activity was increased by more than 40% upon coexpression of $DegP_{S210A}$, DsbA, DsbC, and the chaperone cocktail without much arrest on cell growth. Nevertheless, there was minimum effect on reducing the formation of inclusion bodies by coexpression of periplasmic folding factors.

MBP appeared to be the most effective tag in solubilizing PalB and inclusion bodies were minimally observed upon overexpression of MBP-PalB. The specific PalB activity could be further improved without growth arrest by coexpression of periplasmic folding factors except FkpA (Figure 6.4). In particular, inclusion bodies were hardly detectable upon coexpression of $DegP_{S210A}$ or DsbA.

While the specific PalB activity reached a high level of 82 U/liter/OD_{600} when FLAG-PalB was overexpressed, inclusion bodies at a significant amount were observed (Figure 6.5). Compared to the single-plasmid system, coexpression of DegP resulted in a slight increase in the specific PalB activity with a concomitant decrease in the amount of inclusion bodies. Coexpression of $DegP_{S210A}$, on the other hand, resulted in a slight increase in the specific PalB activity but with a significant increase in the amount of inclusion bodies. The results suggest that the DegP protease activity could be responsible for in-vivo cleaning up misfolded FLAG-PalB. In general, coexpression of periplasmic folding factors except FkpA improved the culture performance. In particular, the specific PalB activity reached 129

U/liter/OD_{600} without growth arrest upon DsbA coexpression, resulted in a high volumetric PalB activity of 360 U/liter

6.4 Discussion

Up to now, it is well recognized that overexpression of bioactive PalB in *E. coli* remains challenging primarily due to intracellular misfolding which demolishes its bioactivity. Though the oxidative periplasm theoretically should be more favorable than the reduced cytoplasm for PalB expression, the functional expression was ineffective due to the instability of the periplasmic gene products (Chapter 5). In this study, four fusion tags were demonstrated to be effective to enhance functional expression of PalB in the periplasm with varied efficacies. All these PalB fusions expressed PalB enzyme activity, implying that downstream processing to remove these tags is not required for preparing biocatalysts. In addition, the functional expression could be further improved by coexpression of periplasmic folding factors.

DsbA and DsbC play crucial roles in disulfide bond formation in the periplasm of *E. coli* [40]. With certain chaperone activities, the two Dsb-family members were demonstrated to be effective fusion tags by increasing the solubility of target proteins, such as bovine enterokinase catalytic subunit [246], palmitoyl-tailed helper cytotoxic T-lymphocyte chimeric epitope of respiratory syncytical virus [248], and yellow fluorescent proteins [72]. While both DsbA and DsbC tags appeared to be effective fusion partners for functional expression of PalB in the periplasm, formation of inclusion bodies was observed. It has been reported that 40-kDa MBP, encoded by the *malE* gene from *E. coli*, can act as an excellent solubility enhancer by enabling the fusion partners to fold into bioactive conformations [86, 137-140], both in the cytoplasm (e.g. N-terminal inhibitory domain of human tissue inhibitor of metalloproteinases-2 [139] and yellow fluorescent proteins [72]) and the periplasm (e.g. RNaseA [137]). In this study, not only PalB activity reached high levels with minimum growth arrest but also the formation of inclusion bodies was minimized upon the overexpression of MBP-PalB, implying that MBP was very effective in solubilizing PalB for functional expression. However, it appears that the function expression could still be limited by the relatively large molecular size of the MBP tag. In contrast, FLAG is a small fusion tag with demonstrative applications in protein expression and purification [263, 264]. In this study, FLAG was the most effective tag for functional expression of PalB. Formation of inclusion bodies was accompanied by the overexpression even though the specific PalB activity was high. The amount of inclusion bodies could be reduced by coexpression of several periplasmic folding factors to improve the functional expression.

It has been widely reported that several periplasmic folding factors can interact with secreted proteins to enhance functional expression via improved translocation and/or polypeptide stabilization [88,

96, 173, 174, 245]. DegP is a periplasmic heat-shock protein with both protease and chaperone activities [183], but primarily acts as a serine protease for breakdown of aberrant periplasmic proteins arising upon extracytoplasmic stress [181, 182]. $DegP_{S210A}$ is a mutant variant with DegP protease activity being inactivated but chaperone activity being retained [183]. Comparing the contributing effects from DegP and $DegP_{S210A}$ in reducing inclusion body formation, DegP appeared to play a role as either protease (for DsbC-PalB and FLAG-PalB) or chaperone (for DsbA-PalB and MBP-PalB) in enhancing functional expression of PalB. FkpA is another periplasmic heat-shock protein with both cis/trans peptidyl-prolyl isomerase (PPIase) and chaperone activities [189, 190, 192] and is capable of reducing the σ^E-dependent response induced by misfolded proteins in the periplasm [179]. However, the effect of FkpA coexpression on improving functional expression of PalB was observable only for DsbC-PalB, but not other PalB-fusions investigated here. Many eukaryotic proteins, including PalB, contain disulfide bonds which often mediate the stabilization of folding conformation. It would be advantageous for these proteins to be expressed in the oxidative periplasm of *E. coli* to facilitate disulfide bond formation. With certain types of chaperone activities for the two Dsb-family members, DsbA catalyzes direct bridging of disulfide bonds in newly translocated proteins, whereas DsbC assists subsequent rearrangement of aberrant disulfide bonds [40]. In this study, coexpression of DsbA appeared to be the most effective in reducing the formation of inclusion bodies for all the four PalB-fusions, implying that functional expression of PalB could be limited by initial bridging of disulfide bonds. On the other hand, coexpression of chaperone cocktail (e.g., SurA, FkpA, DsbA, and DsbC) appears to improve functional expression of PalB for all the fusions by increasing PalB activities though formation of inclusion bodies was not in particular reduced.

6.5 Abstract of the submitted paper

Functional expression of heterologous *Pseudozyma antarctica* lipase B (PalB) in the periplasm of *Escherichia coli* was explored using four fusion tags, i.e. DsbC, DsbA, maltose binding protein (MBP), and FLAG in the sequence of increasing expression efficacy. Amongst these fusion tags, FLAG and MBP appear to be the most effective ones in terms of boosting enzyme activity and enhancing solubility of PalB, respectively. Overexpression of these PalB fusions often resulted in concomitant formation of insoluble inclusion bodies. Coexpression of a selection of periplasmic folding factors, including DegP (and its mutant variant of $DegP_{S210A}$), FkpA, DsbA, DsbC, and a cocktail of SurA, FkpA, DsbA, and DsbC, could improve the expression performance. Coexpression of DsbA appeared to be the most effective in reducing the formation of inclusion bodies for all the four PalB fusions, implying that functional expression of PalB could be limited by initial bridging of disulfide bonds. Culture performance

was optimized by overexpressing FLAG-PalB with DsbA coexpression, resulting in a high volumetric PalB activity of 360 U/liter.

Keywords: *Escherichia coli*, folding factor, fusion tag, gene expression, lipase, recombinant protein production

Chapter 7
Functional Display PalB on the *E. coli* Cell Surface

7.1 Introduction

As described before, whether the PalB was expressed in the cytoplasm or periplasm of *E. coli*, finally the cells need to be collected, lysed, and/or purified before PalB is used as an enzyme to catalyze ester-related reactions. Preferably, it could save the effort and the cost for PalB purifications if the enzyme was secreted into the extracellular medium or displayed on the cell-surface. Cell surface display is a novel technique [91] to display peptides or proteins on the surface of gram-negative [114, 265-267] and gram-positive [268, 269] bacteria, yeast [270], or even mammalian cells [271] by linking them to surface anchoring motifs. Bacterial surface display has found many applications in biotechnology and medicine [91, 272], e.g., recombinant enzymes displayed on the bacterial surface and can thus be used as biofactories [270, 273, 274], drug delivery of vaccines, antigens or antibodies [275].

In order to functional display protein on the *E. coli* cell-surface, usually cell surface displaying protein (passenger) needs to be fused to a signal peptide and an anchoring motif (carrier) for the journey of translocation of two membranes of *E. coli*. EstA is an autotransporter protein which consists of an N-terminal domain harboring the catalytic activity and a C-terminal domain forming a β-barrel-like structure inserted into the bacterial outer membrane which mediates the translocation of the N-terminal domain (Figure 7.1) [114, 269]. pEst100-FSCut contains the coding sequence for the signal peptide of *E. coli* alkaline phosphatase (*phoA*) under control of the P_{lac} promoter, followed by the coding sequence of a linker peptide, which for inserting the passenger protein, an E-tag epitope, and EstA. YFP and EYFP were functionally displayed on *E. coli* cell surface by using this displaying vector in this lab recently [118, 276].

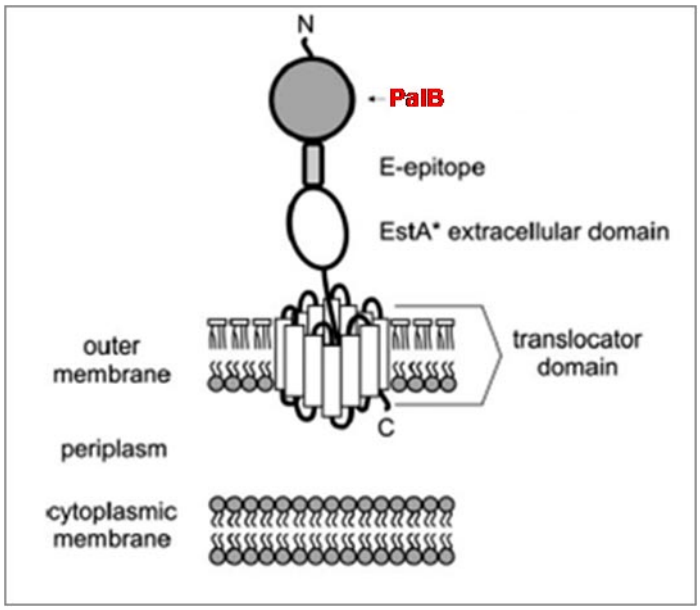


Figure 7.1 Schematic drawing of the autotransporter EstA construct used in this study for the displaying PalB on *E. coli* cell surface (modified after [114, 269]).

The displayed protein's sizes seemed not a big problem, it can be from 10 to 15 amino acids-long antigens for vaccines [275], or full size protein of up to 613 amino acids [114]. However, the cell bearing severe stresses become a challenge for applying display technology, especially for gram negative bacteria, e.g., *E. coli* [118]. There are also many unclear factors which influence functional display, e.g., the passenger protein is unfolded [117], partially folded or completely folded [277, 278] before translocated out of the periplasm of *E. coli*; or the displayed protein is tightly anchored, or cleaved after translocation and released in the extracellular milieu [269].

Displaying PalB on the cell surface was an experiment for proof of concept, we could not do a widely searching for the best expression system, but displaying a plasmid, which we had in this lab. In this study, *palB* was constructed into the display vector pEst100-FSCut and PalB displaying on *E. coli* cell surface was conducted. Upon the predictable cell stresses due to the PalB displaying and disulfide bond effect on the PalB expression, coexpression of various periplasmic folding factors were also investigated.

7.2 Materials and Methods

7.2.1 Strains and Plasmids

The strains, plasmids and oligonucleotides used in this study are summarized in Table 7.1 and briefly described here. JM109 was used as the host strain for PalB cell surface display and XL10-Gold® was used for cloning host. Molecular cloning protocols and the chemicals were used as described in Chapter 3.

Table 7.1 Strains, plasmids, and oligonucleotides.

Strains, plasmid, and oligonucleotides	Relevant genotype or phenotype^	Source and reference
E. coli		
ER2925	*Ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 glaT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10) TetS endA1 rpsL136 dam13::Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	NEB [279]
DH5α	F'/*(ϕ80 lacZ ΔM15) Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (r_K^-, m_K^+) phoA supE44 λ-thi-1 gyrA96 relA1*	Lab stock [257]
JM109	F′ *traD36 proA$^+$B$^+$ lacI^q Δ(lacZ)M15/ Δ(lac-proAB) glnV44 e14$^-$ gyrA96 recA1 relA1 endA1 thi hsdR17*	Lab stock [63]
XL10-Gold®	Tet^R Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* Hte [F' *pro*AB *lacI*^{q}ZΔ*M*15 Tn10 (Tet^R) Amy Cm^R]	Stratagene
P. antarctica		
ATCC 32657	The source of *palB* gene	ATCC

Plasmid		
pACYC177	A cloning vector, Ori (pACYC184), Kn^R, Cm^R	Lab stock [280]
pARDegP	P_{araB}::*degP*, Ori (pACYC184), Cm^R	This lab [187]
pARDsbA	P_{araB}::*dsbA*, Ori (pACYC184), Cm^R	This study
pARDsbC	P_{araB}::*skp*, Ori (pACYC184), Cm^R	This study
pARFkpA	P_{araB}::*fkpA*, Ori (pACYC184), Cm^R	This lab [195]
pARSkp	P_{araB}::*skp*, Ori (pACYC184), Cm^R	This lab [118]
pARSurA	P_{araB}::*surA*, Ori (pACYC184), Cm^R	This lab
pEst100-FSCut	P_{lac}::*cut*::*estA*, Ori (pBR322), Cm^R	H. Kolmer [114]
pESTKnN	P_{lac}::*palB*::*estA*, Ori (pBR322), Kn^R	This study
pESTCmN	P_{lac}::*palB*::*estA*, Ori (pBR322), Cm^R	This study
Oligonucleotide		
P12	5'-CC*GGCCCAGCCGGCC*CTACCTTCCGGTTCGGACCCTGCC-3'	This study
P20	5'-*CT*GGCCCCCGAGGCCGGGGTGACGATGCCGGAGCAG-3'	This study

^ Designed restriction sites are underlined and the introduced mutations are in italic.

The plasmid, pESTCmN and pESTKnN, were used to display PalB on *E. coli* cell surface. To construct the plasmid pESTCmN, the *palB* gene was PCR-amplified using *P. antarctica* chromosomal DNA as the template, P12 and P20 as the primers, and *Pfu* polymerase (Stratagene). The PCR products were cloned into pPCRScript® (Stratagene). DNA sequencing was performed to ensure that no mutations occurred during PCR. The *palB*-containing DNA fragment was purified and digested with *Sfi*I, gel extracted and ligated to the *Sfi*I digested fragment of pEst100-FSCut (H. Kolmer [114]) to obtain pESTCmN in which *palB* under the control of P_{lac} promoter and the plasmid had a Cm^R marker. In order

to make it compatible with the periplasmic folding factors which are also Cm^R marker, the Cm^R marker was replaced by Kn^R. To do so, Kn fragment was obtained from pACYC177 [280] by digesting with *Stu*I to make a blunt cut at both ends. In order to avoid the *Dcm* methylation at the two *Stu*I sites, the plasmid pACYC177 was isolated from ER2925 [279], a *dam* mutant strain. pESTCmN was digested by *Sca*I which located in the CDS fragment. The 1.3-kb Kn^R cassette and 6.8-kb pESTCmN were gel-extracted respectively and were ligated together to form pESTKnN (Figure 7.2). The coexpression periplasmic folding factors containing a selection of folding gene(s) were derivatives of pAR3 [233]. pARDegP [187], pARDsbA (this study, Chapter 5), pARDsbC (this study, Chapter 5), pARFkpA [195], pARSkp [118], pARSurA (lab stock) contain the gene of *degP*, *dsbA*, *dsbC*, *fkpA*, *skp*, *surA*, respectively, fused with the *araB* promoter. Plasmid pESTKnN has a pBR322 replication origin and a Kn^R marker and is, therefore, compatible with all the folding-factor-gene-containing plasmids having a pACYC184 replication origin and a Cm^R marker.

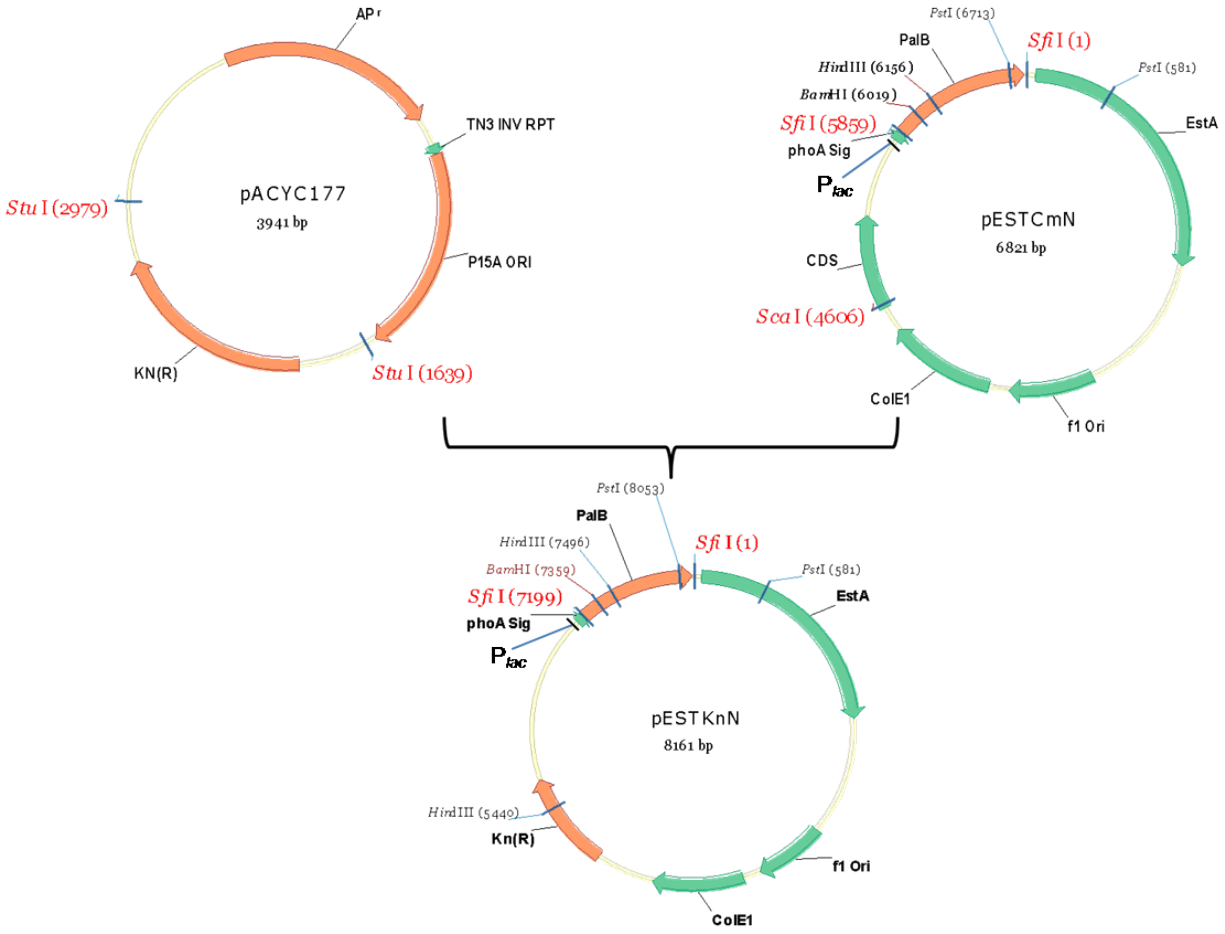


Figure 7.2 Depict of construction of PalB cell surface display plasmid pESTKnN from pESTCmN. The *palB* gene was under the control of P_{lac} promoter and inserted between signal peptide *phoA* and carrier *estA*. f1 Ori, f1 replication origin; colE1, ColE1 replication origin; P15A ORI, P15A replication origin; cat, chloramphenicol resistance marker; Kn(R), kanamycin resistance marker; Ap^r, ampicillin resistance marker. *Bam*HI, *Hin*dIII, *Pst*I, *Sca*I, *Sfi*I, and *Stu*I are restriction enzymes for DNA digestion.

7.2.2 Cultivation

The cultivation procedure was identical with the procedure described in Chapter 3, except the overnight culture was incubated at 32 °C and the culture was incubated at 28 °C after inoculated with 1 ml of overnight seed culture.

7.2.3 Sample Preparations

Three culture samples equivalent to 1.0 OD_{600} unit each were centrifuged at 6000 × g for 5 min at 4 °C.

For preparing the sample to estimate the lipase activity, the supernatant from the culture sample tube and the pellet was resuspended in 1 ml of PBS buffer were used as the supernatant (1.0 OD_{600} unit) and whole cell sample (1.0 OD_{600} unit), respectively.

For immunogold labeling, the pellet from a culture sample tube was resuspended in 100 µl PBS buffer (1.0 OD_{600} unit). A formvar-coated nickel grid was placed with the coating side facing down on the surface of a drop (100 µl) of the cell suspension on a parafilm sheet for approximately 10 min. The grid was blotted dry and placed on a drop of PBS with 0.3% skim milk powder as a blocking agent for 20 min. The grid was incubated in the blocking agent containing the primary antibody (e.g., anti-PalB rabbit, diluted in 1/50) for 1 h. The grid was washed with drops of the blocking agent 3 times for 2 min each. The grid was incubated on a drop of the secondary antibody (e.g., goat anti-rabbit IgG conjugated with gold from Sigma, diluted in 1/50 in DI water) for 1 h. The grid was washed completely with drops of PBS 3 times for 2 min each, and then washed with DI water 4 times. The grid was stained with 2% aqueous uranyl acetate for 1 min and washed with DI water for 1 min. The grid was air-dried and was ready for observation under a transmission electron microscope (TEM). All the steps were followed by intermediate blot drying of grids on a filter paper.

7.2.4 Analytical Methods

PalB assay followed exactly the procedures as described in Chapter 3. For the qualitative visualization of PalB activity in the tributyrin plate the samples of supernatant and whole cells were used at 25 µl (0.025 OD_{600} unit) and 5 µl (0.005 OD_{600} unit) respectively. For the pNPP assay the sample of whole cells was used at 100 µl (0.1 OD_{600} unit).

Immunogold labeling was observed under TEM (Philips CM10, Eidenhoven, Netherlands) in the Biology Department at University of Waterloo.

7.3 Results

7.3.1 PalB Displaying on Cell surface of *E. coli*

PalB displaying on cell surface of *E. coli* was conducted by using DH5α (pESTCmN) and JM109 (pESTKnN) as the host/vector systems under various IPTG inducer concentration conditions and the results are summarized in Table 7.2 and Figure 7.3 and 7.4. The cells had been stressed even without IPTG supplementation. After induction the stress situation became much worse, and cell growth was significantly inhibited due to physiological deterioration. Especially for JM109 (pESTKnN) system, the cell densities stayed at the same low value with various IPTG concentration, indicating that the cell could not bear the PalB display with even a very mild IPTG induction. It seemed that the DH5α system was better than the JM109 system, not only the cell density was higher during the low IPTG concentrations (Table 7.2), but also the PalB activity was better for whole cell and lower for supernatant samples indicating the PalB was not cleaved and released into medium (Figure 7.3). Unfortunately, due to the low PalB activity in the pNPP assay and the limitation of the measurement, the spectrophotometric reading at OD_{410}could not be matched well with the PalB activity (data not shown). The result of TEM (Figure 7.4) confirms that PalB was successfully displayed on *E. coli* cell surface albeit at a relatively low level. In addition, the pESTCmN has the same resistance marker (Cm^R) with the plasmids for expressing the periplasmic folding factors, therefore, the following coexpression of folding factors were only studied in the JM109 (pESTKnN) display system.

Table 7.2 Comparison of the performances of PalB displayed on cell surface of DH5α (pESTCmN) or JM109 (pESTKnN) under various IPTG inducer concentrations at 28 °C

IPTG (mM)	0	0.05	0.1	0.2	0.5	1
Cell Density (OD_{600})		**1&6**	**2&7**	**3&8**	**4&9**	**5&10**
DH5α	2.98±0.10*					
DH5α (pESTCmN)	2.35±0.10	1.72±0.02	1.35±0.08	1.11±0.10	1.04±0.05	0.99±0.10
JM109	2.93±0.10					
JM109 (pESTKnN)	1.36±0.05	1.08±0.04	1.09±0.07	1.10±0.09	1.09±0.05	1.07±0.08

* Indicates the range of OD_{600} variance of samples taken from three different flasks for the same expression system.

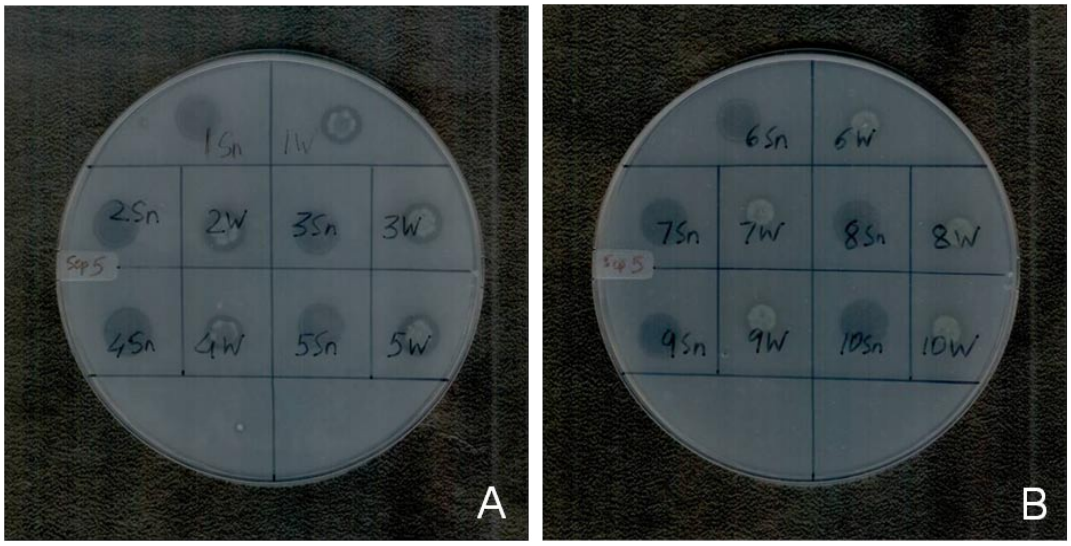


Figure 7.3 Qualitative visualization of PalB activity by using tributyrin plate. The area and intensity of the halo corresponding to the lysate sample correlates with the PalB expression level. The samples of PalB displaying on *E. coli* cell surface at various IPTG inducer concentration conditions were presented for DH5α (pESTCmN) system (Panel A) and JM109 (pESTKnN) system (Panel B). IPTG inducer concentration supplementated in the systems were 1&6/ 0.05 mM; 2&7/ 0.1 mM; 3&8/ 0.2 mM; 4&9/ 0.5 mM; and 5&10/ 1 mM. Sn, Supernatant sample; W, whole cell sample.

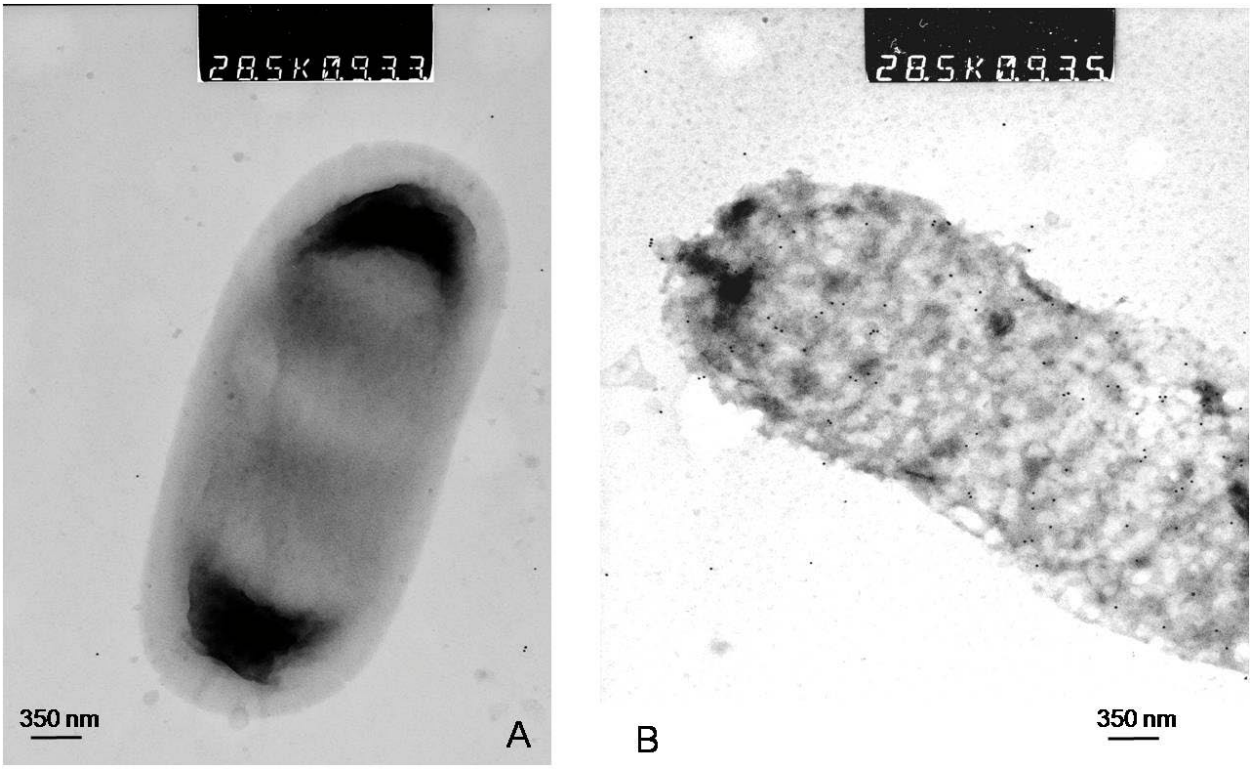


Figure 7.4 Results of TEM for immunogold labeling of displayed PalB on *E. coli* cell-surface. The black dots represent the gold particles conjugated with the displayed PalB. Panel A/ JM109; Panel B/ JM109 (pESTKnN) induced with 0.05 mM IPTG. A typical picture was taken after visualizing the overall images of the cells under TEM.

7.3.2 Coexpression of Periplasmic Folding Factors Improved PalB Display Performance

Using JM109 (pESTKnN) as the host/vector system for studying the improvement of PalB displaying on cell surface with coexpression of the periplasmic folding factors were conducted, and the PalB display performances are summarized in Table 7.3 and Figure 7.5 and 7.6. Compared to the individual display systems, the cells became unhealthy to some extent after induction at 0.05 mM IPTG and/or 1 g/liter arabinose supplementation at 28 °C. However, the folding factors in the test, except Skp, showed significant improvement of cell survival (Figure 7.5) and growth (Table7.3) after IPTG-induction. The TEM results in Figure 7.6 showed the improvement of PalB display with coexpression of the folding factors. It did not show any enhancement for PalB display with coexpression of DegP (Panel E in Figure

7.6) and Skp (Panel F in Figure 7.6), so good cell surfaces were kept for the cells in these two systems. The presence of SurA (Panel A in Figure 7.6) not only improved the display, but also the cell physiology, which indicated by more gold dots on the surface and an almost healthy cell surface. A better PalB display was shown for coexpression of FkpA (Panel B in Figure 7.6) and DsbC (Panel D in Figure 7.6), but poor cell surface. The best PalB display was shown with DsbA coexpression. Cell growth was recovered and individual cell appeared to be more robust for display PalB which indicated by high density of cell surface displayed PalB and a healthy cell surface (Panel C in Figure 7.6).

Table 7.3 PalB display performances with coexpression of the periplasmic folding factors

Strain		Cell Optical Density (OD_{600})	
		No-induction	After induction*
1	JM109(pEstKnN)	1.356±0.048 #	1.093±0.048
2	JM109(pEstKnN, pARDegP)	2.170±0.052	1.490±0.045
3	JM109(pEstKnN, pARFkpA)	1.804±0.054	1.203±0.053
4	JM109(pEstKnN, pARSurA)	2.195±0.045	1.505±0.095
5	JM109(pEstKnN, pARSkp)	2.235±0.050	0.935±0.045
6	JM109(pEstKnN, pARDsbA)	1.956±0.049	1.455±0.100
7	JM109(pEstKnN, pARDsbC)	2.075±0.052	1.420±0.045

* 0.05 mM IPTG and/or 1 g/liter arabinose was supplementated in the system for induction of PalB display and/or coexpression of folding factors, respectively, at 28 °C.

the range of OD_{600} variance of samples taken from three different flasks for the same expression system.

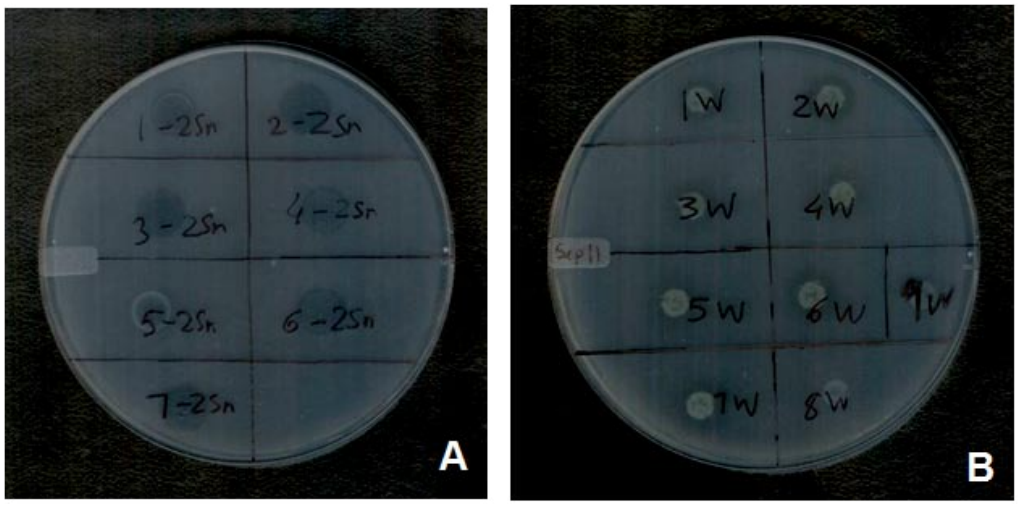
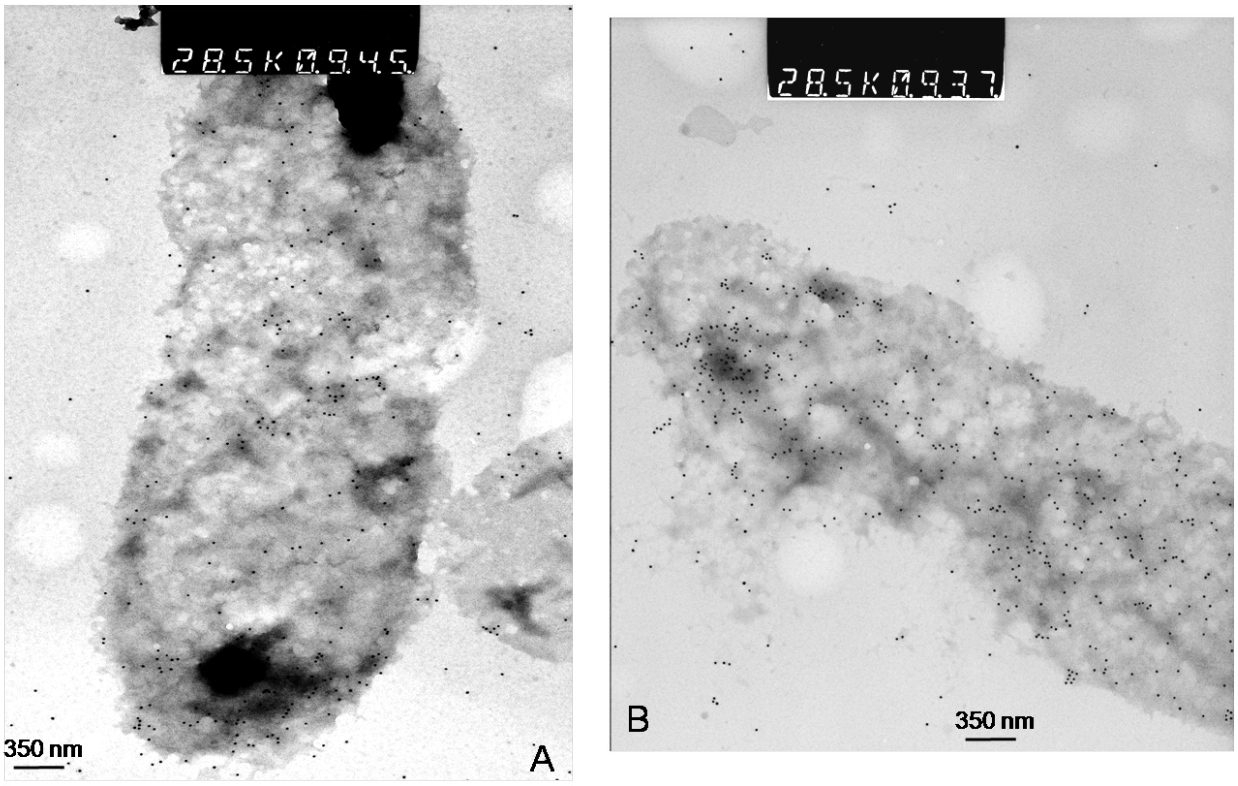
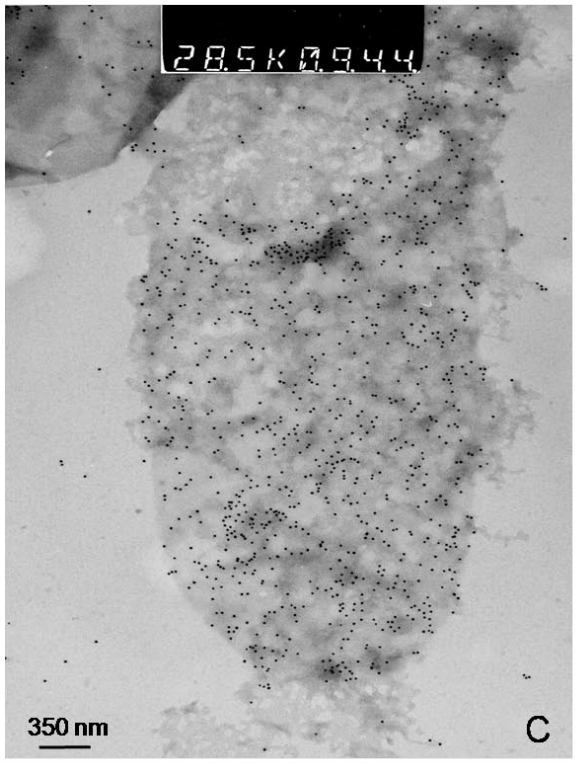
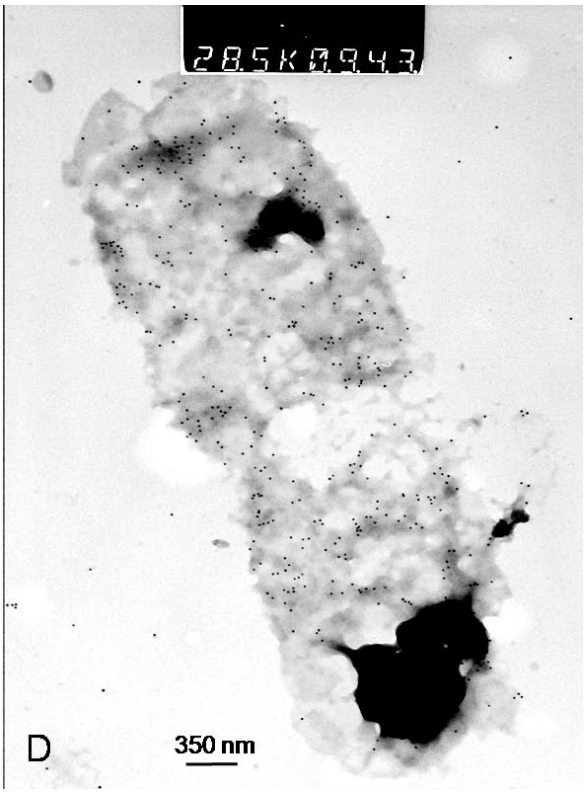


Figure 7.5 Qualitative visualization of PalB activity by using tributyrin plate. The area and intensity of the halo corresponding to the lysate sample correlates with the PalB expression level. The samples of PalB displaying on *E. coli* cell surface with various periplasmic folding factors (Table 7.3) were presented: Supernatant sample (Panel A), and whole cell sample (Panel B). The coexpressed folding factors were 1/ -, control; 2/ DegP; 3/ FkpA; 4/ SurA; 5/ Skp; 6/ DsbA; and 7/ DsbC. 8/ JM109and 9/ DH5α in Panel B, respectively, were for negative control.


28.5K 0.945
350 nm
A
28.5K 0.937
B
350 nm


28.5K 0.944
350 nm
C


28.5K 0.943
D
350 nm

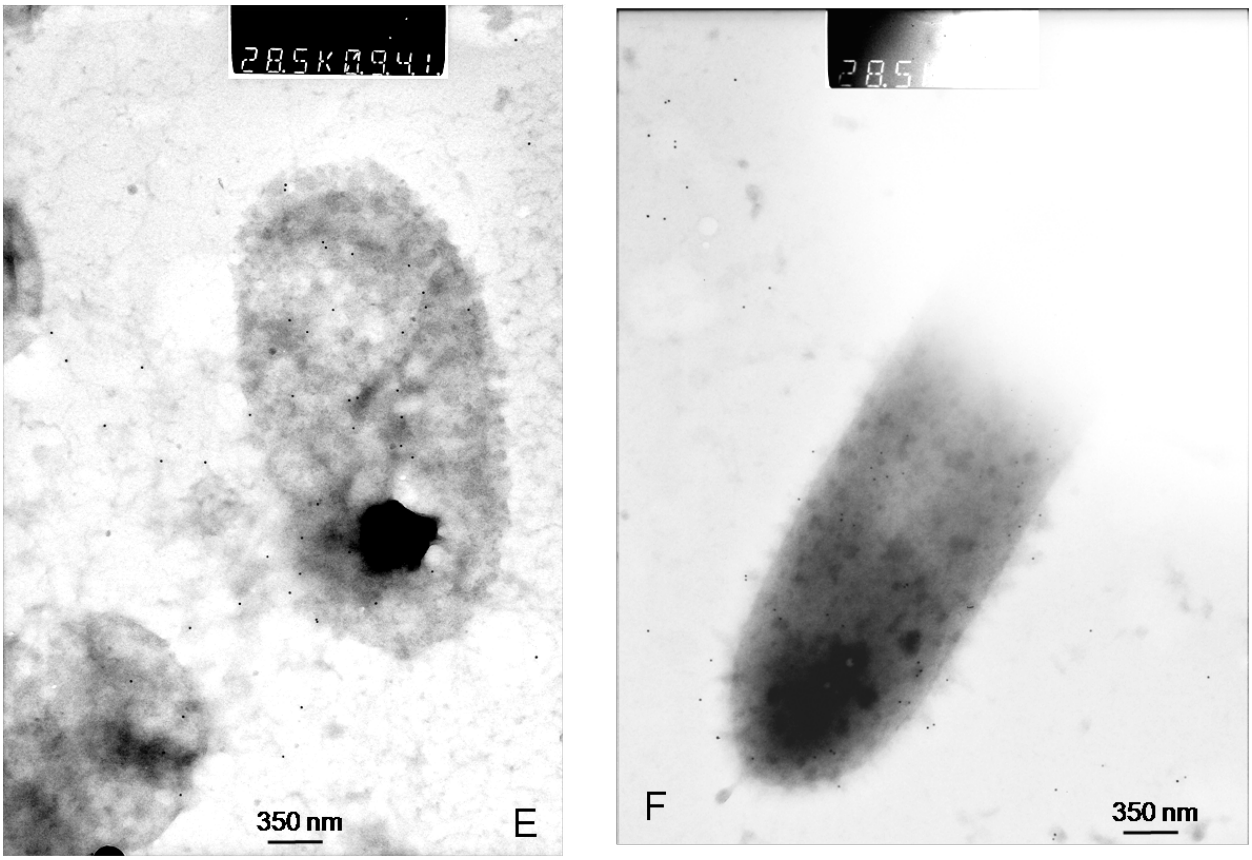


Figure 7.6 Results of TEM for immunogold labeling of displayed PalB on *E. coli* cell surface using JM109 (pESTKnN) as the displaying host/vector system with coexpressed various periplasmic folding factors: Panel A/ SurA; Panel B/ FkpA; Panel C: DsbA; Pane D: DsbC; Pane E: DegP; and Pane F: Skp. The culture was supplemented with 0.1 mM IPTG and 1 g/liter arabinose. Note that the density of the black dots, representing the gold particles conjugated with the displayed PalB. A typical picture was taken after visualizing the overall images of the cells under TEM.

7.4 Discussion

Although it looked like the host strain was sensitive, e.g., JM109 was poorer than DH5α, it is the fact that displaying PalB on *E. coli* surface resulted in a detrimental effect on cell physiology, indicating in growth arrest and poor display performance. The main reasons are (i) *E. coli* is not a good secretion host [38]; and (ii) in order to reach its destination, the cell surface displaying protein needs to maintain the export-competent form (no spontaneous folding) for efficiently crossing two membranes (the cytoplasmic and periplasmic membrane) [117]. The first obstacle could be minimized by choosing a signal peptide and coexpression the periplasmic folding factor. The second one is much harder because it does not depend on the carrier but the passenger protein properties. For example, the disulfide bonds could not be formed without the assistance of Dsb-family protein, e.g., DsbA [40, 105]. In fact, about this "unfolded and devoid of disulfide bonds passenger before translocation" hypothesis was debated and proved wrong [277, 278]. Fortunately, we had proved that although the disulfide bond formation was not critical for PalB activity (Chapter 3), the disulfide bond formation did improve PalB activity (Chapter 4, 5, and 6). The PalB was functionally displayed in *E. coli* cell surface and highly improved by DsbA coexpression indicating PalB was formed 3-D structure before it was secreted out of *E. coli*. It may need to further approve the PalB really displayed on the cell surface by conducting experiments such as trypsin digestion or EDTA and SDS sensitivity tests, but a positive feedback for cell surface displayed PalB used as a biocatalyst had given in a study conducted by a master student in this lab [276]. In addition, the PalB cell surface displaying on *E. coli* is an experiment proving a concept in this study. The presences of DsbA and DsbC showed the positive effects on the PalB expressions in the cytoplasm of a mutant *E. coli* strain, in the periplasm of the regular *E. coli* host, and cell surface displayed PalB on *E. coli*, indicating disulfide bond is, if not critical, essential for functional PalB formation. In this chapter, it was proved again that the presence of DsbC was not crucial as DsbA because the rather simple disulfide bond formation in PalB, e.g., three distant consecutive disulfide bonds (even may not be needed all of them).

Skp is a well characterized periplasmic chaperone that interacts with unfolded proteins translocated into the periplasm via the Sec secretion system [179] and is under the control of both σ^E and Cpx stress response systems [178]. Skp has been shown to improve the folding of recombinant proteins or antibody fragments in the periplasm of *E. coli* [111, 203-205]. Skp showed a good enhancer for YFP display on the cell surface of *E. coli* host JM109 [118]. However, the effect of Skp on the increase in the solubility and prevention of the premature folding of periplasmic carrier-passenger fusions could not resolve PalB folding problem, which needed specific assistance for disulfide bonds formation.

DegP is an inducible serine protease activity for breakdown of aberrant periplasmic proteins arising upon extracytoplasmic stresses [181, 182] and its expression can be activated via both σ^E and Cpx pathways [121]. It is reputed for improving the periplasmic protein expression by relieving the extracytoplasmic stress upon overexpression of several periplasmic gene products, such as alkaline phosphatase (PhoA) [184], DsbA'-PhoA [259], MalS [183], OmpF mutants [260], maltose-binding protein (MBP or MalE) variants [179, 261], penicillin acylase [187], and PalB expression in the periplasm of *E. coli* (in Chapter 5 and 6 in this study). However, until now there is no evidence of a direct effect of DegP in the productive folding of OMPs *in vivo* and *in vitro* [173]. That is not surprised for seeing the cell physiology improvement but not any enhancement of PalB cell surface display when coexpressed with DegP.

Probably due to both FkpA and SurA having chaperone as well as PPIase activity and relate to σ^E cell tress response, they showed similar [179], but different effects [173]. Coexpression of FkpA, but not SurA, increased the amount of fusion protein displayed on the phage and improved single chain antibody fragment expression in the periplasm [193, 194]. *SurA* null mutants showed decreased levels of properly folded OMPs [208, 209], however, FkpA did not showed specific assistance for the folding of OMPs [173]. They were believed to be most important folding factors in the periplasm of *E. coli* [258]. The best assistant for EYFP cell surface display was reported from FkpA coexpression among the six investigated folding factors, e.g., DegP, DegQ, DegS, SurA, DsbA and DsbC [276]. In this study, the presence of both FkpA and SurA showed much similar effect on the cell surface displayed PalB in *E. coli*, but SurA seemed more efficient for releasing cell stresses and preventing the cell burst.

In a summary, cell surface displayed PalB caused severe cell stresses on *E. coli*. The whole cell biocatalyst concept is impressive, but maybe other displaying hosts, e.g., gram-positive bacteria, or yeast, are more attractive for cell surface display. However, this topic is beyond the scope of this study and will not be discussed further.

Chapter 8
Original Contributions and Recommendations

8.1 Original Contributions to Research

Identified Limiting Factors of PalB Expression in the Cytoplasm of *E. coli*

Lipase B from *Pseudozyma antarctica* (PalB) was actively expressed in the reduced cytoplasm of a regular *Escherichia coli* strain of BL21(DE3), implying that disulfide bond formation was not strictly required for functional expression. However, the expression was ineffective and was primarily limited by the formation of PalB inclusion bodies and growth arrest, both of which were associated with PalB misfolding and deteriorated physiology. The culture performance in terms of cell growth and PalB expression level could be significantly improved by simultaneous coexpression of multiple chaperones of TF and GroEL/ES, but not by individual coexpression of either one of them. It was proposed that the two chaperones mediating the early stage and late stage of cytoplasmic PalB folding would be required simultaneously for boosting both the overall PalB synthesis rate and the cytoplasmic folding efficiency.

Addition of Fusion Tag improving the solubility but only DsbA tag improved bioactivity of PalB

The cytoplasmic expression of PalB ended up with insoluble inclusion bodies and the unstable PalB showed the direct expression of the PalB in the periplasm requires improvement. By the addition of fusion tag at the N-terminus of PalB, the cell's physiology and the solubility and bioactivity of PalB were surprisingly improved in both the oxidized cytoplasm and periplasm. In addition, the PalB fusions (e.g., DsbA-PalB in the oxidized cytoplasm and Flag-PalB in the periplasm) showed high bioactivities. This offers a possibility for reducing the cost of biocatalyst production by (i) directly using the PalB fusion as enzyme, or (ii) based on the tag's affinity property to search for carriers in the immobilization process.

Folding Factors and Fusion Tag Rescuing Unstable PalB and Enhancing its Expression in the Periplasm of *E. coli*

Functional expression of recombinant PalB in *E. coli* was explored. While PalB was stably expressed in the cytoplasm, most of the expressed gene product aggregated in cells as inactive inclusion bodies. In contrast, PalB was extremely unstable when expressed in the periplasm, also leading to poor expression performance. Such unstable PalB can be rescued either by the coexpression of several periplasmic folding factors, such as DegP, FkpA, DsbA, and DsbC, or by adding fusion tag. As a result, the performance for

functional PalB expression in the periplasm was significantly improved. The overall improvement reached 40-fold.

PalB Displayed on Cell surface and its Potential Use in the Whole Cell Immobilization

When PalB displayed on *E. coli* cell surface, the cell physiology seriously deteriorated resulting in growth arrest, cell lysis and poor display performance. Due to the cultivation condition being too harsh for cell surface display (0.05 mM IPTG, 28 °C, 200 rpm, same as other shake-flasks experiments), all the folding factors showed limited improvements of either the physiology or in the density of surface-displayed PalB. Amongst the folding factors tested, DsbA showed much better assistance than others in improving the density of surface-displayed PalB. After immobilization of the whole surface-displayed PalB cells, the lipase activity was demonstrated by the lipase enzymatic assay reaction and reached a half of the commercial immobilized PalB. This illustrated the feasibility of whole cell PalB immobilization.

Disulfide bonds' functions in PalB

Disulfide bond(s) may not be critical for PalB activity, but disulfide bond formation enhanced PalB formation and thus improved PalB activity, which was enhanced by the coexpression of DsbA with PalB or addition DsbA-tag to the N-terminus of PalB in the oxidized cytoplasm or the periplasm.

8.2 Recommendations

Regarding further improvements in engineering a robust strain and optimizing the cultivation process, future studies are recommended on PalB which include the following topics.

Genetic Engineering PalB Suitable for Expression in *E. coli*

Although *palB* was expressed in *E. coli* systems, all the various strategies used could not effectively overcome either low expression, or huge amount of inclusion bodies due to PalB natural property, e.g., cold lipase, three disulfide bonds and a quite hydrophobic surface. Based on systematic studies, the disulfide bonds were suspected to constrain PalB bioactivity. Due to technical limitations, the effect of the Dsb catalysts on the PalB expression was not completed in this study. After obtaining the plasmid DsbABCD set [212, 245], it is necessary to screen in detail the effect of Dsb catalyst on PalB expression. Removing useless disulfide bond(s), or at least two of them (e.g., N-terminus and C-terminus ones, far from the enzymatic active site) could improve PalB expression in *E. coli*. In addition changing some amino acids on the surface of PalB could reduce its hydrophobicity and also improve PalB expression. These studies can be done by site-directed mutagenesis.

Some lipase-producing bacteria express lipase with a specific intermolecular chaperone Lif (lipase-specific foldase) for lowering the energetic barriers during lipase folding [281]. One should try to find another way for improving the PalB expression in *E. coli*.

Genetic Screening of PalB Suitable for Biodiesel Production

Although the catalytic mechanism for transesterification has not been studied extensively, it is believed to be similar to that for hydrolysis [282] except that the water molecule is replaced by a short-chain alcohol. It has been found in biodiesel production that one of the substrates e.g., methanol, tends to inhibit or denature immobilized PalB, resulting in a reduced product yield and even a permanent damage of biocatalysts [11]. In practical application, methanol is slowly or stepwise added to the bioreactor to minimize this inhibition effect [11]. Another substrate, methyl acetate, has been used as an acyl acceptor to avoid the above methanol effect [283], but the cost of this substrate might be prohibitive for large-scale production. Methods for pretreatment or regeneration of biocatalysts were also proposed [23]. From a process viewpoint, it will be of economical benefit to genetic engineer a robust *palB* that is resistant to methanol inhibition.

The exact mechanism of methanol inhibition of PalB is unclear. On the basis of its protein structure [221, 222, 228], a restricted entrance into the active site was proposed for its substrate specificity and high degree of stereospecificity. It is believed that the presence of methanol blocks the primary substrate (e.g., triglyceride) from entering the active site. On the other hand, there are several microbial lipases that are more methanol-resistant though their transesterification activities might not be as effective as that of PalB. For example, methanolysis using *C. rugosa* and *P. fluorescens* lipases can be activated by the surrounding water, implying that water can potentially prevent the inactivation of these lipases by methanol [10]. Also, lipases from *P. cepacia* [10] and *Candida parapsilosis* [284] activated by a minimum amount of water activity exhibit a potential resistance to methanol upon methanolysis. Analyzing the protein sequences of these lipases would shed light on developing genetic strategies for designing methanol-resistant lipases. Gene shuffling involving the wild-type or mutant PalB [285] and the above methanol-resistant lipases can be explored by constructing hybrid lipase variants with the desired properties for biodiesel production. Digesting and reassembling the DNA region corresponding to lipase B genes will be conducted to generate random mutations. In addition, error-prone polymerase chain reaction (PCR) will be used to generate mutations particularly in the domain near the substrate binding site. Directed evolution [286] will also be possible for searching PalB variants resistant to methanol inhibition and it will be critical to establish an effective system for high-throughput screening of lipase

mutants. Similar genetic approaches were previously made for improving the thermal inactivation for PalB [27, 227].

Optimization of Cultivation Condition for PalB Production

Cultivation condition is very important for developing a high yield process. By using a bioreactor with fully controlled pH, dissolved oxygen (DO), temperature, the cell stress from the environmental effects will be mitigated. Various culture parameters, such as medium recipe, pH, dissolved oxygen (DO), temperature, and induction conditions could be evaluated to optimize culture performance for simultaneous high-level PalB gene expression and high-cell-density cultivation using developed host/vector systems.

A product oriented approach is proposed for modeling the cell specific productivity of *E. coli*. In order to get the feasibility evaluation, a simple, time honored modeling approach could be applied [287]. This model relates the cell specific productivity of PalB, $(\frac{1}{X}\frac{dP}{dt})$, to the specific growth rate and biomass concentration as follows:

$$\frac{1}{X}\frac{dP}{dt} = \alpha\,\mu + \beta \qquad (1)$$

where:

X = biomsss concentration (g/liter)

P = PalB concentration (g/liter)

α = growth associated productivity constant

μ = specific growth rate (h^{-1})

β = non-growth associated productivity constant

Immobilization of PalB Fusion or Cell surface Displayed PalB

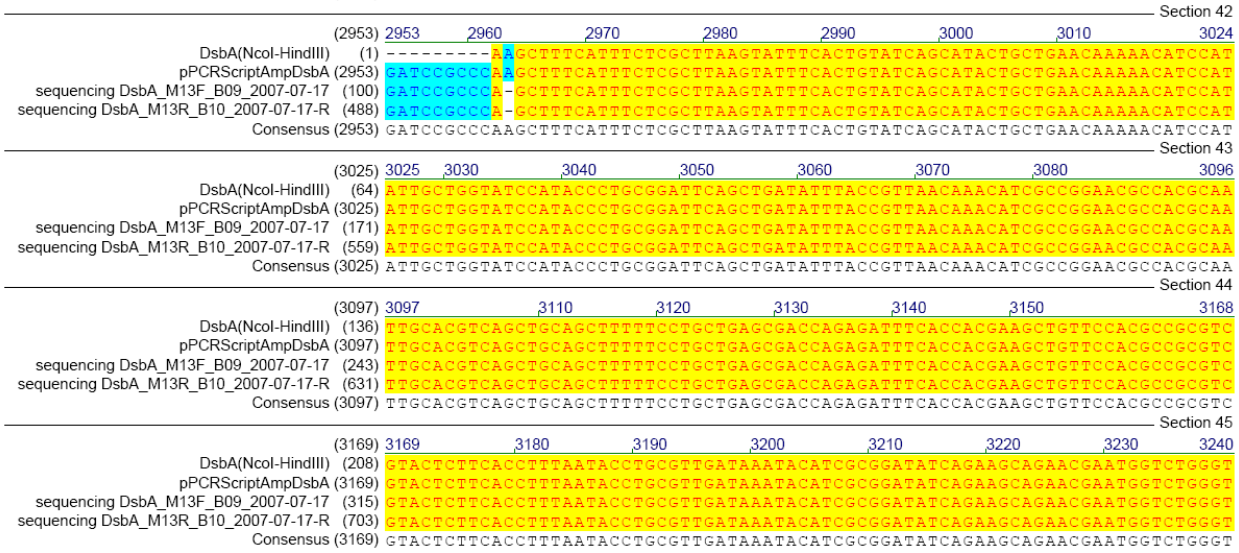
Reducing manufacturing and processing costs is a big challenge for biocatalysts competing with chemical catalysts. From the results obtained in this study, it is possible to reduce the cost by directly using PalB fusion or cell surface displayed PalB in an immobilization process. The study of the enzyme immobilization as cell surface displayed PalB started recently in this lab [276] but further research is needed.

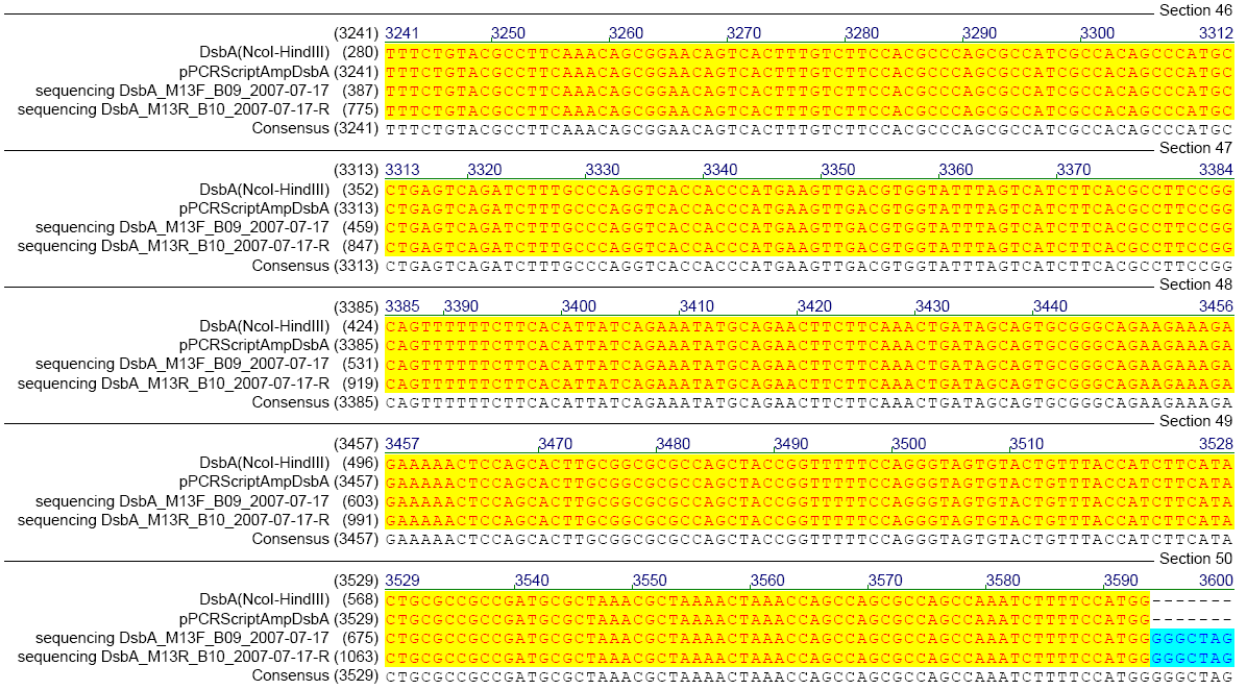
Appendix A

DNA Sequencing Results

A1. DNA Sequencing for pPCRScriptAmpDsbA

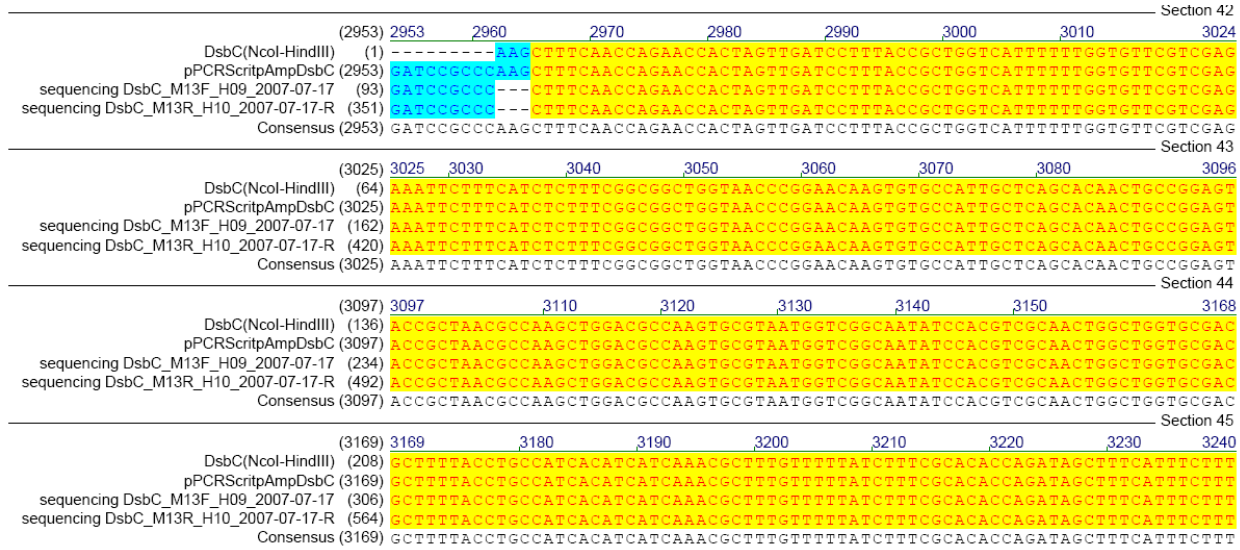
An “A” at 3’-end of HindIII side was missed in pPCRScriptAmpDsbA, but it doesn’t affect the translation of DsbA in the expression vector pARDsbA.

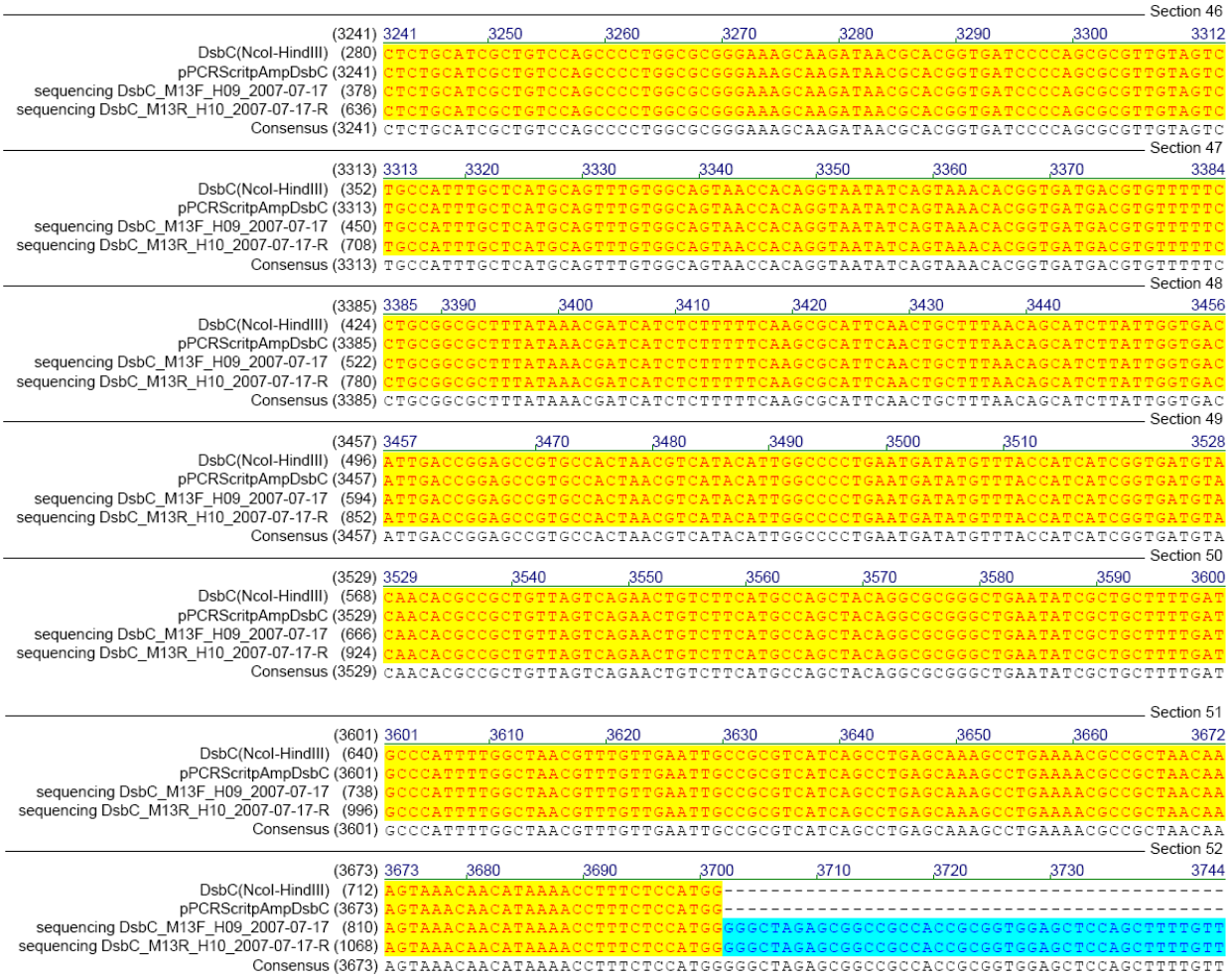




A2. DNA Sequencing for pPCRScriptAmpDsbC

pPCRScriptAmpDsbC is missing AAG at 3'-end of HindIII side, but doesn't affect the translation of DsbC in the expression vector pARDsbC.



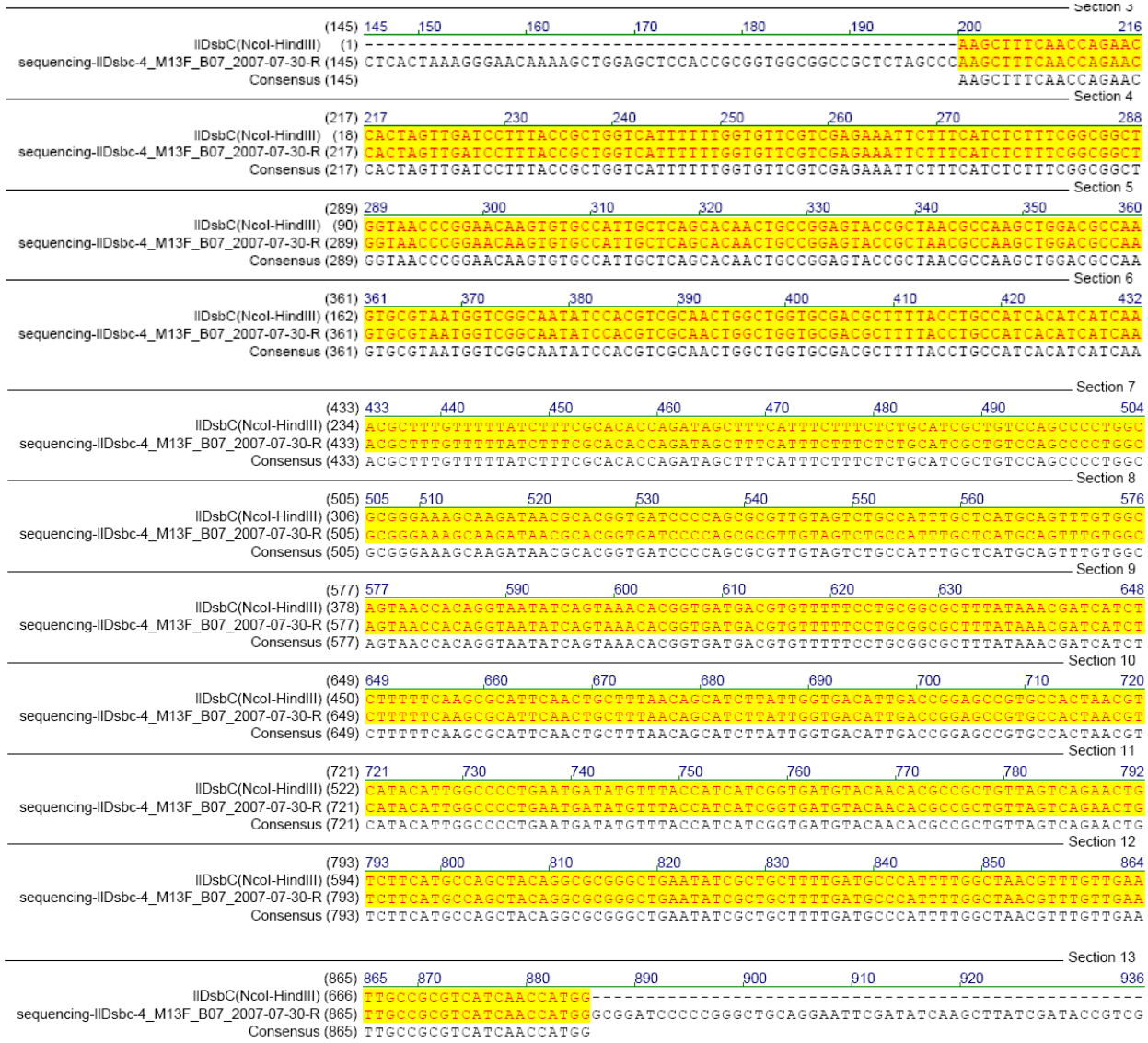

Section 46
(3241) 3241 3250 3260 3270 3280 3290 3300 3312
DsbC(NcoI-HindIII) (280) CTCTGCATCGCTGTCCAGCCCCTGGCGCGGGAAAGCAAGATAACGCACGGTGATCCCCAGCGCGTTGTAGTC
pPCRScritpAmpDsbC (3241) CTCTGCATCGCTGTCCAGCCCCTGGCGCGGGAAAGCAAGATAACGCACGGTGATCCCCAGCGCGTTGTAGTC
sequencing DsbC_M13F_H09_2007-07-17 (378) CTCTGCATCGCTGTCCAGCCCCTGGCGCGGGAAAGCAAGATAACGCACGGTGATCCCCAGCGCGTTGTAGTC
sequencing DsbC_M13R_H10_2007-07-17-R (636) CTCTGCATCGCTGTCCAGCCCCTGGCGCGGGAAAGCAAGATAACGCACGGTGATCCCCAGCGCGTTGTAGTC
Consensus (3241) CTCTGCATCGCTGTCCAGCCCCTGGCGCGGGAAAGCAAGATAACGCACGGTGATCCCCAGCGCGTTGTAGTC
Section 47
(3313) 3313 3320 3330 3340 3350 3360 3370 3384
DsbC(NcoI-HindIII) (352) TGCCATTTGCTCATGCAGTTTGTGGCAGTAACCACAGGTAATATCAGTAAACACGGTGATGACGTGTTTTTC
pPCRScritpAmpDsbC (3313) TGCCATTTGCTCATGCAGTTTGTGGCAGTAACCACAGGTAATATCAGTAAACACGGTGATGACGTGTTTTTC
sequencing DsbC_M13F_H09_2007-07-17 (450) TGCCATTTGCTCATGCAGTTTGTGGCAGTAACCACAGGTAATATCAGTAAACACGGTGATGACGTGTTTTTC
sequencing DsbC_M13R_H10_2007-07-17-R (708) TGCCATTTGCTCATGCAGTTTGTGGCAGTAACCACAGGTAATATCAGTAAACACGGTGATGACGTGTTTTTC
Consensus (3313) TGCCATTTGCTCATGCAGTTTGTGGCAGTAACCACAGGTAATATCAGTAAACACGGTGATGACGTGTTTTTC
Section 48
(3385) 3385 3390 3400 3410 3420 3430 3440 3456
DsbC(NcoI-HindIII) (424) CTGCGGCGCTTTATAAACGATCATCTCTTTTTCAAGCGCATTCAACTGCTTTAACAGCATCTTATTGGTGAC
pPCRScritpAmpDsbC (3385) CTGCGGCGCTTTATAAACGATCATCTCTTTTTCAAGCGCATTCAACTGCTTTAACAGCATCTTATTGGTGAC
sequencing DsbC_M13F_H09_2007-07-17 (522) CTGCGGCGCTTTATAAACGATCATCTCTTTTTCAAGCGCATTCAACTGCTTTAACAGCATCTTATTGGTGAC
sequencing DsbC_M13R_H10_2007-07-17-R (780) CTGCGGCGCTTTATAAACGATCATCTCTTTTTCAAGCGCATTCAACTGCTTTAACAGCATCTTATTGGTGAC
Consensus (3385) CTGCGGCGCTTTATAAACGATCATCTCTTTTTCAAGCGCATTCAACTGCTTTAACAGCATCTTATTGGTGAC
Section 49
(3457) 3457 3470 3480 3490 3500 3510 3528
DsbC(NcoI-HindIII) (496) ATTGACCGGAGCCGTGCCACTAACGTCATACATTGGCCCCTGAATGATATGTTTACCATCATCGGTGATGTA
pPCRScritpAmpDsbC (3457) ATTGACCGGAGCCGTGCCACTAACGTCATACATTGGCCCCTGAATGATATGTTTACCATCATCGGTGATGTA
sequencing DsbC_M13F_H09_2007-07-17 (594) ATTGACCGGAGCCGTGCCACTAACGTCATACATTGGCCCCTGAATGATATGTTTACCATCATCGGTGATGTA
sequencing DsbC_M13R_H10_2007-07-17-R (852) ATTGACCGGAGCCGTGCCACTAACGTCATACATTGGCCCCTGAATGATATGTTTACCATCATCGGTGATGTA
Consensus (3457) ATTGACCGGAGCCGTGCCACTAACGTCATACATTGGCCCCTGAATGATATGTTTACCATCATCGGTGATGTA
Section 50
(3529) 3529 3540 3550 3560 3570 3580 3590 3600
DsbC(NcoI-HindIII) (568) CAACACGCCGCTGTTAGTCAGAACTGTCTTCATGCCAGCTACAGGCGCGGGCTGAATATCGCTGCTTTTGAT
pPCRScritpAmpDsbC (3529) CAACACGCCGCTGTTAGTCAGAACTGTCTTCATGCCAGCTACAGGCGCGGGCTGAATATCGCTGCTTTTGAT
sequencing DsbC_M13F_H09_2007-07-17 (666) CAACACGCCGCTGTTAGTCAGAACTGTCTTCATGCCAGCTACAGGCGCGGGCTGAATATCGCTGCTTTTGAT
sequencing DsbC_M13R_H10_2007-07-17-R (924) CAACACGCCGCTGTTAGTCAGAACTGTCTTCATGCCAGCTACAGGCGCGGGCTGAATATCGCTGCTTTTGAT
Consensus (3529) CAACACGCCGCTGTTAGTCAGAACTGTCTTCATGCCAGCTACAGGCGCGGGCTGAATATCGCTGCTTTTGAT
Section 51
(3601) 3601 3610 3620 3630 3640 3650 3660 3672
DsbC(NcoI-HindIII) (640) GCCCATTTTGGCTAACGTTTGTTGAATTGCCGCGTCATCAGCCTGAGCAAAGCCTGAAAACGCCGCTAACAA
pPCRScritpAmpDsbC (3601) GCCCATTTTGGCTAACGTTTGTTGAATTGCCGCGTCATCAGCCTGAGCAAAGCCTGAAAACGCCGCTAACAA
sequencing DsbC_M13F_H09_2007-07-17 (738) GCCCATTTTGGCTAACGTTTGTTGAATTGCCGCGTCATCAGCCTGAGCAAAGCCTGAAAACGCCGCTAACAA
sequencing DsbC_M13R_H10_2007-07-17-R (996) GCCCATTTTGGCTAACGTTTGTTGAATTGCCGCGTCATCAGCCTGAGCAAAGCCTGAAAACGCCGCTAACAA
Consensus (3601) GCCCATTTTGGCTAACGTTTGTTGAATTGCCGCGTCATCAGCCTGAGCAAAGCCTGAAAACGCCGCTAACAA
Section 52
(3673) 3673 3680 3690 3700 3710 3720 3730 3744
DsbC(NcoI-HindIII) (712) AGTAAACAACATAAAACCTTTCTCCATGG----------
pPCRScritpAmpDsbC (3673) AGTAAACAACATAAAACCTTTCTCCATGG----------
sequencing DsbC_M13F_H09_2007-07-17 (810) AGTAAACAACATAAAACCTTTCTCCATGGGGGCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTT
sequencing DsbC_M13R_H10_2007-07-17-R (1068) AGTAAACAACATAAAACCTTTCTCCATGGGGGCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTT
Consensus (3673) AGTAAACAACATAAAACCTTTCTCCATGGGGGCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTT

A3. DNA Sequencing for pPCRScriptAmpllDsbA

PCR product of *lldsbA* in pPCRScriptAmpllDsbA is sequenced correctly.

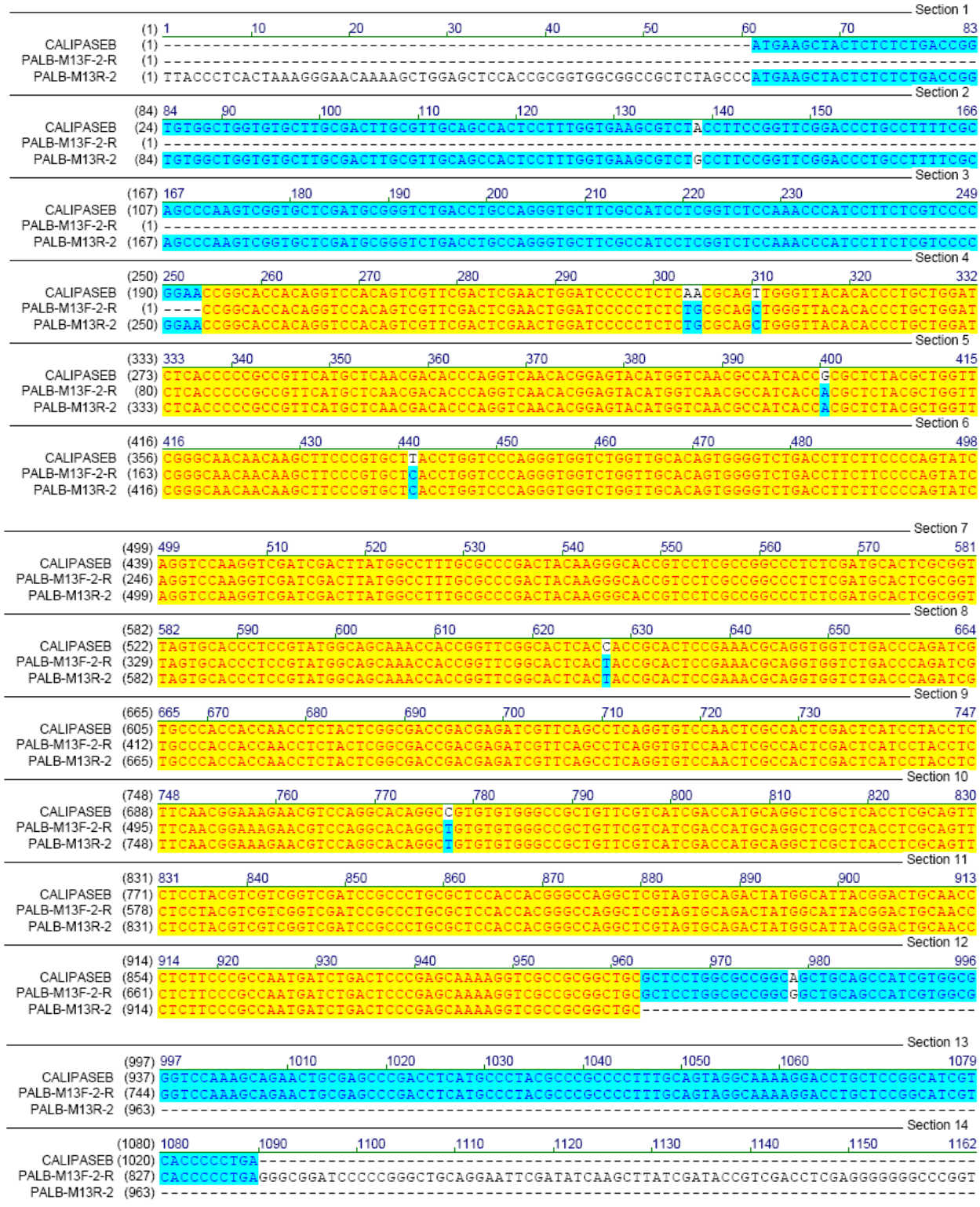
A4. DNA Sequencing for pPCRScriptAmpllDsbC

PCR product of *lldsbC* in pPCRScriptAmpllDsbC is sequenced correctly.



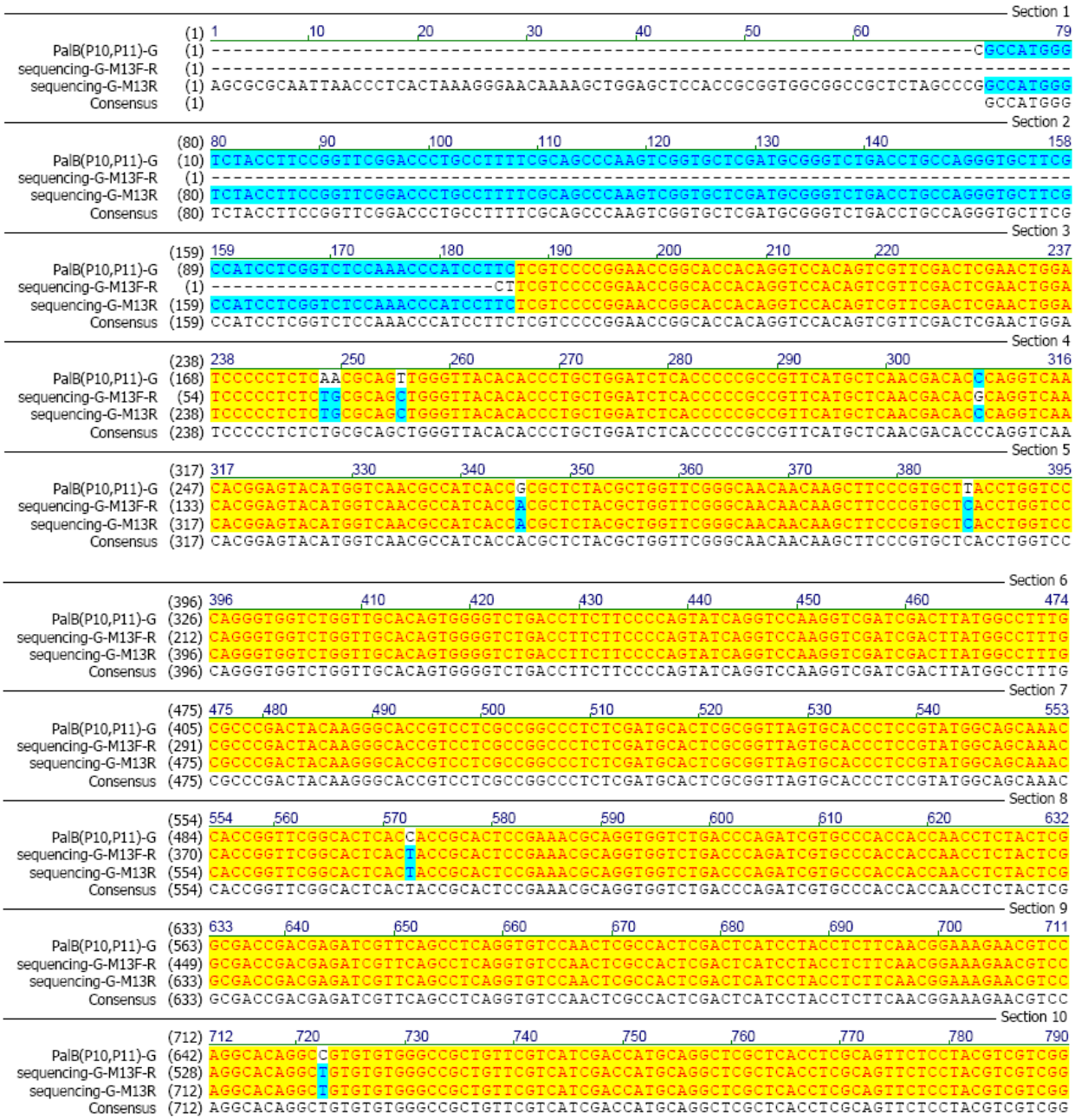
A5. DNA Sequencing for PCR Product of *palB* without Restriction Sites

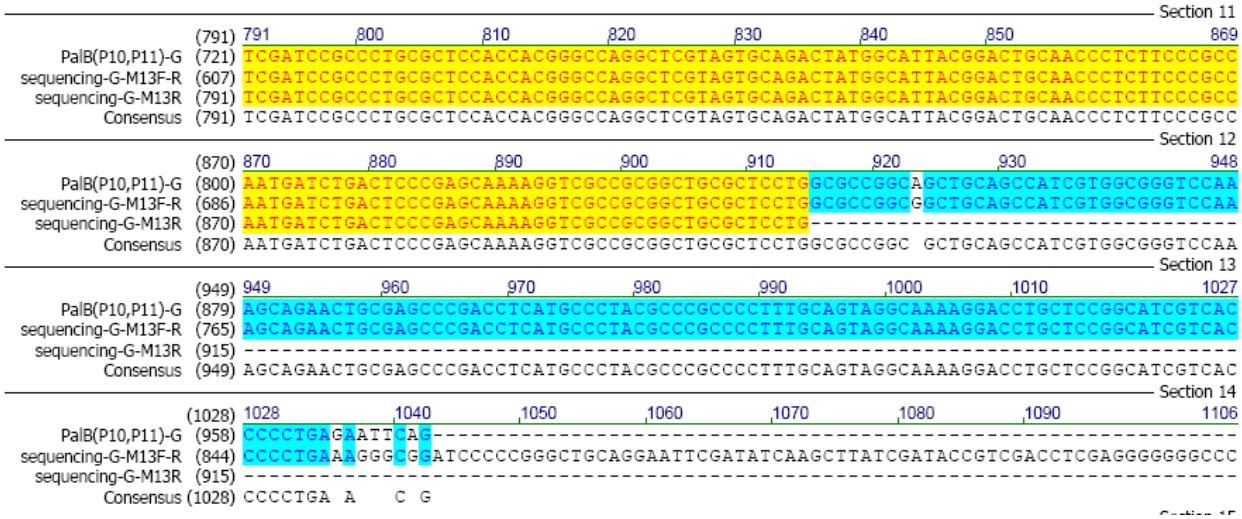
There were 9-codon differences between PCR product of *palB* and the documented sequence. But comparing the different PCR products for different constructions, all these differences are almost the same. Therefore, these "mutations" might come from either the host ATCC 32457, or the typos in the documentation. Even if these mutations were real, they could not change amino acids because they are all silent mutations.


Section 1
CALIPASEB (1) ----------ATGAAGCTACTCTCTCTGACCGG
PALB-M13F-2-R (1) ----------
PALB-M13R-2 (1) TTACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGCCCATGAAGCTACTCTCTCTGACCGG
Section 2
CALIPASEB (24) TGTGGCTGGTGTGCTTGCGACTTGCGTTGCAGCCACTCCTTTGGTGAAGCGTCTACCTTCCGGTTCGGACCCTGCCTTTTCGC
PALB-M13F-2-R (1) ----------
PALB-M13R-2 (84) TGTGGCTGGTGTGCTTGCGACTTGCGTTGCAGCCACTCCTTTGGTGAAGCGTCTGCCTTCCGGTTCGGACCCTGCCTTTTCGC
Section 3
CALIPASEB (107) AGCCCAAGTCGGTGCTCGATGCGGGTCTGACCTGCCAGGGTGCTTCGCCATCCTCGGTCTCCAAACCCATCCTTCTCGTCCCC
PALB-M13F-2-R (1) ----------
PALB-M13R-2 (167) AGCCCAAGTCGGTGCTCGATGCGGGTCTGACCTGCCAGGGTGCTTCGCCATCCTCGGTCTCCAAACCCATCCTTCTCGTCCCC
Section 4
CALIPASEB (190) GGAACCGGCACCACAGGTCCACAGTCGTTCGACTCGAACTGGATCCCCCTCTCAACGCAGTTGGGTTACACACCCTGCTGGAT
PALB-M13F-2-R (1) ----CCGGCACCACAGGTCCACAGTCGTTCGACTCGAACTGGATCCCCCTCTCTGCGCAGCTGGGTTACACACCCTGCTGGAT
PALB-M13R-2 (250) GGAACCGGCACCACAGGTCCACAGTCGTTCGACTCGAACTGGATCCCCCTCTCTGCGCAGCTGGGTTACACACCCTGCTGGAT
Section 5
CALIPASEB (273) CTCACCCCCGCCGTTCATGCTCAACGACACCCAGGTCAACACGGAGTACATGGTCAACGCCATCACCGCGCTCTACGCTGGTT
PALB-M13F-2-R (80) CTCACCCCCGCCGTTCATGCTCAACGACACCCAGGTCAACACGGAGTACATGGTCAACGCCATCACCACGCTCTACGCTGGTT
PALB-M13R-2 (333) CTCACCCCCGCCGTTCATGCTCAACGACACCCAGGTCAACACGGAGTACATGGTCAACGCCATCACCACGCTCTACGCTGGTT
Section 6
CALIPASEB (356) CGGGCAACAACAAGCTTCCCGTGCTTACCTGGTCCCAGGGTGGTCTGGTTGCACAGTGGGGTCTGACCTTCTTCCCCAGTATC
PALB-M13F-2-R (163) CGGGCAACAACAAGCTTCCCGTGCTCACCTGGTCCCAGGGTGGTCTGGTTGCACAGTGGGGTCTGACCTTCTTCCCCAGTATC
PALB-M13R-2 (416) CGGGCAACAACAAGCTTCCCGTGCTCACCTGGTCCCAGGGTGGTCTGGTTGCACAGTGGGGTCTGACCTTCTTCCCCAGTATC
Section 7
CALIPASEB (439) AGGTCCAAGGTCGATCGACTTATGGCCTTTGCGCCCGACTACAAGGGCACCGTCCTCGCCGGCCCTCTCGATGCACTCGCGGT
PALB-M13F-2-R (246) AGGTCCAAGGTCGATCGACTTATGGCCTTTGCGCCCGACTACAAGGGCACCGTCCTCGCCGGCCCTCTCGATGCACTCGCGGT
PALB-M13R-2 (499) AGGTCCAAGGTCGATCGACTTATGGCCTTTGCGCCCGACTACAAGGGCACCGTCCTCGCCGGCCCTCTCGATGCACTCGCGGT
Section 8
CALIPASEB (522) TAGTGCACCCTCCGTATGGCAGCAAACCACCGGTTCGGCACTCACCACCGCACTCCGAAACGCAGGTGGTCTGACCCAGATCG
PALB-M13F-2-R (329) TAGTGCACCCTCCGTATGGCAGCAAACCACCGGTTCGGCACTCACTACCGCACTCCGAAACGCAGGTGGTCTGACCCAGATCG
PALB-M13R-2 (582) TAGTGCACCCTCCGTATGGCAGCAAACCACCGGTTCGGCACTCACTACCGCACTCCGAAACGCAGGTGGTCTGACCCAGATCG
Section 9
CALIPASEB (605) TGCCCACCACCAACCTCTACTCGGCGACCGACGAGATCGTTCAGCCTCAGGTGTCCAACTCGCCACTCGACTCATCCTACCTC
PALB-M13F-2-R (412) TGCCCACCACCAACCTCTACTCGGCGACCGACGAGATCGTTCAGCCTCAGGTGTCCAACTCGCCACTCGACTCATCCTACCTC
PALB-M13R-2 (665) TGCCCACCACCAACCTCTACTCGGCGACCGACGAGATCGTTCAGCCTCAGGTGTCCAACTCGCCACTCGACTCATCCTACCTC
Section 10
CALIPASEB (688) TTCAACGGAAAGAACGTCCAGGCACAGGCCGTGTGTGGGCCGCTGTTCGTCATCGACCATGCAGGCTCGCTCACCTCGCAGTT
PALB-M13F-2-R (495) TTCAACGGAAAGAACGTCCAGGCACAGGCTGTGTGTGGGCCGCTGTTCGTCATCGACCATGCAGGCTCGCTCACCTCGCAGTT
PALB-M13R-2 (748) TTCAACGGAAAGAACGTCCAGGCACAGGCTGTGTGTGGGCCGCTGTTCGTCATCGACCATGCAGGCTCGCTCACCTCGCAGTT
Section 11
CALIPASEB (771) CTCCTACGTCGTCGGTCGATCCGCCCTGCGCTCCACCACGGGCCAGGCTCGTAGTGCAGACTATGGCATTACGGACTGCAACC
PALB-M13F-2-R (578) CTCCTACGTCGTCGGTCGATCCGCCCTGCGCTCCACCACGGGCCAGGCTCGTAGTGCAGACTATGGCATTACGGACTGCAACC
PALB-M13R-2 (831) CTCCTACGTCGTCGGTCGATCCGCCCTGCGCTCCACCACGGGCCAGGCTCGTAGTGCAGACTATGGCATTACGGACTGCAACC
Section 12
CALIPASEB (854) CTCTTCCCGCCAATGATCTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGCAGCTGCAGCCATCGTGGCG
PALB-M13F-2-R (661) CTCTTCCCGCCAATGATCTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGCGGCTGCAGCCATCGTGGCG
PALB-M13R-2 (914) CTCTTCCCGCCAATGATCTGACTCCCGAGCAAAAGGTCGCCGCGGCTGC----------
Section 13
CALIPASEB (937) GGTCCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCTTTGCAGTAGGCAAAAGGACCTGCTCCGGCATCGT
PALB-M13F-2-R (744) GGTCCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCTTTGCAGTAGGCAAAAGGACCTGCTCCGGCATCGT
PALB-M13R-2 (963) ----------
Section 14
CALIPASEB (1020) CACCCCCTGA----------
PALB-M13F-2-R (827) CACCCCCTGAGGGCGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGT
PALB-M13R-2 (963) ----------

A6. DNA Sequencing for pPCRScriptAmpG

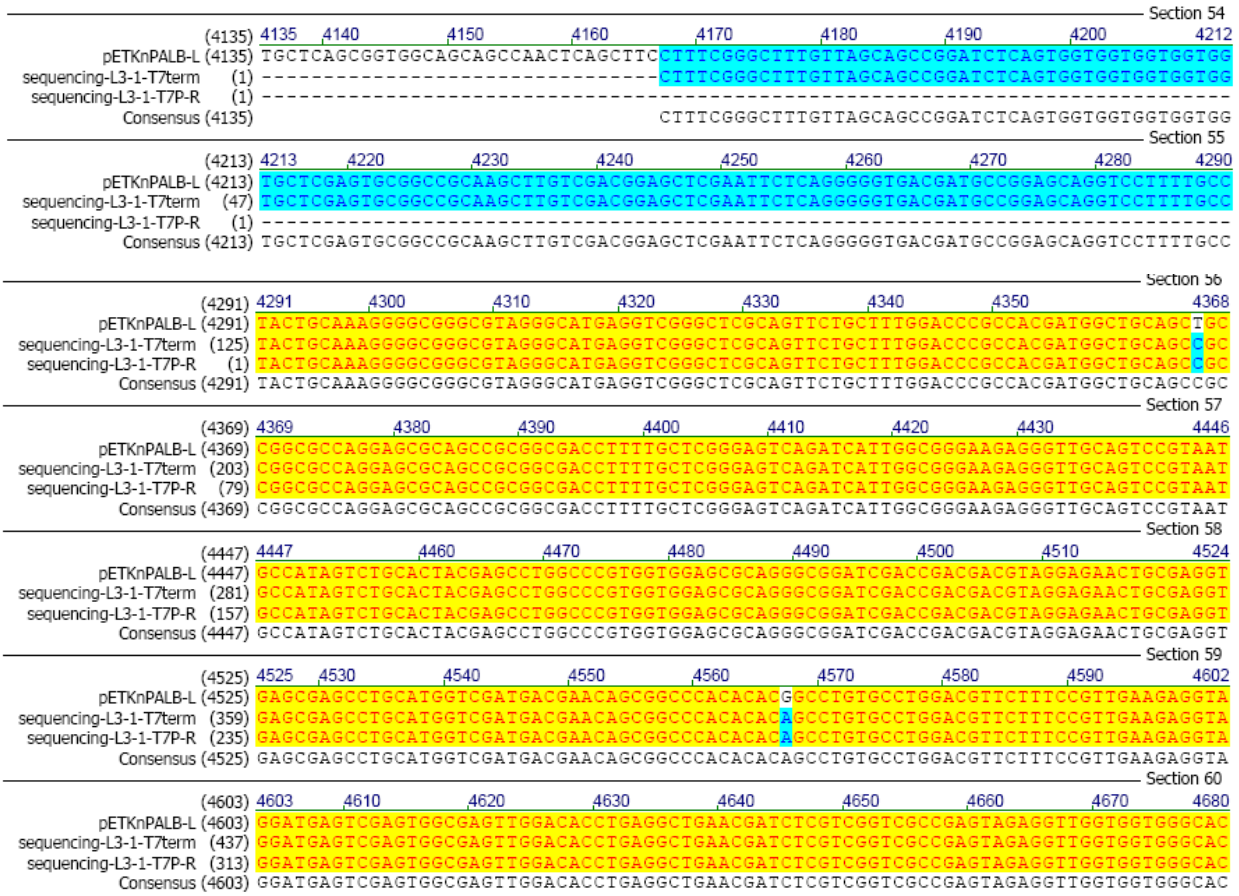
Sequencing is OK, even though the 3' end is wrong, which does not affect the *palB (g)* in frame and the stop codon for translation termination.

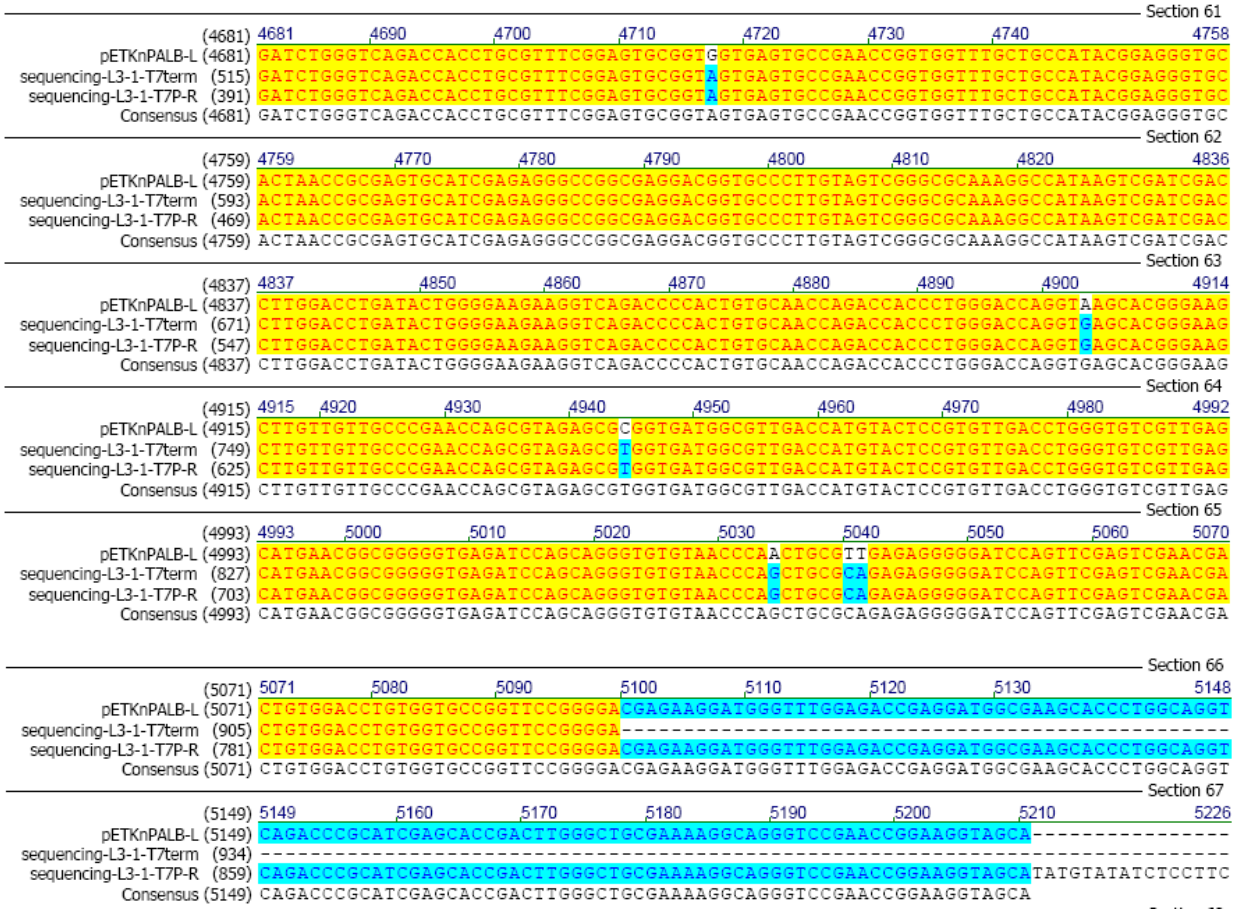




A7. DNA Sequencing for pETknL

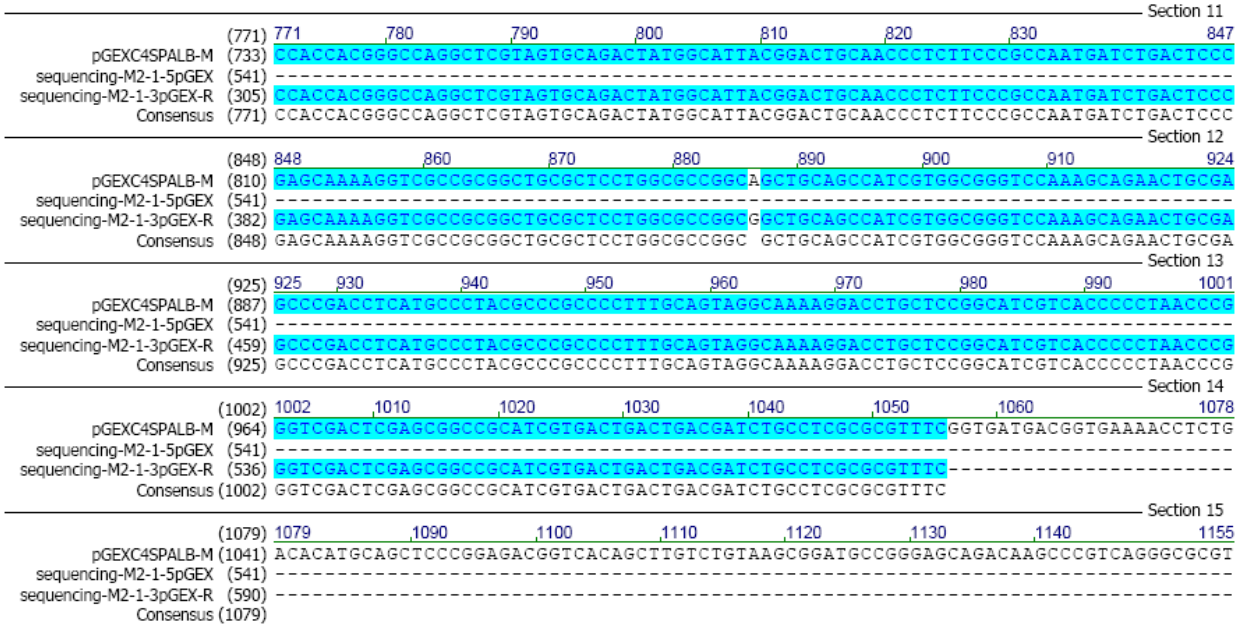
PCR product of *palB*(*l*) is sequenced correctly.




Section 61
(4681) 4681 4690 4700 4710 4720 4730 4740 4758
pETKnPALB-L (4681) GATCTGGGTCAGACCACCTGCGTTTCGGAGTGCGGTGGTGAGTGCCGAACCGGTGGTTTGCTGCCATACGGAGGGTGC
sequencing-L3-1-T7term (515) GATCTGGGTCAGACCACCTGCGTTTCGGAGTGCGGTAGTGAGTGCCGAACCGGTGGTTTGCTGCCATACGGAGGGTGC
sequencing-L3-1-T7P-R (391) GATCTGGGTCAGACCACCTGCGTTTCGGAGTGCGGTAGTGAGTGCCGAACCGGTGGTTTGCTGCCATACGGAGGGTGC
Consensus (4681) GATCTGGGTCAGACCACCTGCGTTTCGGAGTGCGGTAGTGAGTGCCGAACCGGTGGTTTGCTGCCATACGGAGGGTGC
Section 62
(4759) 4759 4770 4780 4790 4800 4810 4820 4836
pETKnPALB-L (4759) ACTAACCGCGAGTGCATCGAGAGGGCCGGCGAGGACGGTGCCCTTGTAGTCGGGCGCAAAGGCCATAAGTCGATCGAC
sequencing-L3-1-T7term (593) ACTAACCGCGAGTGCATCGAGAGGGCCGGCGAGGACGGTGCCCTTGTAGTCGGGCGCAAAGGCCATAAGTCGATCGAC
sequencing-L3-1-T7P-R (469) ACTAACCGCGAGTGCATCGAGAGGGCCGGCGAGGACGGTGCCCTTGTAGTCGGGCGCAAAGGCCATAAGTCGATCGAC
Consensus (4759) ACTAACCGCGAGTGCATCGAGAGGGCCGGCGAGGACGGTGCCCTTGTAGTCGGGCGCAAAGGCCATAAGTCGATCGAC
Section 63
(4837) 4837 4850 4860 4870 4880 4890 4900 4914
pETKnPALB-L (4837) CTTGGACCTGATACTGGGGAAGAAGGTCAGACCCCACTGTGCAACCAGACCACCCTGGGACCAGGTAAGCACGGGAAG
sequencing-L3-1-T7term (671) CTTGGACCTGATACTGGGGAAGAAGGTCAGACCCCACTGTGCAACCAGACCACCCTGGGACCAGGTGAGCACGGGAAG
sequencing-L3-1-T7P-R (547) CTTGGACCTGATACTGGGGAAGAAGGTCAGACCCCACTGTGCAACCAGACCACCCTGGGACCAGGTGAGCACGGGAAG
Consensus (4837) CTTGGACCTGATACTGGGGAAGAAGGTCAGACCCCACTGTGCAACCAGACCACCCTGGGACCAGGTGAGCACGGGAAG
Section 64
(4915) 4915 4920 4930 4940 4950 4960 4970 4980 4992
pETKnPALB-L (4915) CTTGTTGTTGCCCGAACCAGCGTAGAGCGCGGTGATGGCGTTGACCATGTACTCCGTGTTGACCTGGGTGTCGTTGAG
sequencing-L3-1-T7term (749) CTTGTTGTTGCCCGAACCAGCGTAGAGCGTGGTGATGGCGTTGACCATGTACTCCGTGTTGACCTGGGTGTCGTTGAG
sequencing-L3-1-T7P-R (625) CTTGTTGTTGCCCGAACCAGCGTAGAGCGTGGTGATGGCGTTGACCATGTACTCCGTGTTGACCTGGGTGTCGTTGAG
Consensus (4915) CTTGTTGTTGCCCGAACCAGCGTAGAGCGTGGTGATGGCGTTGACCATGTACTCCGTGTTGACCTGGGTGTCGTTGAG
Section 65
(4993) 4993 5000 5010 5020 5030 5040 5050 5060 5070
pETKnPALB-L (4993) CATGAACGGCGGGGGGTGAGATCCAGCAGGGTGTGTAACCCAACTGCGTTGAGAGGGGGATCCAGTTCGAGTCGAACGA
sequencing-L3-1-T7term (827) CATGAACGGCGGGGGGTGAGATCCAGCAGGGTGTGTAACCCAGCTGCGCAGAGAGGGGGATCCAGTTCGAGTCGAACGA
sequencing-L3-1-T7P-R (703) CATGAACGGCGGGGGGTGAGATCCAGCAGGGTGTGTAACCCAGCTGCGCAGAGAGGGGGATCCAGTTCGAGTCGAACGA
Consensus (4993) CATGAACGGCGGGGGGTGAGATCCAGCAGGGTGTGTAACCCAGCTGCGCAGAGAGGGGGATCCAGTTCGAGTCGAACGA
Section 66
(5071) 5071 5080 5090 5100 5110 5120 5130 5148
pETKnPALB-L (5071) CTGTGGACCTGTGGTGCCGGTTCCGGGGACGAGAAGGATGGGTTTGGAGACCGAGGATGGCGAAGCACCCTGGCAGGT
sequencing-L3-1-T7term (905) CTGTGGACCTGTGGTGCCGGTTCCGGGGA----------
sequencing-L3-1-T7P-R (781) CTGTGGACCTGTGGTGCCGGTTCCGGGGACGAGAAGGATGGGTTTGGAGACCGAGGATGGCGAAGCACCCTGGCAGGT
Consensus (5071) CTGTGGACCTGTGGTGCCGGTTCCGGGGACGAGAAGGATGGGTTTGGAGACCGAGGATGGCGAAGCACCCTGGCAGGT
Section 67
(5149) 5149 5160 5170 5180 5190 5200 5210 5226
pETKnPALB-L (5149) CAGACCCGCATCGAGCACCGACTTGGGCTGCGAAAAGGCAGGGTCCGAACCGGAAGGTAGCA----------
sequencing-L3-1-T7term (934) ----------
sequencing-L3-1-T7P-R (859) CAGACCCGCATCGAGCACCGACTTGGGCTGCGAAAAGGCAGGGTCCGAACCGGAAGGTAGCATATGTATATCTCCTTC
Consensus (5149) CAGACCCGCATCGAGCACCGACTTGGGCTGCGAAAAGGCAGGGTCCGAACCGGAAGGTAGCA

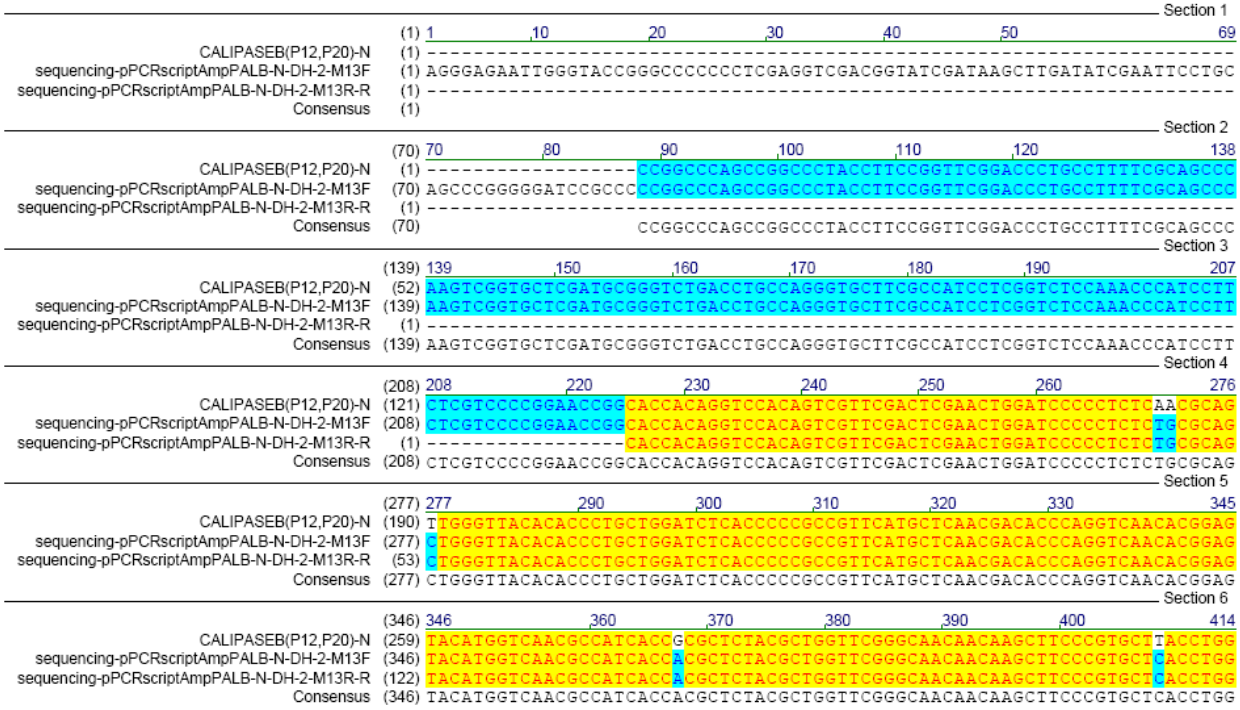
A8. DNA Sequencing for pGEXC4S-M

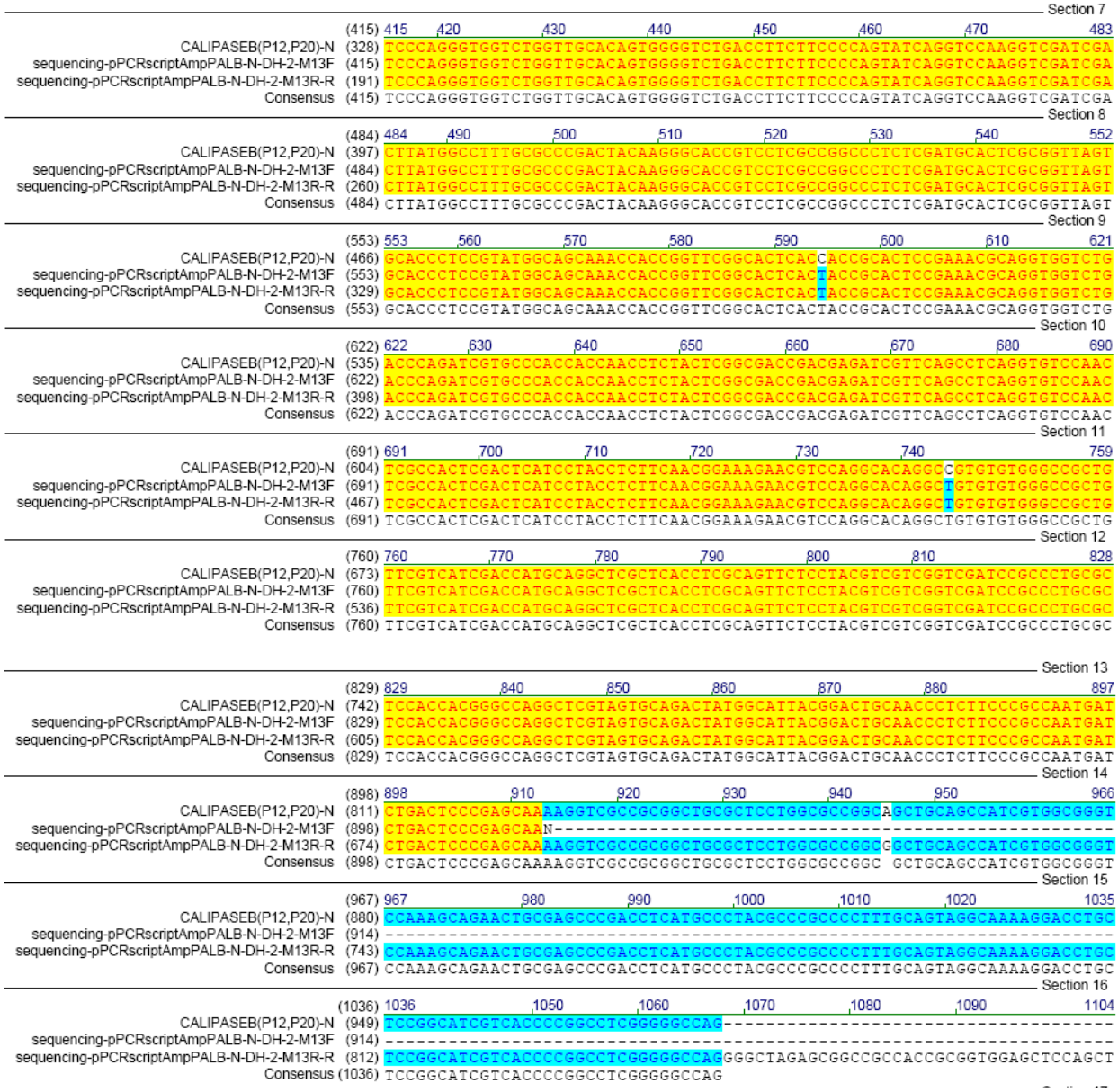
PCR product *palB*(*m*) is sequenced correctly.



A9. DNA Sequencing for pPCRScriptAmpN

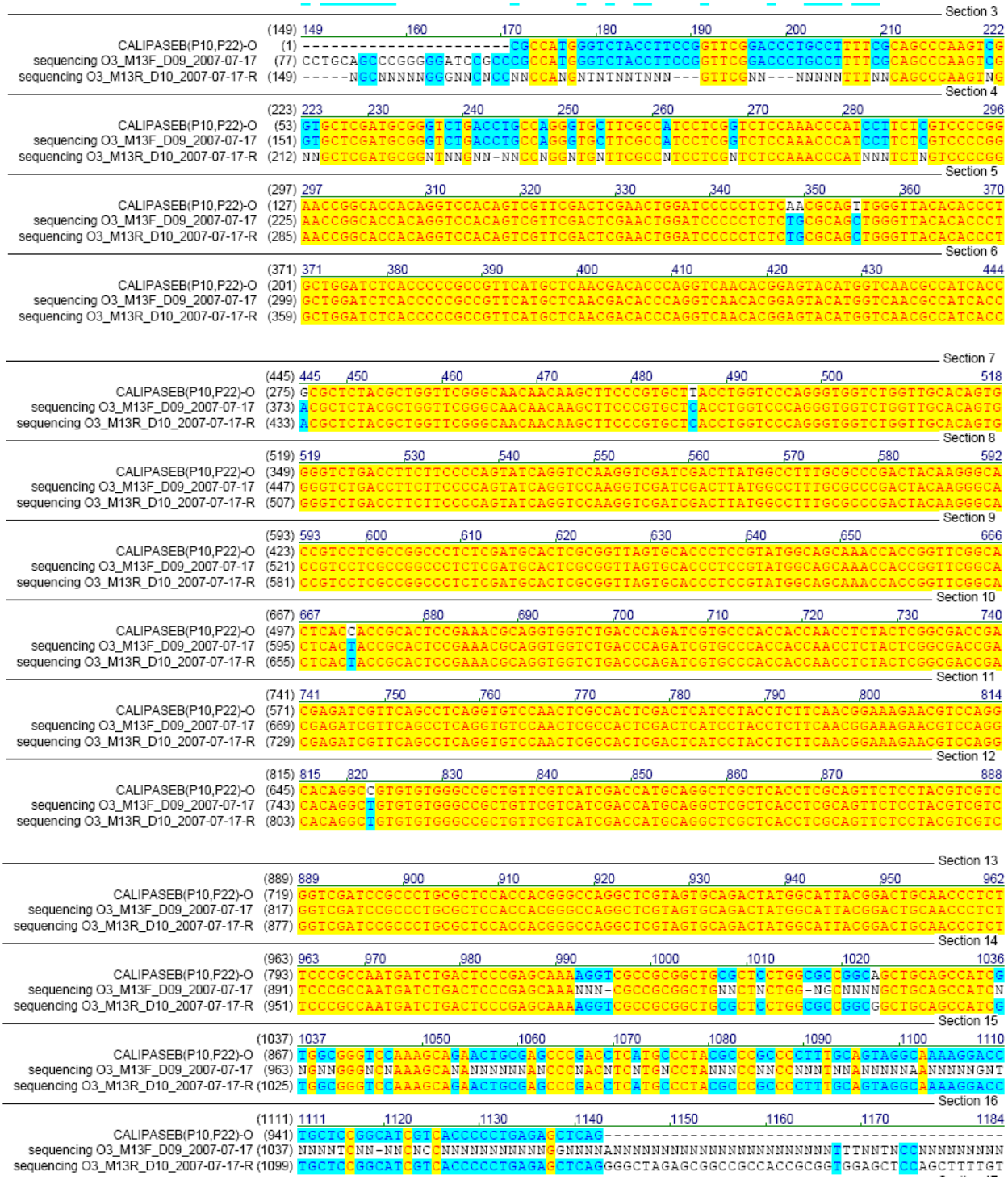
PCR product of *palB*(*n*) is sequenced correctly.




Section 7
(415) 415 420 430 440 450 460 470 483
CALIPASEB(P12,P20)-N (328) TCCCAGGGTGGTCTGGTTGCACAGTGGGGTCTGACCTTCTTCCCCAGTATCAGGTCCAAGGTCGATCGA
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (415) TCCCAGGGTGGTCTGGTTGCACAGTGGGGTCTGACCTTCTTCCCCAGTATCAGGTCCAAGGTCGATCGA
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (191) TCCCAGGGTGGTCTGGTTGCACAGTGGGGTCTGACCTTCTTCCCCAGTATCAGGTCCAAGGTCGATCGA
Consensus (415) TCCCAGGGTGGTCTGGTTGCACAGTGGGGTCTGACCTTCTTCCCCAGTATCAGGTCCAAGGTCGATCGA
Section 8
(484) 484 490 500 510 520 530 540 552
CALIPASEB(P12,P20)-N (397) CTTATGGCCTTTGCGCCCGACTACAAGGGCACCGTCCTCGCCGGCCCTCTCGATGCACTCGCGGTTAGT
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (484) CTTATGGCCTTTGCGCCCGACTACAAGGGCACCGTCCTCGCCGGCCCTCTCGATGCACTCGCGGTTAGT
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (260) CTTATGGCCTTTGCGCCCGACTACAAGGGCACCGTCCTCGCCGGCCCTCTCGATGCACTCGCGGTTAGT
Consensus (484) CTTATGGCCTTTGCGCCCGACTACAAGGGCACCGTCCTCGCCGGCCCTCTCGATGCACTCGCGGTTAGT
Section 9
(553) 553 560 570 580 590 600 610 621
CALIPASEB(P12,P20)-N (466) GCACCCTCCGTATGGCAGCAAACCACCGGTTCGGCACTCACCACCGCACTCCGAAACGCAGGTGGTCTG
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (553) GCACCCTCCGTATGGCAGCAAACCACCGGTTCGGCACTCACTACCGCACTCCGAAACGCAGGTGGTCTG
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (329) GCACCCTCCGTATGGCAGCAAACCACCGGTTCGGCACTCACTACCGCACTCCGAAACGCAGGTGGTCTG
Consensus (553) GCACCCTCCGTATGGCAGCAAACCACCGGTTCGGCACTCACTACCGCACTCCGAAACGCAGGTGGTCTG
Section 10
(622) 622 630 640 650 660 670 680 690
CALIPASEB(P12,P20)-N (535) ACCCAGATCGTGCCCACCACCAACCTCTACTCGGCGACCGACGAGATCGTTCAGCCTCAGGTGTCCAAC
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (622) ACCCAGATCGTGCCCACCACCAACCTCTACTCGGCGACCGACGAGATCGTTCAGCCTCAGGTGTCCAAC
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (398) ACCCAGATCGTGCCCACCACCAACCTCTACTCGGCGACCGACGAGATCGTTCAGCCTCAGGTGTCCAAC
Consensus (622) ACCCAGATCGTGCCCACCACCAACCTCTACTCGGCGACCGACGAGATCGTTCAGCCTCAGGTGTCCAAC
Section 11
(691) 691 700 710 720 730 740 759
CALIPASEB(P12,P20)-N (604) TCGCCACTCGACTCATCCTACCTCTTCAACGGAAAGAACGTCCAGGCACAGGCCGTGTGTGGGCCGCTG
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (691) TCGCCACTCGACTCATCCTACCTCTTCAACGGAAAGAACGTCCAGGCACAGGCTGTGTGTGGGCCGCTG
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (467) TCGCCACTCGACTCATCCTACCTCTTCAACGGAAAGAACGTCCAGGCACAGGCTGTGTGTGGGCCGCTG
Consensus (691) TCGCCACTCGACTCATCCTACCTCTTCAACGGAAAGAACGTCCAGGCACAGGCTGTGTGTGGGCCGCTG
Section 12
(760) 760 770 780 790 800 810 828
CALIPASEB(P12,P20)-N (673) TTCGTCATCGACCATGCAGGCTCGCTCACCTCGCAGTTCTCCTACGTCGTCGGTCGATCCGCCCTGCGC
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (760) TTCGTCATCGACCATGCAGGCTCGCTCACCTCGCAGTTCTCCTACGTCGTCGGTCGATCCGCCCTGCGC
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (536) TTCGTCATCGACCATGCAGGCTCGCTCACCTCGCAGTTCTCCTACGTCGTCGGTCGATCCGCCCTGCGC
Consensus (760) TTCGTCATCGACCATGCAGGCTCGCTCACCTCGCAGTTCTCCTACGTCGTCGGTCGATCCGCCCTGCGC
Section 13
(829) 829 840 850 860 870 880 897
CALIPASEB(P12,P20)-N (742) TCCACCACGGGCCAGGCTCGTAGTGCAGACTATGGCATTACGGACTGCAACCCTCTTCCCGCCAATGAT
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (829) TCCACCACGGGCCAGGCTCGTAGTGCAGACTATGGCATTACGGACTGCAACCCTCTTCCCGCCAATGAT
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (605) TCCACCACGGGCCAGGCTCGTAGTGCAGACTATGGCATTACGGACTGCAACCCTCTTCCCGCCAATGAT
Consensus (829) TCCACCACGGGCCAGGCTCGTAGTGCAGACTATGGCATTACGGACTGCAACCCTCTTCCCGCCAATGAT
Section 14
(898) 898 910 920 930 940 950 966
CALIPASEB(P12,P20)-N (811) CTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGCAGCTGCAGCCATCGTGGCGGGT
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (898) CTGACTCCCGAGCAAN----------
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (674) CTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGCGGCTGCAGCCATCGTGGCGGGT
Consensus (898) CTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGC GCTGCAGCCATCGTGGCGGGT
Section 15
(967) 967 980 990 1000 1010 1020 1035
CALIPASEB(P12,P20)-N (880) CCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCCTTTGCAGTAGGCAAAAGGACCTGC
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (914) ----------
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (743) CCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCCTTTGCAGTAGGCAAAAGGACCTGC
Consensus (967) CCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCCTTTGCAGTAGGCAAAAGGACCTGC
Section 16
(1036) 1036 1050 1060 1070 1080 1090 1104
CALIPASEB(P12,P20)-N (949) TCCGGCATCGTCACCCCGGCCTCGGGGGCCAG----------
sequencing-pPCRscriptAmpPALB-N-DH-2-M13F (914) ----------
sequencing-pPCRscriptAmpPALB-N-DH-2-M13R-R (812) TCCGGCATCGTCACCCCGGCCTCGGGGGCCAGGGGCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCT
Consensus (1036) TCCGGCATCGTCACCCCGGCCTCGGGGGCCAG

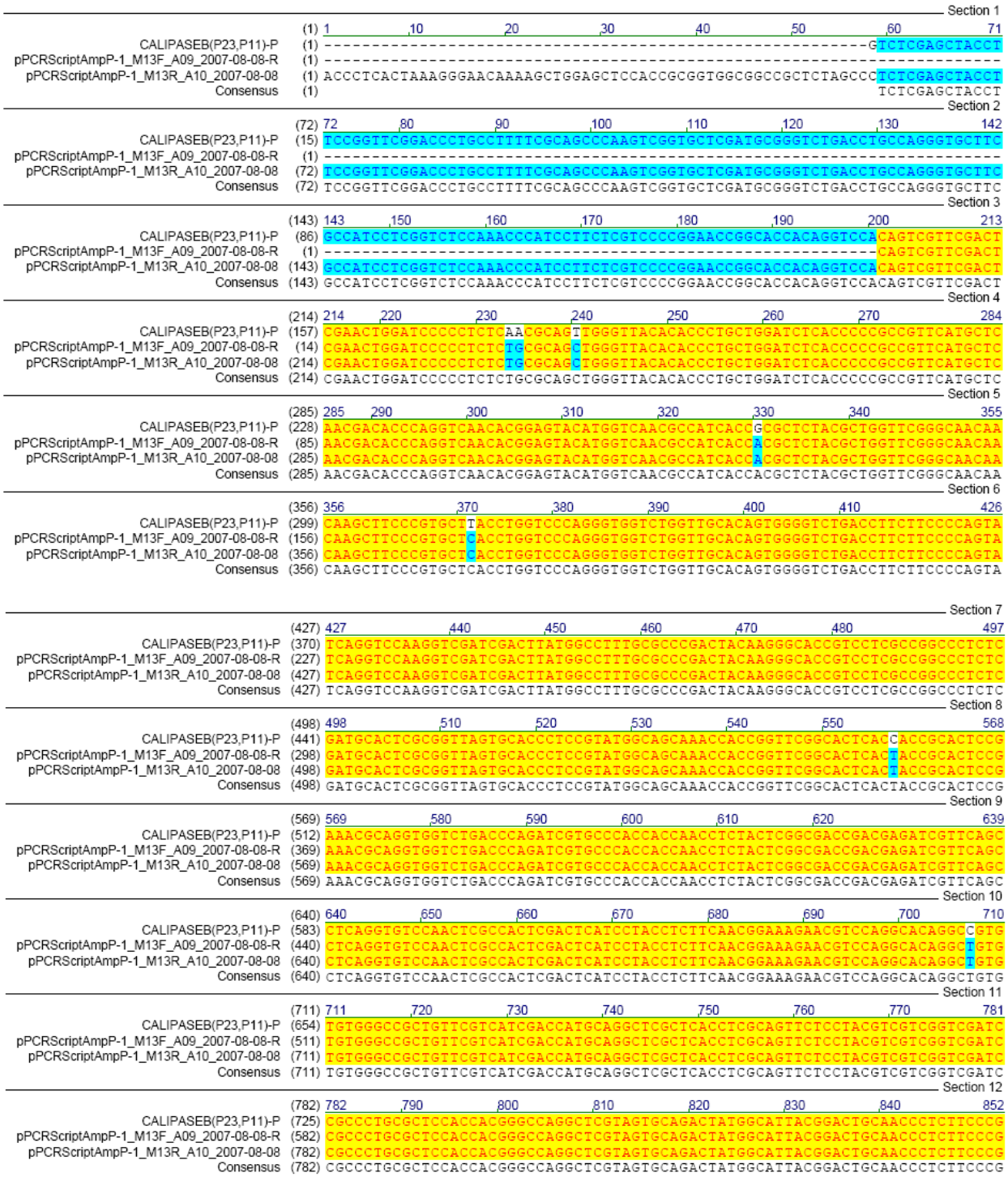
A10. DNA Sequencing for pPCRScriptCmO

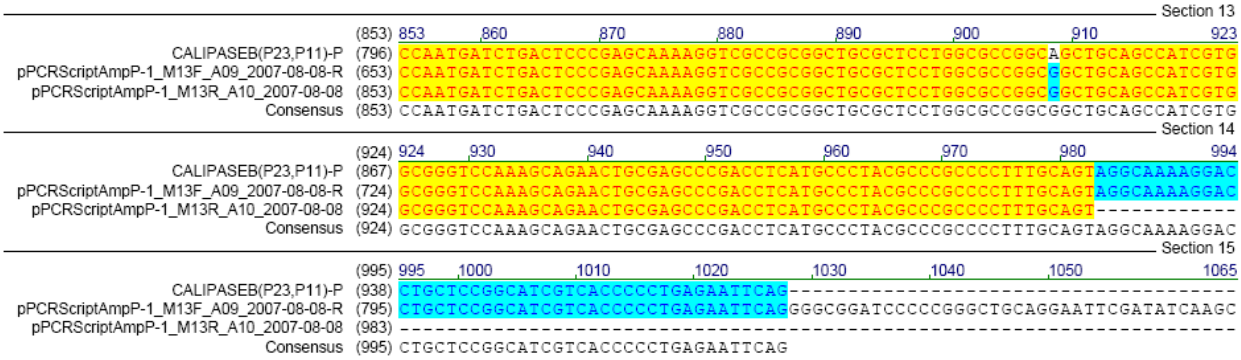
PCR product of *palB*(*o*), is sequenced correctly.



A11. DNA Sequencing for pPCRScriptAmpP

PCR product of *palB*(*p*) is sequenced correctly.




Section 13
(853) 853 860 870 880 890 900 910 923
CALIPASEB(P23,P11)-P (796) CCAATGATCTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGCAGCTGCAGCCATCGTG
pPCRScriptAmpP-1_M13F_A09_2007-08-08-R (653) CCAATGATCTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGCGGCTGCAGCCATCGTG
pPCRScriptAmpP-1_M13R_A10_2007-08-08 (853) CCAATGATCTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGCGGCTGCAGCCATCGTG
Consensus (853) CCAATGATCTGACTCCCGAGCAAAAGGTCGCCGCGGCTGCGCTCCTGGCGCCGGCGGCTGCAGCCATCGTG
Section 14
(924) 924 930 940 950 960 970 980 994
CALIPASEB(P23,P11)-P (867) GCGGGTCCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCTTTGCAGTAGGCAAAAGGAC
pPCRScriptAmpP-1_M13F_A09_2007-08-08-R (724) GCGGGTCCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCTTTGCAGTAGGCAAAAGGAC
pPCRScriptAmpP-1_M13R_A10_2007-08-08 (924) GCGGGTCCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCTTTGCAGT----------
Consensus (924) GCGGGTCCAAAGCAGAACTGCGAGCCCGACCTCATGCCCTACGCCCGCCCCTTTGCAGTAGGCAAAAGGAC
Section 15
(995) 995 1000 1010 1020 1030 1040 1050 1065
CALIPASEB(P23,P11)-P (938) CTGCTCCGGCATCGTCACCCCCTGAGAATTCAG----------
pPCRScriptAmpP-1_M13F_A09_2007-08-08-R (795) CTGCTCCGGCATCGTCACCCCCTGAGAATTCAGGGGCGGATCCCCCGGGCTGCAGGAATTCGATATCAAGC
pPCRScriptAmpP-1_M13R_A10_2007-08-08 (983) ----------
Consensus (995) CTGCTCCGGCATCGTCACCCCCTGAGAATTCAG

Appendix B

DNA Gel Results for PalB Expression Plasmid

In order to construct PalB expression plasmid, different PCR products of *palB*, named G~P, respectively, were required to fit into the specific expression vectors. PCR product of *palB*, G, was also used for PalB entry vector pENTR-G.

No.	Name	Primer (forward, reverse)	Restriction enzyme site	Notes
1	**G**	P10, P11	*Nco*I-*Eco*RI	The 3'-end was wrong, but does not affect protein expression
2	**L**	P17, P11	*Nde*I-*Eco*RI	Sequenced correctly
3	**I**	P14, P11	*Xma*I-*Eco*RI	
4	**M**	P15, P19	*Eco*RI-*Xma*I	Sequenced correctly
5	**N**	P12, P20	*Sfi*I-*Sfi*I	Sequenced correctly
6	**H**	P12, P13	*Sfi*I-*Sfi*I	
7	**O**	P10, P22	*Nco*I-*Sac*I	Sequenced correctly
8	**P**	P23, P11	*Xho*I-*Eco*RI	Sequenced correctly

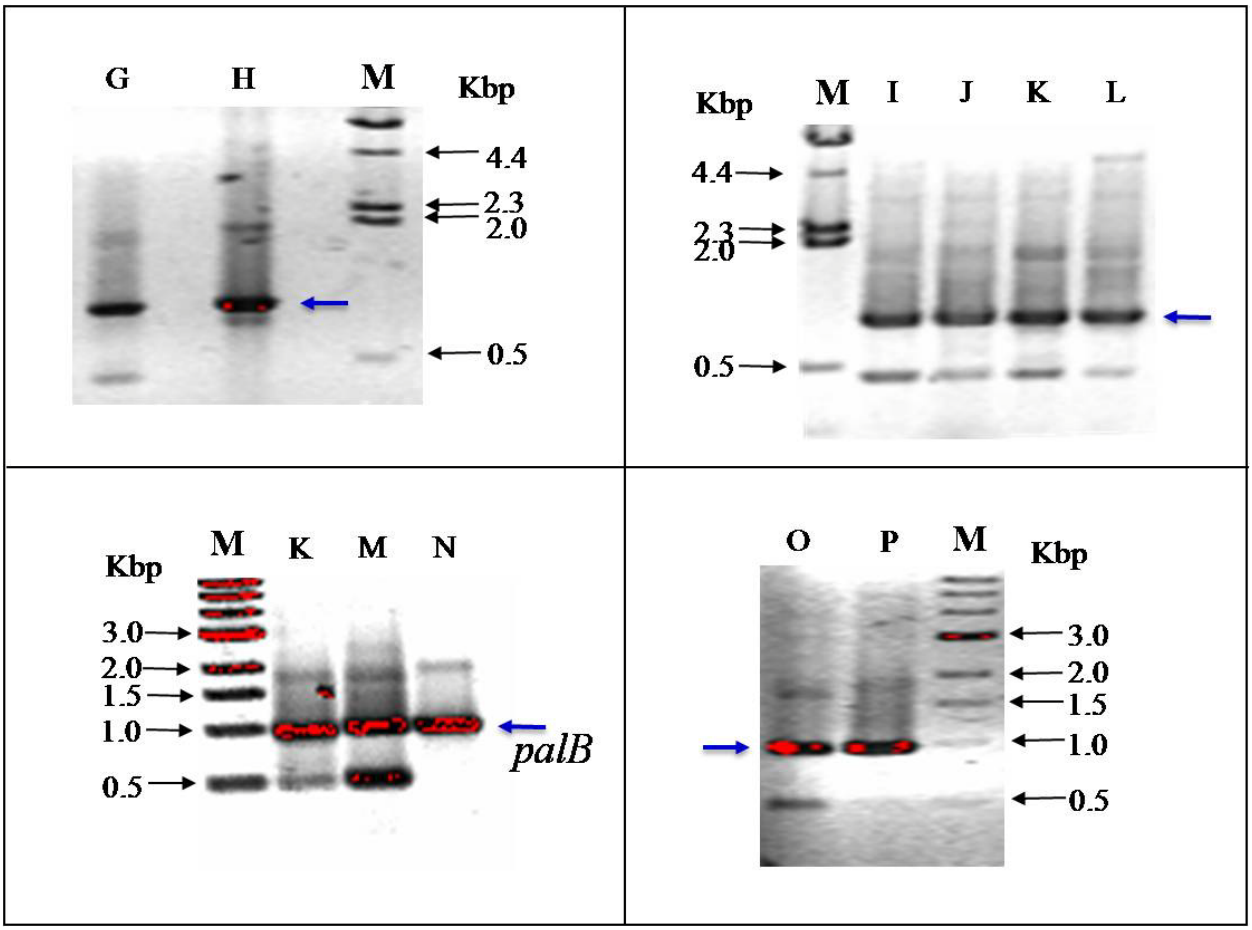


Figure B.1 PCR products of *palB* which used for construction of PalB expression plasmid in *E. coli* expression systems

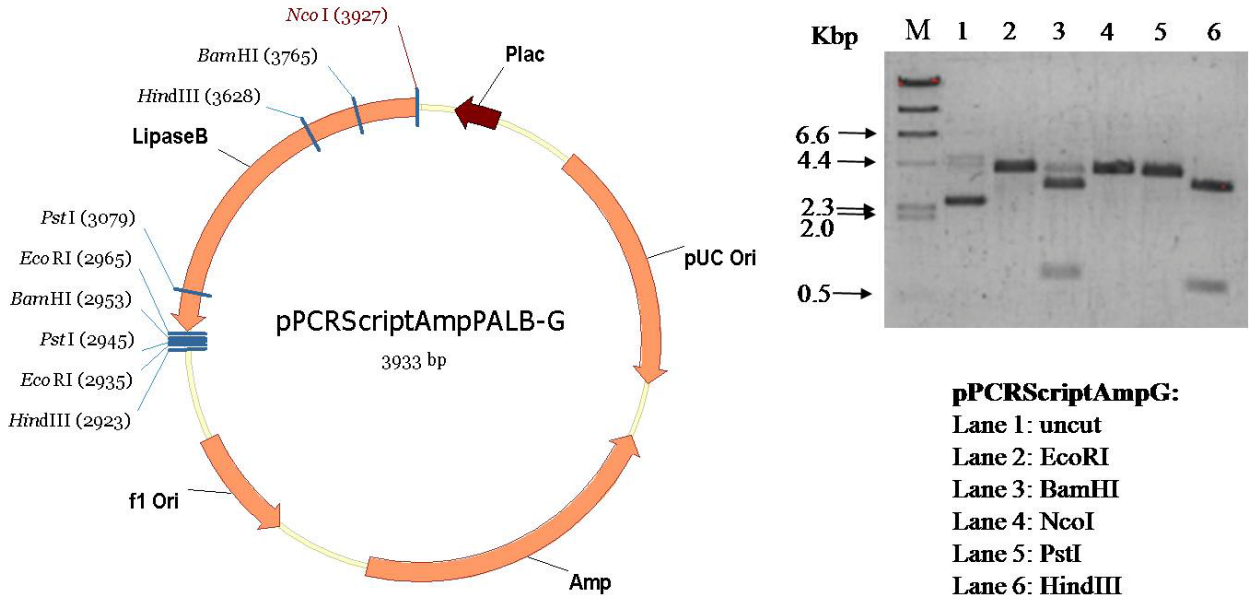


Figure B.2 Restriction analysis of pPCRScriptAmpG

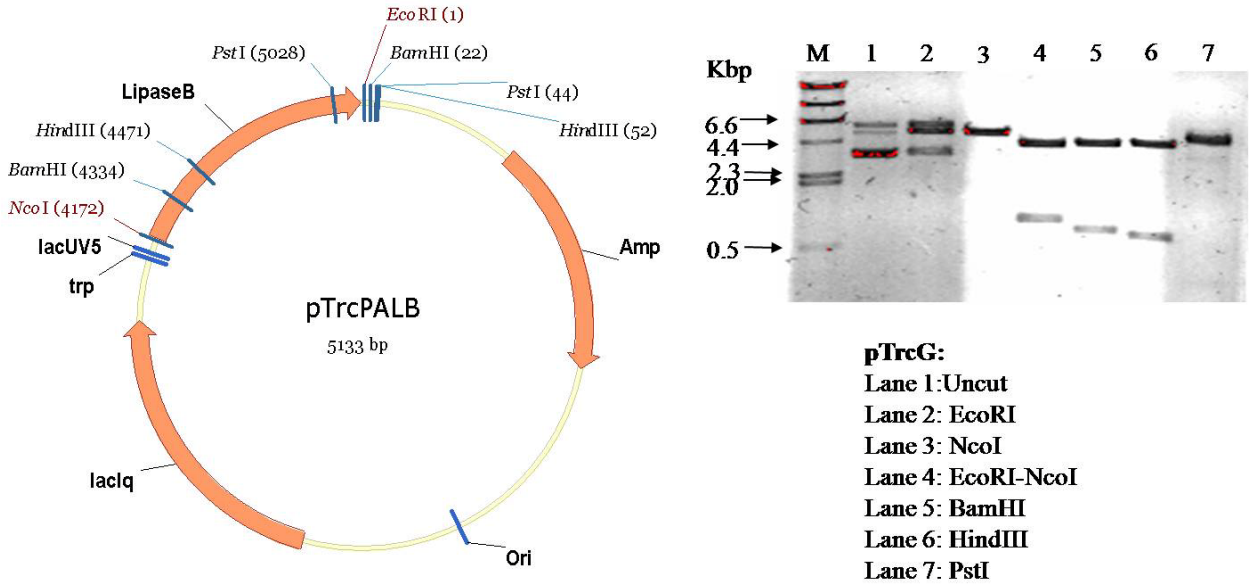


Figure B.3 Restriction analysis of pTrcG

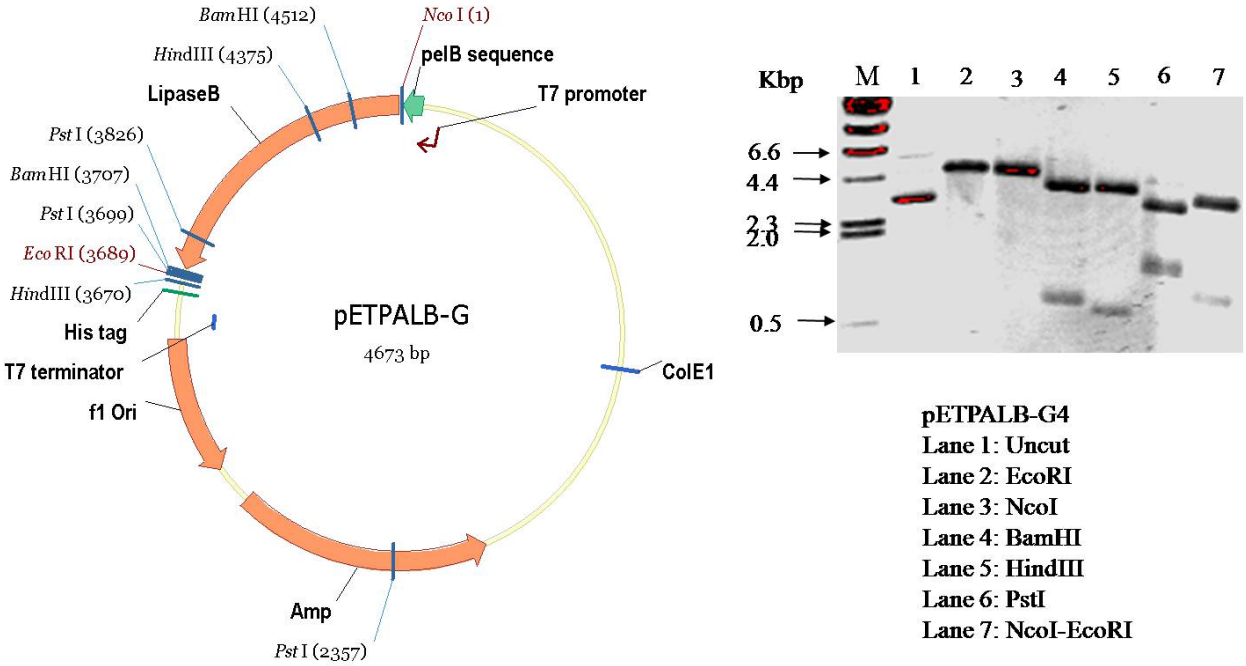


Figure B. 4 Restriction analysis of pETG

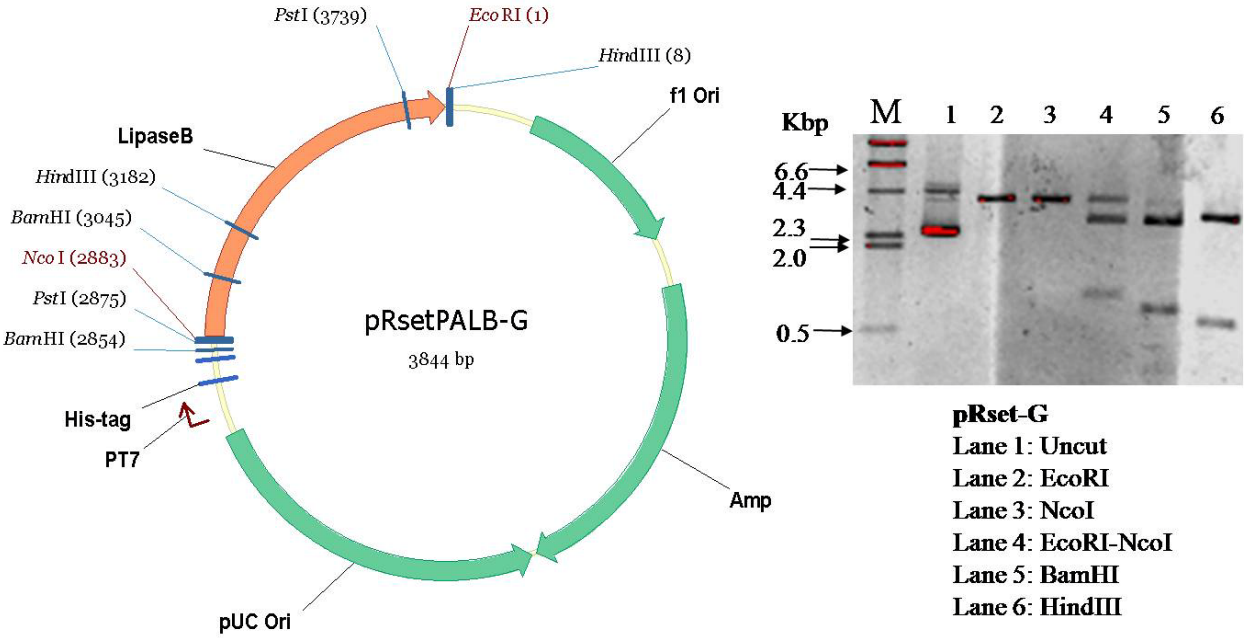
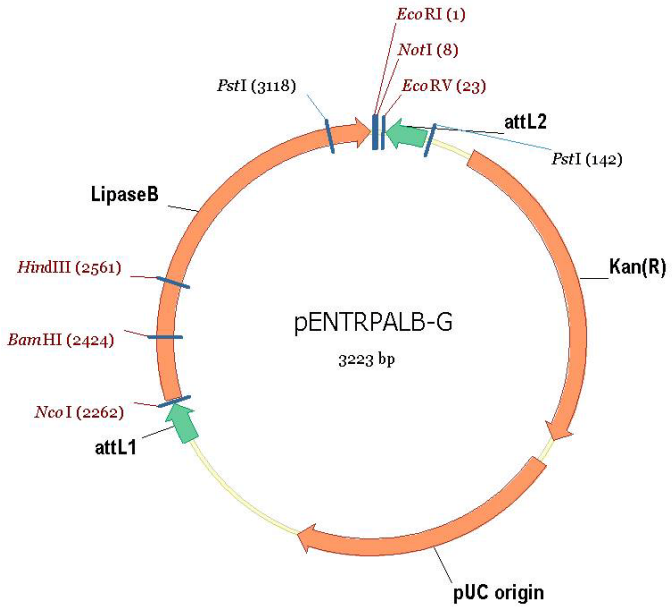
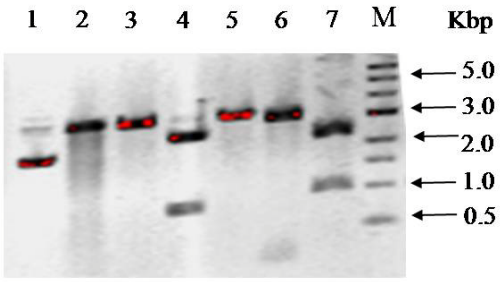


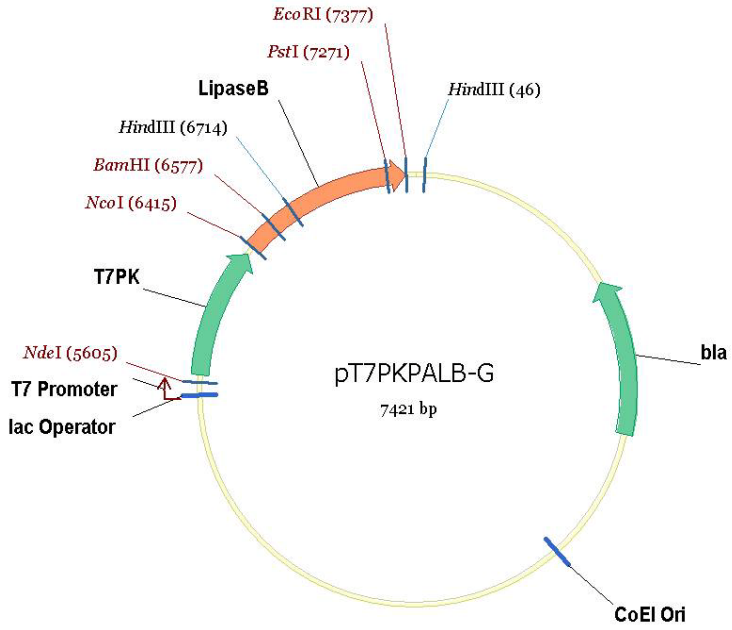
Figure B.5 Restriction analysis of pRset-G

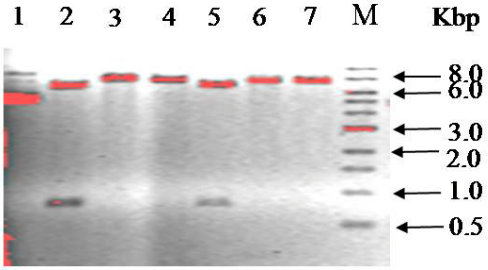




pENTRG
Lane 1:uncut
Lane 2: EcoRI
Lane 3: NcoI
Lane 4: BamHI
Lane 5: HindIII
Lane 6: PstI
Lane 7: EcoRI-NcoI

Figure B.6 Restriction analysis of pENTRG





pT7PKPALB-G-2
Lane 1: uncut
Lane 2: BamHI
Lane 3: EcoRI
Lane 4: NcoI
Lane 5: HindIII
Lane 6: PstI
Lane 7: NdeI

Figure B.7 Restriction analysis of pT7PK-G

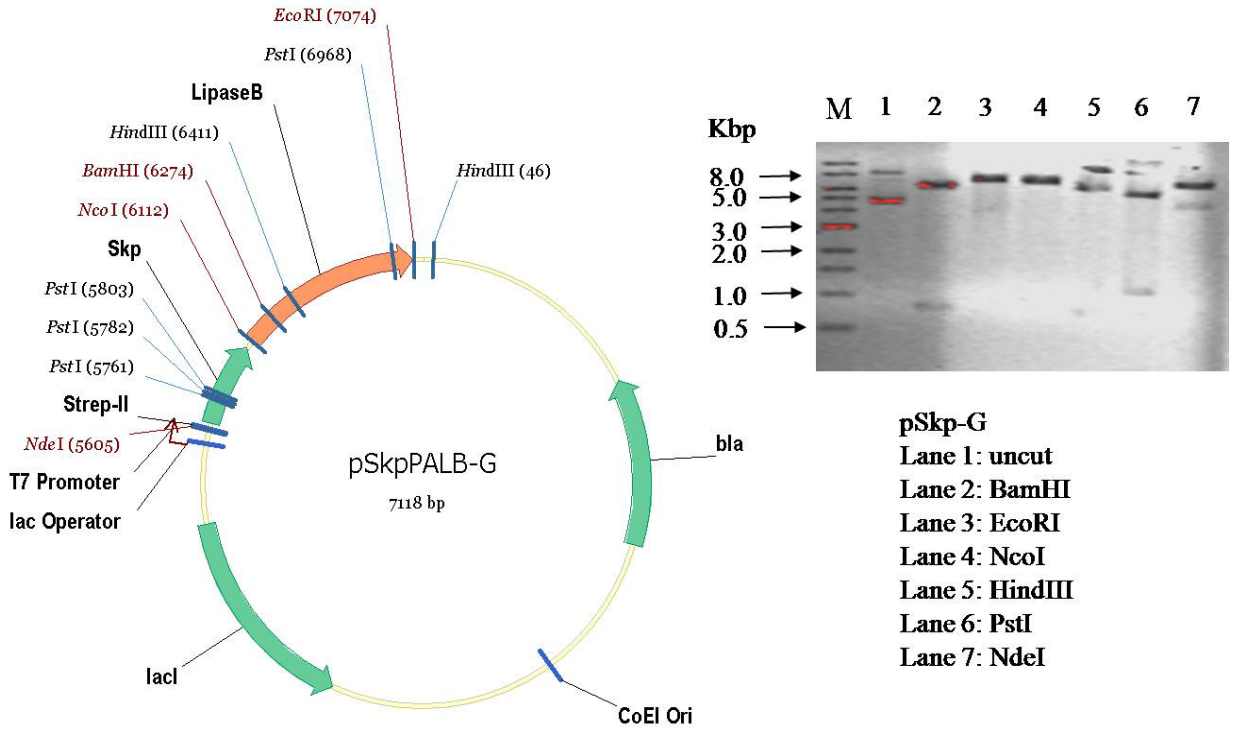


Figure B.8 Restriction analysis of pSkp-G

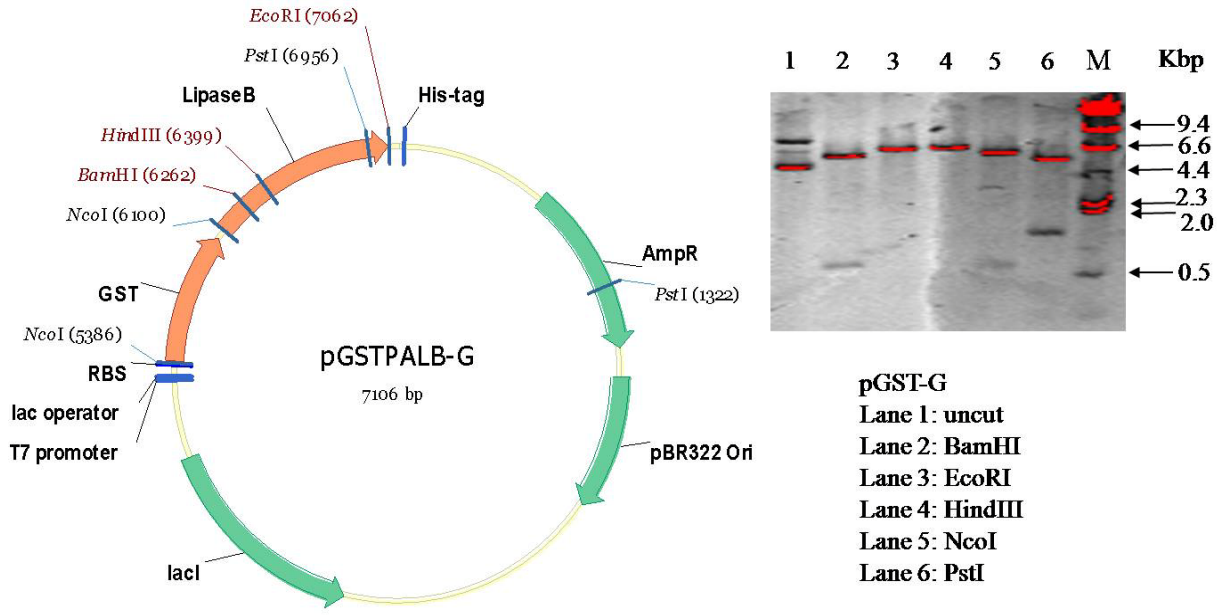


Figure B.9 Restriction analysis of pGST-G

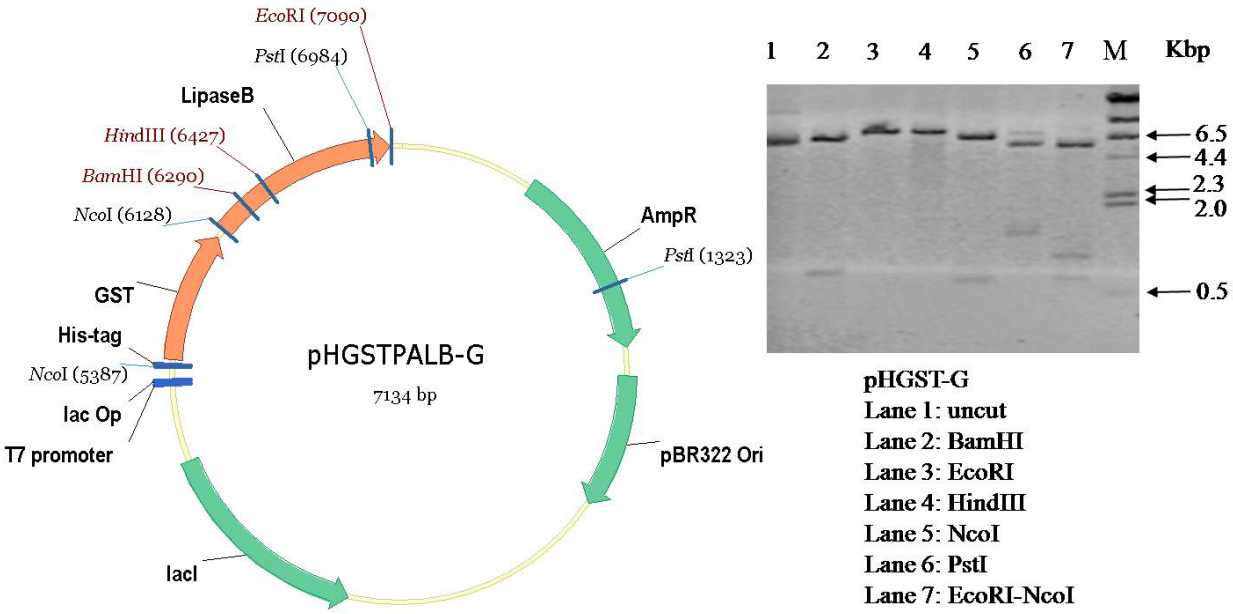


Figure B.10 Restriction analysis of pHGST-G

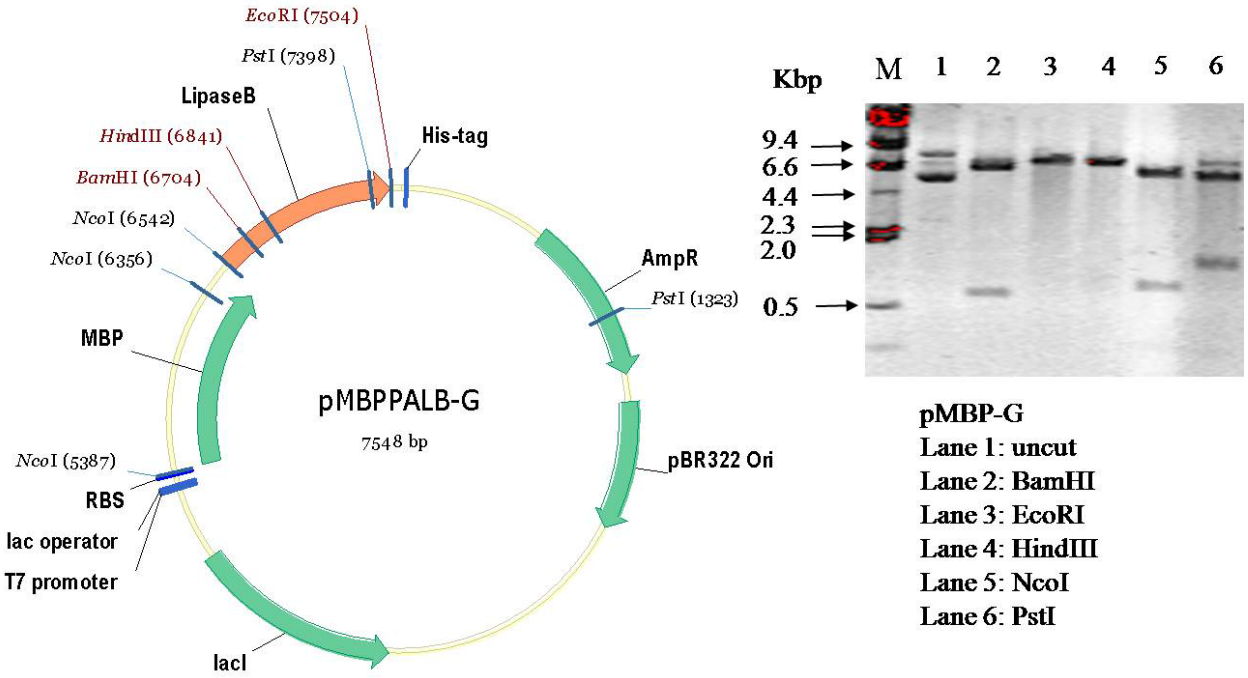


Figure B.11 Restriction analysis of pMBP-G

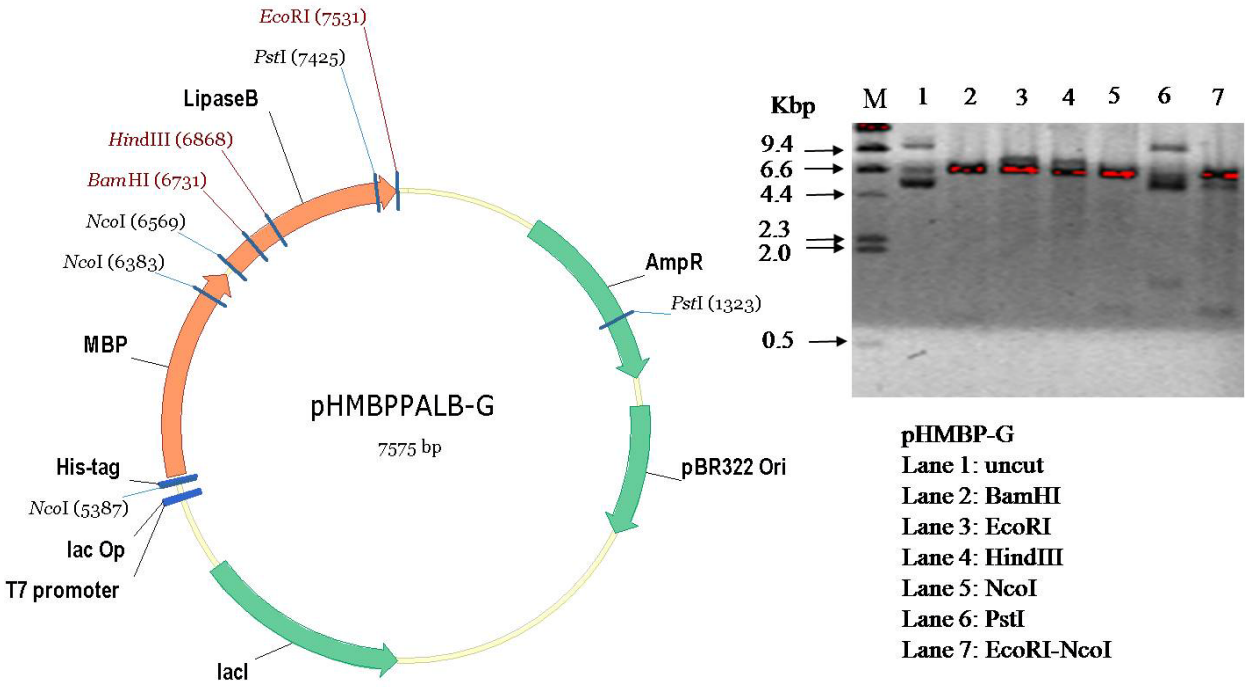


Figure B.12 Restriction analysis of pHMBP-G

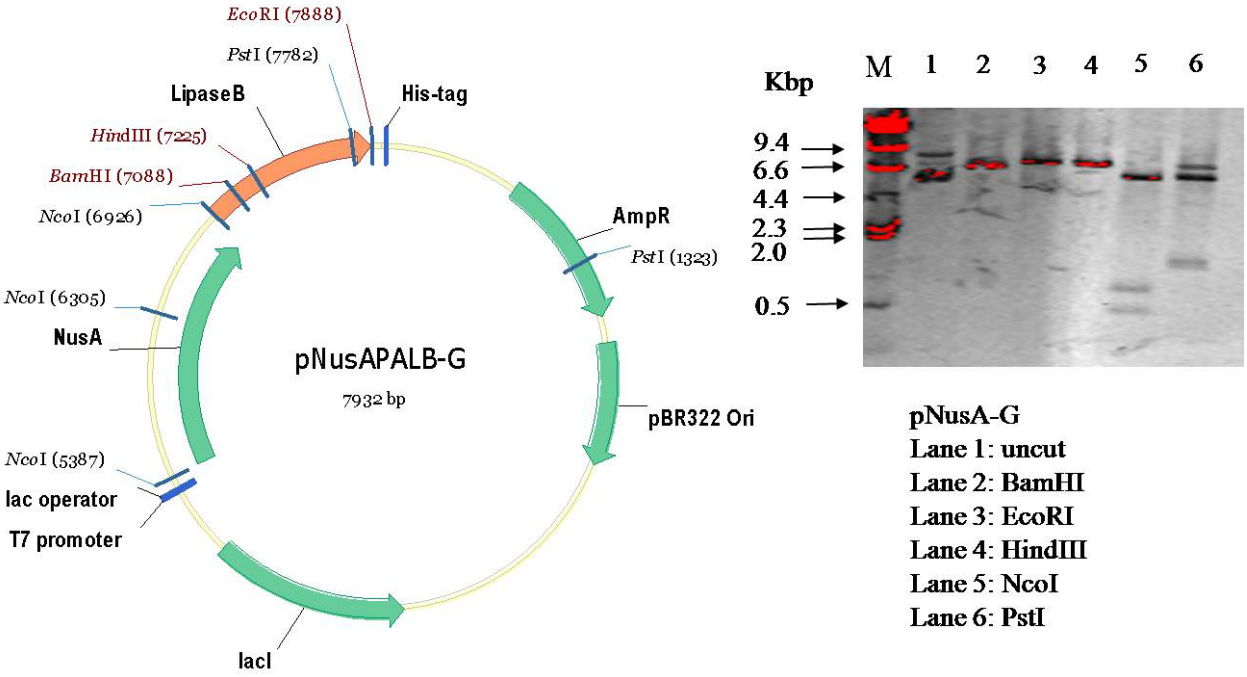


Figure B.13 Restriction analysis of pNusA-G

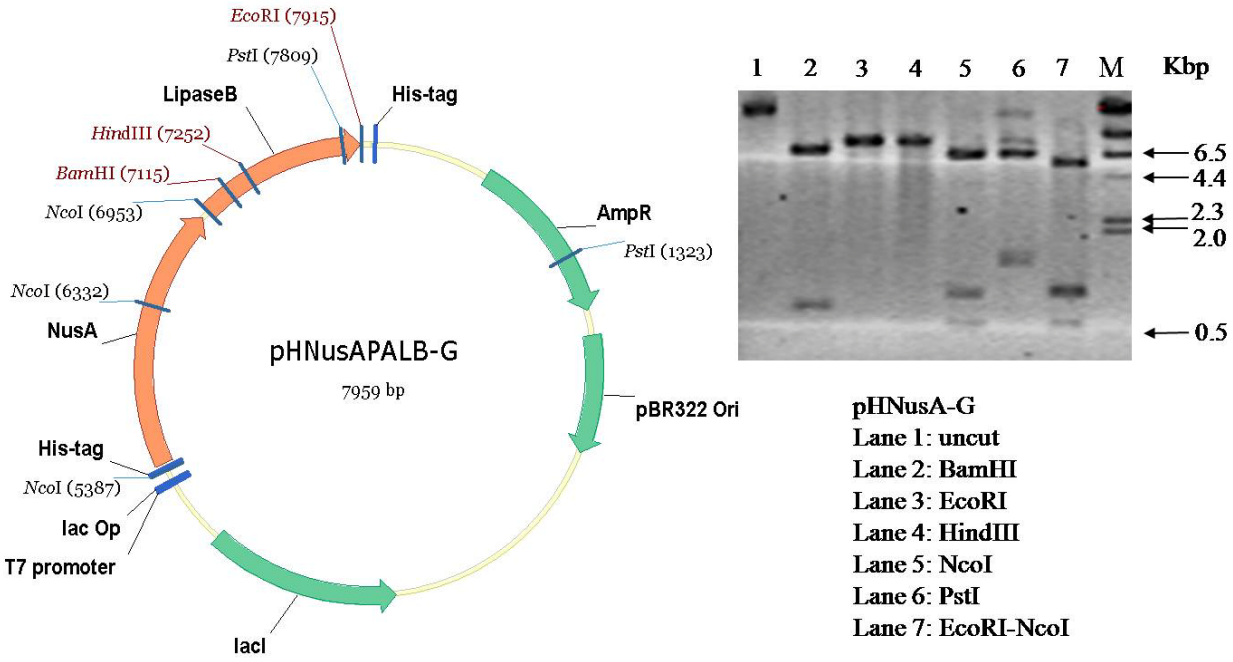


Figure B.14 Restriction analysis of pHNusA-G

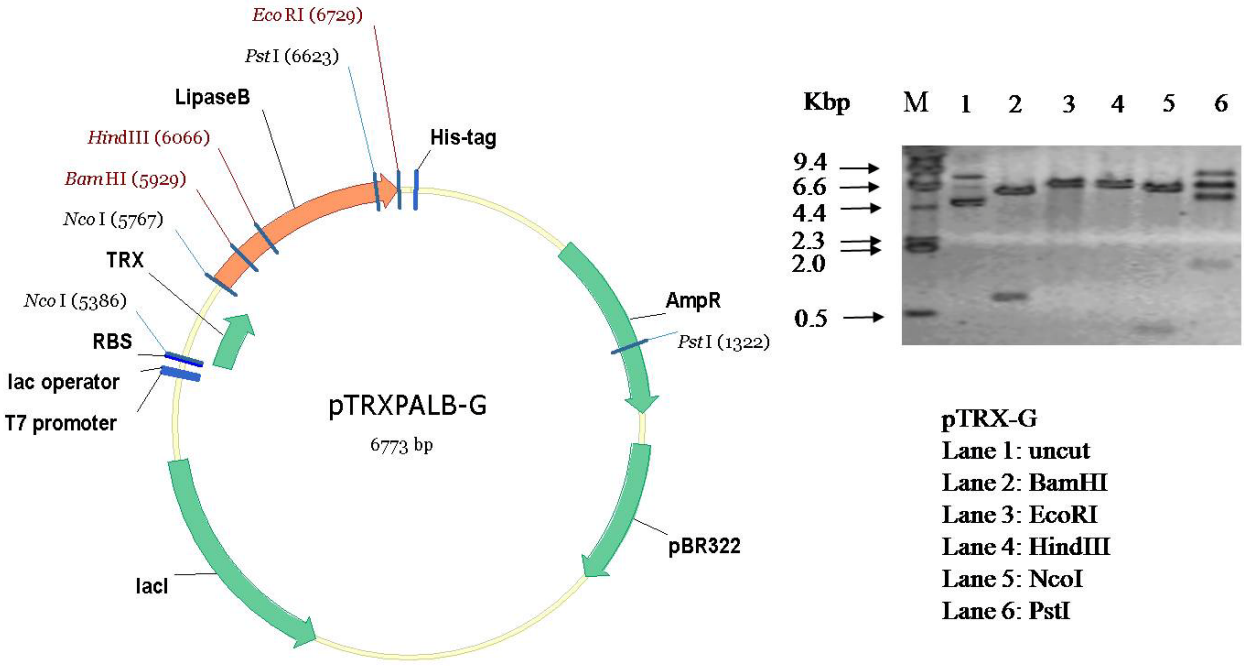
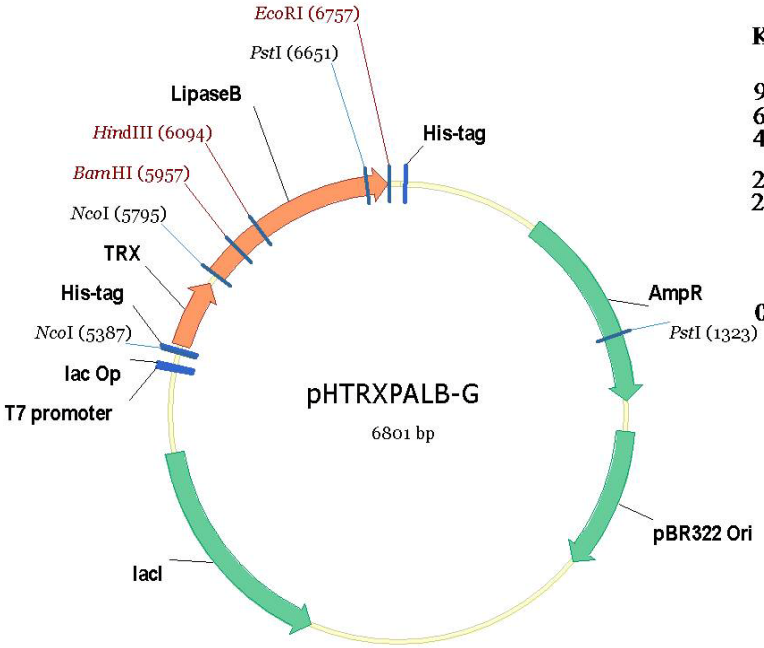
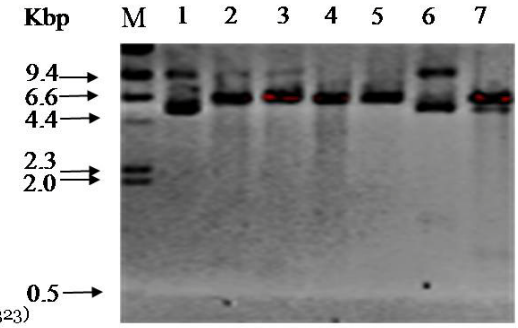


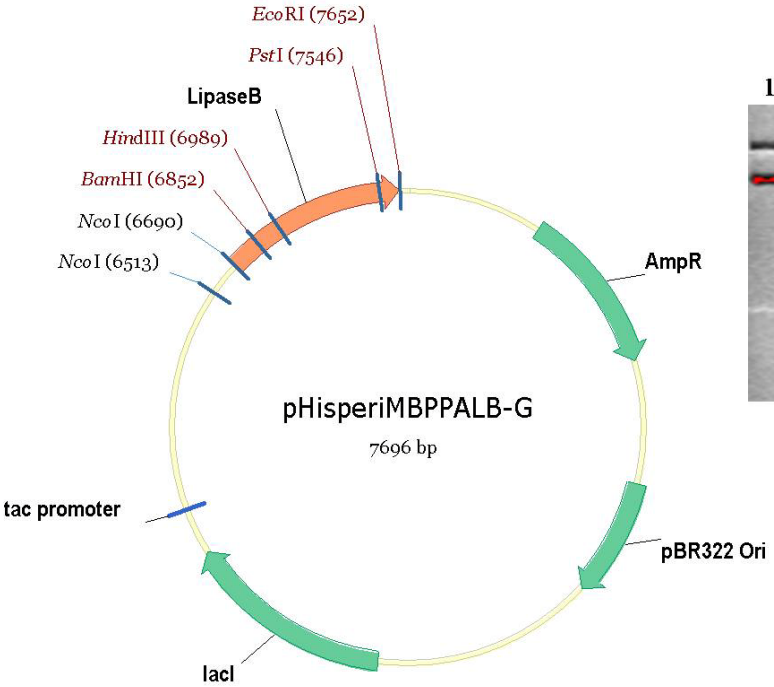
Figure B.15 Restriction analysis of pTRX-G

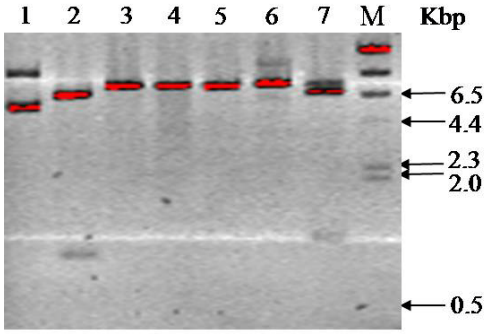




pHTRX-G
Lane 1: uncut
Lane 2: BamHI
Lane 3: EcoRI
Lane 4: HindIII
Lane 5: NcoI
Lane 6: PstI
Lane 7: EcoRI-NcoI

Figure B.16 Restriction analysis of pHTRX-G





pHisperiMBP-G
Lane 1: uncut
Lane 2: BamHI
Lane 3: EcoRI
Lane 4: HindIII
Lane 5: NcoI
Lane 6: PstI
Lane 7: EcoRI-NcoI

Figure B.17 Restriction analysis of pHisperiMBP-G

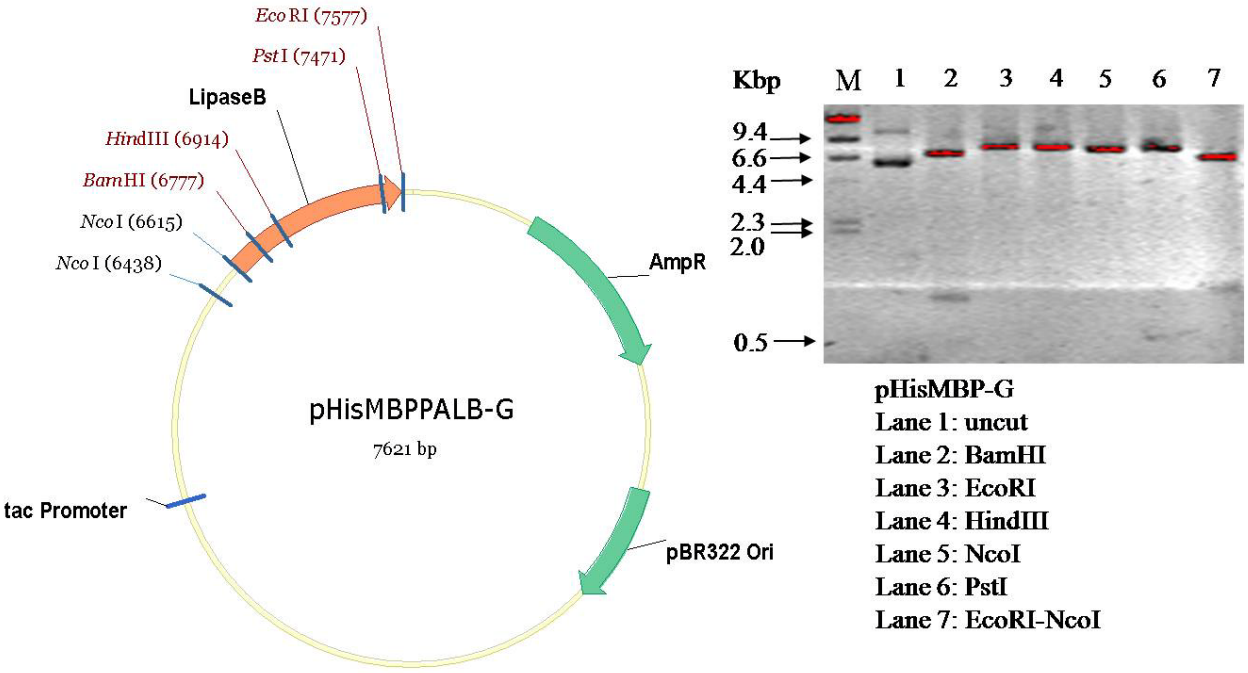


Figure B.18 Restriction analysis of pHisMBP-G

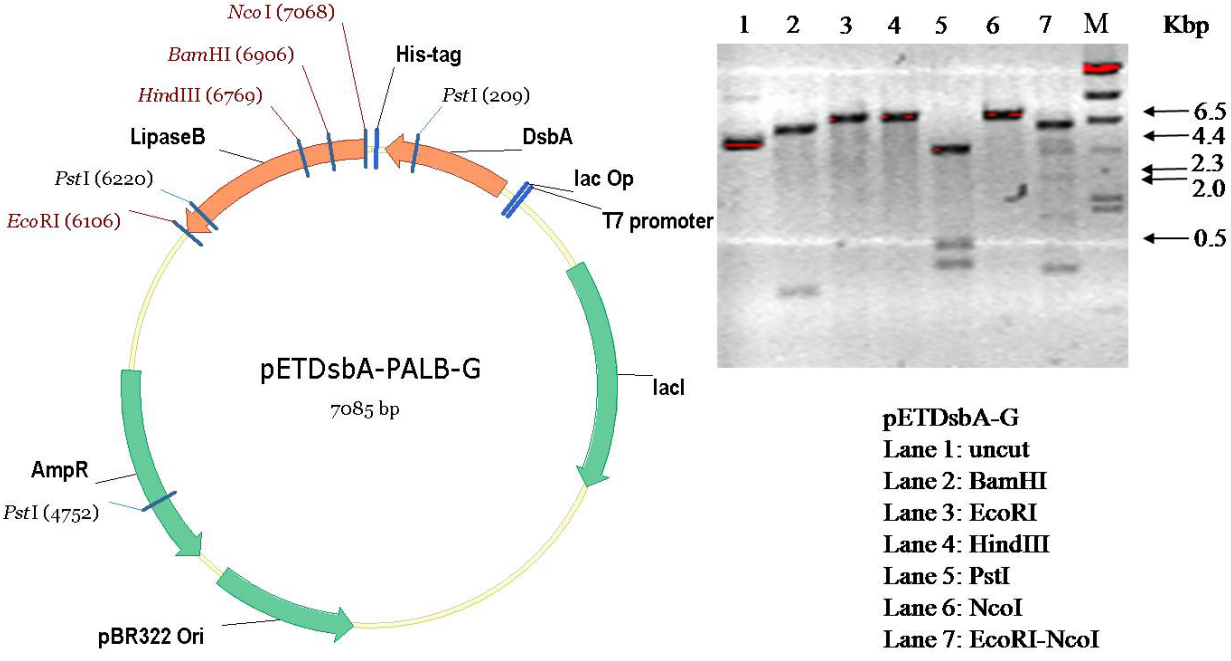


Figure B.19 Restriction analysis of pETDsbA-G

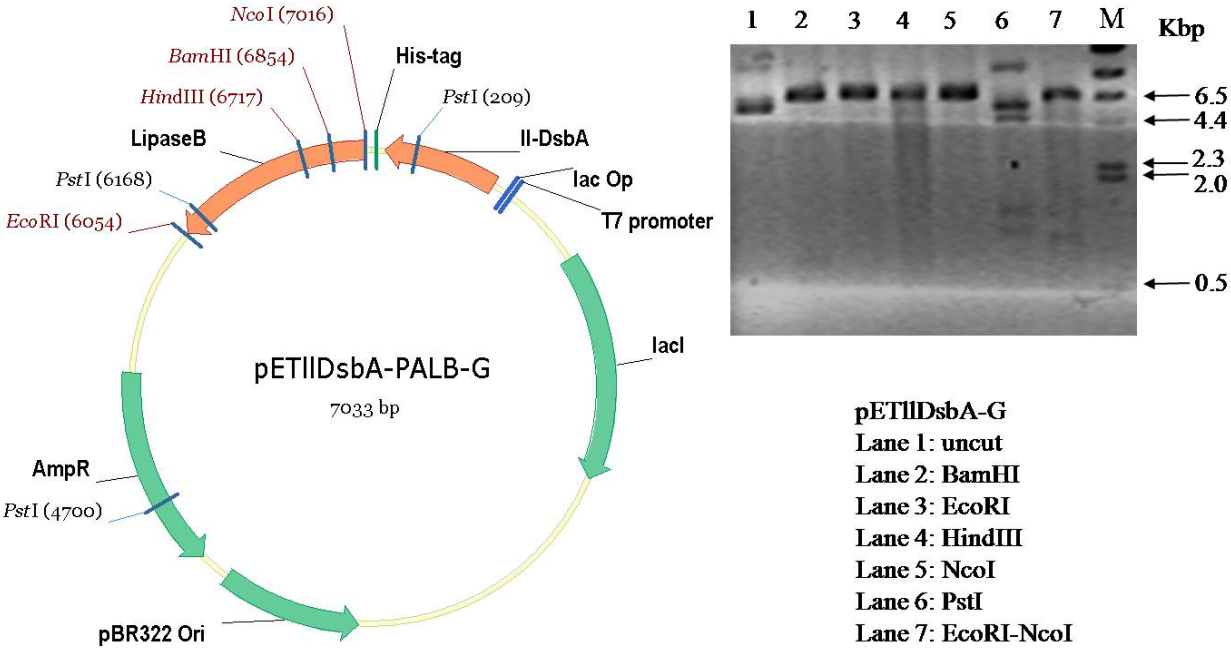


Figure B.20 Restriction analysis of pETIIDsbA-G

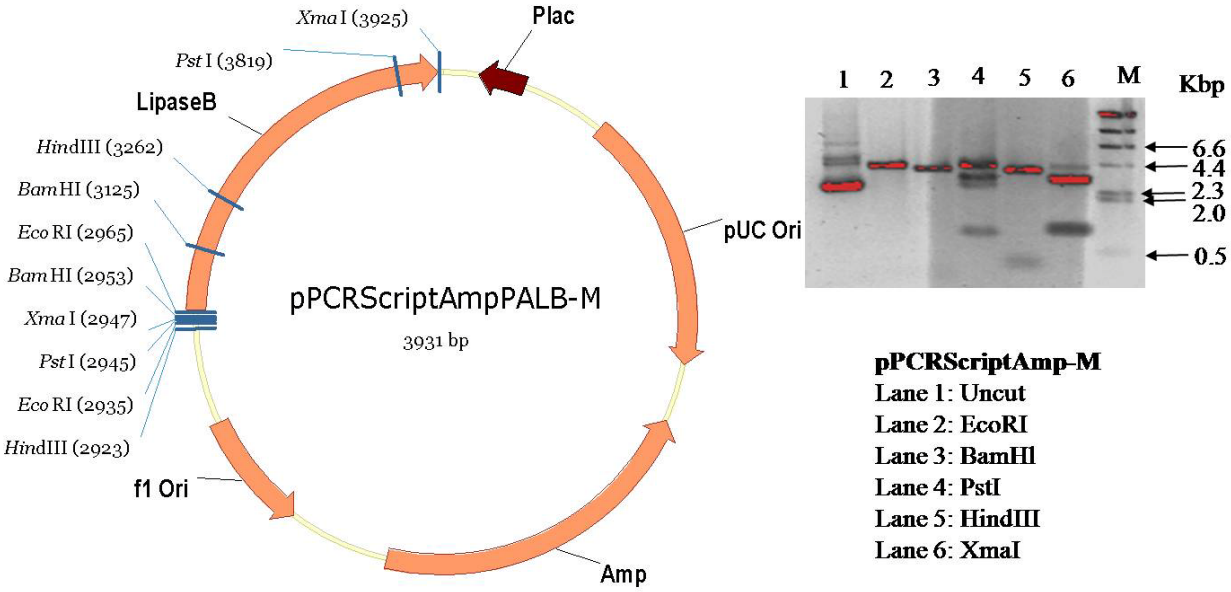


Figure B.21 Restriction analysis of pPCRSciptAmpM

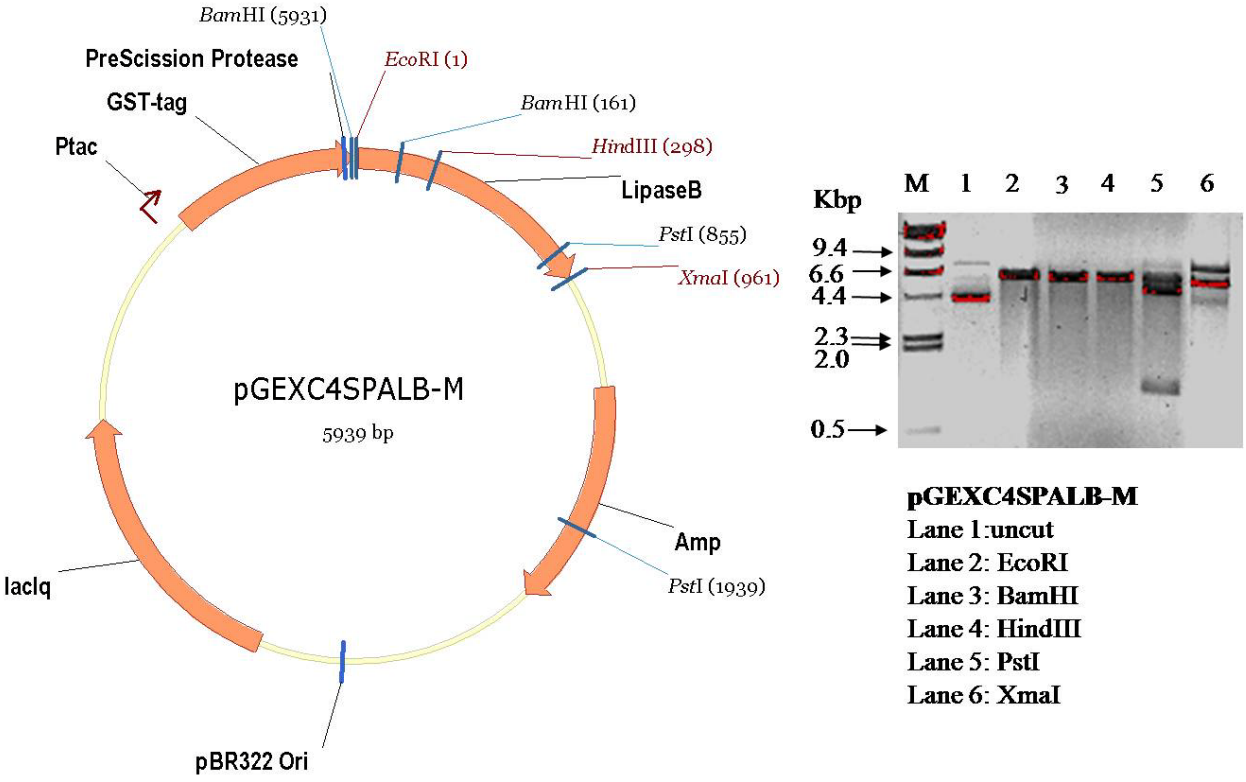


Figure B.22 Restriction analysis of pGEXC4SM

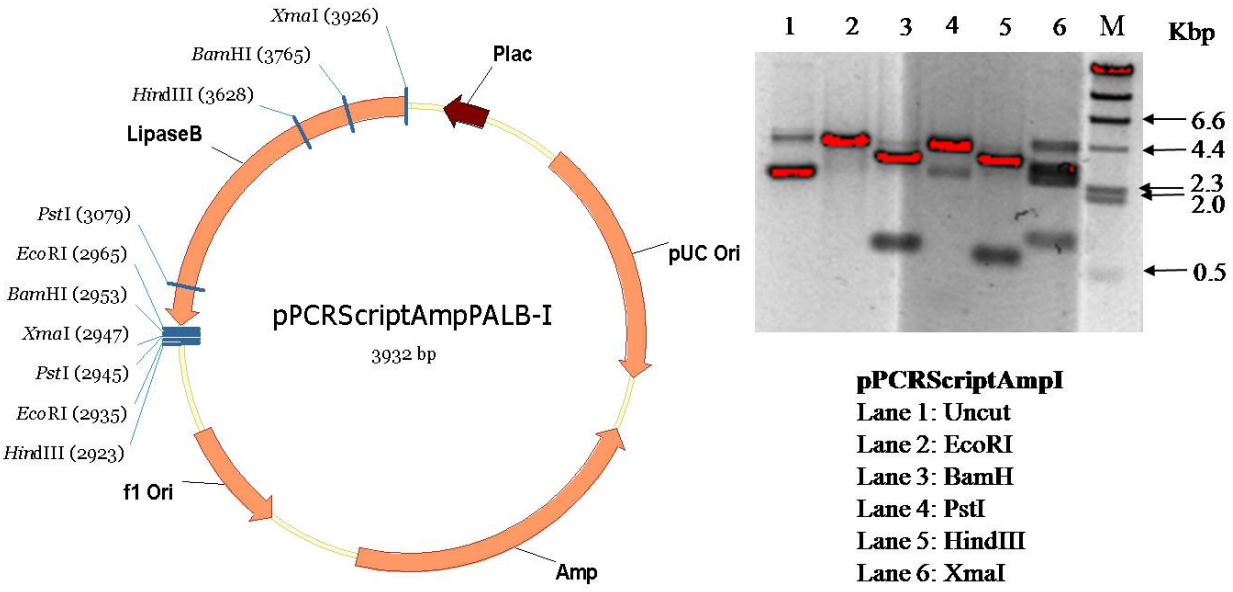


Figure B.23 Restriction analysis of pPCRScriptAmpI

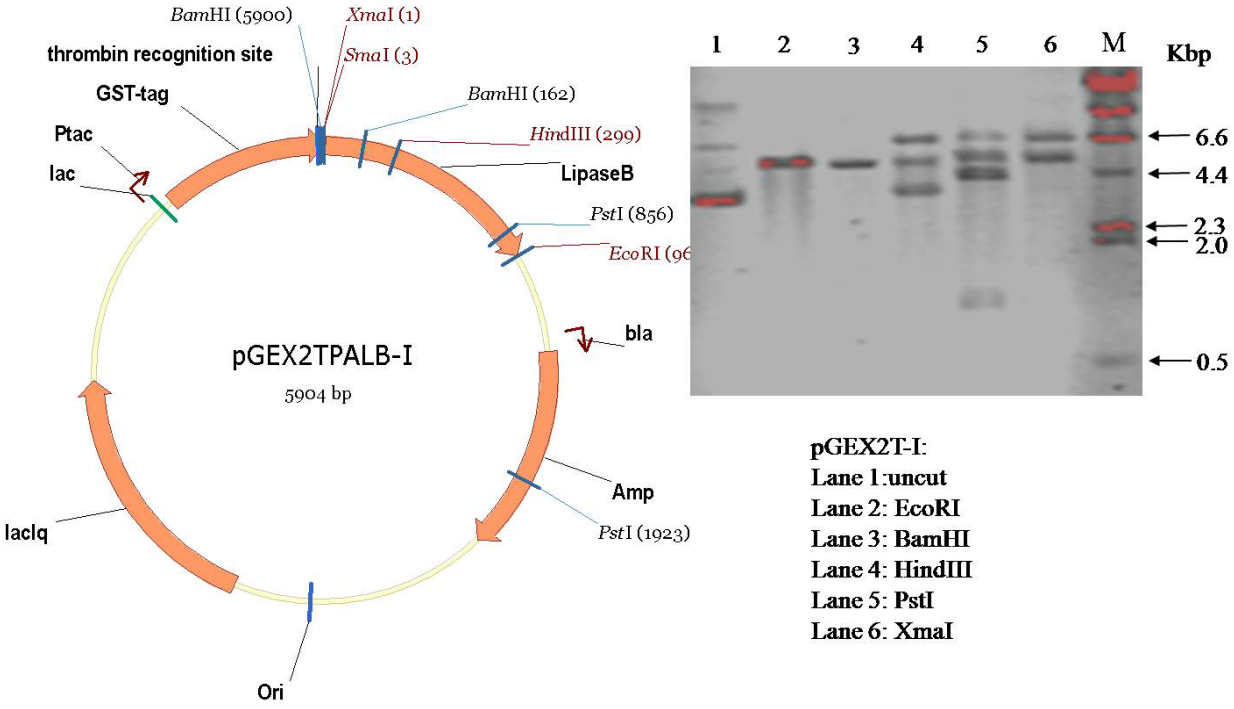


Figure B.24 Restriction analysis of pGEX2TI

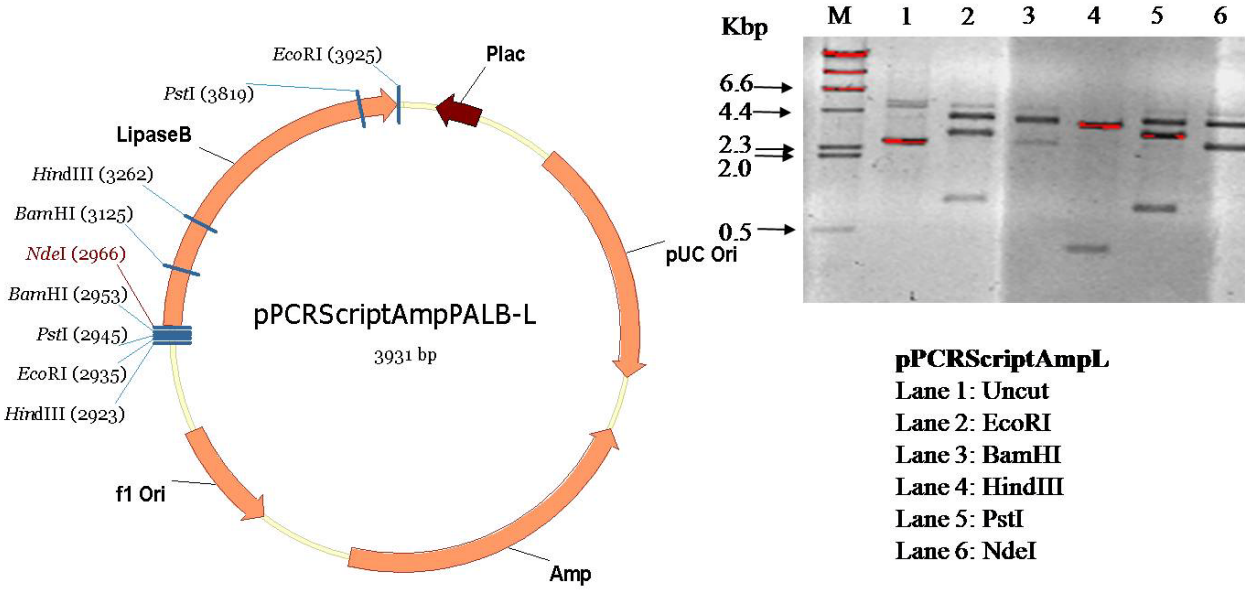


Figure B.25 Restriction analysis of pPCRScriptAmpL

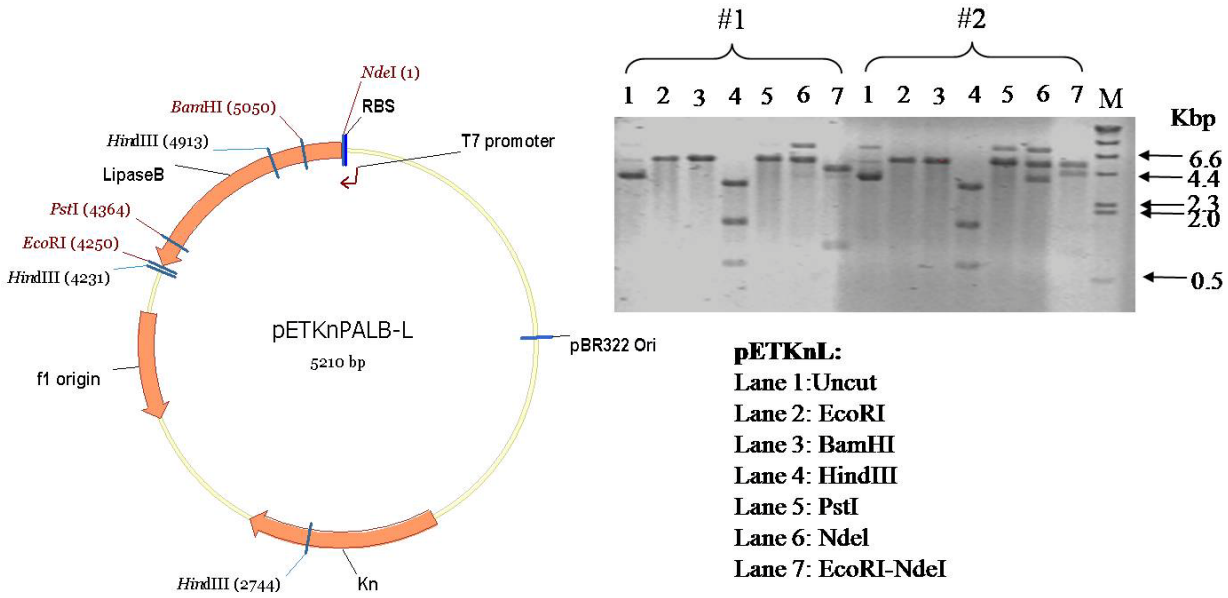


Figure B.26 Restriction analysis of pETKnL

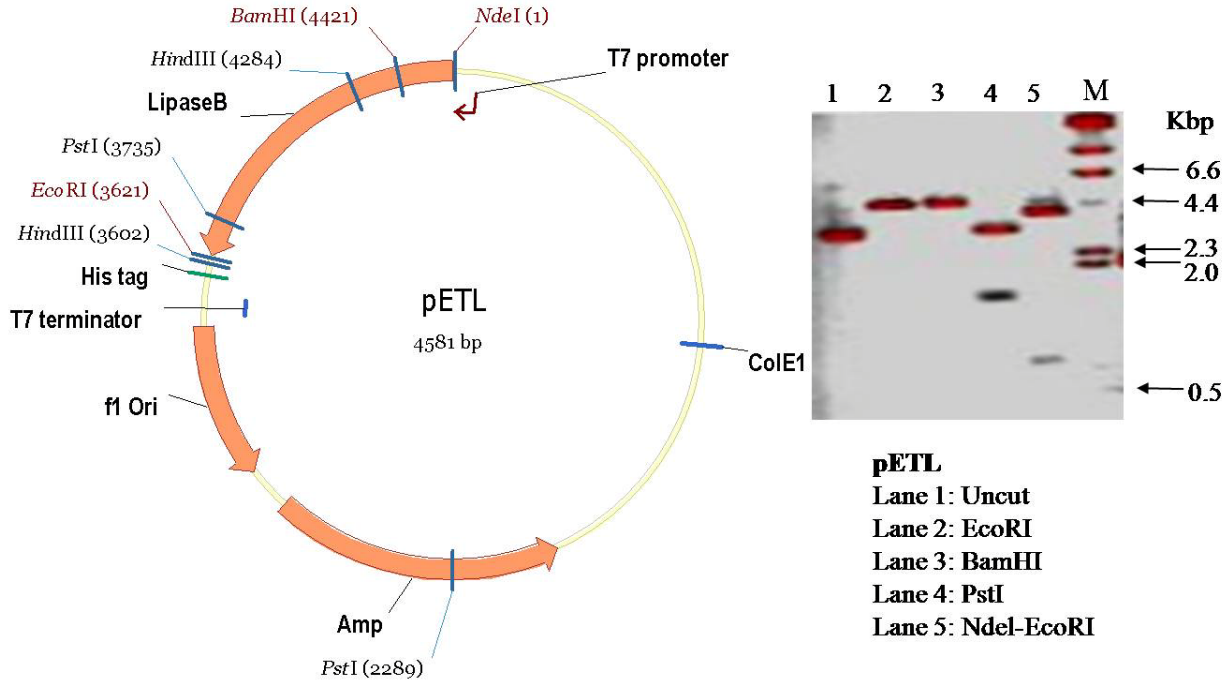


Figure B.27 Restriction analysis of pETL

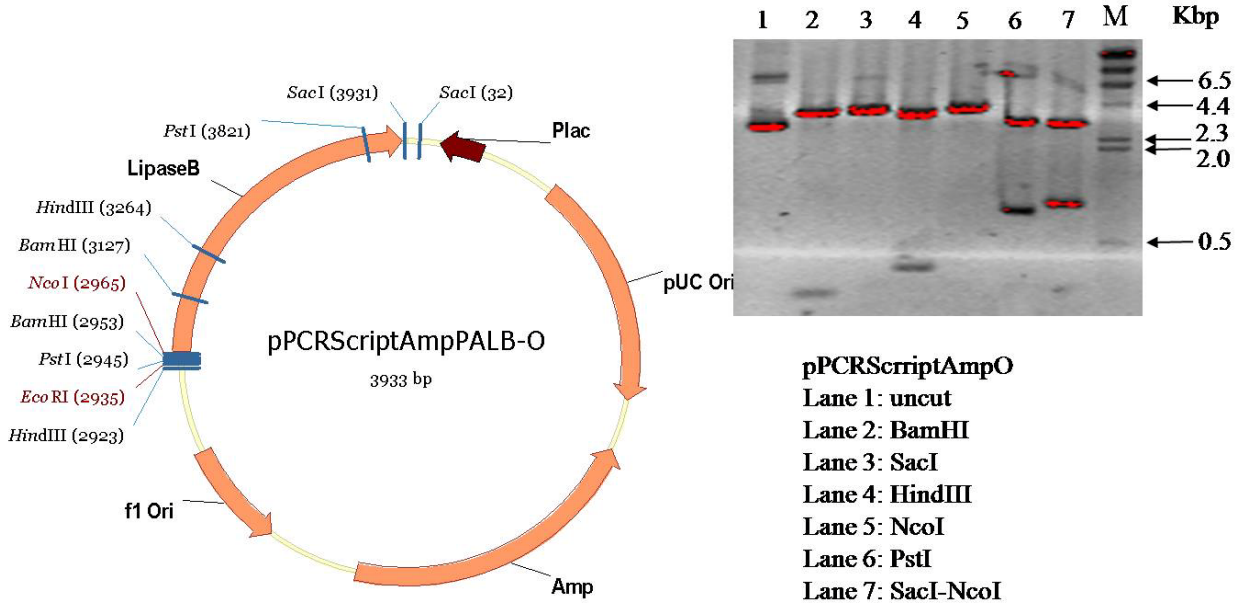


Figure B.28 Restriction analysis of pPCRScriptAmpO

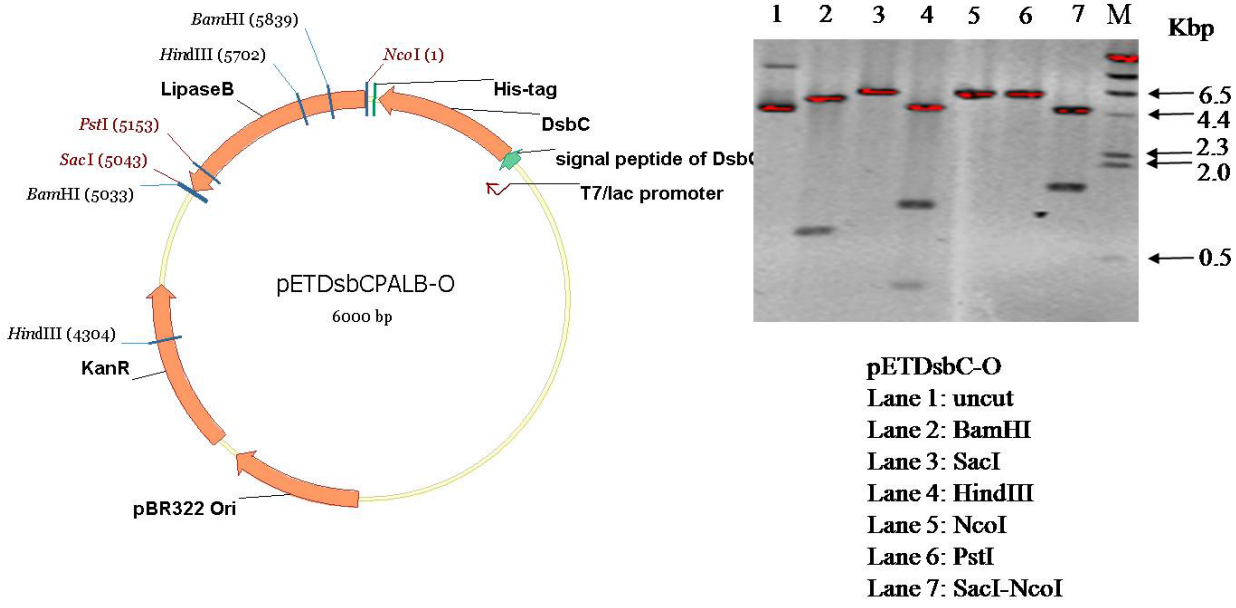


Figure B.29 Restriction analysis of pETDsbC-O

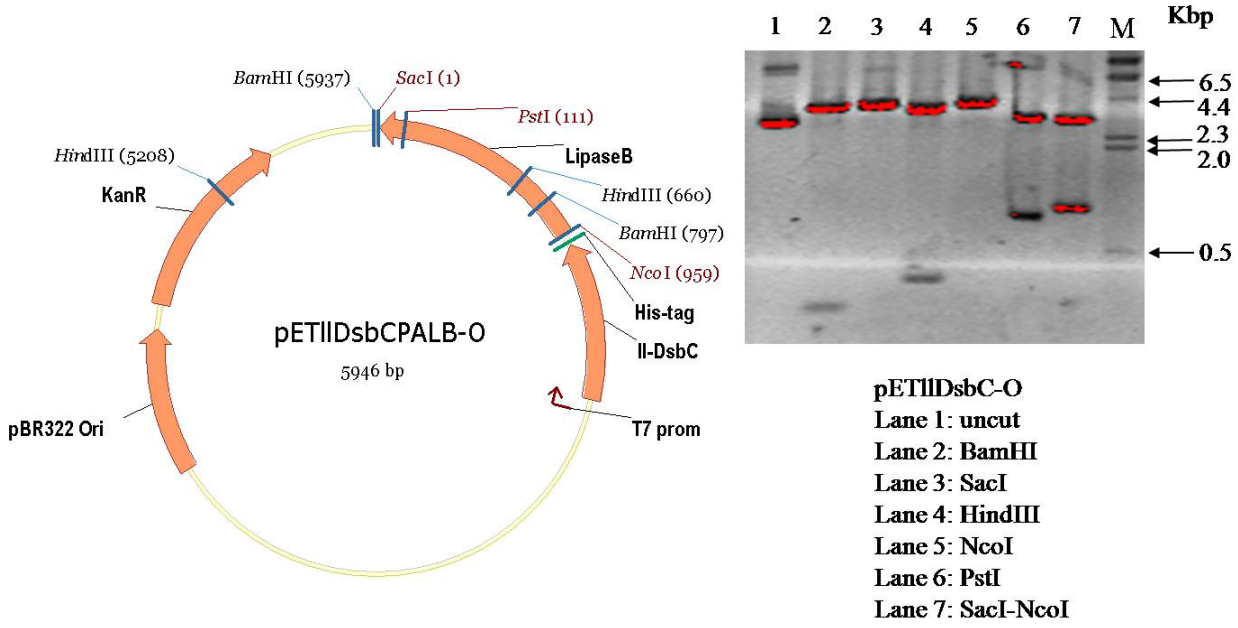


Figure B.30 Restriction analysis of pETIIDsbC-O

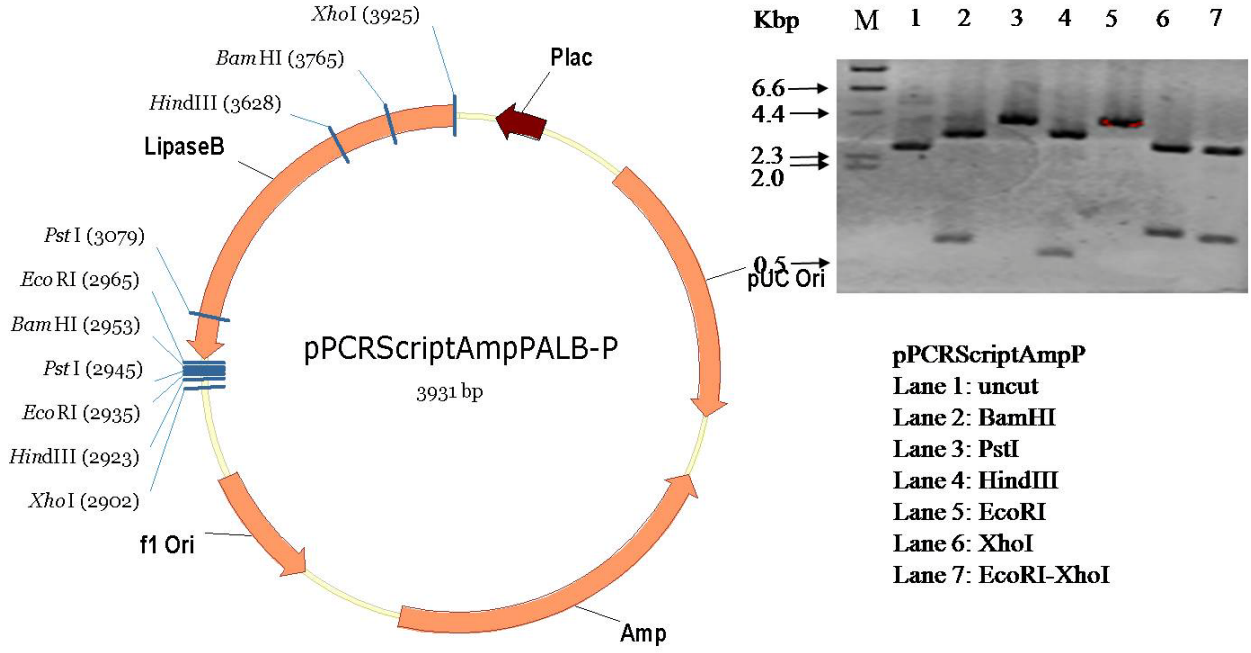


Figure B.31 Restriction analysis of pPCRScriptAmpP

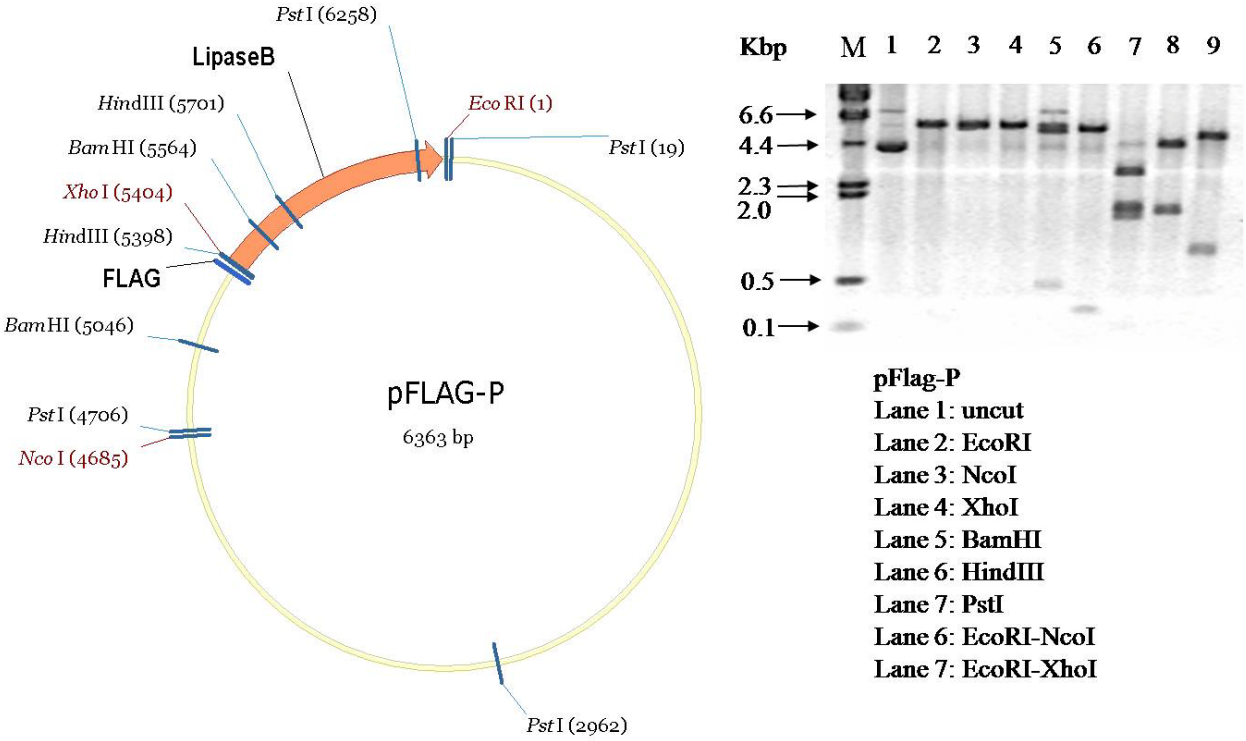


Figure B.32 Restriction analysis of pFlag-P

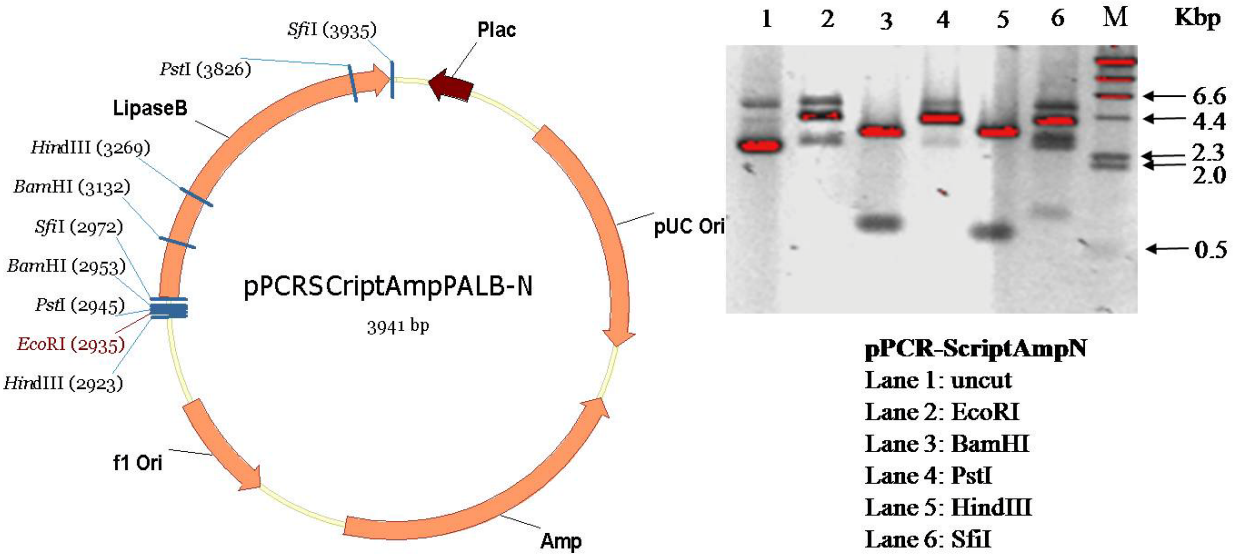


Figure B.33 Restriction analysis of pPCRScriptAmpN

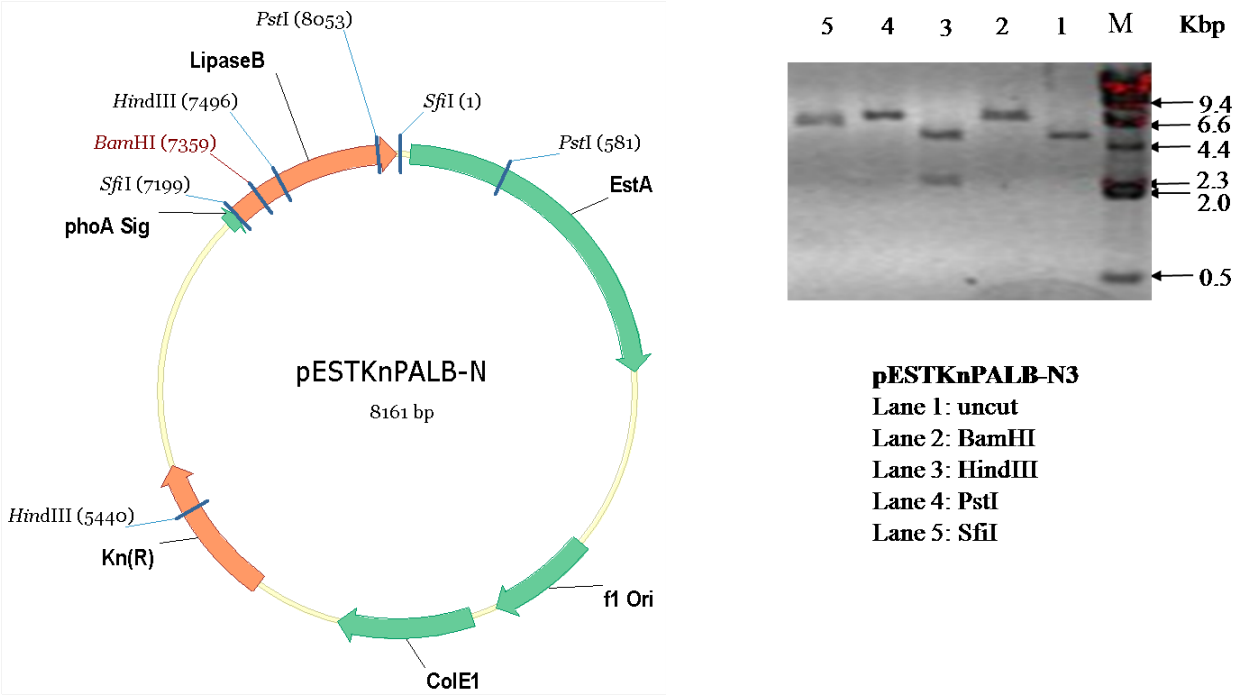


Figure B.34 Restriction analysis of pESTKnN

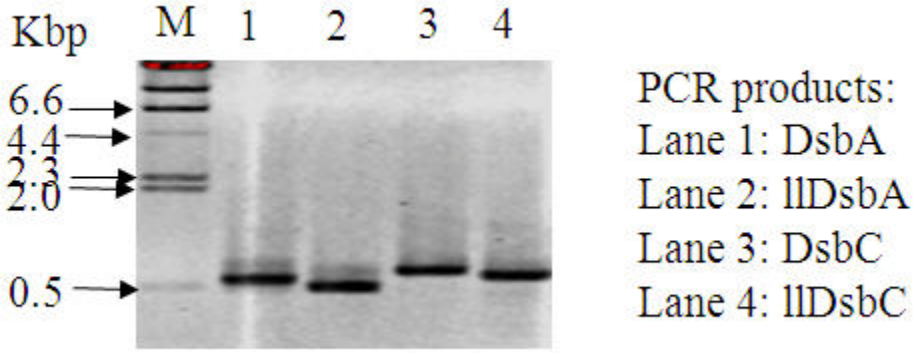


Figure B.35 PCR products for construction of chaperones pARDsbA, pARllDsbA, pARDsbC, pARllDsbC

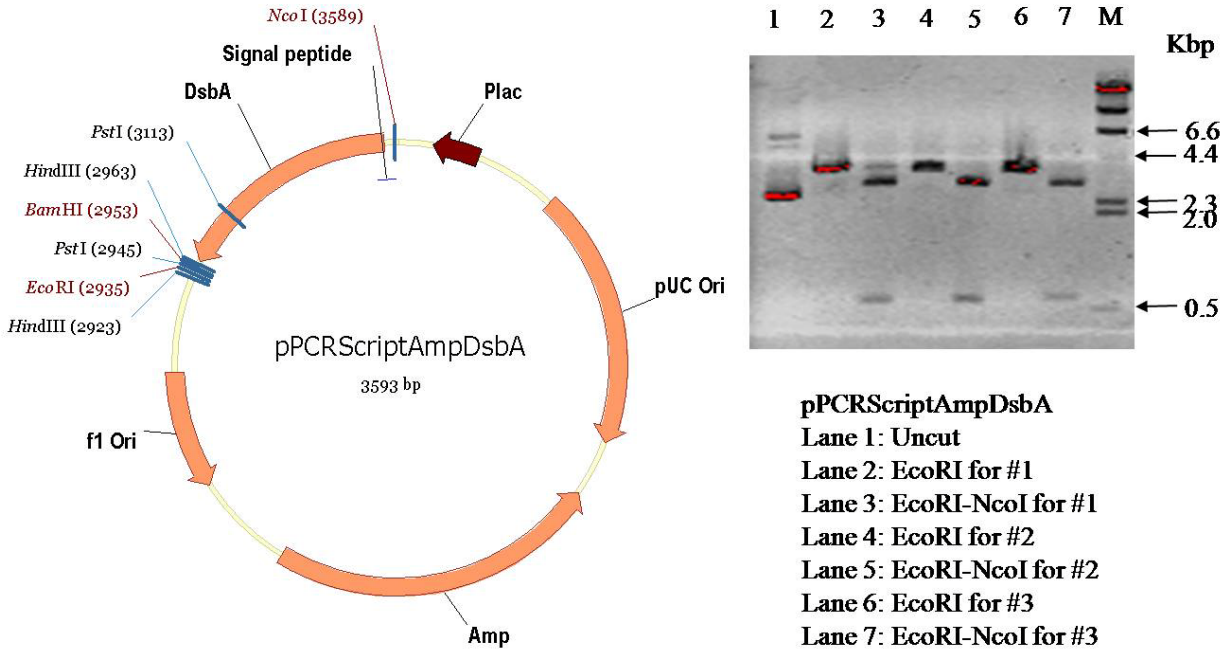


Figure B.36 Restriction analysis of pPCRscriptAmpDsbA

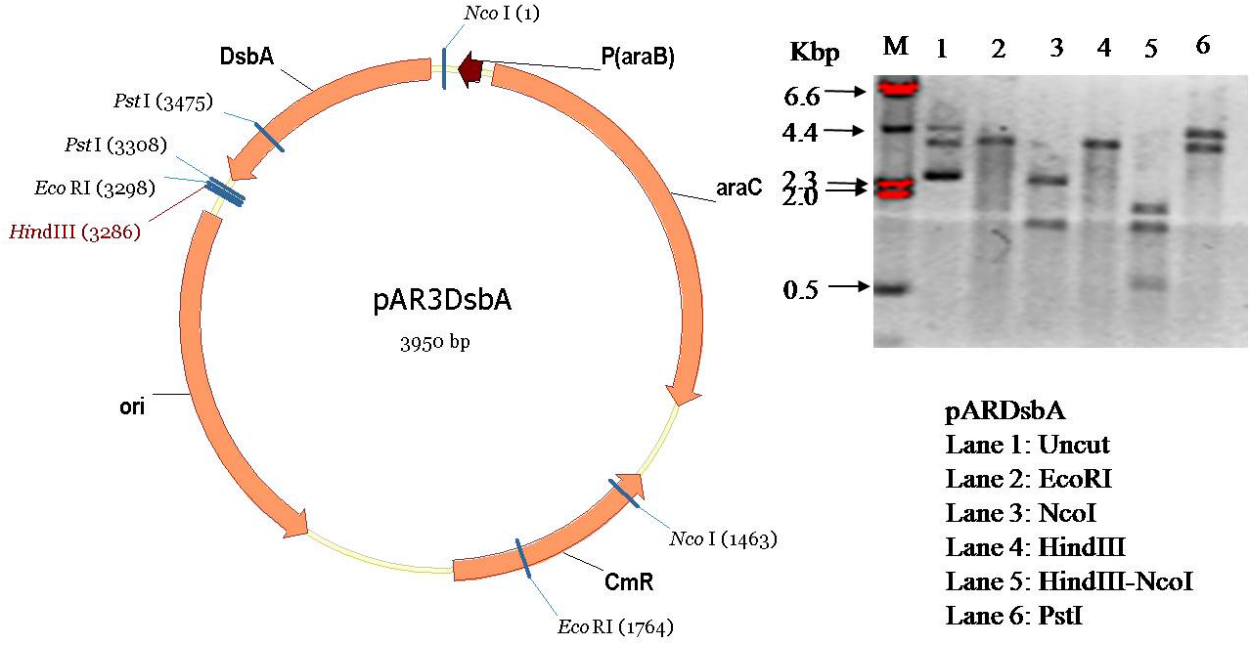


Figure B.37 Restriction analysis of pARDsbA

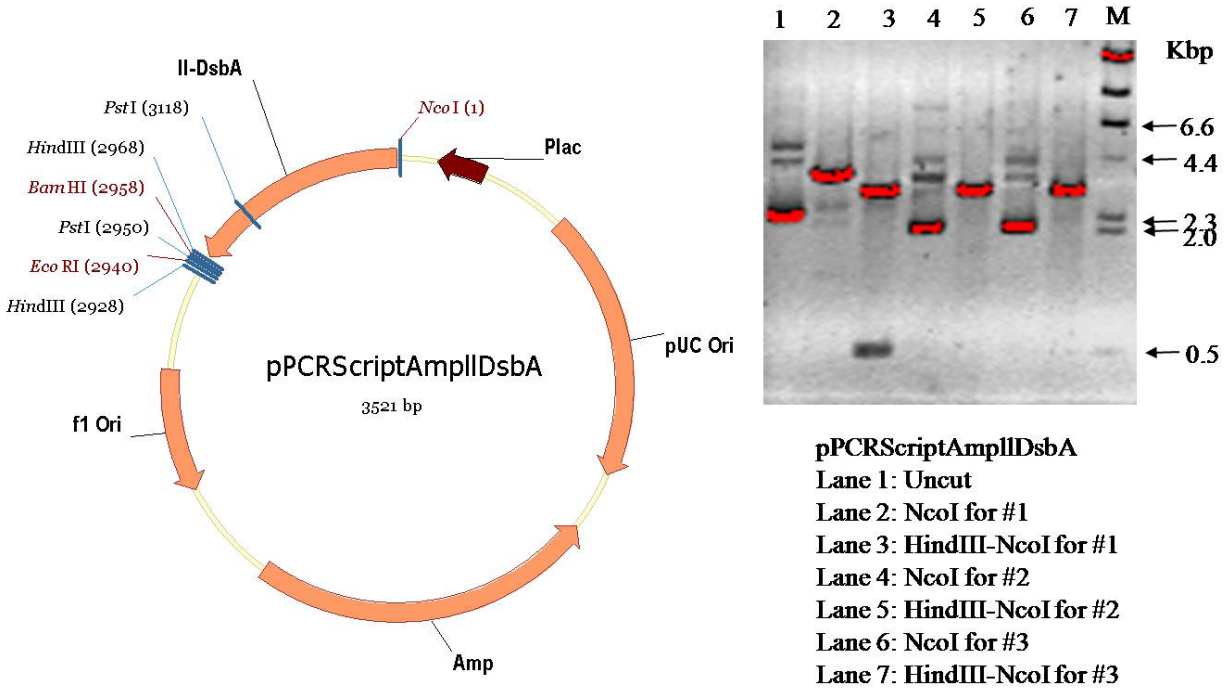


Figure B.38 Restriction analysis of pPCRScriptAmpIIDsbA

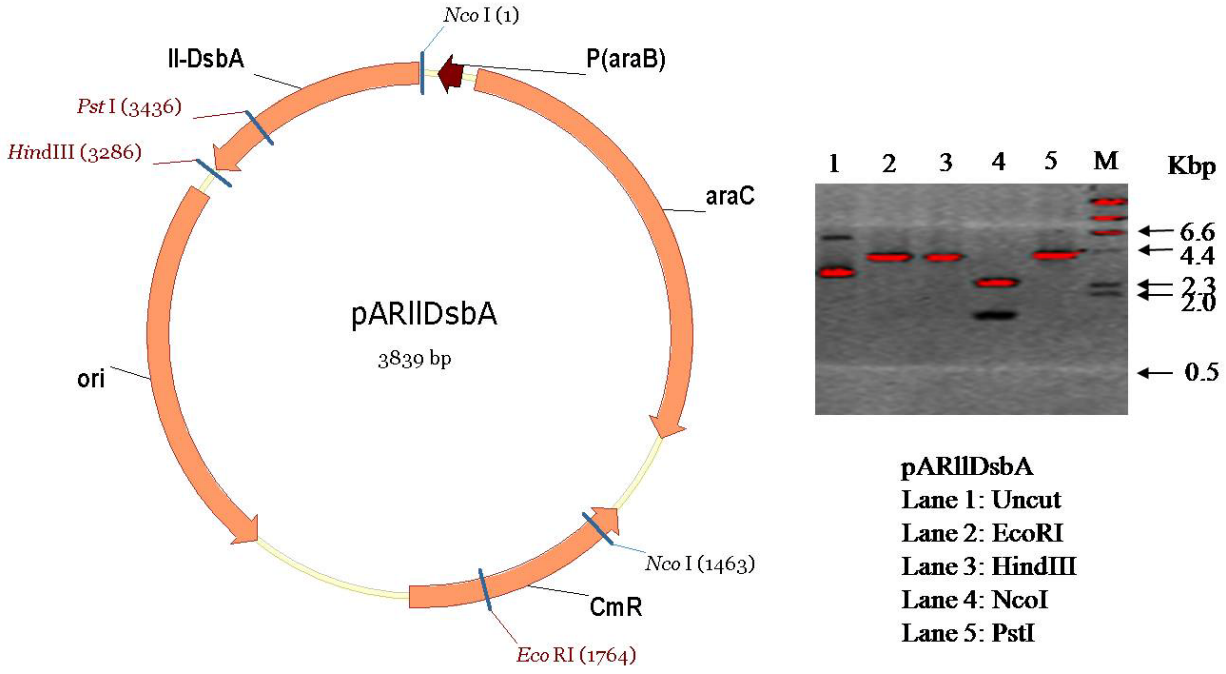


Figure B.39 Restriction analysis of pARIIDsbA

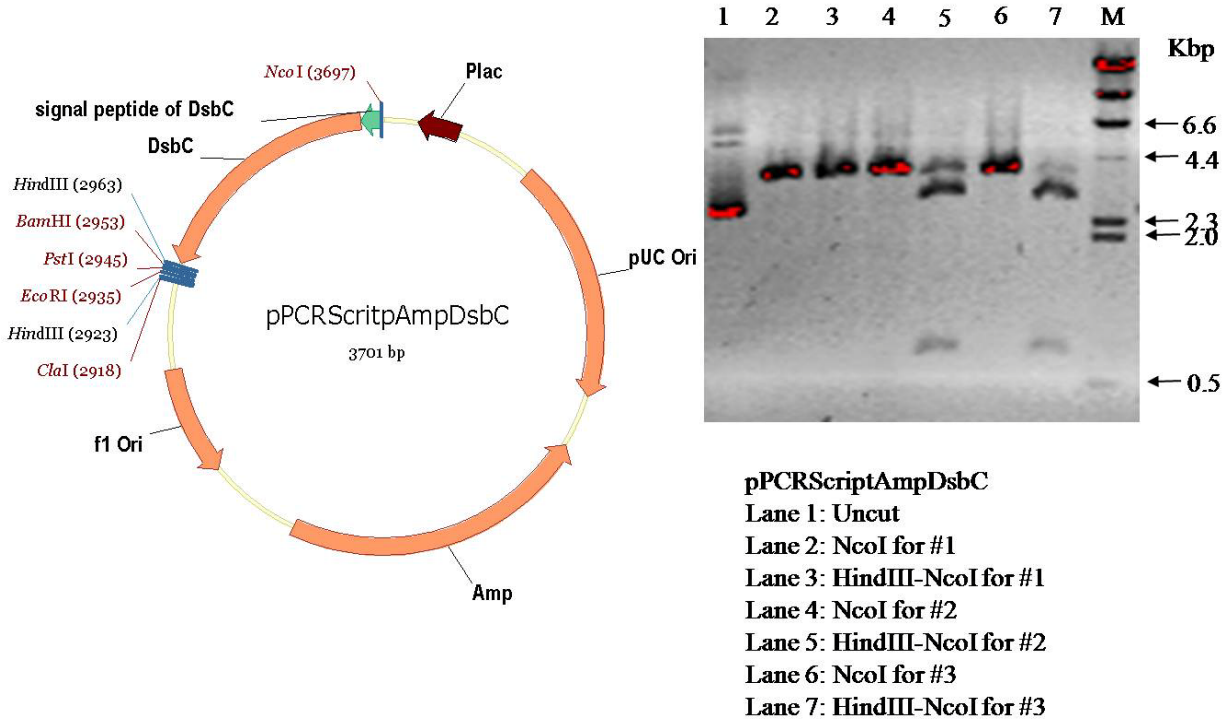


Figure B.40 Restriction analysis of pPCRScriptAmpDsbC

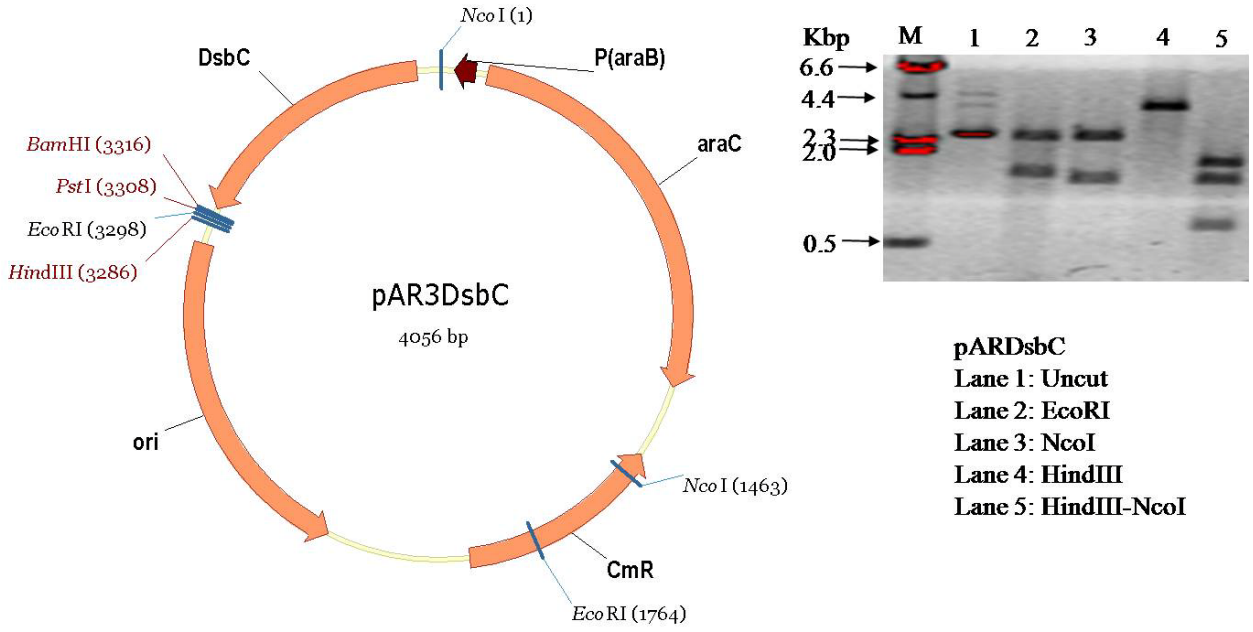


Figure B.41 Restriction analysis of pARDsbC

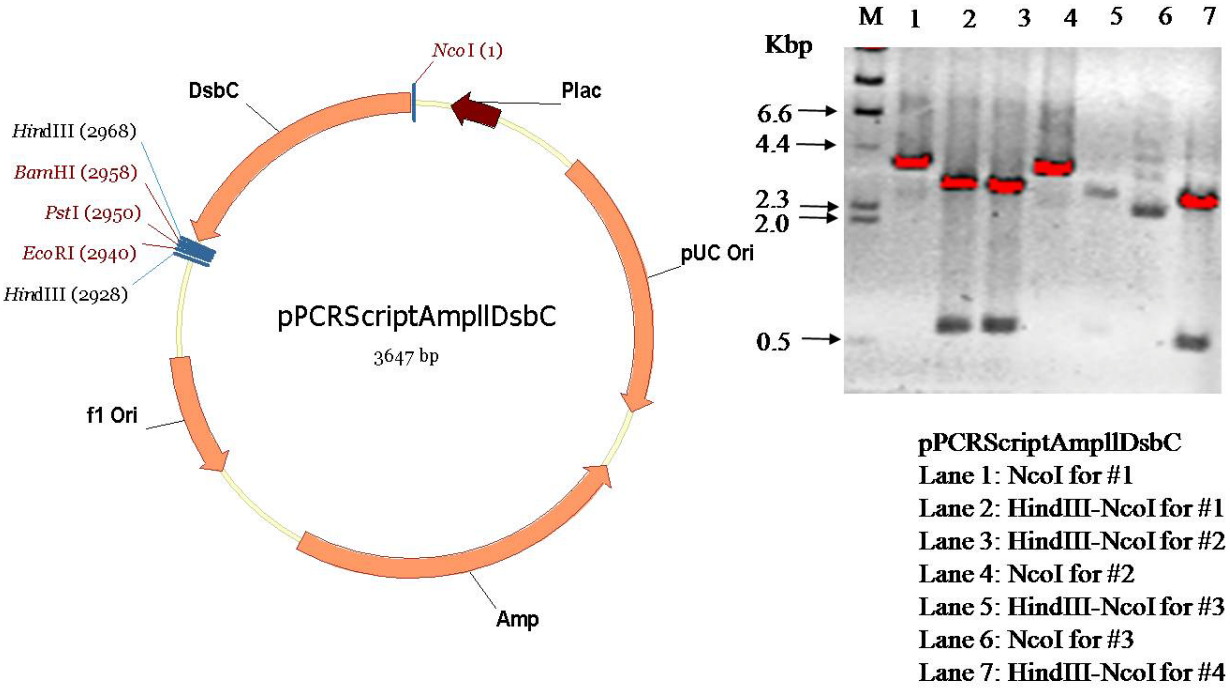


Figure B.42 Restriction analysis of pPCRScriptAmpIIDsbC

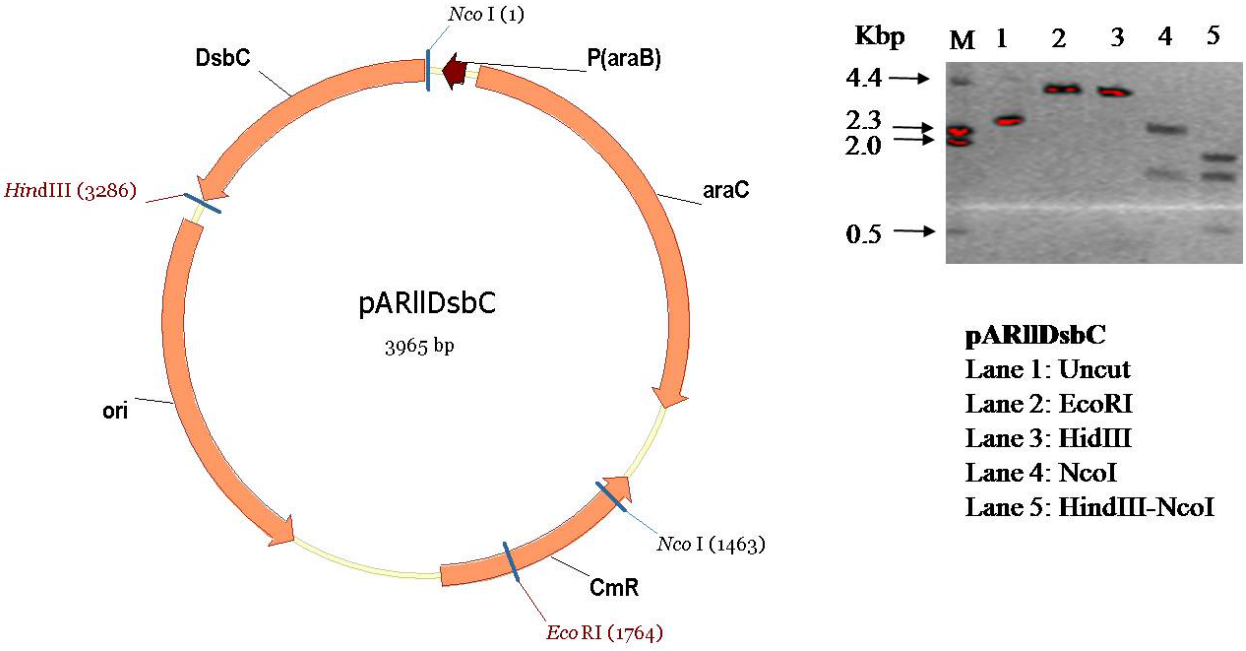


Figure B.43 Restriction analysis of pARIIDsbC

Appendix C

Protocols for Molecular Cloning

C1. Isolation of Genomic DNA of *Pseudozyma antarctica*

from strain ATCC 32657 (DNeasy® Tissue Kit Catalog no. 69504, Qiagen)

1. Harvest ATCC 32657 cells from 1 ml of overnight culture (at 30 °C, 220 rpm) by centrifuging for 10 min at 5000 × g, at room temperature. Discard the supernatant.
2. Resuspend the cells in 600 µl of 1 M sorbitol buffer. Add 200 units lyticase and incubate at 30 °C for 30 min.
3. Spin down spheroplasts at 300 × g for 10 min.
4. Resuspend the spheroplasts in 180 µl of Buffer ATL. Add 20 µl of proteinase K, vortex, and incubate at 55 °C in a shaking water-bath until the cells are completely lysed.
5. Vortex for 15 s. Add 200 µl of Buffer AL to the sample, vortex, and incubate at 70 °C for 10 min.
6. Add 200 µl of ethanol (96~100%) to the sample, and vortex.
7. Pipette the mixture from Step 6 into the DNeasy Mini-spin column placed in a 2 ml collection tube. Centrifuge at 6000 × g for 1 min. Discard the flow-through with the collection tube.
8. Place the DNeasy column in a new 2 ml collection tube, add 500 µl of Buffer AW1, and centrifuge at 6000 × g for 1 min. Discard the flow-through with the collection tube.
9. Place the DNeasy in a new 2 ml collection tube, add 500 µl of Buffer AW2, and centrifuge at 15000 × g for 5 min to dry the DNeasy membrane. Discard the flow-through with the collection tube.
10. Place the DNeasy column in a clean 1.5 ml microcentrifuge tube, and pipette 200 µl of Buffer AE directly onto the membrane. Incubate at room temperature for 1 min, and then centrifuge at 6000 × g for 1 min to elute.
11. Use the Nanodrop® spectrophotometer to measure the DNA's concentration, label the sample's name and concentration on the surface of the tube, and store it in -20 °C freezer.

C2. DNA Restriction Digestion

1. For analysis purpose, set up the reaction mixture in a microcentrifuge tube as follows:

DI Water	To make up the mixture to 10 µl		
Reaction Buffer	10 X	1 X	1 µl

DNA	50-500 ng/µl	50~100 ng/µl	y µl
Restriction Enzyme	1000 U/µl[#]	~1 U	1 µl[#]

[#] The restriction enzymes should be diluted with the appropriate diluents buffer to 1 U/µl.

* For most of the commercial enzymes, 1 Unit of enzyme is defined as the amount required digesting 1µg of DNA in 1 h at 37 °C.

To extract DNA from agarose gel, the amount of DNA can go up to 1 µg per 20 µl of total solution, the amount of restriction enzyme is about 5~8 U/µg DNA, and the incubation time is about 2 h, or even overnight.

2. Mix the ingredients by pulse-spinning the microcentrifuge tubes.
3. Incubate the reaction solution at 37°C (or other recommended temperature for specific restriction enzymes) for at least 1 hr.

C3. DNA Gel Electrophoresis Analysis

Stock Solution

0.5M EDTA, pH 8.0, stored at room temperature

1. Dissolve 14.61 g EDTA (MW=292.2) into 80ml DI water
2. Adjust the pH to 8.0 with NaOH and Add DI water to the final volume of 100 ml.
3. Autoclave the solution for 25 min.

50X TAE buffer, pH 8.3, stored at room temperature

1. Dissolve 48.44 g Tris Base (MW=121.1) in 150ml DI H_2O
2. Add 11.42 ml of glacial acetic acid and 20 ml of 0.5 M EDTA (pH 8.0) solution
3. Add DI H_2O to the final volume of 200 ml

10 mg/ml Ethidium Bromide (EtBr), stored at room temperature

1. Mix 1 g EtBr with 100 ml DI H_2O by stirring on a magnetic stirrer until the dye has been completely dissolved.
2. Wrap the bottle in aluminum foil.

10X DNA Sample Loading Dye

1. Mix 5 ml glycerol, 25 µl bromophenol blue and 25 µl xylene cyanole FF.
2. Add 1 X TAE buffer to the final volume of 10 ml.

Procedures

1. Add 6 ml 50 X TAE buffer into 294 ml DI H_2O (300 ml of 1 X TAE buffer). For gel extraction purposes, adjust the pH of running TAE buffer with 3M Sodium Acetate buffer (pH 5.2) to less than 7.5. Otherwise don't need to adjust pH before use.
2. In separate beaker, add 0.225 g of agarose into 30 ml of 1 X TAE buffer.
3. Assemble the UVTP gel tray on the gel caster properly.
4. Boil the agarose solution. Pour the agarose solution (~50 ºC) into the tray with comb pre-inserted.
5. Take out the comb until the gel completely solidified and put the tray into electrophoresis apparatus.
6. Pour 1 X TAE buffer till the level is about 2~4 mm above the gel.
7. Add 1 µl loading buffer to each DNA sample and mix by pulse-spinning.
8. Load DNA samples.

Table 1: Well Volume Capacity

Well Number	Thickness (mm)	Well Width (mm)	Well Volume Capacity (µl)
8	1.50	5.54	15
15	1.50	2.59	10

For higher loading volumes, the wells can be made by combining the tooth of the comb with a tape.

9. Conduct electrophoresis under constant voltage, 100V, for 45 min.

Table 2: DNA Size Migration with Sample Loading Dyes

Agarose Concentration (%)	Xylene Cyanol*	Bromophenol Blue*
0.5~1.5	4~5 Kb	400~500 bp

* During electrophoresis, Xylene Cyanol is blue and Bromophenol Blue is purple

10. Stain gel with 1 µg/ml EtBr for 10 min. and destain gel with DI H_2O for 15 min. The staining solution can be reused several times in a week.
11. Take picture using Kodak DC290 under the UV illuminator.

Table 3: Gel Volume Requirements

Gel Size	0.25 cm thick	0.5 cm thick	0.75 cm thick	1.0 cm thick

7x10 cm	15 ml	30 ml	45 ml	60 ml

Table 4: Gel Concentration Required for DNA Separation

Gel Concentration	DNA Size (Kb)
0.50	1-13
0.75	0.8-12
1.00	0.5-10
1.25	0.4-7
1.50	0.2-3
2-5*	0.01-0.5

*Sieving Agarose such as AmpliSize agarose

Note: This specific instruction is designed for a single, 0.5 cm thick, 7×10 cm Gel with 0.75 w/v% Agarose Concentration, for DNA size 0.8~12 Kb. Gel volume is adjustable depending on the thickness and the size of gel required (Table 3). The agarose concentration depends on the size of the DNA fragment analyzed (Table 4). The weight of the agarose required depends on the above two factors.

Reference

Sub-Cell GT Agarose Gel Electrophoresis Systems: Instruction Manual, Bio-Rad.

C4. DNA Agarose Gels Extraction and Purification

(QIAEX II agarose gels extraction protocol)

1. Excise the DNA band under the UV illuminator from the agarose gel with a clean sharp cutter. Chop it into small pieces and put into a weighted colorless tube.
2. Get the net gel slice weight. Add 300 µl of Buffer QX1 to 100 mg of gel for the DNA fragments 0.1~4 kb. For the DNA fragment > 4 kb, add extra 200 µl of DI H_2O.
3. Add 30 µl of QIAEX II to 2~10 µg DNA sample. Incubate at 50 °C water-bath for 10 min to solubilize the agarose and bind the DNA. Mix by vortexing every 2 min to keep QIAEX II in suspension. The color of the mixture should be yellow. If it is orange or purple, add 10 µl 3 M of NaAc (pH5.0) and mix.

4. Centrifuge the sample at 12000 × g for 1 min and remove supernatant with a pipette.
5. Resuspend the pellet with 500 µl of Buffer QX1 and mix by vortexing. Centrifuge the sample at 12000 × g for 1 min and remove supernatant with a pipette.
6. Resuspend the pellet with 500 µl of Buffer PE and mix by vortexing. Centrifuge the sample at 12,000 × g for 1 min and remove supernatant with a pipette.
7. Repeat step 6 again.
8. Air-dry the pellet for 10-15 min or until the pellet turns white.
9. Add 20 µl of 10 mM Tris-HCl (pH 8.5) or DI H_2O and resuspend by vortexing. Incubate at room temperature for 5 min. If the DNA fragments 4~10 kb, incubate at 50 °C water-bath for 5 min.
10. Centrifuge for 1 min. carefully pipette the supernatant into a clean tube. This DNA fragment is ready for ligation right away or store it at -20 °C freezer.

C5. Ligation Reaction for Constructing Expression Plasmid from PCR Product

1) Digest the plasmid DNA and the insert DNA with the appropriate restriction enzymes. Isolate the desired fragments by gel electrophoresis, and purify them by gel extraction.
2) Set up the ligation reaction as follows:

Component	Amount	Volume used (µl)
Insert DNA	0.1-1µg *	x
Vector DNA	0.05-1µg *	y
10X Ligation Buffer	-	1.0
T4 DNA Ligase Enzyme	0.1 Weiss Units ©	0.2 µl
DI H_2O	To make up the reaction volume to 10 µl.	

*****It is recommended to use equimolar amount of vector and insert DNA for cohesive end ligation.
©One Weiss Unit is defined as the amount of enzyme that catalyzes the exchange of 1 nmole of P32 from pyrophosphate into ATP in 20 minutes at 37 °C. One Weiss Unit corresponds to 60 cohesive end units as defined by New England Biolabs.

3) Mix the components in a pulse spinner.
4) Incubate the reaction mixture 1~4 hours at room temperature, or overnight at 16 °C.
5) Set up two additional control reactions that contain

i) the vector DNA alone and

ii) the insert DNA alone.

6) Take 1 µl of the sample for transformation by electroporation to the *E. coli* host competent cell DH5α.

C6. Gateway® Technology for Constructing Expression Vector

1) Add the following components to a 1.5 ml microcentrifuge tube at room temperature:

Entry clone	(50~80 ng)	~1 µl
Destination vector	(100 ng)	~1 µl
TE buffer, pH 8.0	To make up the mixture volume to 4 µl	

2) Thaw the LR Clonase™ II enzyme mix on ice and vortex it briefly.

3) Add 1 µl of enzyme mix to the sample in 1), vortex it briefly, and spin down.

4) Incubate the reaction mixture at room temperature for 1 h.

5) Add 0.5 µl of the proteinase K solution to the reaction mixture to terminate the reaction by incubating the sample at 37 ℃ for 10 min.

6) Take 1 µl of the sample for transformation by electroporation to the competent cell DH5α.

Background of Gateway® technology:

The traditional cloning, generally speaking, is one PCR product corresponding to one expression vector. It is very tedious and time-consuming if cloning one gene in a variety of expression vectors for systematically studying protein expression and/or functional analysis, such like *palB* in this study. Hartley *et al.* first invented a new DNA cloning technique that uses *in vitro* site-specific recombination to accomplish the directional cloning of PCR product and the subsequent automatic subcloning of the DNA segment into new vector backbone at high efficiency. The resulting subclone maintains orientation and reading frame register, allowing N- and C-terminal fusions to be created for optimized protein expression [288].

Invitrogen commercialized this invention to Gateway® technology. The entry clone is a plasmid containing the specific insert gene flanked by *attL1-gene-attL2*, which serve as a recognition site for the LR Clonase™ enzymes. The destination vector carries the selection *ccdB* gene flanked by *attR1-ccdB-attR2*. In the LR reaction, the entry clone and the destination vector are incubated together with the LR Clonase™ mix, resulting in the expression vector. The Gateway® system consists of two selection

schemes. The entry clone is Kn^R. The destination vector is Ap^R and Cm^R with *ccdB* gene, which inhibits growth of *E. coli*. Therefore transformants selected by Ap^R, not Ap^R+Cm^R, will be the expression vector.

In this study, a standard restriction digestion with a following ligation using a Gateway® containing *attL* sites, pENTR, was used to construct PalB entry clone.

C7. Transformation of Plasmid DNA into *E. coli* Strain

(Modified after the protocol from Bio-Rad)

There are two methods for transformation of plasmid DNA into *E. coli* strain, chemical transformation and electroporation. For chemical transformation, cells are grown to mid-log phase, harvested and treated with divalent cations such as $CaCl_2$. Cells treated in such a way are named to be competent. To chemically transform cells, the competent cells are mixed with the plasmid DNA, on ice, followed by a brief heat shock. Then, cells are incubated with rich medium and allowed to express the antibiotic resistant gene for 30-60 min prior to plating.

For electroporation, cells are also grown to mid-log phase but are then washed extensively with 10% of glycerol to eliminate all salts. Aliquot 40 µl competent cells into 1.5 ml of microcentrifuge tube and store them at -80 °C freezer. To electroporate DNA into cells, the competent cells are mixed with the 50~100 ng DNA and then pipetted into a cooled plastic cuvette containing electrodes. A short electric pulse, about 2400 volts/cm, is applied to the cells causing small holes in the membrane through which the DNA enters. The cells are then incubated with 1 ml of SOS broth before plating. Incubation time depends on copy number, high copy number needs less time. Spread 100~200 µl of the incubated cells on an agar plate supplemented with antibiotics when necessary and incubate the plate at 37 °C for overnight.

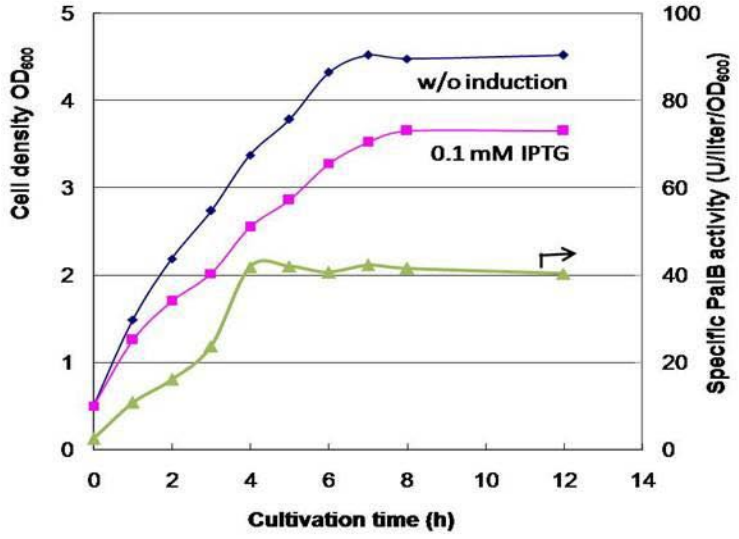
Note: the success of the electroporation depends greatly on the purity of the plasmid DNA, especially on its salt content. Impure solutions might cause a small explosion (known as arcing), in which case the bacteria are dead. In general, the efficiency of transformation is much higher by electroporation. Therefore, it is always the best choice for plasmid DNA transformation using ligation mixture. For regular plasmid DNA transformation by electroporation, the plasmid extracted from kit is too high concentrated and needs to be diluted. Otherwise, too many cells will cover all the plate to find a single colony.

Appendix D

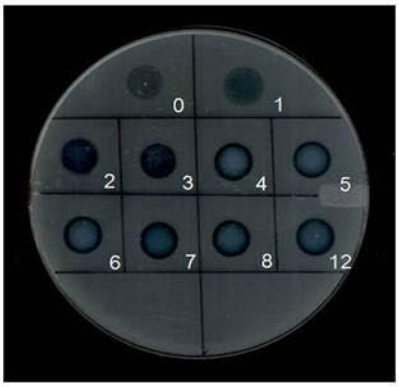
E. coli Growth Curve and PalB Activity along the Course of Cultivation

According to the procedure for cultivation, the sample was taken and cell density was measured along the course of cultivation for both non-induction and 0.1 mM IPTG induction culture. For the induction culture, the sample was also taken for monitoring the PalB activity using both tributyrin plate and pNPP enzymatic reaction following the standard protocol given in 3.2.3. The cultivation was conducted using BL21(DE3) (pHisperiMBP-G) expression system under 28 °C 200 rpm. The inducer concentration was 0.1 mM IPTG. The results were summarized in Figure D. The cells have reached the stationary phase 8 h later after induction. Due to the overexpression of PalB, cells in the induction system were arrested and OD_{600} was always less than non-induction system. The difference reached the maximum about 20% when both systems arrived at stationary phase. The PalB activity was exponentially increased during the first 4 hour after induction and then stopped. That indicated that the sample taken at 4 h after induction when the cultivation was conducted at 28 C is the best point. This conclusion is also confirmed by using tributyrin plate (Figure DB). The specific PalB activity reached 40 U/liter/OD_{600} which is comparable with the result in Figure 6.4B.

A



B



Visualization of PalB activity along the course of cultivation time
Strain: BL21(DE3) (pHisperiMBP-G)
Cultivation condition: 28°C, 200rpm
0: non-induction sample
1~12: 0.1 mM IPTG after induction, h

Figure D. *E. coli* growth curves of the non-induction and the induction PalB expression system BL21(DE3) (pHisperiMBP-G) and PalB activity along the course of cultivation after induction. Panel A: quantitative PalB assay results by using pNPP enzymatic reaction. The reaction was conducted in a shaker at 37 °C, 200 rpm. Panel B: visualization of PalB activity on tributirin plate. The plate loaded 20 µl of lysate was incubated in an incubater at 37 °C for 5h.

Appendix E

Coefficient Number for PalB Assay

PalB enzymatic assay was conducted according the protocol described in 3.2.3. PalB activity was calculated from the OD_{410} reading by using the following equations:

$$PalBActivity_{extracellular} = \frac{a*(OD_{410(t=t)} - OD_{410(t=0)})*V_{reaction}}{V_x*t}*1000 \quad \text{U/liter} \quad \textbf{(1)}$$

$$PalBActivity_{intracellular} = \frac{a*(OD_{410(t=t)} - OD_{410(t=0)})*V_{reaction}*V_{PB}}{V_x*V_{sample}*t}*1000 \quad \text{U/liter} \quad \textbf{(2)}$$

$$PalBActivity_{volumetric} = PalBActivity_{extracellular} + PalBActivity_{intracellular} \quad \text{U/liter} \quad \textbf{(3)}$$

$$PalBActivity_{specific} = \frac{PacActivity_{Volumetric}}{OD_{600}} \quad \text{U/liter/OD}_{600} \quad \textbf{(4)}$$

Where a is the slope of the standard curve for pNP concentration to OD_{410} reading, a = 133.4 (**Error! Reference source not found.**)

$V_{reaction}$ is volume of enzymatic reaction (ml),

V_x is volume of enzyme (ml),

V_{sample} is sample volume taken from culture (ml)

V_{PB} is volume of phosphate buffer in which the cell pellet was resuspended (ml),

t is enzymatic reaction time (min),

OD_{600} is cell's optical density.

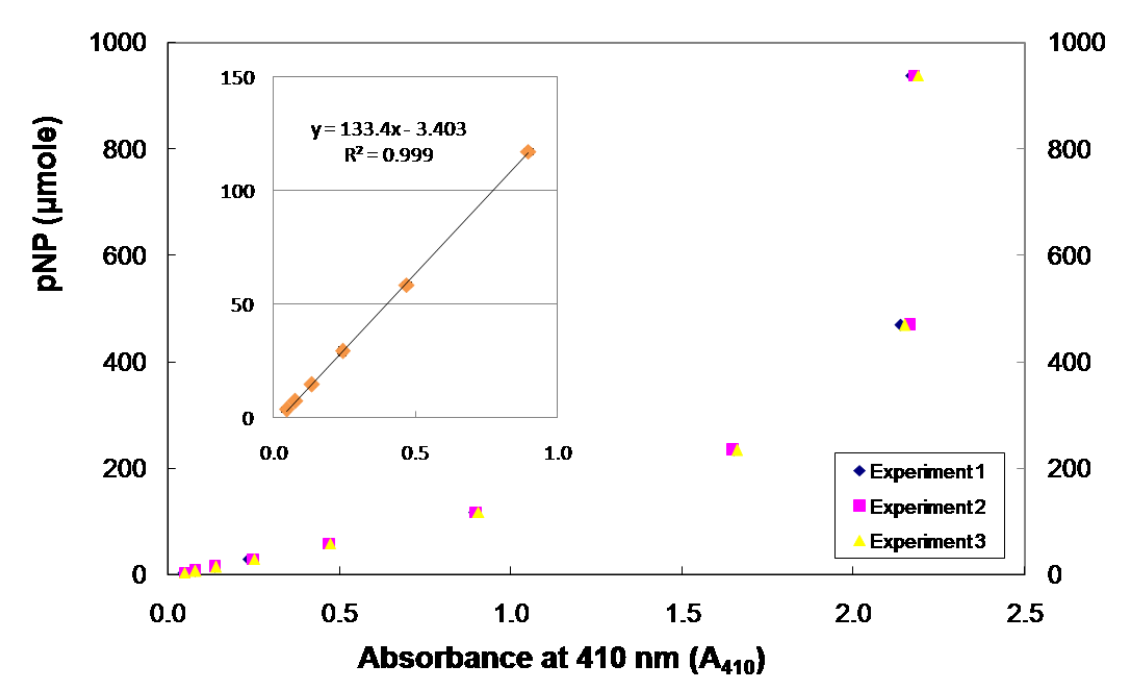


Figure E. Standard curve relating pNP concentration to its absorbance at 410 nm (A_{410}) reading.

References

1. Dresselhaus, MS; Thomas, IL (2001) Alternative energy technologies. Nature, 414: 332-337.
2. Gerpen, JV (2005) Biodiesel processing and production. Fuel Process. Technol., 86: 1097-1107.
3. Srivastava, A; Prasad, R (2000) Triglycerides-based diesel fuels. Renew. Sust. Energ. Rev., 4: 111-133.
4. Pinto, AC; Guarieiro, LLN; Rezende, MJC; Ribeiro, NM; Torres, EA; Lopes, WA; Pereira, PAD; de Andrade, JB (2005) Biodiesel: An overview. J. Brazil. Chem. Soc., 16: 1313-1330.
5. Bunkyakiat, K; Maknee, S; Saqangkeaw, R; Ngamprasertsith, S (2006) Continuous production of biodiesel via transesterification from vegetable oils in supercritical methanol. Energ. Fuel., 20: 812-817.
6. Demirbas, A (2003) Biodiesel fuels from vegetable oils via catalytic and non-catalytic supercritical alcohol transesterifications and other methods: a survey. Energ. Convers. Manage., 44: 2093-2109.
7. Komers, KS, R.; Machek, J.; and Skopal F. (2001) Biodiesel from rapeseed oil, methanol and KOH 3. Analysis of composition of actual reaction mixture. Eur. J. Lipid Sci. Tech., 103: 363-371.
8. Demirbas, A (2002) Biodiesel from vegetable oils via transesterification in supercritical methanol. Engerg. Convers. Manage., 43: 2349-2356.
9. Oliveira, D; Luccio, dD; Faccio, C; Rosa, CD; Bender, JP; Lipke, N; Amroginski, C; Dariva, C; Oliveira, JVd (2005) Optimization of alkaline transesterification of soybean oil and castor oil for biodiesel production. Appl. Biochem. Biotechnol., 122: 553-560.
10. Kaieda, M; Samukawa, T; Kondo, A; Fukuda, H (2001) Effect of methanol and water contents on production of biodiesel fuel from plant oil catalyzed by various lipases in a solvent-free system. J. Biosci. Bioeng., 91: 12-15.
11. Shimada, Y; Watanabe, Y; Sugihara, A; Tominaga, Y (2002) Enzymatic alcoholysis for biodiesel fuel production and application of the reaction to oil processing. J. Mol. Catal. B, 17: 133-142.
12. Chang, H-M; Liao, H-F; Lee, C-C; Shieh, C-J (2005) Optimized synthesis of lipase-catalyzed biodiesel by Novozym 435. J. Chem. Tech. Biotechnol., 80: 307-312.
13. Sharmaa, R; Chistib, Y; Banerjee, UC (2001) Production, purification, characterization, and applications of lipases. Biotech. Adv., 19: 627-662.
14. Linko, YY; Lamsa, M; Wu, XY; Uosukainen, E; Seppala, J; Linko, P (1998) Biodegradable products by lipase biocatalysis. J. Biotechnol., 66: 41-50.

15. Iso, M; Chen, BX; Eguchi, M; Kudo, T; Shrestha, S (2001) Production of biodiesel fuel from triglycerides and alcohol using immobilized lipase. J. Mol. Catal. B, 16: 53-58.
16. Soumanou, MM; Bornscheuer, UT (2003) Improvement in lipase-catalyzed synthesis of fatty acid methyl esters from sunflower oil. Enzyme Microb. Tech., 33: 97-103.
17. Deng, L; Xu, XB; Haraldsson, GG; Tan, TW; Wang, F (2005) Enzymatic production of alkyl esters through alcoholysis: A critical evaluation of lipases and alcohols. J. Am. Oil Chem. Soc., 82: 341-347.
18. Matsumoto, T; Takahashi, S; Kaieda, M; Ueda, M; Tanaka, A; Fukuda, H; Kondo, A (2001) Yeast whole-cell biocatalyst constructed by intracellular overproduction of Rhizopus oryzae lipase is applicable to biodiesel fuel production. Appl. Microbiol. Biotechnol., 57: 515-520.
19. Xu, YY; Du, W; Zeng, J; Liu, DH (2004) Conversion of soybean oil to biodiesel fuel using lipozyme TL IM in a solvent-free medium. Biocatal. Biotransfor., 22: 45-48.
20. Jaeger, KE; Reetz, MT (1998) Microbial lipases form versatile tools for biotechnology. Trend. Biotechnol., 16: 396-403.
21. Jaeger, KE; Eggert, T (2002) Lipases for biotechnology. Curr. Opin. Biotechnol., 13: 390-397.
22. Boekhout, T (1995) *Psedudozyma* Bandomi emend. Boekhout, A genus for yeast-like anamorphs of ustilaginales. J. Gen. Appl. Microbiol., 41: 359-366.
23. Chen, JW; Wu, WT (2003) Regeneration of immobilized *Candida antarctica* lipase for transesterification. J. Biosci. Bioeng., 95: 466-469.
24. Lai, CC; Zullaikah, S; Vali, SR; Ju, YH (2005) Lipase-catalyzed production of biodiesel from rice bran oil. J. Chem. Tech. Biot., 80: 331-337.
25. Hoegh, I; Patkar, S; Halkier, T; Hansen, M (1995) Two lipases from *Candida antarctica*: cloning and expression in *Aspergillus oryzae.* Can. J. Bot., 73: S869-S875.
26. Rotticci-Mulder, JC; Gustavsson, M; Holmquist, M; Hult, K; Martinelle, M (2001) Expression in *Pichia pastoris* of *Candida antarctica* lipase B and lipase B fused to a cellulose-binding domain. Protein Expr. Purif., 21: 386-392.
27. Zhang, N; Suen, W-C; Windsor, W; Xiao, L; Madison, V; Zaks, A (2003) Improving tolerance of *Candida antarctica* lipase B towards irreversible thermal inactivation through directed evolution. Protein Eng. Des. Sel., 16: 599-605.
28. Blank, K; Morfill, J; Gumpp, H; E., GH (2006) Functional expression of *Candida antarctica* lipase B in *Escherichia coli*. J. Biotechnol., 125: 474-483.

29. Liu, D; Schmid, RD; Rusnak, M (2006) Functional expression of *Candida antarctica* lipase B in the *Escherichia coli* cytoplasm - a screening system for a frequently used biocatalyst. Appl. Microbiol. Biotechnol., 72: 1024-1032.
30. Eiteman, MA; Altman, E (2006) Overcoming acetate in *Escherichia coli* recombinant protein fermentations. Trend. Biotechnol., 24: 530-536.
31. Andersen, DC; Krummen, L (2002) Recombinant protein expression for therapeutic applications. Curr. Opin. Biotech., 13: 117-123.
32. Schmidt, FR (2004) Recombinant expression systems in the pharmaceutical industry. Appl. Microbiol. Biotechnol., 65: 363-372.
33. Walsh, G (2006) Biopharmaceutical benchmarks 2006. Nat. Biotechnol., 24: 769-776.
34. Kurland, CG; Dong, HJ (1996) Bacterial growth inhibition by overproduction of protein. Mol. Microbiol., 21: 1-4.
35. Choi, JH; Keum, KC; Lee, SY (2006) Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chem. Eng. Sci., 61: 876-885.
36. Aldor, IS; Krawitz, DC; Forrest, W; Chen, C; Nishihara, JC; Joly, JC; Champion, KM (2005) Proteomic profiling of recombinant *Escherichia coli* in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. Appl. Environ. Microbiol., 71: 1717-1728.
37. Lee, SY (1996) High cell density culture of *Escherichia coli*. Trend. Biotechnol., 14: 98-105.
38. Makrides, SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol. Rev., 60: 512-538.
39. Choi, JH; Jeong, KJ; Kim, SC; Lee, SY (2000) Efficient secretory production of alkaline phosphatase by high cell density culture of recombinant *Escherichia coli* using the *Bacillus sp* endoxylanase signal sequence. Appl. Microbiol. Biotechnol., 53: 640-645.
40. Kadokura, H; Katzen, F; Beckwith, J (2003) Protein disulfide bond formation in prokaryotes. Annu. Rev. Biochem., 72: 111-135.
41. Chou, CP (2007) Protein overexpression in *Escherichia coli*: Engineering bugs to make them more productive in protein synthesis and more robust when being cultured. Bioscience World, 10:
42. Jana, S; Deb, JK (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. Appl. Microbiol. Biotechnol., 67: 289-298.
43. Baneyx, F; Mujacic, M (2004) Recombinant protein folding and misfolding in *Escherichia coli*. Nat. Biotechnol., 22: 1399-1408.

44. Jonasson, P; Liljeqvist, S; Nygren, P-A; Stahl, S (2002) Genetic design for facilitated production and recovery of recombinant proteins in *Escherichi coli*. Biotechnol. Appl. Bioc., 35: 91-105.
45. Kues, U; Stahl, U (1989) Replication of plasmids in gram-negative bacteria. Microbiol. Rev., 53: 491-516.
46. Badyakina, AO; Nesmeyanova, MA (2005) Biogenesis and secretion of overproduced protein in recombinant strains of *Escherichia coli*. Process Biochem., 40: 509-518.
47. Nordstrom, K; Uhlin, BE (1992) Runaway-replication plasmids as tools to produce large quantities of proteins from cloned genes in bacteria. Bio/Technology, 10: 661-666.
48. Togna, AP; Shuler, ML; Wilson, DB (1993) Effects of plasmid copy number and runaway plasmid replication on overproduction and excretion of β-lactamase from *Escherichia coli*. Biotechnol. Prog., 9: 31-39.
49. Ricci, JCD; Hernandez, ME (2000) Plasmid effects on *Escherichia coli* metabolism. Crit. Rev. Biotechnol., 20: 79-108.
50. Wang, Z; Xiang, L; Shao, J; Wegrzyn, A; Wegrzyn, G (2006) Effects of the presence of ColE1 plasmid DNA in *Escherichia coli* on the host cell metabolism. Microbial Cell Factories, 5: 34.
51. Betenbaugh, MJ; Dhurjati, P (1990) Effects of promoter induction and copy number amplification on cloned gene expression and growth of recombinant cell cultures. Ann. NY Acad. Sci., 111-120.
52. Neubauer, P; Lin, HY; Mathiszik, B (2003) Metabolic load of recombinant protein production: Inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*. Biotechnol. Bioeng., 83: 53-64.
53. Grabherr, R; Nilsson, E; Striedner, G; Bayer, K (2002) Stabilizing plasmid copy number to improve recombinant protein production. biotechnol. Bioeng., 77: 142-147.
54. Jones, KL; Kim, S-W; Keasling, JD (2000) Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab. Eng., 2: 328-338.
55. Seo, J-H; Bailey, JE (1985) Effects of recombinant plasmid content on growth properties and cloned gene product formation in *Escherichia coli*. Biotechnol. Bioeng., 27: 1668-1674.
56. Bartolome, B; Jubete, Y; Martinez, E; de la Cruz, F (1991) Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene, 102: 75-78.
57. Castanie, M; Berges, H; Oreglia, J; Prere, M; Fayet, O (1997) A set of pBR322-compatible plasmids allowing the testing of chaperone-assisted folding of proteins overexpressed in *Escherichia coli*. Anal. Biochem., 254: 150-152.

58. Kim, BG; Shuler, ML (1991) Kinetic analysis of the effects of plasmid multimerization on segregational instability of ColE1 type plasmids in *Escherichia coli* B/r. Biotechnol. Bioeng., 37: 1076-1086.
59. Yang, W; Zhang, L; Lu, ZG; Tao, W; Zhao, ZH (2001) A new method for protein coexpression in *Escherichia coli* using two incompatible plasmids. Protein Expr. Purif., 22: 472-478.
60. Velappan, N; Sblattero, D; Chasteen, L; Pavlik, P; Bradbury, ARM (2007) Plasmid incompatibility: more compatible than previously thought? Protein Eng. Des. Sel., 20: 309-313.
61. Sambrook, J; Fritsch, EF; Maniatis, T, *Molecular cloning: A laboratory manual*. 1989, New York, USA: Cold Spring Harbor Laboratory Press.
62. Sorensen, H; Mortensen, K (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J. Biotechnol., 115: 113-128.
63. Yanisch-Perron, C; Vieira, J; Messing, J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene, 33: 103-119.
64. DeBoer, HA; Shepard, HM, *Strategies for optimizing foreign gene expression in Escherichia coli*, in *Genes: Structure and expression*, A.M. Kroon, Editor. 1983, John Wiley & Sons Ltd. p. 205-248.
65. Russell, DR; Bennett, GN (1982) Cloning of small DNA fragments containing the *Escherichia coli* tryptophan operon promoter and operator. Gene, 17: 9-18.
66. Tacon, W; Carey, N; Emtage, S (1980) The construction and characterization of plasmid vectors suitable for the expression of all DNA phages under the control of the *Escherichia coli* tryptophan promoter. Mol. Gen. Genet., 177: 427-438.
67. Xu, Y; Hsieh, M; Narayanan, N; Anderson, WA; Scharer, JM; Moo-Young, M; Chou, CP (2005) Cytoplamic overexpression, folding, and processing of penicillin acylase precursor in *Escherichia coli*. Biotechnol. Prog., 21: 1357-1365.
68. Xu, Y; Weng, C; Narayanan, N; Hsieh, M; Anderson, WA; Scharer, JM; Moo-Young, M; Chou, CP (2005) Chaperone-mediated folding and maturation of the penicillin acylase precursor in the cytoplasm of *Escherichia coli*. Appl. Environ. Microbiol., 71: 6247-6253.
69. Narayanan, N; Hsieh, M-Y; Xu, Y; Chou, CP (2006) Arabinose-induction of *lac*-derived promoter systems for penicillin acylase production in *Escherichia coli*. Biotechnol. Prog., 22: 617-625.
70. Lechmann, M; Kremmer, E; Sticht, H; Steinkasserer, A (2002) Overexpression, purification, and biochemical characterization of the extracellular human CD83 domain and generation of monoclonal antibodies. Protein Expr. Purif., 24: 445-452.

71. Xu, Y; Rosenkranz, S; Weng, C-L; Scharer, JM; Moo-Young, M; Chou, CP (2006) Characterization of the T7 promoter system for expressing penicillin acylase in *Escherichia coli*. Appl. Microbiol. Biotechnol., 72: 529-536.
72. Dümmler, A; Lawrence, A-M; de Marco, A (2005) Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. Microbial Cell Factories, 4: 34.
73. Vasina, JA; Baneyx, F (1996) Recombinant protein expression at low temperatures under the transcriptional control of the major *Escherichia coli* cold shock promoter *cspA*. Appl. Environ. Microbiol., 62: 1444-1447.
74. TaKaRa, *Cold shock expression system pCold DNA, Manual, pp1-9*. 2003.
75. Mayer, M (1995) A new set of useful cloning and expression vectors derived from pBlueScript. Gene, 163: 41-46.
76. Nishihara, K; Kanemori, M; Yanagi, H; Yura, T (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. Appl. Environ. Microbiol., 66: 884-889.
77. Narayanan, N; Xu, Y; Chou, CP (2006) High-level gene expression for recombinant penicillin acylase production using the *araB* promoter system in *Escherichia coli*. Biotechnol. Prog., 22: 1518-1523.
78. Skerra, A (1994) Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. Gene, 151: 131-135.
79. Terpe, K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol., 72: 211-222.
80. Chou, C-H; Aristidou, AA; Meng, S-Y; Bennett, GN; San, K-Y (1995) Characterization of a pH-inducible promoter system for high-level expression of recombinant proteins in *Escherichia coli*. Biotechnol. Bioeng., 47: 186-192.
81. de Smit, MH; van Duin, J (1990) Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. Proc. Nat. Acad. Sci. U. S. A., 87: 7668-7672.
82. Simmons, LC; Yansura, DG (1996) Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. Nat. Biotechnol., 14: 629-634.
83. Thomas, DY; Dubuc, G; Narang, S (1982) *Escherichia coli* plasmid vectors containing synthetic translational initiation sequences and ribosome binding sites fused with the *lacZ* gene. Gene, 19: 211-219.

84. Izard, JW; Kendall, DA (1994) Signal peptides: exquisitely designed transport promoters. Mol. Microbiol., 13: 765-773.
85. Toivonen, JM; Manjiry, S; Tooouraille, S; Alziari, S; O'Dell, KMC; Howard, T; Jacobs, HTD (2003) Gene dosage and selective expression modify phenotype in a Drosophila model of human mitochondrial disease. Mitochondrion, 3: 83-96.
86. Terpe, K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol., 60: 523-533.
87. Waugh, DS (2005) Making the most of affinity tags. Trend. Biotechnol., 23: 316-320.
88. Choi, JH; Lee, SY (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl. Microbiol. Biotechnol., 64: 625-635.
89. Cornelis, P (2000) Expressing genes in different *Escherichia coli* compartments. Curr. Opin. Biotech., 11: 450-454.
90. Georgiou, G; Segatori, L (2005) Preparative expression of secreted proteins in bacteria: status report and future prospects. Curr. Opin. Biotech., 16: 538-545.
91. Lee, SY; Choi, JH; Xu, Z (2003) Microbial cell-surface display. Trend. Biotechnol., 21: 45-52.
92. Baneyx, F (1999) Recombinant protein expression in *Escherichia coli*. Curr. Opin. Biotech., 10: 411-421.
93. Sorensen, H; Mortensen, K (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microbial Cell Factories, 4: 1.
94. LaVallie, ER; DiBlasio, EA; Kovacic, S; Grant, KL; F., SP; M., MJ (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. Bio/Technology, 11: 187-193.
95. Weickert, MJ; Pagratis, M; Curry, SR; Blackmore, R (1997) Stabilization of apoglobin by low-temperature increases yield of soluble recombinant hemoglobin in *Escherichia coli*. Appl. Environ. Microbiol., 63: 4313-4320.
96. Wall, JG; Pluckthun, A (1995) Effects of overexpressing folding modulators on the *in vivo* folding of heterologous proteins in *Escherichia coli*. Curr. Opin. Biotech., 6: 507-516.
97. Strandberg, L; Enfors, S-O (1991) Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. 57: 1669-1674.
98. Murby, M; Samuelsson, E; Nguyen, TN; Mignard, L; Power, U; H. Binz; Uhle´n, M; Ståhl., S (1995) Hydrophobicity engineering to increase solubility and stability of a recombinant protein from respiratory syncytial virus. Eur. J. Biochem., 230: 38-44.

99. Dale, GE; Broger, C; Langen, H; D'Arcy, A; Stuber, D (1994) Improving protein solubility through rationally designed amino acid replacements: Solubilization of the trimethoprim-resistant type SI dihydrofolate reductase. Protein Eng., 7: 933-939.
100. Sandee, D; Tungpradabkul, S; Kurokawa, Y; Fukui, K; Takagi, M (2005) Combination of Dsb coexpression and an addition of sorbitol markedly enhanced soluble expression of single-chain Fv in *Escherichia coli*. Biotechnol. Bioeng., 91: 418-424.
101. Choi, JH; Lee, SJ; Lee, SJ; Lee, SY (2003) Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling. Appl. Environ. Microbiol., 69: 4737-4742.
102. Sugimoto, S; Yokoo, Y; Hatakeyama, N; Yotsuji, A; Teshiba, S; Hagino, H (1991) Higher culture pH is preferable for inclusion body formation of recombinant salmon growth hormone in *Escherichia coli*. Biotechnol. Lett., 13: 385-388.
103. Yee, L; Blanch, HW (1992) Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. Bio/Technology, 10: 1550-1556.
104. Bessette, PH; Aslund, F; Beckwith, J; Georgiou, G (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. Proc. Nat. Acad. Sci. U. S. A., 96: 13703-13708.
105. Bardwell, JCA (1994) Building bridges: disulphide bond formation in the cell. Mol. Microbiol., 14: 199-205.
106. Bardwell, JCA (2004) The dance of disulfide formation. Nat. Struct. Mol. Biol., 11: 582-583.
107. Robinson, AS; King, J (1997) Disulphide-bonded intermediate on the folding and assembly pathway of a non-disulphide bonded protein. Nat. Struct. Biol., 4: 450-455.
108. Derman, AI; Prinz, WA; Belin, D; Beckwith, J (1993) Mutations that allow disulfide bone formation in the cytoplasm of *Escherichia coli*. Science, 262: 1744-1747.
109. Åslund, F; Beckwith, J (1999) The thioredoxin superfamily: redundancy, specificity, and gray-area genomics. J. Bacteriol., 181: 1375-1379.
110. Jurado, P; Ritz, D; Beckwith, J; Lorenzo, Vd; Ferna´ndez, LA (2002) Production of functional single-chain Fv antibodies in the cytoplasm of *Escherichia coli*. J. Mol. Biol., 320: 1-10.
111. Levy, R; Weiss, R; Chen, G; Iverson, BL; Georgiou, G (2001) Production of correctly folded Fab antibody fragment in the cytoplasm of *Escherichia coli trxB gor* mutants via the coexpression of molecular chaperones. Protein Expr. Purif., 23: 338-347.

112. Mergulhao, FJM; Summers, DK; Monteiro, GA (2005) Recombinant protein secretion in *Escherichia coli*. Biotechnol. Adv., 23: 177-202.
113. Huang, H-C; Sherman, MY; Kandror, O; Goldberg, AL (2001) The molecular chaperone DnaJ is required for the degradation of a soluble abnormal protein in *Escherichia coli*. J. Biol. Chem., 276: 3920-3928.
114. Becker, S; Theile, S; Heppeler, N; Michalczyk, A; Wentzel, A; Wilhelm, S; Jaeger, KE; Kolmar, H (2005) A generic system for the *Escherichia coli* cell-surface display of lipolytic enzymes. FEBS Lett., 579: 1177-1182.
115. Benhar, I (2001) Biotechnological applications of phage and cell display. Biotechnol. Adv., 19: 1-33.
116. Wu, ML; Tsai, CY; Chen, TH (2006) Cell surface display of Chi92 on Escherichia coli using ice nucleation protein for improved catalytic and antifungal activity. FEMS Microbiol. Lett., 256: 119-125.
117. Oliver, DC; Huang, G; Nodel, E; Pleasance, S; Fernandez, RC (2003) A conserved region within the *Bordetella pertussis* autotransporter BrkA is necessary for folding of its passenger domain. Mol. Microbiol., 47: 1367-1383.
118. Narayanan, N; Chou, CP (2008) Physiological improvement to enhance *Escherichia coli* cell-surface display via reducing extracytoplasmic stress. Biotechnol. Prog., 24: 293-301.
119. Connolly, L; De Las Penas, A; Alba, BM; Gross, CA (1997) The response to extracytoplasmic stress in *Escherichia coli* is controlled by partially overlapping pathways. Gene. Dev., 11: 2012-2021.
120. Danese, PN; Silhavy, TJ (1997) The σ^E and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. Gene. Dev., 11: 1183-1193.
121. Raivio, TL; Silhavy, TJ (1999) The σ^E and Cpx regulatory pathways: overlapping but distinct envelope stress responses. Curr. Opin. Microbiol., 2: 159-165.
122. Ruiz, N; Silhavy, TJ (2005) Sensing external stress: watchdogs of the *Escherichia coli cell* envelope. Curr. Opin. Microbiol., 8: 122-126.
123. Villaverde, A; Carrio, MM (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol. Lett., 25: 1385-1395.
124. Carrio, MM; Corchero, JL; Villaverde, A (1998) Dynamics of *in vivo* protein aggregation: building inclusion bodies in recombinant bacteria. FEMS Microbiol. Lett., 169: 9-15.

125. Choe, WS; Nian, R; Lai, WB (2006) Recent advances in biomolecular process intensification. Chem. Eng. Sci., 61: 886-906.
126. Butt, TR; Edavettal, SC; Hall, JP; Mattern, MR (2005) SUMO fusion technology for difficult-to-expression proteins. Protein Expr. Purif., 43: 1-9.
127. Hoffmann, F; van den Heuvel, J; Zidek, N; Rinas, U (2004) Minimizing inclusion body formation during recombinant protein production in *Escherichia coli* at bench and pilot plant scale. Enzyme Microb. Tech., 34: 235-241.
128. Dong, H; Nilsson, L; Kurland, CG (1995) Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J. Bacteriol., 177: 1497-1504.
129. Gill, RT; Valdes, JJ; Bentley, WE (2000) A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. Metab. Eng., 2: 178-189.
130. Chou, CP (2007) Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. Appl. Microbiol. Biotechnol., 76: 521-532.
131. Arechaga, I; Miroux, B; Runswick, MJ; Walker, JE (2003) Over-expression of *Escherichia coli* F_1F_0-ATPase subunit a is inhibited by instability of the *uncB* gene transcript. FEBS Lett., 547: 97-100.
132. Gualerzi, CO; Pon, CL (1990) Initiation of mRNA translation in prokaryotes. Biochemistry, 29: 5881-5889.
133. Tessier, L-H; Sondermeyer, P; Faure, T; Dreyer, D; Benavente, A; Villeval, D; Courtney, M; Lecocq, J-P (1984) The influence of mRNA primary and secondary structure on human IFN-γ gene expression in *E. coli*. Nucleic Acids Res., 12: 7663-7675.
134. Betton, J-M; Sassoon, N; Hofnung, M; Laurent, M (1998) Degradation versus aggregation of misfolded maltose-binding protein in the periplasm of *Escherichia coli*. J. Biol. Chem., 273: 8897-8902.
135. Wickner, S; Maurizi, MR; Gottesman, S (1999) Posttranslational quality control: folding, refolding, and degrading proteins. Science, 286: 1888-1893.
136. Goldberg, AL (2003) Protein degradation and protection against misfolded or damaged proteins. Nature, 426: 895-899.
137. Nallamsetty, S; Austin, BP; Penrose, KJ; Waugh, DS (2005) Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. Protein Sci., 14: 2964-2971.

138. Hammarstrom, M; Woestenenk, EA; Hellgren, N; Hard, T; Berglund, H (2006) Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. J. Struct. Funct. Genornics, 7: 1-14.
139. Kapust, RB; Waugh, DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci., 8: 1668-1674.
140. Shih, YP; Kung, WM; Chen, JC; Yeh, CH; Wang, AHJ; Wang, TF (2002) High-throughput screening of soluble recombinant proteins. Protein Sci., 11: 1714-1719.
141. Liu, Y; Zhao, TJ; Yan, YB; Zhou, HM (2005) Increase of soluble expression in *Escherichia coli* cytoplasm by a protein disulfide isomerase gene fusion system. Protein Expr. Purif., 44: 155-161.
142. Tsunoda, Y; Sakai, N; Kikuchi, K; Katoh, S; Akagi, K; Miura-Ohnuma, J; Tashiro, Y; Murata, K; Shibuya, N; Katoh, E (2005) Improving expression and solubility of rice proteins produced as fusion proteins in *Escherichia coli*. Protein Expr. Purif., 42: 268-277.
143. Esposito, D; Chatterjee, DK (2006) Enhancement of soluble protein expression through the use of fusion tags. Curr. Opin. Biotech., 17: 353-358.
144. Schrödel, A; Marco, Ad (2005) Characterization of the aggregates formed during recombinant protein expression in bacteria. BMC Biochem., 6: 10.
145. Nygren, P-A; Stahl, S; Uhlen, M (1994) Engineering proteins to facilitate bioprocessing. Trend. Biotechnol., 12: 184-188.
146. De Marco, V; Stier, G; Blandin, S; de Marco, A (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. biochem. Bioph. Res. Co., 322: 766-771.
147. Chatterjee, DK; Esposito, D (2005) Enhanced soluble protein expression using two new fusion tags. Protein Expr. Purif., 46: 122-129.
148. Fink, AL (1999) Chaperone-mediated protein folding. Physiol. Rev., 79: 425-449.
149. Hartl, FU; Hayer-Hartl, M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science, 295: 1852-1858.
150. Hesterkamp, T; Hauser, S; Lutcke, H; Bukau, B (1996) *Escherichia coli* trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. Proc. Nat. Acad. Sci. U. S. A., 93: 4437-4441.
151. Scholz, C; Stoller, G; Zarnt, T; Fischer, G; Schmid, FX (1997) Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. EMBO J., 16: 54-58.
152. Ferbitz, L; Maier, T; Patzelt, H; Bukau, B; Deuerling, E; Ban, N (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature, 431: 590-596.

153. Lakshmipathy, SK; TomicC, S; Kaiser, hM; Chang, H-C; Genevaux, P; Georgopoulos, C; Barral, JM; Johnson, AE; Hartl, FU; Etchells, SA (2007) Identification of nascent chain interaction sites on trigger factor. J. Biol. Chem., 282: 12186-12193.
154. Hartl, FU (2007) Chaperone-assisted protein folding in the cytosol. FASEB J., 21: A153-A153.
155. Teter, SA; Houry, WA; Ang, D; Tradler, T; Rockabrand, D; Fischer, G; Blum, P; Georgopoulos, C; Hartl, FU (1999) Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell, 97: 755-765.
156. Blum, P; Ory, J; Bauernfeind, J; Krska, J (1992) Physiological consequences of DnaK and DnaJ overproduction in *Escherichia coli*. J. Bacteriol., 174: 7436-7444.
157. Kandror, O; Sherman, M; Moerschell, R; Goldberg, AL (1997) Trigger factor associates with GroEL *in vivo* and promotes its binding to certain polypeptides. J. Biol. Chem., 272: 1730-1734.
158. Deuerling, E; Patzelt, H; Vorderwulbecke, S; Rauch, T; Kramer, G (2003) Trigger factor and DnaK possess overlapping substrate pools and binding specificities. Mol. Microbiol., 47: 1317-1328.
159. Ranson, NA; White, HE; Saibil, HR (1998) Chaperonins. Biochem. J., 333: 233-242.
160. Thomas, JG; Baneyx, F (1996) Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overexpressing heat-shock proteins. J. Biol. Chem., 271: 11141-11147.
161. Buchner, J (1996) Supervising the fold: functional principles of molecular chaperones. FASEB J., 10: 10-19.
162. Kandror, O; Sherman, M; Goldberg, A (1999) Rapid degradation of an abnormal protein in *Escherichia coli* proceeds through repeated cycles of association with GroEL. J. Biol. Chem., 274: 37743-37749.
163. Kandror, O; Sherman, M; Rhode, M; Goldberg, AL (1995) Trigger factor is involved in GroEL-dependent protein degradation in *Escherichia coli* and promotes binding of GroEL to unfolded proteins. EMBO J., 14: 6021-6027.
164. Clark, PL; King, J (2001) A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early *in vitro* refolding intermediates. J. Biol. Chem., 276: 25411-25420.
165. Alba, BM; Gross, CA (2004) Regulation of the *Escherichia coli* σ^E-dependent envelope stress response. Mol. Microbiol., 52: 613-619.
166. Gamer, J; Multhaup, G; Tomoyasu, T; McCarty, JS; Rudiger, S; Schonfeld, HJ; Schirra, C; Bujard, H; Bukau, B (1996) A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones

regulates activity of the *Escherichia coli* heat-shock transcription factor sigma (32). EMBO J., 15: 607-617.

167. Phillips, GJ; Silhavy, TJ (1990) Heat-shock proteins DnaK and GroEL facilitate export of LacZ hybrid proteins in *E. coli.* Nature, 344: 882-884.
168. Straus, DB; Walter, WA; Gross, CA (1990) DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of σ^{32}. Gene. Dev., 4: 2202-2209.
169. Tilly, K; McKittrick, N; Zylick, M; Georgopoulos, C (1983) The *dnaK* protein modulates the heat-shock response of *Escherichia coli.* Cell, 34: 641-646.
170. Wild, J; Altman, E; Yura, T; Gross, CA (1992) DnaK and DnaJ heat shock proteins participate in protein export in *Escherichia coli.* Gene. Dev., 6: 1165-1172.
171. Schlieker, C; Bukau, B; Mogk, A (2002) Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: Implications for their applicability in biotechnology. J. Biotechnol., 96: 13-21.
172. Guisbert, E; Herman, C; Lu, CZ; Gross, CA (2004) A chaperone network controls the heat shock response in *E. coli.* Gene. Dev., 18: 2812-2821.
173. Mogensen, JE; Otzen, DE (2005) Interactions between folding factors and bacterial outer membrane proteins. Mol. Microbiol., 57: 326-346.
174. Duguay, AR; Silhavy, TJ (2004) Quality control in the bacterial periplasm. Biochim. Biophys. Acta., 1694: 121-134.
175. Erickson, JW; Gross, CA (1989) Identification of the σ^{E} subunit of *Escherichia coli* RNA Polymerase - a second alternate σ factor involved in high-temperature gene expression. Gene. Dev., 3: 1462-1471.
176. Mecsas, J; Rouviere, PE; Erickson, JW; Donohue, TJ; Gross, CA (1993) The activity of σ^{E}, an *Escherichia coli* heat-inducible sigma factor, is modulated by expression of outer membrane proteins. Gene. Dev., 7: 2618-2628.
177. Danese, PN; Silhavy, TJ (1998) CpxP, a stress-combative member of the Cpx regulon. J. Bacteriol., 180: 831-839.
178. Dartigalongue, C; Missiakas, D; Raina, S (2001) Characterization of the *Escherichia coli* σ^{E} regulon. J. Biol. Chem., 276: 20866-20875.
179. Missiakas, D; Betton, JM; Raina, S (1996) New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. Mol. Microbiol., 21: 871-884.

180. Lipinska, B; Zylicz, M; Georgopoulos, C (1990) The Htra (Degp) protein, essential for *Escherichia Coli* survival at high-temperatures, is an endopeptidase. J. Bacteriol., 172: 1791-1797.
181. Strauch, KL; Johnson, K; Beckwith, J (1989) Characterization of *degP*, a gene required for proteolysis in the cell envelope and essential for growth of *Escherichia coli* at high temperature. J. Bacteriol., 171: 2689-2696.
182. Kolmar, H; Waller, PR; Sauer, RT (1996) The DegP and DegQ periplasmic endoproteases of *Escherichia coli*: Specificity for cleavage sites and substrate conformation. J. Bacteriol., 178: 5925-5929.
183. Spiess, C; Beil, A; Ehrmann, M (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. Cell, 97: 339-347.
184. Kadokura, H; Kawasaki, H; Yoda, K; Yamasaki, M; Kitamoto, K (2001) Efficient export of alkaline phosphatase overexpressed from a multicopy plasmid requires *degP*, a gene encoding a periplasmic protease of *Escherichia coli*. J. Gen. Appl. Microbiol., 47: 133-141.
185. Chen, C; Snedecor, B; Nishihara, JC; Joly, JC; McFarland, N; Andersen, DC; Battersby, JE; Champion, KM (2004) High-level accumulation of a recombinant antibody fragment in the periplasm of *Escherichia coli* requires a triple-mutant (*degP prc spr*) host strain. Biotechnol. Bioeng., 85: 463-474.
186. CastilloKeller, M; Misra, R (2003) Protease-deficient DegP suppresses lethal effects of a mutant OmpC protein by its capture. J. Bacteriol., 185: 148-154.
187. Pan, K-L; Hsiao, H-C; Weng, C-L; Wu, M-S; Chou, CP (2003) Roles of DegP in prevention of protein misfolding in the periplasm upon overexpression of penicillin acylase in *Escherichia coli*. J. Bacteriol., 185: 3020-3030.
188. Lin, W-J; Huang, S-W; Chou, CP (2001) DegP coexpression minimizes inclusion body formation upon overproduction of recombinant penicillin acylase in *Escherichia coli*. Biotechnol. Bioeng., 73: 484-492.
189. Arie, JP; Sassoon, N; Betton, JM (2001) Chaperone function of FkpA, a heat shock prolyl isomerase, in the periplasm of *Escherichia coli*. Mol. Microbiol., 39: 199-210.
190. Ramm, K; Pluckthun, A (2000) The periplasmic *Escherichia coli* peptidylprolyl *cis*, *trans*-isomerase FkpA - II. Isomerase-independent chaperone activity *in vitro*. J. Biol. Chem., 275: 17106-17113.

191. Ramm, K; Pluckthun, A (2001) High enzymatic activity and chaperone function are mechanistically related features of the dimeric *E. coli* peptidyl-prolyl-isomerase FkpA. J. Mol. Biol., 310: 485-498.
192. Saul, FA; Arie, JP; Normand, BVL; Kahn, R; Betton, JM; Bentley, GA (2004) Structural and functional studies of FkpA from *Escherichia coli*, a *cis/trans* peptidyl-prolyl isomerase with chaperone activity. J. Mol. Biol., 335: 595-608.
193. Bothmann, H; Pluckthun, A (2000) The periplasmic *Escherichia coli* peptidylprolyl *cis, trans*-isomerase FkpA - I. Increased functional expression of antibody fragments with and without *cis*-prolines. J. Biol. Chem., 275: 17100-17105.
194. Zhang, Z; Song, LP; Fang, M; Wang, F; He, D; Zhao, R; Liu, J; Zhou, ZY; Yin, CC; Lin, Q; Huang, HL (2003) Production of soluble and functional engineered antibodies in *Escherichia coli* improved by FkpA. Biotechniques, 35: 1032-1042.
195. Wu, MS; Pan, KL; Chou, CP (2007) Effect of heat-shock proteins for relieving physiological stress and enhancing the production of penicillin acylase in *Escherichia coli*. Biotechnol. Bioeng., 96: 956-966.
196. Korndorfer, IP; Dommel, MK; Skerra, A (2004) Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture. Nat. Struct. Mol. Biol., 11: 1015-1020.
197. Schlapschy, M; Dommel, MK; Hadian, K; Fogarasi, M; Korndorfer, IP; Skerra, A (2004) The periplasmic *E. coli* chaperone Skp is a trimer in solution: biophysical and preliminary crystallographic characterization. J. Biol. Chem., 385: 137-143.
198. Schafer, U; Beck, K; Muller, M (1999) Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. J. Biol. Chem., 274: 24567-24574.
199. de Cock, H; Schafer, U; Potgeter, M; Demel, R; Muller, M; Tommassen, J (1999) Affinity of the periplasmic chaperone Skp of *Escherichia coli* for phospholipids, lipopolysaccharides and non-native outer membrane proteins - Role of Skp in the biogenesis of outer membrane protein. Eur. J. Biochem., 259: 96-103.
200. Bulieris, PV; Behrens, S; Holst, O; Kleinschmidt, JH (2003) Folding and insertion of the outer membrane protein OmpA is assisted by the chaperone Skp and by lipopolysaccharide. J. Biol. Chem., 278: 9092-9099.

201. Chen, R; Henning, U (1996) A periplasmic protein (Skp) of *Escherichia coli* selectively binds a class of outer membrane proteins. Mol. Mcriobiol., 19: 1287-1294.
202. Harms, N; Koningstein, G; Dontje, W; Muller, M; Oudega, B; Luirink, J; de Cock, H (2001) The early interaction of the outer membrane protein PhoE with the periplasmic chaperone Skp occurs at the cytoplasmic membrane. J. Biol. Chem., 276: 18804-18811.
203. Bothmann, H; Pluckthun, A (1998) Selection for a periplasmic factor improving phage display and functional periplasmic expression. Nat. Biotechnol., 16: 376-380.
204. Hayhurst, A; Harris, WJ (1999) *Escherichia coli* Skp chaperone coexpression improves solubility and phage display of single-chain antibody fragments. Protein Expr. Purif., 15: 336-343.
205. Mavrangelos, C; Thiel, M; Adamson, PJ; Millard, DJ; Nobbs, S; Zola, H; Nicholson, IC (2001) Increased yield and activity of soluble single-chain antibody fragments by combining high-level expression and the Skp periplasmic chaperonin. Protein Expr. Purif., 23: 289-295.
206. Sklar, JG; Wu, T; Kahne, D; Silhavy, TJ (2007) Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. Gene. Dev., 21: 2473-2484.
207. Behrens, S; Maier, R; Cock, Hd; X.Schmid, F; A.Gross, C (2001) The SurA periplasmic PPIase lacking its parvulin domains functions *in vivo* and has chaperone activity. EMBO J., 20: 285-294.
208. Lazar, SW; Kolter, R (1996) SurA assists the folding of *Escherichia coli* outer membrane proteins. J. Bacteriol., 178: 1770-1773.
209. Hennecke, G; Nolte, J; Volkmer-Engert, R; Schneider-Mergener, J; Behrens, S (2005) The periplasmic chaperone SurA exploits two features characteristic of integral outer membrane proteins for selective substrate recognition. J. Biol. Chem., 280: 23540-23548.
210. Rietsch, A; Beckwith, J (1998) The genetics of disulfide bond metabolism. Annu. Rev. Genet., 32: 59-94.
211. Sone, M; Akiyana, Y; Ito, K (1997) Differential *in vivo* roles played by DsbA and DsbC in the formation of protein disulfide bonds. J. Biol. Chem., 272: 10349-10352.
212. Kurokawa, Y; Yanagi, H; Yura, T (2001) Overproduction of bacterial protein disulfide isomerase (DsbC) and its modulator (DsbD) markedly enhances periplasmic production of human nerve growth factor in *Escherichia coli*. J. Biol. Chem., 276: 14393-14399.
213. Qiu, J; Swartz, JR; Georgiou, G (1998) Expression of active human tissue-type plasminogen activator in *Escherichia coli*. Appl. Environ. Microbiol., 64: 4891–4896.
214. Wülfing, C; Rappuoli, R (1997) Efficient production of heat-labile enterotoxin mutant proteins by overexpression of *dsbA* in a *degP*-deficient *Escherichia coli* strain. Arch. Microbiol., 167: 280-283.

215. Joly, JC; Leung, WS; Swartz, JR (1998) Overexpression of *Escherichia coli* oxidoreductases increases recombinant insulin-like growth factor-I accumulation. Proc. Nat. Acad. Sci. U. S. A., 95: 2773-2777.
216. Missiakas, D; Georgopoulos, C; Raina, S (1994) The *Escherichia coli dsbC (xprA)* gene decodes a periplasmic protein involved in disulfide bond formation. EMBO J., 13: 2013-2020.
217. Missiakas, D; Schwager, F; Raina, S (1995) Identification and characterization of a new disulfide isomerase-like protein (DsbD) in *Escherichia coli*. EMBO J., 14: 3415-3424.
218. Chen, J; Song, J-L; Zhang, S; Wang, Y; Cui, D-F; Wang, C-C (1999) Chaperone activity of DsbC. J. Biol. Chem., 274: 19601-19605.
219. Joly, JC; Swartz, JR (1994) Protein folding activities of *Escherichia coli* protein disulfide isomerase. Biochemistry, 33: 4231-4236.
220. Ollis, D; Cheah, E; Cygler, M; Dijkstra, B; Frolow, F; Franken, S; Harel, M; Remington, S; Silman, I; Schrag, J (1992) The alpha/beta hydrolase fold. Protein Eng., 5: 197-211.
221. Uppenberg, J; Hansen, MT; Patkar, S; Jones, TA (1994) The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica*. Structure, 2: 293-308.
222. Uppenberg, J; Öhrner, N; Norin, M; Hult, K; Kleywegt, GJ; Patkar, S; Waagen, V; Anthonsen, T; Jones., TA (1995) Crystallographic and molecular modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols. Biochemistry, 34: 16838-16851.
223. Secundo, F; Carrea, G; Soregaroli, C; Varinelli, D; Morrone, R (2001) Activity of different *Candida antarctica* lipase B formulations in organic solvents. Biotechnol. Bioeng., 73: 157-163.
224. Rotticci, D; Rotticci-Mulder, JC; Denman, S; Norin, T; Hult, K (2001) Improved enantioselectivity of a lipase by rational protein engineering. Chem Bio Chem, 2: 766-770.
225. Holmquist, M; Heffner, F; NORIN, T; Hult, K (1996) A structural basis for enantioselective inhibition of *Candida rugosa* lipase by long-chain aliphatic alcohols. Protein Sci., 5: 83-88.
226. Ottosson, J; Hult, K (2001) Influence of acyl chain length on the enantioselectivity of *Candida antarctica* lipase B and its thermodynamic components in kinetic resolution of sec-alcohols. J. Mol. Catal. B, 11: 1025-1028.
227. Suen, W-C; Zhang, N; Xiao, L; Madison, V; Zaks, A (2004) Improved activity and thermostability of *Candida antarctica* lipase B by DNA family shuffling. Protein Eng. Des. Sel., 17: 133-140.
228. Uppenberg, J; Patkar, S; T., B; Jones, TA (1994) Crystallization and preliminary X-ray studies of lipase B from *Candida antarctica*. J. Mol. Biol., 235: 790-792.

229. Raza, S; Fransson, L; Hult, K (2001) Enantioselectivity in *Candida antarctica* lipase B: A molecular dynamics study. Protein Sci., 10: 329-338.
230. Rotticci-Mulder, JC, *Expression and mutagenesis studies of Candida antarctica lipase B*, in *Department of Biotechnology*. 2003, Royal Institute of Technology AlbaNova University Centre: SE-106 91 Stockholm, Sweden.
231. Magnusson, A, *Rational redesign of Candida antarctica lipase B*, in *department of Biochemistry*. 2005, Royal Institute of Technology School of Biotechnology: Stockholm, Sweden.
232. Thomas, JG; Ayling, A; Baneyx, F (1997) Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli.* Appl. Biochem. Biotechnol., 66: 197-238.
233. Perez-Perez, J; Gutierrez, J (1995) An arabinose-inducible expression vector, pAR3, compatible with ColE1-derived plasmids. Gene, 158: 141-142.
234. Kok, RG; Christoffels, M; Vosman, B; Hellingwerf, KJ (1993) Growth-phase-dependent expression of the lipolytic system of Acinetobacter calcoaceticus BD413: cloning of a gene encoding one of the esterases. J. Gen. Microbiol., 139: 2329-2342.
235. Towbin, H; Staehelin, T; Gordon, J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Nat. Acad. Sci. U. S. A., 76: 4350-4354.
236. LaVallie, ER; McCoy, JM (1995) Gene fusion expression systems in *Escherichia coli.* Curr. Opin. Biotech., 6: 501-506.
237. Missiakas, D; Raina, S (1997) Protein folding in the bacterial periplasm. J. Bacteriol., 179: 2465-2471.
238. Prinz, WA; Aslund, F; Holmgren, A; Bechwith, J (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. J. Biol. Chem., 272: 15661-15667.
239. Mössner, E; Huber-Wunderlich, M; Rietsch, A; Beckwith, J; Glockshuber, R; Åslund, F (1999) Importance of redox potential for the *in vivo* function of the cytoplasmic disulfide reductant thioredoxin from *Escherichia coli*. J. Biol. Chem., 274: 25254-25259.
240. Ritz, D; Patel, H; Doan, B; Zheng, M; Aslund, F; Storz, G; Beckwith, J (2000) Thioredoxin 2 is involved in the oxidative stress response in *Escherichia coli*. J. Biol. Chem., 275: 2505-2512.
241. Tzvetkov, N; Breuer, P; Boteva, R (2006) Cysteine-free glutathione-S-transferase as a tool for thiol-specific labeling of proteins. Biotechniques, 40: 145-146.

242. Busso, D; Delagoutte-Busso, B; Moras, D (2005) Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli.* Anal. Biochem., 343: 313-321.
243. Messens, J; Collet, J-F; Belle, KV; Brosens, E; Loris, R; Wyns, L (2007) The oxidase DsbA folds a protein with a nonconsecutive disulfide. J. Biol. Chem., 282: 31302-31307.
244. Jeong, KJ; Lee, SY (2000) Secretory Production of Human Leptin in *Escherichia coli.* Biotechnol. Bioeng., 67: 398-407.
245. Kurokawa, Y; Yanagi, H; Yura, T (2000) Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in *Escherichia coli.* Appl, Environ. Microbiol., 66: 3960-3965.
246. Collins-Racie, LA; McColgan, JM; K, LG; DiBlasio-Smith, EA; McCoy, JM; LaVallie, ER (1995) Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA. Bio. Technol., 13: 982-987.
247. Winter, J; Neubauer, P; Glockshuber, R; Rudolph, R (2000) Increased production of human proinsulin in the periplasmic space of *Escherichia coli* by fusion to DsbA. J. Biotechnol., 84: 175-185.
248. Fan, C-f; Zeng, R-h; Sun, C-y; Mei, X-g; Wang, Y-f; Liu, Y (2004) Fusion of DsbA to the N-terminus of CTL chimeric epitope, F/M2:81-95, of respiratory syncytical virus prolongs protein- and virus-specific CTL responses in Balb/c mice. Vaccine, 23: 2869-2875.
249. Dong, H; Xu, X; Deng, M; Yu, H; Zhao, H; Song, H; Geng, Y (2007) Expression and bioactivity of recombinant segments of human perforin. Biochem. Cell Biol., 85: 203-208.
250. Zhang, Y; Olsen, DR; Nguyen, KB; Olson, PS; Rhodes, ET; Mascarenhas, D (1998) Expression of eukaryotic proteins in soluble form in *Escherichia coli.* Protein Expr. Purif., 12: 159-165.
251. Lloyd, JS; Bhambra, A; Murrell, JC; Dalton, H (1997) Inactivation of the regulatory protein B of soluble methane monooxygenase from *Methylococcus capsulatus* (Bath) by proteolysis can be overcome by a Gly to Gln modification. Eur. J. Biochem., 248: 72-79.
252. Lukacin, R; Groning, I; Schiltz, E; Britsch, L; Matern, U (2000) Purification of recombinant 3b-hydroxylase from Petunia hybrida and assignment of the primary site of proteolytic degradation. Arch. Biochem. Biophys., 375: 364-370.
253. Hong, YR; Mullaney, JM; Black, LW (1995) Protection from proteolysis using a T4:T7-RNAP phage expression-packaging-processing system. Gene, 162: 5-11.

254. Mironova, R; Niwa, T; Hayashi, H; Dimitrova, R; Ivanov, I (2001) Evidence for non-enzymatic glycosylation in *Escherichia coli*. Mol. Microbiol., 39: 1061-1068.
255. Gottesman, S; Wickner, S; Maurizi, MR (1997) Protein quality control: triage by chaperones and proteases. Gene. Dev., 11: 815-823.
256. Phillips, TA; Vanbogelen, RA; Neidhardt, FC (1984) *Lon* gene product of *Escherichia coli* is a heat-shock protein. J. Bacteriol., 159: 283-287.
257. Woodcock, DM; Crowther, PJ; Doherty, J; Jefferson, S; Decruz, E; Noyerweidner, M; Smith, SS; Michael, MZ; Graham, MW (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucleic Acids Research, 17: 3469-3478.
258. Schlapschy, M; Grimm, S; Skerra, A (2006) A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in *Escherichia coli*. Protein Eng. Des. Sel., 19: 385-390.
259. Guigueno, A; Belin, P; Boquet, P (1997) Defective export in *Escherichia coli* caused by DsbA'-PhoA hybrid proteins whose DsbA' domain cannot fold into a conformation resistant to periplasmic proteases. J. Bacteriol., 179: 3260-3269.
260. Misra, R; Castillokeller, M; Deng, M (2000) Overexpression of protease-deficient DegP(S210A) rescues the lethal phenotype of *Escherichia coli* OmpF assembly mutants in a *degP* background. J. Bacteriol., 182: 4882-4888.
261. Betton, JM; Boscus, D; Missiakas, D; Raina, S; Hofnung, M (1996) Probing the structural role of an alpha beta loop of maltose- binding protein by mutagenesis: Heat-shock induction by loop variants of the maltose-binding protein that form periplasmic inclusion bodies. J. Mol. Biol., 262: 140-150.
262. Wickner, W; Driessen, AJM; Hartl, F-U (1991) The enzymology of protein translocation across the *Escherichia coli* plasma membrane. Annu. Rev. Biochem., 60: 101-124.
263. Kaberdin, VR; Miczak, a; Jakobsen, JS; Lin-Chao, S; McDowall, KJ; Gabain, AV (1998) The endoribonucleolytic N-terminal half of *Escherichia coli* RNase E is evolutionarily conserved in *Synechocystis sp.* and other bacteria but not the C-terminal half, which is sufficient for degradosome assembly. Proc. Nat. Acad. Sci. U. S. A., 95: 11637-11642.
264. Razis, AFA; Ismail, EN; Hambali, Z; Abdullah, MNH; Ali, AM; Lila, MAM (2006) The periplasmic expression of recombinant human epidermal growth factor (hEGF) in *Escherichia coli*. Asia Pac. J. Mol. Biol., 14: 41-45.

265. Li, L; Kang, DG; Cha, HJ (2004) Functional display of foreign protein on surface of *Escherichia coli* using N-terminal domain of ice nucleation protein. Biotechnol. Bioeng., 85: 214-221.
266. Wentzel, A; Christmann, A; Adams, T; Kolmar, H (2001) Display of passenger proteins on the surface of *Escherichia coli* K-12 by the enterohemorrhagic *E. coli* intimin EaeA. J. Bacteriol., 183: 7273-7284.
267. Jose, J; Kramer, J; Klauser, T; Pohlner, J; Meyer, TF (1996) Absence of periplasmic DsbA oxidoreductase facilitates export of cysteine-containing passenger proteins to the *Escherichia coli* cell surface via the Iga_{β} autotransportor pathway. Gene, 178: 107-110.
268. Georgiou, G; Poetschke, HL; Stathopoulos, C; Francisco, JA (1993) Practical applications of engineering gram-negative bacterial cell surfaces. Trend. Biotechnol., 11: 6-10.
269. Rutherford, N; Mourez, M (2006) Surface display of proteins by gram-negative bacterial autotransporters. Microbial Cell Factories, 5: 22.
270. Matsumoto, T; Fukuda, H; Ueda, M; Tanaka, A; Kondo, A (2002) Construction of yeast strains with high cell surface lipase activity by using novel display systems based on the Flo1p flocculation functional domain. Appl. Environ. Microbiol., 68: 4517-4522.
271. Kiermer, V (2005) Mammalian cell put on a display. Nature, 2: 160-161.
272. Samuelson, P; Gunneriusson, E; Nygren, PA; Stahl, S (2002) Display of protein on bacteria. J. Biotechnol., 96: 129-154.
273. Jung, HC; Kwon, SJ; Pan, JG (2006) Display of a thermostable lipase on the surface of a solvent-resistant bacterium, *Pseudomonas putida* GM730, and its applications in whole-cell biocatalysis. BMC Biotechnol., 6: 23.
274. Wilhelm, S; Rosenau, F; Becker, S; Buest, S; Hausmann, S; Kolmar, H; Jaeger, KE (2007) Functional cell-surface display of a lipase-specific chaperone. ChemBioChem, 8: 55-60.
275. Sergeeva, A; Kolonin, MG; Molldrem, JJ; Pasqualini, R; Arap, W (2006) Display technologies: Application for the discovery of drug and gene delivery agents. Adv. Drug. Deliver. Rev., 58: 1622-1654.
276. Oertle, E, *Biodiesel production: bioprocess development of an artificial biofilm using bio-encapsulation of an E. coli strain genetically engineered for lipase cell-surface display*, in *Chemical Engineering*. 2008, University of Waterloo: Waterloo.
277. Skillman, KM; Barnard, TJ; Peterson, JH; Ghirlando, R; Bernstein, HD (2005) Efficient secretion of a folded protein domain by a monomeric bacterial autotransporter. Mol. Microbiol., 58: 945-58.

278. Veiga, E; de Lorenzo, V; Fernandez, LA (2004) Structural tolerance of bacterial autotransporters for folded passenger protein domains. Mol. Microbiol., 52: 1069-1080.
279. Palmer, BR; Marinus, MG (1994) The *dam* and *dcm* strains of *Escherichia coli* - a review. Gene, 143: 1-12.
280. Rose, RE (1988) The nucleotide sequence of pACYC177. Nucleic Acids Res., 16: 356.
281. Rosenau, F; Tommassen, J; Jaeger, K-E (2004) Lipase-specific foldases. ChemBioChem, 5: 152-161.
282. Jaeger, KE; Dijkstra, BW; Reetz, MT (1999) Bacterial biocatalysts: molucular biology, three-dimensional structures, and biotechnological applications of lipases. Annu. Rev. Microbiol., 53: 315-351.
283. Xu, YY; Du, W; Liu, DH; Zeng, J (2003) A novel enzymatic route for biodiesel production from renewable oils in a solvent-free medium. Biotechnol. Lett., 25: 1239-1241.
284. Neugnot, V; Moulin, G; Dubreucq, E; Bigey, F (2002) The lipase/acyltransferase from *Candida parapsilosis* - Molecular cloning and characterization of purified recombinant enzymes. Eur. J. Biochem., 269: 1734-1745.
285. Stemmer, W (1994) DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. Proc. Nat. Acad. Sci. U. S. A., 91: 10747-10751.
286. Yuan, L; Kurek, I; English, J; Keenan, R (2005) Laboratory-directed protein evolution. Microbiol. Mol. Biol. Rev., 69: 373-392.
287. Luedeking, R; Piret, EL (1959) A kinetic study of the lactic acid fermentation. Batch process at controlled pH. Biochem. Microbiol. Tech. Eng., 1: 393-412.
288. Hartley, JL; Temple, GF; Brasch, MA (2000) DNA cloning using *in vitro* site-specific recombination. Genome Research, 10: 1788-1795.