Advanced Genomic Engineering Strategy based on Recombineering Protocols to "Tailor" *Escherichia coli* Strains

by

Karan Sukhija

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Chemical Engineering

Waterloo, Ontario, Canada, 2011

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

A systematic approach based on bacteriophage λ (λ Red) and flippase-flippase recognition targets (FLP-FRT) recombinations was proposed for genomic engineering of *Escherichia coli*. For demonstration purposes, DNA operons containing heterologous genes (i.e. *pac* encoding *E. coli* penicillin acylase and *palB2* encoding *Pseudozyma antarctica* lipase B mutant) engineered with regulatory elements, such as strong/inducible promoters (i.e. P_{trc} and P_{araB}), operators, and ribosomal binding sites, were integrated into the *E. coli* genome at designated locations (i.e. *lacZYA*, *dbpA*, and *lacI-mhpR* loci) either as a gene replacement or gene insertion using various antibiotic selection markers (i.e. kanamycin and chloramphenicol) under various genetic backgrounds (i.e. HB101 and DH5α). The expression of the inserted foreign genes was subject to regulation using appropriate inducers [Isopropyl β-D-1-thiogalactopyranoside (IPTG) and arabinose] at tuneable concentrations. The developed approach has paved an effective way to "tailor" plasmid-free *E. coli* strains with desired genotypes suitable for various biotechnological applications, such as biomanufacturing and metabolic engineering.

Acknowledgements

I am indebted to my supervisors, Dr. C. P. Chou, Dr. Duane Chung and Dr. Murray Moo-Young for their encouragement, guidance, patience and support during my research and study at University of Waterloo. They have allowed me to learn at my own pace, and I am very grateful for their patience. I feel that their different perspectives have provided me with the insight in research that very few have. Their enthusiasm for research has motivated me to continue a career in this field.

I would like to thank to the members of the committee for my thesis: Dr. Marc Aucoin, Dr. Jeno Scharer, Dr. C. P. Chou, Dr. Duane Chung, and Dr. Murray Moo-Young for agreeing to serve as my committee members, as well as providing constructive comments on my thesis.

It was an honor to begin this journey with my colleague and close friend, Michael Pyne. He has provided me with valuable advice both on personal and professional levels. I am also grateful for the support of my labmates, Niju, Shreyas, Victoria, Kajan, Dr. Abedi, Steve, David, Val, Lu, Fred, Bahram and Saideh, as well as my co-op students, Saad, Yo Ha Na, and Bensen.

The support I have received from my friends over the last two years has been invaluable. I would like to thank Nicole, Nimesh, Fawaz, Sandeep, Scott and my EBC friends; Jackson, Lanky, Mihir, Modi, Munish, Murtaza, Niket, Rajan, Suraj, and Zamir for helping me keep my sanity during the more stressful periods.

This thesis would not have been possible without the unconditional love and support from my family. I am grateful to my sister, Gunit, for believing in me, and always encouraging me to reach my goals. Thank you, Mom and Dad, for the values you have taught me. Your guidance and prayers through out the years that has helped me achieve success.

This work was supported in part by Natural Sciences and Engineering Research Council (NSERC) and the Canada Research Chair (CRC) program of Canada.

Dedication

To my Parents

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Overview

Escherichia coli (E. coli) has been long recognized as a versatile host system for manufacturing of both ribosomal and non-ribosomal products primarily due to the availability of various powerful biotechnologies associated with this microorganism. The extensive development of genetic tools and protocols enables the manipulation and engineering of *E. coli* strains suitable for industrial applications. Among them, recombinant DNA technology, developed in the 1970s, has opened an avenue for introducing foreign genes via shuttle vectors, such as plasmids, into *E. coli* host cells for gene expression. The plasmid-based over-expression of genes in *E. coli* has proven to be useful for recombinant protein production (*1*). In addition to protein overproduction, recombinant DNA technology has been commonly used to graft foreign genes into *E. coli* host cells for metabolic engineering purposes, such as the production of non-natural metabolites (*2-4*). However, there are intrinsic limitations associated with these plasmid-based applications. For example, the energy required to maintain and propagate a plasmid comes from cellular resources of the host. Furthermore, high-copy number plasmids could lead to unnatural and deleterious dosing of the grafted genes, causing significant stress to host cells (*5*). Such energy related stress to the cell is generally referred to as metabolic burden. On the other hand, recombinant cells could be genetically unstable provided the grafted foreign genes are carried by a low-copy number plasmid. To address these technical issues, integrating foreign genes into the *E. coli* genome appears to be a plausible solution.

In addition to traditional protocols for manipulating the *E. coli* genome, such as random mutagenesis, conjugation, transduction, etc., consistent yet low-level interest has been shown in genome-based expression platforms (*6, 7*). However, early systems suffered several technical drawbacks and limitations, including labor-intensive, time-consuming, dependent upon specific mutant strains, only applicable to suicide or non-replicative plasmids, and difficult to target to a specific chromosomal locus. Examples include plasmid integration in replication-deficient *polA* mutants (*8*), random Tn*1545*-mediated plasmid insertion (*9*), chromosomal integration of linear substrates in *recBCD* (*10*), *recBC sbcB* (*11*), or *recD* (*12*) mutants, FLP recombination of suicide plasmids targeted to a single chromosomal FRT site (i.e. the FLIRT system) (*13*), and suicide plasmid integration and cointegrate resolution via I-*Sce*I meganuclease-stimulated double strand break (*14*).

Recently, the advent of genomic editing strategies based upon the recombination systems from the *Rac* prophage (i.e. RecET) (*15*) or the bacteriophage λ Red operon (i.e. Exo, Bet, and Gam) (*16, 17*) allow rapid, universal, and specific genomic manipulation. Basically, an exogenous DNA fragment carrying two homologous arms on both ends and containing an antibiotic marker gene flanked by two FRT sites is prepared by PCR-amplification of a defined region in a template vector. The PCR product is then transformed into *E. coli* cells and undergoes site-specific homologous recombination mediated by either RecET or λ Red proteins in order to replace the targeted gene. The antibiotic marker is subsequently removed from the genome by FLP-FRT recombination to generate an antibiotic marker-less mutant derivative. A comprehensive *E. coli* mutant library, known as the Keio collection (*18*); www.shigen.nig.ac.jp/ecoli/strain/top/top.jsp), was made by using the λ Red recombination protocol to inactivate various non-essential genes.

Almost all genomic engineering protocols have been developed for the purpose of inactivating chromosomal genes. The capacity for integrating heterologous genes into *E. coli* genome to extensively "tailor" *E. coli* strains for various biotechnological applications is worth exploring.

1.2 Research Objectives

The overall objectives of this thesis are as follows:

i) Demonstrate a systematic approach based on λ Red and FLP-FRT recombinations to integrate heterologous genes into the *E. coli* genome through either gene replacement or gene insertion.
ii) Demonstrate that the expression of the integrated genes can be tunable by controlling the inducer concentration in the medium.

1.3 Outline of thesis

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

Chapter 1 contains an introduction to the thesis, including an overview of plasmid-based expression of genes and the limitations. The hypothesis, objectives and the scope the thesis are also given in this chapter.

Chapter 2 reviews commonly used methods for the genetic engineering of *E. coli* for gene deletions, insertions and replacements. This chapter also includes a brief introduction of gene

expression systems and proteins used to demonstrate advanced recombineering method developed in this study.

Chapter 3 describes the methods used to construct various plasmids and strains used in this study, and the analytical methods used to confirm and characterize the mutant strains derived in this study.

Chapter 4 presents the results of the gene replacement and gene insertion experiments, as well as the results from studies on the tunable expression of these genes.

Chapter 5 summarizes the proposed gene knock-in method and discusses technical aspects associated with it.

Chapter 6 provides a conclusion for this report, and recommendations for future projects.

Chapter 2
Literature Review

2.1 *Escherichia coli*: The most widely used host for biotechnological and pharmaceutical applications

Numerous important factors must be considered before choosing a host for the heterologous expression of a protein, including cell growth characteristics, gene expression levels, post-translational modifications, and government regulations regarding therapeutic proteins (*19, 20*). Many well-established expression systems exist, including those derived from yeast (*21*, *22*), bacteria (*20, 23*, *24*), fungi (*25, 26*), insect cells (*27*) and mammalian cells (*28, 29*).

With a well-characterized genome, a number of cloning and expression vectors and mutant strains, and the ability to grow rapidly to a high density on minimal media, *E. coli* became a model organism in biology since its discovery in 1919 (*20*). In fact, the creation of the first recombinant DNA by Stanley Cohen and Herbert Boyer in 1973, which is considered as the birth of biotechnology by many, was first established in this bacterium (*30*). The production of human insulin as the first functional recombinant protein from *E. coli* in 1980 was another major milestone (*31, 32*). Since then, many therapeutic proteins have been produced by recombinant strains of *E. coli*. Between 2003 and 2006, 9 out of 31 biopharmaceuticals approved by the United States Environmental Protection Agency have been produced in *E. coli* (34). 60% of recombinant genes reported in journals catalogued in PubMed between 1995 and 2009 have been expressed in *E. coli* (35). Despite a few disadvantages, such as the lack of post-translational processing abilities in this bacterium required for many eukaryotic proteins, lack of protein secretion and very low disulfide bond formation within transcribed proteins (*20*), *E. coli* retains its popularity as one of the most versatile hosts for recombinant protein production (1).

The genetic engineering method studied here has been developed in *E. coli* because of its use as a model organism, and its application in many areas of biology.

2.2 *E. coli* expression vectors

Bacterial plasmids are double-stranded closed circular DNA molecules, either naturally occurring in bacteria or man-made. An expression vector is an artificial bacterial plasmid used for expressing a gene encoding a protein of interest within a bacterial cell. Man-made vectors consist of all features

required for the maintenance of the plasmid within the bacterial cell, and the expression of the gene of interest. These features are summarized in Table 2.1.

Table 2.1: Components of Expression Vectors

Expression vector component	Function	Example of component
Origin of replication (ori)	An ori is a sequence of DNA where replication is initiated. This sequence also determines the vector copy number in a cell. Cells with plasmids with low copy number ori, such as pMB1(*33*), typically have 25-50 plasmid copies per cell. Higher copy numbers, between 150-200 plasmid copies per cell, can be achieved using a high copy number ori such as the pUC ori (*34*).	pUC ori, pMB1 ori, pACYC ori, pSC101 ori
Selectable marker	Selectable markers are antibiotic resistance genes required for the maintenance of the plasmid in the cell. These genes also allow for the selection of cells containing the plasmid. Under selective conditions, only cells containing selectable markers can survive.	Ampicillin resistance (Ap^r), Chloramphenicol resistance (Cm^r), Kanamycin resistence (Kn^r)
Ribosomal Binding Site (RBS)	The ribosome binds to the RBS to initiate protein translation. This sequence is six base pairs long and is usually 8 nucleotides upstream of the start codon (AUG) in bacteria. In prokaryotes, the RBS is referred to as the Shine-Dalgarno sequence (*20*).	Consensus sequence of AGGAGG.
Promoter and regulatory elements	Promoters are short regulatory sequences found upstream of the RBS which bind the transcription complex to initiate transcription of a gene. Prokaryotic promoters consist of two short nucleotide sequences at 10 and 35 nucleotides upstream from the transcription start site, and these sequences are called -10 and -35 elements. The conserved sequences of the -10 and -35 elements are TATAAT and TTGACAT respectively, although only about three out of six base pairs are found in a given promoter (*20*). Promoters can be grouped into constitutive and inducible promoters. Constitutive promoters allow for continues transcription of a gene, whereas inducible promoters are activation by chemicals (such as Isopropyl β-D-1-thiogalactopyranoside or lactose) (*20, 24, 35, 36*), temperature change (*37*), or other physical factors. Ideal promoters for recombinant protein production should be	*Trc* promoter (P_{trc}), *araB* promoter (P_{araB}), *T7* promoter (P_{T7}), *lac* promoter (P_{lac})

	inducible, strong enough for high transcription rate, and have very little basal activity when uninduced. Other regulatory elements might include, operator regions, and genes encoding for repressors and transcription terminators. Repressors inhibit transcription in the absence of an inducer by binding to a sequence between the promoter and genes of the operon. This sequence is called the operator region (*24, 38, 39*). For example, the lacI repressor inhibits transcription from *lac*-derived promoters by binding to the operator sequence within this operon. Inducers cause an allosteric change in the repressor molecule such that the repressor is unable to bind to the operator sequences. Thus, inducers allow for transcription from these promoters. In the case for lac-derived promoters, the lacI repressor molecule is allosterically modified by IPTG or allolactose, which allows for transcription from these promoters (*40*). Transcription terminators are sequences at the end of a gene that end transcription. Two types of transcription terminators are available for *E. coli* vectors. With intrinsic transcription terminators, the transcribed mRNA forms a hairpin structure, disrupting the mRNA-DNA-RNA polymerase complex. The second type is the Rho-dependent terminators, where a RNA helicase protein called Rho factor disrupts the mRNA-DNA-RNA polymerase, thus ending transcription (*20*).	
Multiple Cloning Site (MCS)	Also called a polylinker, the MCS is a short sequence within the expression vector consisting of up to 20 single-cutting restriction sites. These restriction sites are used to sub clone the gene encoding the protein of interest into the vector. The MCS is usually downstream of the promoter region such that the promoter can drive the transcription of the sub cloned gene (*20*).	The MCS is composed of various restriction sites. Some of the commonly used restriction enzymes to sub clone a gene into the pUC19 MCS, for instance, are *EcoRI*, *HindIII*, *BamHI*, and *PstI*.

2.2.1 Recombinant protein production from pET expression vectors

Several expression vectors exist for recombinant protein production in *E. coli*. Of these, pET expression vectors based on the T7 promoter (P_{T7}) are few of the most commonly used expression vectors (*41, 42*). This expression system is used within host strains of *E. coli* with a chromosomally

integrated T7 RNA polymerase gene regulated by the *lacUV5* promoter. Upon induction by IPTG, the T7 RNA polymerase is expressed from the genome. The T7 RNA polymerase binds specifically to the P_{T7} region, and promotes high level transcription of the recombinant proteins. Up to 50% of the total cellular protein content is reached after only a few hours of induction by IPTG (36) (*20*, *42*).

It is necessary for the *E. coli* host strain to contain a chromosomally integrated gene encoding the T7 RNA polymerase gene (*20*). The DE3 strains of *E. coli*, such as the *E. coli* BL21 (DE3) strains, are commonly used with this system since they contain the λ prophage carrying the T7 RNA polymerase and *lacI*^{q} genes. In this study, bacteriophage λ based genomic engineering was attempted in *E. coli* BL21 (DE3). However, no transformants were derived via electroporation or P1 phage transduction for unknown reasons. Thus, other expression systems and compatible host strains were explored.

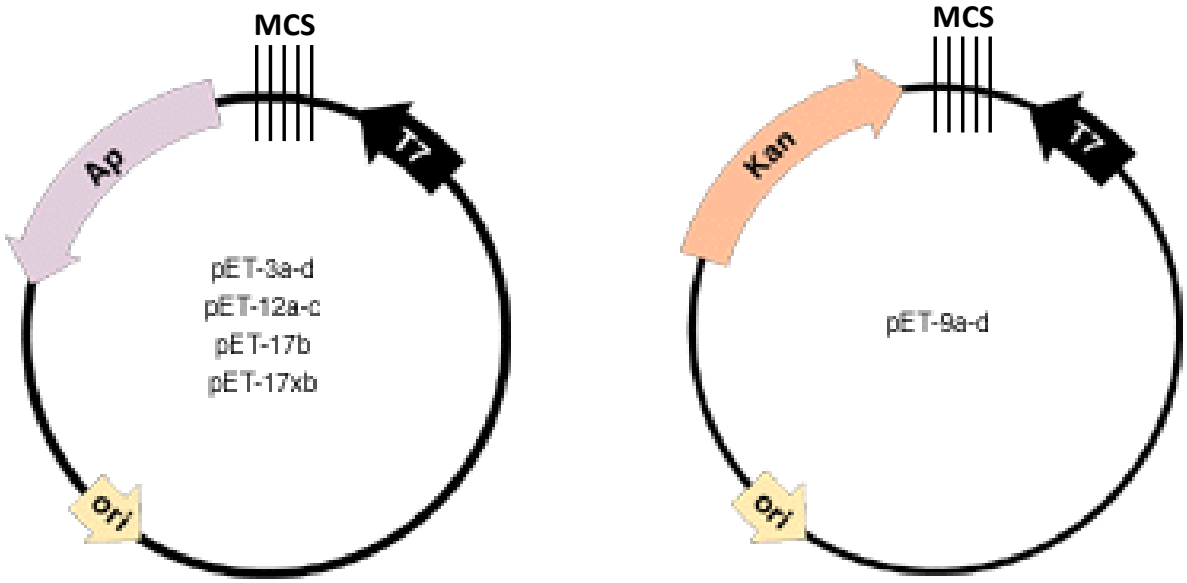


Figure 2.1 pET vectors. The gene encoding protein of interest is sub cloned into the multiple cloning site (MCS) downstream of the T7 promoter (T7). The antibiotic resistance genes, either ampicillin resistance gene (Ap) or Kanamycin resistence gene (Kn), are used for selection of *E. coli* transformants carrying these plasmids. The pBR322 origin of replication (ori) is used in these plasmids. Images are modified from www.novagen.com/sharedimages/technicalliterature/7_tb055.pdf.

2.2.2 Recombinant protein production using regulatory elements from *E. coli lac* operon

This operon consists of three genes, *lacZ, lacY* and *lacA* (*lacZYA*) that encode for β-galactosidase, lactose permease and transacetylase respectively enzymes necessary for the utilization of lactose. The P_{lac} and operator genes are directly upstream of the *lacZYA* genes. The *lacI* gene, encoding for the transcriptional repressor molecule, is about 120 nucleotides upstream of the *lacZYA* genes (Figure 2.2) (*20, 24, 40, 43*).

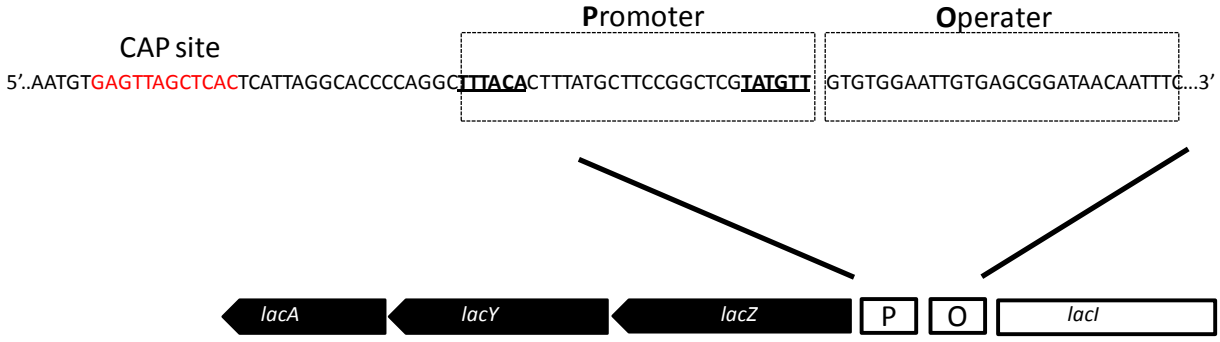


Figure 2.2 *E. coli lac* operon. The *lacI* repressor, promoter (P) and operator (O) genes are found upstream of the *lacZYA* genes. Binding of the CAP molecule to the CAP site increases in the presence of glucose, driving the transcription of the *lacZYA* genes by inducing the *lac* promoter.

In the absence of lactose, the *lacI* repressor binds to the operator sequence, inhibiting the RNA polymerase from binding to the promoter. The lactose metabolite allolactose, IPTG, and other β-galactosides allosterically change the conformation of the repressor molecule such that it is unable to bind to the operator. In the absence of this repressor inhibition, the RNA polymerase binds to the operator sequence and initiates transcription. Thus, lactose and IPTG are able to induce expression of the *lacZYA* genes (*20, 24, 40, 43*).

Expression systems based on the *lac* operon take advantage of the *lac* regulatory apparatus; the promoter (including the operator sequence) and the *lacI* repressor genes. A gene encoding a protein of interest is sub-cloned in an expression vector carrying these regulatory genes, with the P_{lac} upstream of the protein encoding gene, and transformed into an *E. coli* host strain. Thus, high level expression of the recombinant protein is induced by adding a β-galactoside in the media. The most common inducer used for this system is IPTG since it cannot be cleaved by β-galactosidase, is very stable in cells and induces the P_{lac} at very low concentrations (*24*).

The wild-type *lac* operon is also repressed by glucose by the use of the catabolite gene activator (CAP), a protein encoded by the *crp* gene. In the absence of glucose, cyclic AMP (cAMP) is produced. CAP binds to cAMP molecules and the CAP-binding site (Figure 2.2) to enhance expression of the *lac* operon. To facilitate stronger expression of recombinant proteins from expression vectors, the *lacUV5* mutant promoter (P_{lacUV5}) is used. Binding of CAP to the CAP site has relatively little effect on transcription with the *lacUV5* promoter. Thus increased protein expression is achieved in rich fermentation media that contain glucose and other carbon sources that would otherwise inhibit transcription via CAP (*24*).

Significant level of basal expression is still present in the absence of an inducer from both P_{lac} and P_{lacUV5}. To overcome this "leaky" expression, the *lacI* repressor gene can be over-expressed. Alternatively, the $lacI^q$ repressor can be used instead of the wild-type repressor to similarly fine-tune regulation. The $lacI^q$ repressor binds to the operator more tightly than the wild-type repressor, thus also decreasing the basal level of expression. Both methods are equally effective, decreasing basal level expression up to ten folds (*20, 24*).

The expression systems derived from the *lac* operon can be used in virtually any *E. coli* host strain. These systems has been studied extensively and optimized for heterologous protein expression. However, the expression level is still lower relative to other prokaryotic promoters. It has been proposed that this lower expression level is due to the differences between the sequences of the -35 (TTTACA) and -10 (TATGTT) elements of $P_{lac,}$ from the *E. coli* consensus sequences for these elements (TTGACA and TATATT, respectively). More specifically, the G/T difference in the -35 region has detrimental effects on protein expression (*20*). This problem has been alleviated by the engineering of *trp*/*lac*-hybrid promoters, which are reviewed in section 2.2.4.

2.2.3 Recombinant protein production using regulatory elements from *E. coli trp* operon

Similar to *lac*-based expression vectors, expression vectors based on the inducible genetic elements from the *trp* operon have also been design. The *trp* operon encodes for five enzymes that catalyze the biosynthesis of tryptophan (*43*). This operon was the first repressible operon discovered. Whereas P_{lac} is activated by chemicals (allolactose, IPTG), the *trp* promoter (P_{trp}) is repressed by tryptophan. The P_{trp} system includes the repressor encoded by the *trpR* gene. Repression occurs when tryptophan binds to the repressor molecule, which in turn binds to the operator. Thus, addition of tryptophan into the fermentation media effectively blocks transcription (*38, 44*). This promoter can be used for over-expression of heterologous protein in any *E. coli* host background.

2.2.4 Recombinant protein production from *trp/lac*-hybrid promoters

As mentioned in section 2.2.2, expression vectors based on the *lac*-operon yield low levels of recombinant protein expression compared to other expression systems primarily due to the difference between the P_{lac} -35 element sequence and the *E. coli* conserved -35 sequence. To overcome this disadvantage, the P_{trp}/P_{lacUV5}-hybrid promoters were designed.

The *tac* (P_{tac}) and *trc* (P_{trc}) promoters consist of the -35 element of P_{trp}, and -10 element and operator sequence of $P_{lacUV5.}$ Thus, these hybrid promoters maintain the consensus sequence for -35 and -10 elements of other *E. coli* promoters. Expression vectors carrying these promoters also contain of the $lacI^q$ gene since these promoters are repressed by the $lacI^q$ repressor molecule similar to P_{lac} and P_{lacUV5}; therefore, these hybrid promoters can be easily induced by IPTG as well (*36*, *45*, *46*).

The only difference between the two hybrid promoters is the consensus sequence between -35 and -10 elements of each promoter (*39*, *46*). This sequence is 16 nucleotides long in P_{tac} , which is one nucleotide less than that of P_{trc}. However, expression levels of genes by P_{tac} and P_{trc} are similar; - approximately 11 folds higher than P_{lacUV5} expression level, and approximately 3 folds higher than P_{trp} expression level (*36*).

The plasmid pTrc99A (Figure 2.3) and its derivatives consist of the P_{trc} expression system (Table 3.1).

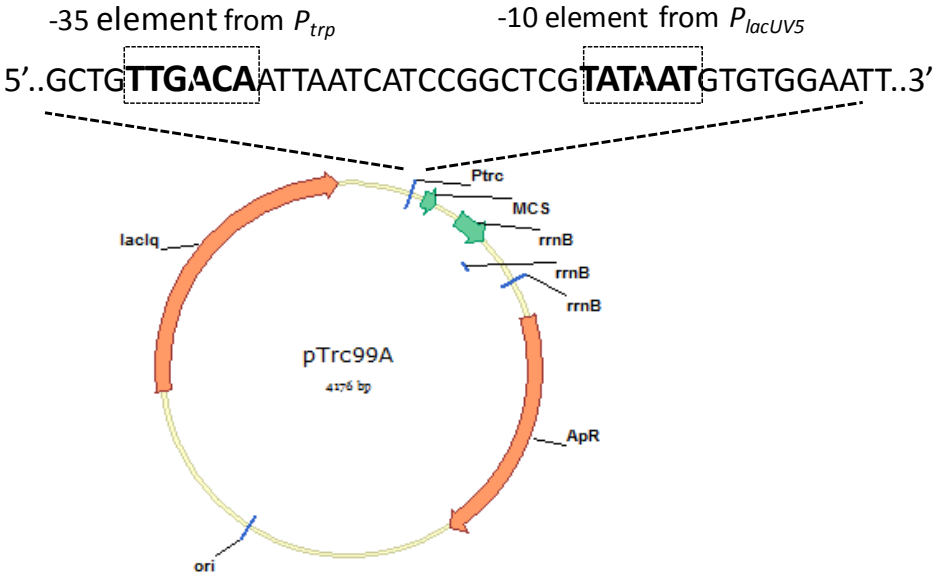


Figure 2.3 pTrc99a plasmid. The vector contains a multiple cloning site (MCS) downstream of the *trc* promoter (Ptrc), a pBR322 origin of replication (ori), *lacI*^{q} repressor gene, an ampicillin resistance gene (ApR), and multiple rho-dependent transcription terminators (rrnB). The *trc* promoter consists of the -35 element of *trp* promoter and -10 element of *lac* promoter.

2.2.5 Recombinant protein production using regulatory elements from *Salmonella typhimurium araBAD* operon

Promoters that respond to chemical inducers other than β-galactosides are necessary for the selective co-expression of multiple heterologous genes from a single host strain. The pAR3 plasmid is a gene expression vector bearing arabinose-inducible genetic elements from the arabinose operon (*araBAD* operon*)* of *Salmonella typhimurium* (*S. typhimurium*) (*35, 43*). This vector allows for differential co-synthesis of enzymes in a single *E. coli* host when used in combination with an IPTG-inducible promoter-based expression vector.

The pAR3 vector was engineered by sub-cloning the *araC* repressor gene, the *araB* promoter (P_{araB}) and operator sequences, and the *araB* ATG start codon in the MCS of the pACYC184 vector (Figure 2.4). In *S. typhimurium* cells, these genetic elements induce the expression of three genes essential for arabinose catabolism in the presence of arabinose; *araB, araA* and *araD* encoding for ribulokinase, arabinose isomerase and ribulose 5-phosphate 4-epimerase, respectively (*43*). Genetic analysis and DNA protection of *araC* repressor protein and gene from *E. coli* and *S. typhimurium* suggest that the *araC* repressor protein acts as both the repressor and activator of transcription of the *araBAD* genes. The repressor molecule binds to two positions on the *araBAD* operator in the presence of arabinose, facilitating the initiation of transcription by RNA polymerase. The repressor molecule binds to two different positions on the *araB* operator in the absence of arabinose, catalyzing the formation of a DNA loop structure to efficiently repressing the transcription of the *araBAD* genes (*47, 48*).

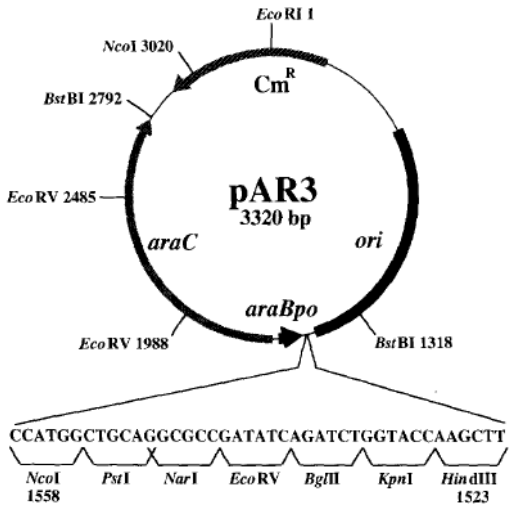


Figure 2.4 pAR3 expression vector. The *araC* repressor gene and *araB* promoter and operator sequences (*araBpo*) from the *S. typhimurium* genome were sub-cloned into the *BstBI* and *NcoI* sites of pACYC184 plasmid to yield the expression plasmid pAR3. This vector also has a pACYC184 origin of replication (ori) and chloramphenicol resistance gene (Cm^R).

2.3 Commonly used genetic engineering strategies

Genetic engineering refers to the manipulation of an organism's genome for the synthesis of economically valuable bio-products, change of hereditary traits, or both. Several innovative strategies have been created for genetic engineering in *E. coli (6, 9, 10, 14, 16, 49-54)*. This section reviews three recently devised and commonly used strategies, and the advantages and disadvantages associated with them.

2.3.1 Transposon-mediated insertions: Tn*5* Transposon technology

Transposons are a type of mobile genetic elements. They are relatively short pieces of DNA that move or replicate by inserting themselves into other pieces of DNA in a bacterial genome. The Tn*5* transposon consists of an antibiotic resistance gene flanked by sequences of terminal repeats and genes encoding transposition functions. Insertional mutagenesis can be achieved by electroporations of plasmids carrying the *Tn5* transposon.

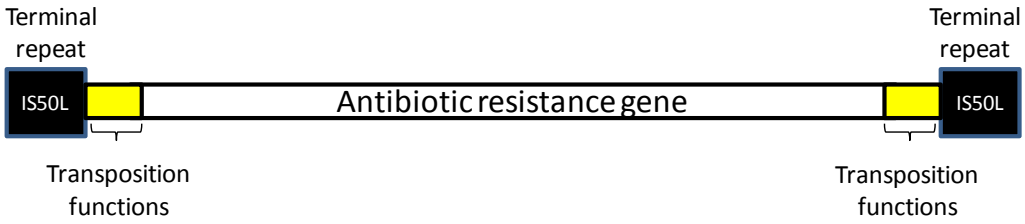


Figure 2.5 A schematic representation of the Tn*5* transposon. The genes encoding the antibiotic resistence gene and transposition functions are shown. IS50L and IS50R are insertion sequences.

Transposons inactivate genes by insertions, thus knocking out the gene. There are several advantages of transposon mutagenesis over other mutagenesis techniques. Bacterial transposons cannot insert into human DNA, making them safer than chemical and ionizing radiation. Although transposition frequency is low, insertional mutagenesis can be selected for by the antibiotic due to the antibiotic resistence gene within the transposon. Moreover, because of the low transposition frequency, transposons usually only lead to one mutation per cell.

Despite the advantages, there are many aspects the limit the use of this technology. Transposons randomly insert themselves into genes. Thus, directed-mutagenesis is not possible using the Tn*5* transposon technology. Also, the transposon can move into other places of the genome after integration into the first site. Finally, this technology is only limited to gene knock outs. Often, insertion of a transposon element in a gene downregulates the expression of downstream genes as well.

2.3.2 FLP recombinase-mediated gene insertions: The FLIRT system

The FLP-mediated DNA Integration and Rearrangements at prearranged genomic Targets (FLIRT) system is based on the flippase (FLP) system from the 2µ circle, a 6.3 kb natural plasmid found in the nuclease of most *Saccharomyces cerevisiae* (*S. cerevisiae*) strains (*13*). In *S. cerevisiae*, the FLP protein recognizes two Flippase Recognition Targets (FRT sites) on the genome, and removes any DNA sequence between these sites by catalyzing homologous recombination using these sites. This recombination removes any genes within that DNA sequence, and leaves behind one FRT site at the site of excision (*13*).

The FLIRT system allows for gene integrations in the *E. coli* genome by employing the yeast FLP/FRT recombination mechanism in *E. coli*. First, FRT sites are introduced at random positions in the *E. coli* genome by a Tn*5* transposon technology. A "suicide" vector containing an FRT site, selectable marker, origin of replication and an exogenous gene of interest is introduced into an *E .coli* strain by electroporation. Regulated expression of the FLP gene from another compatible expression vector provides the FLP protein that mediates the homologous recombination between the FLP sites on the *E. coli* chromosome and the cloning vector, efficiently integrating the suicide vector (including the exogenous genes) into the genome (*13*).

Although the FLIRT method yields mutant *E. coli* strains at relatively high efficiencies, a few drawbacks limit the use of this system. First, integration of antibiotic markers into the chromosome along with the exogenous gene is required to select for desired recombinants. Since there are very few commonly used antibiotic resistance markers used for *E. coli*, this limits the number of genes that can be integrated in the genome. Second, this method is cannot be used for integrating genes at pre-selected sites in the genome since the FRT sites are introduced at random sites using the Tn*5* transposon system. Finally, this long genetic engineering protocol can only be used to insert DNA sequences into the genome. Gene deletion and replacements are not possible with the FLIRT method.

2.3.3 λ-Red mediated gene deletions: Recombineering

Recombineering (recombination-mediated genomic engineering) (*55*) is a popular method used for gene deletion in *E. coli* based on homologous recombination using recombination proteins from the λ phage. The λ-Red genes encoding the three recombination proteins, exo, bet, and gam, are expressed from an arabinose-inducible promoter from an expression plasmid (the pKD6 vector, see Table 3.1). A double-stranded linear DNA fragment, PCR-amplified from template plasmids pKD3 or pKD4 (see Table 3.1), containing sufficient 5' and 3' end homology to the target chromosomal DNA sequence, and an FRT-flanked antibiotic resistance gene, is first transformed into the desired *E. coli* strain bearing the pKD46 plasmid. Induction of the λ-Red genes by arabinose from pKD6 produces the three recombination proteins in the strain. The exo protein is a 5'-3' exonuclease that creates single-stranded overhangs on the linear DNA (*56*). The gam protein prevents degradation of the linear DNA by inhibiting the *E. coli* exonuclease V protein (encoded by the *recBCD*) (*57, 58*) . The bet protein protects these overhangs and directly assists in homologous recombination, efficiently replacing the target gene with the linear DNA sequence (*59*). The mutant strain can be cured of the

pKD46 plasmid by increasing the culturing temperature to above 32ºC since this plasmid contains a temperature sensitive replication of origin (*sc101* ori) (*16*). This final step yields plasmid-free strains of *E. coli* with the desired gene deletion.

Apart from genomic engineering, recombineering has also been used to modify *E. coli* vectors and Bacterial Artificial Chromosomes (BACs) *in vivo* (*60, 61*). Arabinose-inducible expression of the FLP gene from another temperature sensitive plasmid can be used for the FRT/FLP-mediated removal of the antibiotic resistance gene from the integrated sequence to yield marker-less mutant strains (*16*). Removal of the marker gene allows for multiple consecutive gene deletions in *E. coli.*

Recombineering is a versatile tool for deleting DNA sequences from the *E. coli* genome. Other than in the Gene Gorging method (Section 2.3.4), the recombineering method has not been further developed for gene insertions or replacements in *E. coli*, which limits the use of this method for genomic engineering.

A collection *E. coli* K12 strains consisting of single-gene deletions of all non-essential genes has been constructed using this method (*18*). This collection, termed "Keio collection", is available at The Coli Genetic Stock Center at Yale University, New Haven, Connecticut.

2.3.4 I-*SceI* meganuclease-mediated gene replacement: The Gene Gorging method

The gene gorging method is based on the λ-Red recombination proteins (see section 2.3.3) and I-*Sce* endonuclease. In this method, two plasmids are introduced into the host strain: 1) a donor plasmid contains the exogenous gene with 5' and 3' sequences homologous to the target gene and an I-*SceI* endonuclease site, and 2) a recombineering plasmid, designated pACBSR, carrying the λ-Red and I-*SceI* endonuclease genes under the control of P_{araB}. Upon arabinose induction, the I-*SceI* cleaves the donor plasmid, providing the double-stranded linear DNA substrate for the λ-Red proteins. Homologous recombination catalyzed by the λ-Red proteins replaces the target DNA sequence on the chromosome with the exogenous DNA in the linearized donor plasmid. The frequency of integration is relatively high (1-15%) because of presence of multiple copies of linear DNA present in the cells (*62*). Due to this high integration frequency, simple screening methods such as colony PCR can be used for isolation of desired recombinants, eliminating the need for integration of selection markers.

Although this method yields genetically engineered *E. coli* strains without antibiotic resistance genes, the integration frequency is not very consistent. Moreover, this method has had limited or no success in pathogenic and laboratory *E. coli* strains (*53*).

Chapter 3
Materials and Methods

3.1 Bacteria and plasmids

The bacterial strains and plasmids used in this study are listed in Table 3.1. *Pfu* polymerase, restriction enzymes and Quick Ligation kits were purchased from New England Biolabs (Beverly, MA). Plasmid DNA purification, gel extraction and PCR purification were performed using spin column kits purchased from Bio Basic Inc. (Markham, ON). The Micropulser from Bio-Rad (Hercules, CA) was used for plasmid and linear DNA transformation. L-arabinose was purchased from Sigma-Aldrich (St. Louis, MO). All bacterial cultures were grown in lysogeny broth (LB) (5 g/l NaCl, 5 g/l Bacto yeast extract and 10 g/l Bacto tryptone). Media and agar plates were supplemented with 100 ug/ml ampicillin (Ap), 10 ug/ml chloramphenicol (Cm), or 25 ug/ml kanamycin (Kn) resistant transformants when necessary.

Plasmids pTrc100catPAC and pAr4KnPalb were constructed for use as template plasmids for Polymerase Chain Reaction (PCR) (Figure 3.1). The FRT-flanked chloramphenicol resistant (Cm^R) gene was PCR-amplified from pKD3 (*16*) with the *HindIII*-Cmcass.F and *PstI*-Cmcass.R primers (Table 3.1). The amplified DNA fragment was digested with *HindIII* and *PstI* and ligated to the multiple cloning site (MCS) in the similarly digested pTrc99A (*45*) to form pTrc100cat. The *pac* gene was PCR-amplified from pTrcKnPAC2902 (*63*) with the *pac*-Fw and *pac*-Rv primers, and the fragment was digested with *EcoRI* and *PstI* and ligated in the MCS of the similarly digested pTrc100cat to form pTrc100catPAC. Similarly, the FRT-flanked kanamycin resistance (Kn^R) gene was PCR-amplified from pKD4 (*16*). The PCR fragment was digested with *HindIII* and *PstI* and ligated into the MCS of similarly digested pAR3 (*35*) to form pAr4Kn. The *palB2* gene was PCR-amplified from the pETGM2 (*64*), digested with *NdeI* and *KpnI* and ligated into the MCS of similarly digested pAr4Kn to form pAr4KnPalb.

Table 3.1: Strains, Plasmids and Oligonucleotides

Strain	Relevant Genotype	Source & Reference
E. coli		
BL21(DE3)	F⁻ *ompT dcm lon hsdS*$_B$ (r_B^-, m_B^-) *gal λ*(DE3[*lacI ind1 sam7 nin*5 *lacUV5-T7* gene 1])	Lab stock, (*65*)
DH5α	F'*(ϕ80 dlac Δ(lacZ)M15) Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (r_K^-, m_K^+) phoA supE44 lambda-thi-1 gyrA96 relA1*)	Lab stock, (*66*)
DH5αPalB2KnR	DH5α *Δ (dbpA)* P$_{araB}$::*palB2* FRT-*Kn*^{R}-FRT	This study
HB101	F- ¢*(gpt-proA)62 leuB6 glnV44 ara-14 galK2 lacY1* ¢*(mcrC-mrr) rpsL20 xyl-5 mtl-1 recA13*	This study, (*67*)
HB101PAC	HB101 *Δ(lacZYA)* P$_{trc}$*::pac*	This study
HB101PACCmR	HB101 *Δ(lacZYA)* P$_{trc}$*::pac* FRT-*Cm*^{R}-FRT	This study
HB101PACins	HB101 P$_{trc}$*::pac* FRT-*Cm*^{R}-FRT	
HB101PACPalB2	HB101 *Δ(lacZYA)* P$_{trc}$*::pac* P$_{araB}$::*palB2*	This study
HB101PACPalB2KnR	HB101 *Δ(lacZYA)* P$_{trc}$*::pac* P$_{araB}$::*palB2* FRT-*Kn*^{R}-FRT	This study
S. marscescens		
ATCC27117	Penicillin-G resistant, 6-APA-sensitive	ATCC
Plasmid		
pAR3	Expression vector, pACYC184 ori, *P$_{araB}$*, *Cm*^{r}	(*35*)
pAR4Kn	Template vector, pACYC184 ori, *P$_{araB}$*,, *Cm*^{r}, FRT-*Kn*^{R}-FRT	This study
pAR4KnPalB2	Template vector, pACYC184 ori, *P$_{araB}$*::*palB2*, *Cm*^{r}, FRT-*Kn*^{R}-FRT	This study
pCP20	Expression vector**,** pSC101 ori, *P$_{araB}$*:: *flippase***,** *Ap*^{r}	(*16*)
pETGM-2	Expression vector, pBR322 ori, *P$_{T7}$*::*palB2*, *Ap*^{r}	(*64*)
pKD3	Template vector, pR6K ori, *Ap*^{r}, FRT-*Cm*^{R}-FRT	(*16*)
pKD4	Template vector, pR6K ori, *Cm*^{r}, FRT-*Kn*^{R}-FRT	(*16*)
pKD46	Expression vectors, pSC101 ori, *P$_{araB}$*:: γ β exo, *Ap*^{r}	(*16*)
pTrc100cat	Template vector, pBR322 ori, *P$_{trc}$*, *Ap*^{r}, FRT-flanked *Cm*^{r}	This study
pTrc100catPAC	Template vector, pBR322 ori, *P$_{trc}$* :: *pac*, *Ap*^{r}, FRT-*Cm*^{R}-	This study

	FRT	
pTrcKnPAC2902	Expression vector, pBR322 ori, P_{trc}:: *pac*, Ap^r, Kn^r	(*16*)

Oligonucleotides	
cynX-Fw	5’-CGCTTGTTCCTGCGCTTTGTTCAT-3’
fnrS-Fw	5’-TCTTGCAAGTGAATGCAACGTCAAGCGA-3’
H6as-P2as.insertion	5’-cttttcgcggtatggcatgatagcgcccggaagagagtcaaGGGAATAAGGGCGACACGG-3’
H1as-P1s.*lacZYA*.KO	5’-atgcagctggcacgacaggtttcccgactggaaagccCCAATACGCAAACCGCCTC-3’
H5s-P1s.insertion	5’-catcgtataacgttactggtttcacattcaccaccctgaaCCCCATACGCAAACCGCCTC-3’
H3s-P3s.*dbpA*.KO	5’-gctaactttgcgtcgatgaccacgagaatagattgttttgGTCTGCGCGTAATCTCTTGC-3’
H4as-P4as.*dbpA*.KO	5’-gcacccggcaattcaacatttcattattttaataacCCGGCAATAGCGGATCATTT-3’
H2s-P2as.*lacZYA*.KO	5’-ccagccgccacgacgtttggtggaatgtcttttgtgGGGAATAAGGGCGACACGGA-3’
HindIII-Cmcass.F	5’-AAGCTTGTGTAGGCTGGAGCTGCTTC-3’
KpnI-KncassR	5’-GGTACCTGGGAATTGGCCATGGTCG-3’
*KpnI-PalB2*F	5’-GGTACCGATATGAAATACCTGCTGCCGACCG-3’
lacI.mhpR seqF	5’-TATCAACTGGCACGGGAACCGTTA-3’
lacI.mhpR seqR	5’-ACGCGGTTGGGAATGTAATTCAGC-3’
lacI-Rv	5’-ATATCCCGCCGTTAACCACCATCA-3’
*NdeI-PalB2*R	5’-CATATGTTCCTTTCGGGCTTTGTTAGCAGCC-3’
pac-Fw	5’-TACACCATTTGCCTATCCTGGGCT-3’
pac-Rv	5’-TCCTGACGACTGGTTTGGCGAATA-3’
PstI-Cmcass.R	5’-CTGCAGATGGGAATTAGGCATGGTCC -3’
ttcA-Rv	5’-TCAAAGGCATCACCCACGGTTCTGAAGT-3’

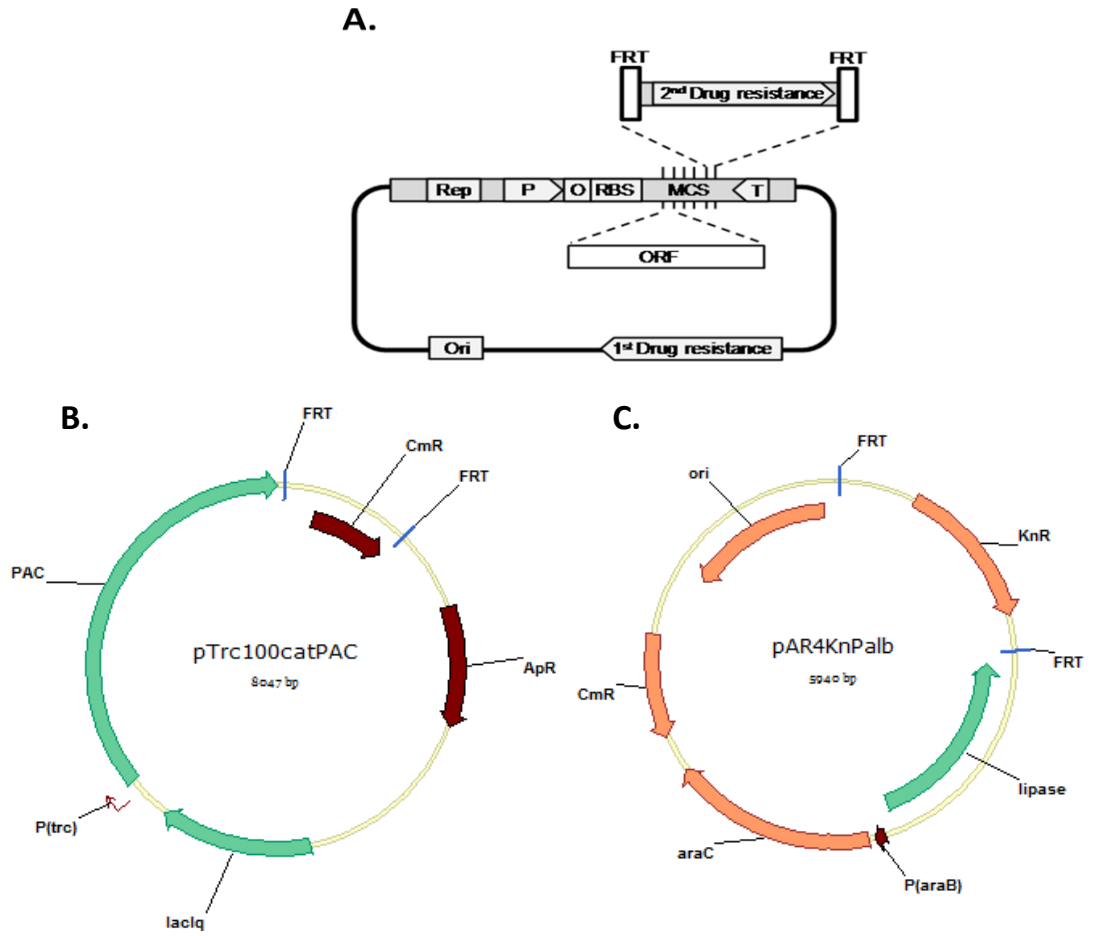


Figure 3.1 Template plasmids. A) Most expression or cloning vectors can be used to derive the template vector. The open reading frame(s) (ORF) of the gene(s) of interest and the antibiotic marker gene (2nd antibiotic marker) flanked with the FRT sites can be cloned into the multiple cloning sites (MCS). Specialized gene regulatory elements, such as an activator or repressor (Rep), promoter (P), operator (O), ribosome binding site (RBS) and terminator (T), can be adopted to regulate the expression of the heterologous gene(s). Alternatively, native regulatory elements of the heterologous gene(s) can be used. B) Plasmid pTrc100catPAC was derived from pTrc99a plasmid. The template plasmid consists of the *trc* promoter (P(trc)) regulated PAC gene (PAC), the repressor gene (lacIq), ampicillin resistance gene (ApR), and FRT sites flanked chloramphenicol resistance gene (CmR). C) Plasmid pAR4KnPalb was derived from pAR3 plasmid. pAR4KnPalb consists of the *araB* promoter (P(araB)) regulated lipase gene (lipase), chloramphenicol resistance gene (CmR), and FRT sites flanked kanamycin resistance gene (KnR).

All strains were constructed using the λ Red and FLP-FRT recombineering protocols reported previously (*16*). The large linear DNA inserts used in the recombineering experiments were amplified from the template plasmids of pTrc100catPAC and pAr4KnPalb using touchdown PCR (*68*), followed by *DpnI* digestion for 2 h to remove template plasmids from the reaction mixure. H2s-P2as.*lacZYA*.KO and H1as-P1s.*lacZYA*.KO primers were used for PCR-amplification of the *pac* operon from pTrc100catPAC. Similarly, H2s-P2s.*dbpA*.KO and H2as-P2as.*dbpA*.KO primers were used to PCR-amplify the *palB2* operon from pAr4KnPalb. HB101PAC is an *E. coli* mutant derived from HB101 by replacing the chromosomal *lacZYA* genes with the *pac* gene. The HB101PACPalB2 and DH5αPalB2 strains were similarly engineered, replacing the chromosomal *dbpA* gene with the *P. antarctica palB2* in HB101PAC and DH5α, respectively. The FLP-FRT recombination was used to remove the antibiotic marker genes on the genome. To construct HB101PACins, the *pac* gene was inserted between the *lacI* and *mhpR* genes on the HB101 genome using the same recombineering protocol.

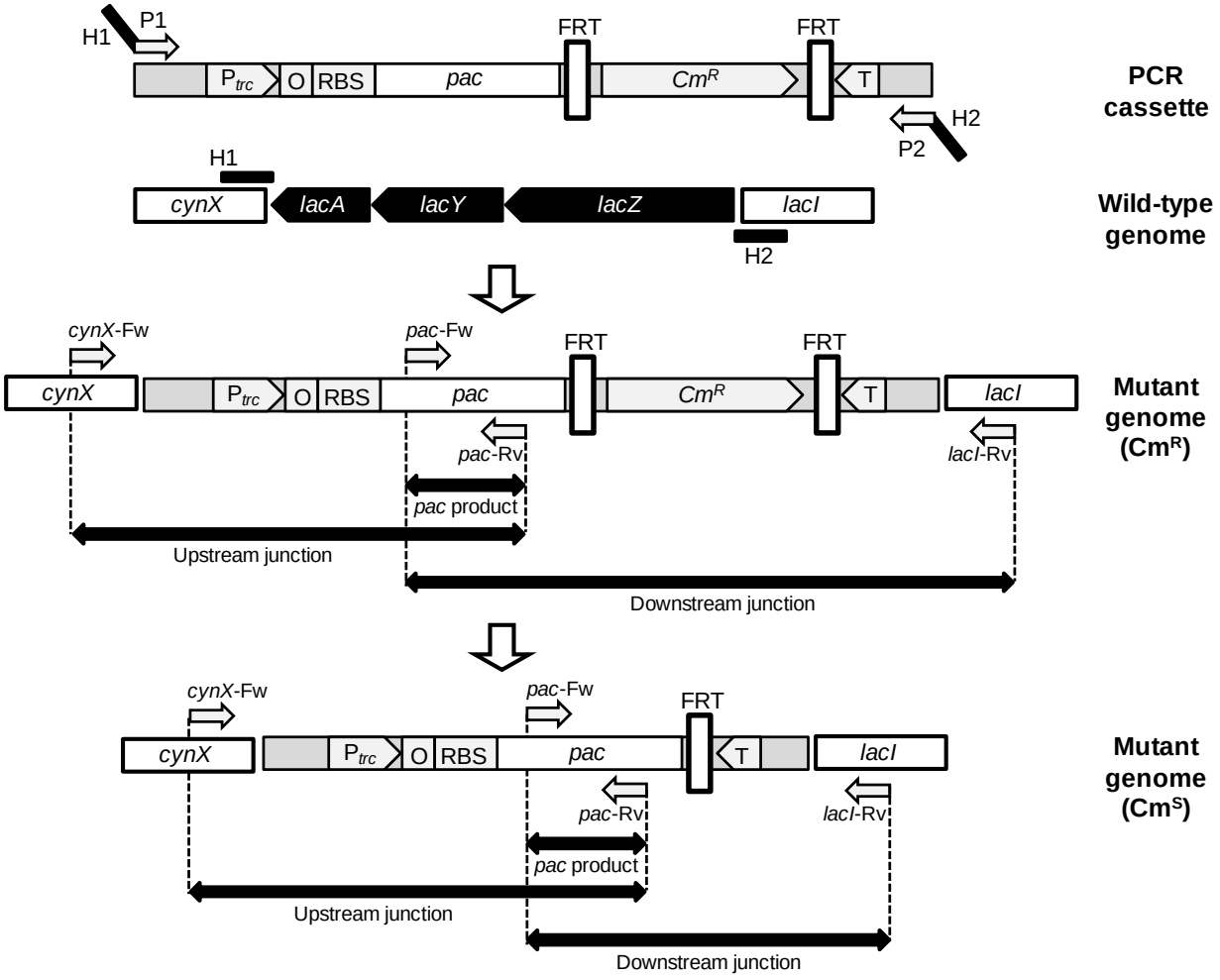


Figure 3.2 Replacement of *lacZYA* with *pac*: The *pac* operon was PCR-amplified based on the template plasmid, pTrc100catPAC using the primer pair (i.e. H1as-P1s.*lacZYA*.KO and H2s-P2as.*lacZYA*.KO) with homology extensions targeting *lacZYA*.

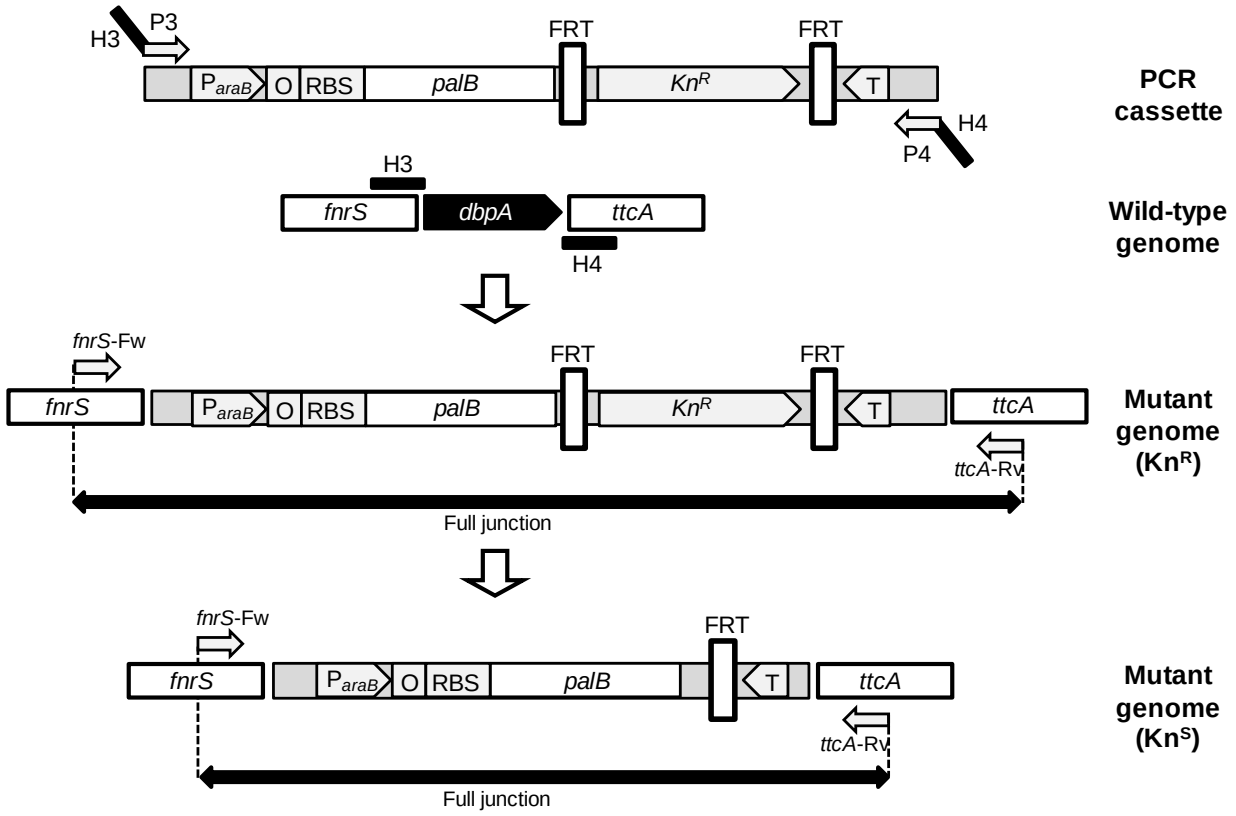


Figure 3.3 Replacement of *dbpA* with *palB2*: The *palB2* operon was PCR-amplified based on the template plasmid, pAr4KnPalB2 using the primer pair (i.e. H3s-P3s.*dbpA*.KO and H4as-P4as.*lacZYA*.KO) with homology extensions targeting *dbpA*.

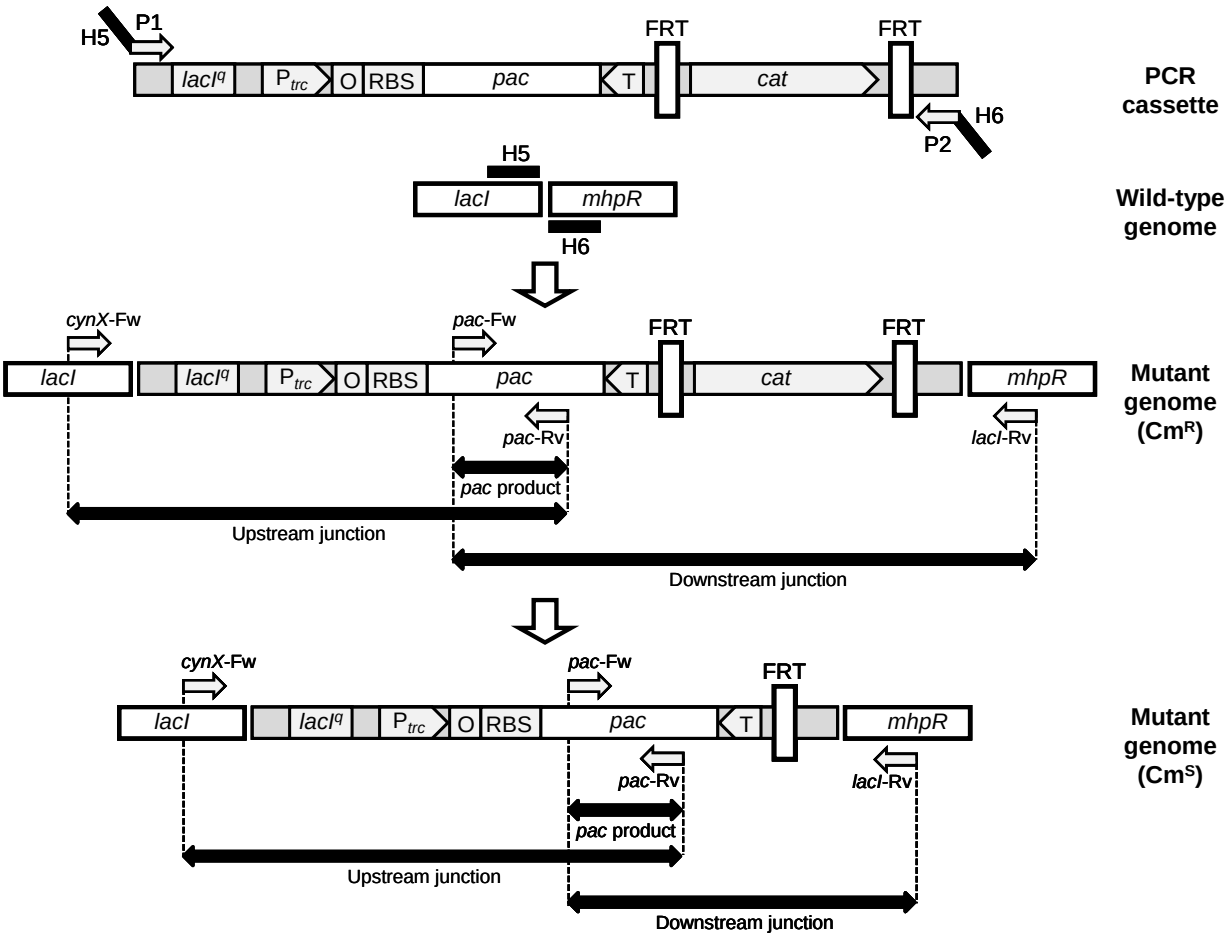


Figure 3.4 Insertion of *pac* into the junction between *lacI* and *mhpR*. The *pac* operon was PCR-amplified based on the template plasmid, pTrc100catPAC using the primer pair (i.e. H5s-P1s.insertion and H6as-P2as.insertion) with homology extensions targeting *lacI*-*mhpR*. H1, H2, H3, H4, H5 and H6 represent homology extensions, whereas P1, P2, P3 and P4 represent priming sites. Various primers, i.e. *cynX*-Fw, *lacI*-Rv, *pac*-Fw, *pac*-Rv, *fnrS*-Fw and *ttcA*-Rv, are used for colony PCR to verify the genotype of recombinant derivatives.

3.2 Analytical methods

Cell extracts were prepared by centrifuging bacterial cultures corresponding to 20 OD units at 2 °C and 6000×g for 15 min, followed by sonication of the resuspended pellet in 1 mL of sodium phosphate buffer (0.05 M, pH 7.4) with an ultrasonic processor (Misonix, Farmingdale, NY). The

supernatant containing the soluble fraction was used as the cell lysate for various enzyme assays. The expression of *pac* was qualitatively visualized by the microbiological screen with *S. marscescens* (*69*). The penicillin acylase activity was assayed using penicillin G as a substrate and the hydrolysis product of 6-aminopenicilanic acid (6-APA) was quantified by a colorimetric method (*70*). The expression of *palB* was qualitatively visualized by the hydrolysis of tributyrin on an agar plate (*71*). The lipase activity was assayed using olive oil as a substrate and the liberated fatty acids were quantified by a pH-stat (*64*). One unit of enzyme activity is defined as the amount of lipase required to liberate one μmole of fatty acid per min. Strains were regenerated three times by re-streaking colonies serially on LB agar plates, and the final regenerated strains were used in colony PCR and phenotype assays.

Chapter 4
Results

4.1 Single gene replacement

Two heterologous genes previously studied in our lab, i.e. *pac* from *E. coli* ATCC11105 and *palB2* from *P. antarctica*, were used for demonstration purposes because they both have an easy protocol for phenotypical screening. In the first demonstration, we conducted the replacement of the *lacZYA* genes with *pac* whose expression was regulated by the *trc* promoter (Figure 3.2). The recombineering protocol (*16*) was used with pTrc100catPAC as a template vector. The 3.2-kb DNA fragment containing the P_{trc}-regulated *pac* gene and FRT-flanked Cm^R marker was first amplified by touchdown PCR in order to avoid amplification of non-specific PCR amplicons. The PCR product flanked with the 36-bp homologous extensions on both termini targeting the *lacZYA* genes was transformed into *E. coli* HB101 for mediating genomic recombination, resulting in a Cm^R recombinant derivative termed as HB101PACCmR. The Cm^R allele in HB101PACCmR was deleted by the FLP-FRT recombination to form a Cm^S recombinant derivative, HB101PAC. The *pac* and *lac* phenotypes of recombinant derivatives were respectively confirmed by the microbiological screen using *S. marscescens* and the blue-white screening using X-gal agar plates (Figure 4.1A and B). Colony PCR using locus-specific primer pairs confirmed the genotypes of recombinant derivatives (Figure 4.2D).

Similarly, we conducted the replacement of the *dbpA* gene, which encodes a non-essential RNA helicase (*72*), with *palB2* whose expression was regulated by the *araB* promoter (Figure 3.3). This gene replacement was carried out in both HB101 and DH5α backgrounds. However, only DH5α derivative is reported here since it is a better host for expressing *palB2* (see further description below). The 2.7-kb DNA fragment containing the P_{araB}-regualted *palB2* gene and FRT-flanked Kn^R marker was amplified by touchdown PCR with pAr4KnPalb as the template vector. The PCR product flanked with 36-bp homologous extension on both termini targeting the *dbpA* gene was transformed in DH5α, generating the Kn^R recombinant derivative termed DH5αPalB2KnR. The Kn^R allele in the resulting strain was deleted using FLP-FRT recombination to form the Kn^S recombinant derivative, DH5αPalB2. The *palB2* phenotype of this final strain was verified by visualizing lipase activity on tributyrin agar plates (Figure 4.1C). Colony PCR using locus-specific primers confirmed the genotype of the recombinant derivatives (Figure 4.2E).

Third generation strain colonies were tested with phenotypic assays and colony PCR for all the recombinant strains to ensure the stability of the integrants. Recombinant colonies were re-streaked and isolated on LB agar as described in Material and Methods (section 3.1).

4.2 Multiple gene replacements

To demonstrate the extensive application of the proposed approach for "tailoring" *E. coli* strains, we conducted multiple gene replacements by respectively replacing the two non-essential alleles (i.e. *lacZYA* and *dbpA*) with two foreign genes (i.e. *pac* and *palB2*, Figure 3.2 and Figure 3.3 respectively). The above-described procedure for replacing the *dbpA* gene with *palB2* was conducted under the genetic background of HB101PAC. As a result, an antibiotic marker-less strain, HB101PACPalB2, was derived and the genotypes and phenotypes were verified by colony PCR using locus-specific primers and various screening methods, respectively (Figure 4.1).

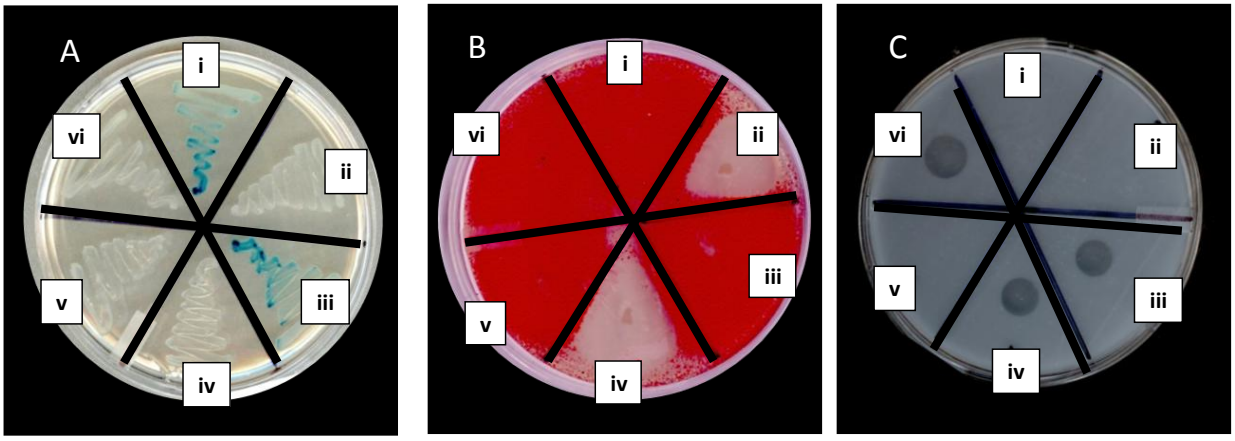


Figure 4.1 Phenotypical verification by various screening plates of recombinant derivatives. (A) Blue/white screening associated with *lacZYA*. (B) *S. marscescens* growth inhibition associated with *pac*. (C) Tributyrin hydrolysis associated with *palB2*. Recombinant derivatives: (i) HB101, (ii) HB101PAC, (iii) HB101PalB2, (iv) HB101PACPalB2, (v) DH5α, and (vi) DH5αPalB2.

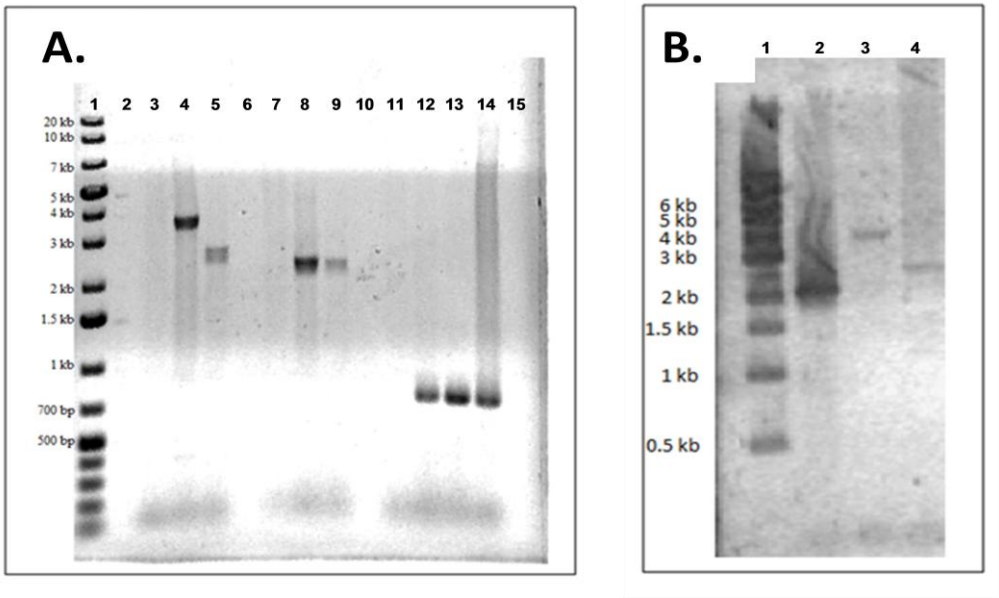


Figure 4.2 Genotypical verification by colony PCR of recombinant derivatives. (A) Colony PCR for *pac* knock-in. The *pac*-Fw and *lacI*-Rv primers were used to verify the presence of the downstream junction containing part of the *pac* gene and the Cm^R gene in the HB101PACCmR strain (lane 4), and to verify the Cm^S genotype of the HB101PAC strain (lane 5). Colony PCR of the wild-type HB101 strain with this primer combination was used as a negative control (lane 1). Presence of the upstream junction containing the P_{trc}-regulated *pac* gene was verified using *cynX*-Fw and *pac*-Rv primers in the HB101PACCmR and HB101PAC strains (lane 8, 9). A wild-type HB101 colony was used as template with this primer combination as a negative control (lane 7). The presence of the *pac* gene was verified using *pac*-specific primers, *pac*-Fw and *pac*-Rv, in the HB101PACCmR and HB101PAC strains (lane 13, 14). HB101 and HB101 containing the pTrcKnPAC2902 colonies were used for colony PCR as negative and positive controls (lane 11, 15), respectively, with the *pac*-specific primers. (B) Colony PCR for *palB2* knock-in. Colony PCR with the *fnrS*-Fw and *ttcA*-Rv primers was used to verify the replacement of the *dbpA* gene with the DNA fragment containing P_{araB}-regulated *palB2* gene and Kn^R gene in the HB101PACPalB2KnR strain (lane 3), and the removal of the Kn^R gene in the HB101PACPalB2 strain (lane4). The HB101PAC colony was used for colony PCR with the same primer set as negative control (lane 2). Please refer to Table 3.1 for the locations of various primers.

4.3 Gene insertions

In addition to gene replacement, we also demonstrated that the proposed approach can be readily used to insert foreign genes into the *E. coli* genome. To do this, the above-described *pac* operon was inserted between the two adjacent genes in *E. coli* HB101 genome, i.e. *lacI*, encoding the *lac* repressor (*73*), and *mhpR*, encoding the hydroxyphenylpropionate regulator protein (*74*), without inactivating any genes on the *E. coli* chromosome. The insertion resulted in the derivation of another antibiotic marker-less strain, HB101PACins. The genotypes and phenotypes were verified using colony PCR and microbiological screen using *S. marscescens*, respectively (Figure 4.3 and Figure 4.4).

The stability of integrated DNA was tested by performing colony PCR with locus specific primers for eight generations (Figure 4.3). These results suggest that this advanced recombineering method yields mutants with the integrated heterologous DNA that is stable for generations.

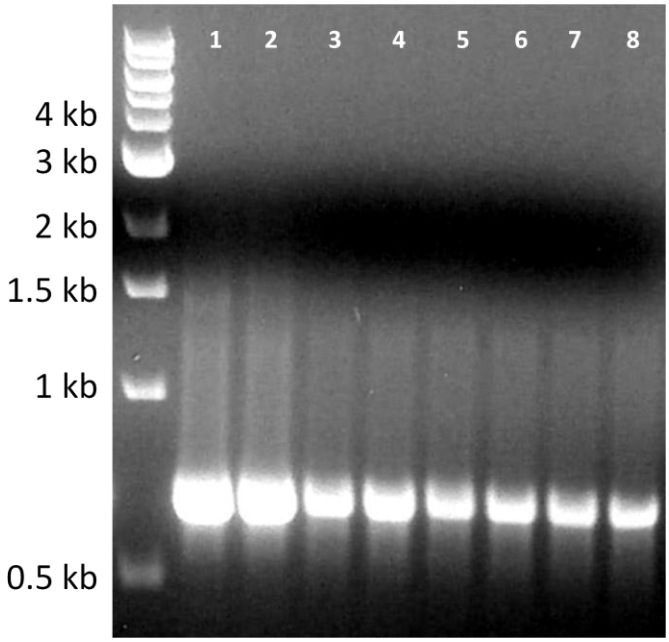


Figure 4.3 Genotypical verification by colony PCR of the HB101PACins derivative. Locus-specific primers were used to amplify 700 bp of the *Cm*^{R} gene in the integrated insert. Lanes 1 to 8 represent colonies from eight generations. Colonies were regenerated by streaking the colony on LB agar.

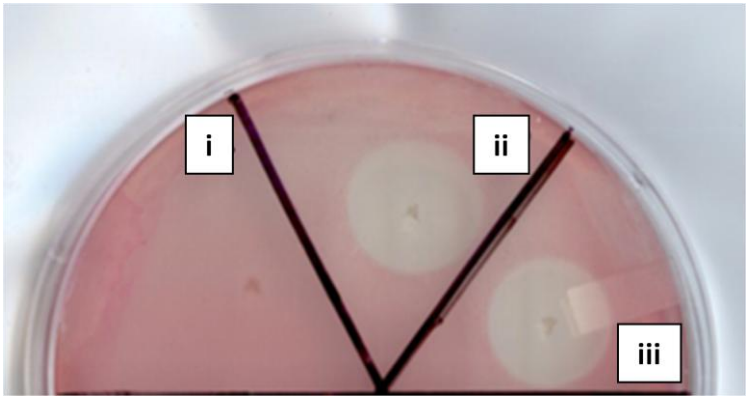


Figure 4.4 Phenotypical verification by the HB101PACins recombinant derivatives using a *S. marscescens* microbiology screen test. The PAC activity is visualized using *S. marscescens* growth inhibition associated with *pac*. Recombinant derivatives: (i) HB101, (ii) HB101PAC, (iii) HB101PACins.

4.4 Regulation of foreign gene expression

The expression of the *pac* and *palB2* genes in the derived antibiotic marker-less strain HB101PACPalB2 was investigated. Since the *trc* and *araB* promoters were used to respectively regulate the expression of *pac* and *palB2*, cultivations supplemented with IPTG and arabinose at various concentrations were conducted. Control cultures without IPTG supplementation and cultures supplemented with IPTG after a four hour growth period were cultured in shake flasks in parallel, and PAC activites of these cultures were analyzed four hours after inoculation. The control culture had a low basal specific PAC activity of approximately 12 U/L/OD_{600}, whereas the specific PAC activity gradually increased for cultures with IPTG supplementation in the range of 0.05~0.5 mM (Figure 4.5). The specific PAC activity reached a saturated level at 61 U/L/OD_{600} when IPTG concentration was at or above 0.5 mM. In other words, the expression of the heterologous *pac* gene, which was integrated onto the *E. coli* genome, could be effectively modulated through IPTG supplementation with an approximately five-fold inducibility in the expression level.

These HB101PACPalB2 culture lysates were also used for conducting the lipase assay. Although the lipase activity was qualitatively detected based on the clear halos developed on tributyrin-agar plates (Figure 4.1C), the activity appeared to be too low to measure using the pH-stat method, even under various induction conditions with arabinose supplementation. It was our previous observation that HB101 was a relatively poor host for expressing the heterologous *palB2* gene, particularly in comparison with DH5α (*75*). To further verify this observation, the expression of *palB2* based on the plasmid pAr4KnPalb in the two hosts of DH5α and HB101 was investigated. The specific lipase activity of the arabinose-induced DH5α (pAr4KnPalb) was more than seven-fold that of HB101 (pAr4KnPalb), i.e. 454 U/L/OD_{600} vs. 61 U/L/OD_{600}. Therefore, we constructed another recombinant derivative similar to HB101PalB2 under the genetic background of DH5α, namely DH5αPalB2KnR harboring the heterologous *palB2* operon in its genome. Using DH5αPalB2KnR, the expression of the *palB2* gene was investigated and the results are summarized in Figure 4.6. The control culture without arabinose supplementation had a low basal specific lipase activity of approximately 22 U/L/OD_{600}, whereas the expression could be induced to reach a specific lipase activity of approximately 76 U/L/OD_{600} when arabinose was supplemented at a concentration higher than 10 g/L. The results suggest the effective modulation of the expression of *palB2*, which was integrated into the *E. coli* chromosome.

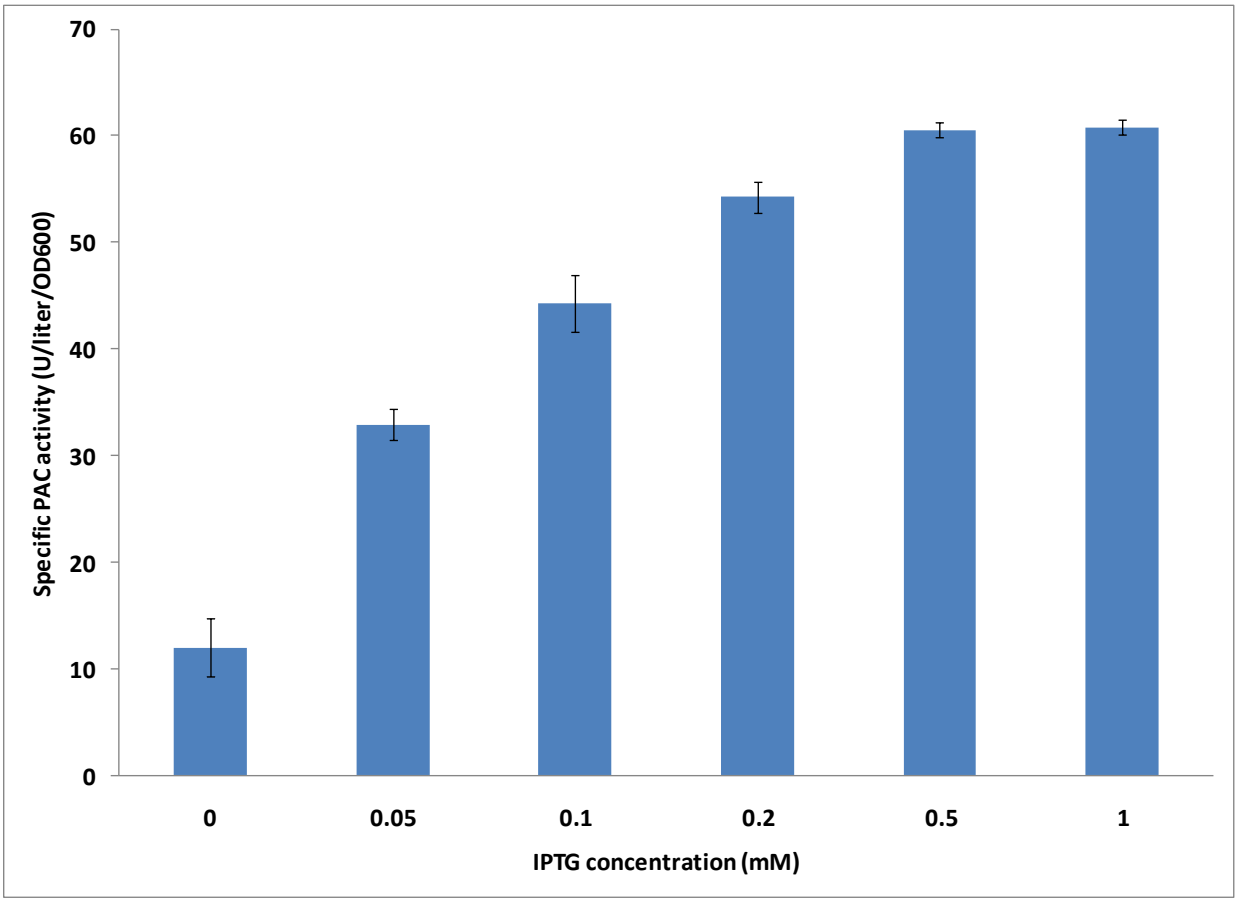


Figure 4.5 Expression of *pac* in HB101PACPalB2 under various induction conditions. PAC activity were assayed as described in the Materials and Methods (section 3.2, (*70*). Cultures grown in parallel were induced with various concentrations IPTG, in triplicate, four hours after inoclulation with 10% (v/v) of overnight culture grown at 37ºC for 16 hours. All cultures were grown at 30ºC for a total of 8 hours, and showed similar growth characteristics. 20 OD600 units of each culture were collected by centrifugation, and cell-free extracts were prepared as described in Material and Methods (section 3.2).

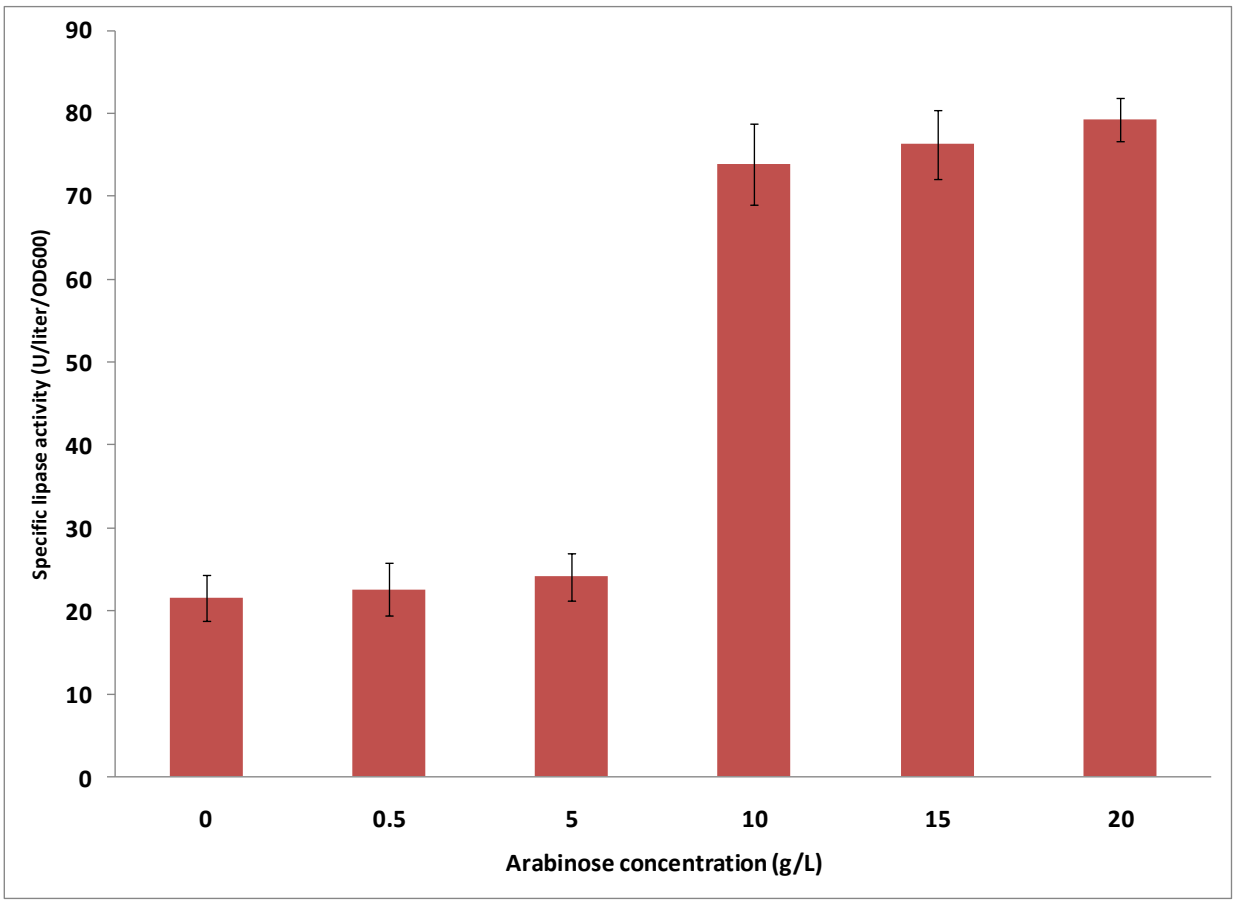


Figure 4.6 Expression from *palB2* gene in DH5αPalB2KnR under various induction conditions. Lipase activities were assayed by the pH-stat method, as described in the Materials and Methods (section 3.2, (*70*). 10% (v/v) of overnight culture grown at 37ºC for 16 hours was used to incolulate sample cultures grown in parallel. The sample cultures were induced with various concentrations L-arabinose, in triplicate, four hours after inoclulation. All cultures were grown at 30ºC for a total of 8 hours, and showed similar growth characteristics. 20 OD_{600} units of each culture were collected by centrifugation, and cell-free extracts were prepared as described in Material and Methods (section 3.2).

Cell densities of plasmid-bearing HB101 strains and the recombineered HB101 strains were collected during fermentation and plotted to compare the growth characteristics of these strains (Figure 4.7). These results show that the recombineered HB101 strains can reach higher cell densities than plasmid-bearing HB101 strains.

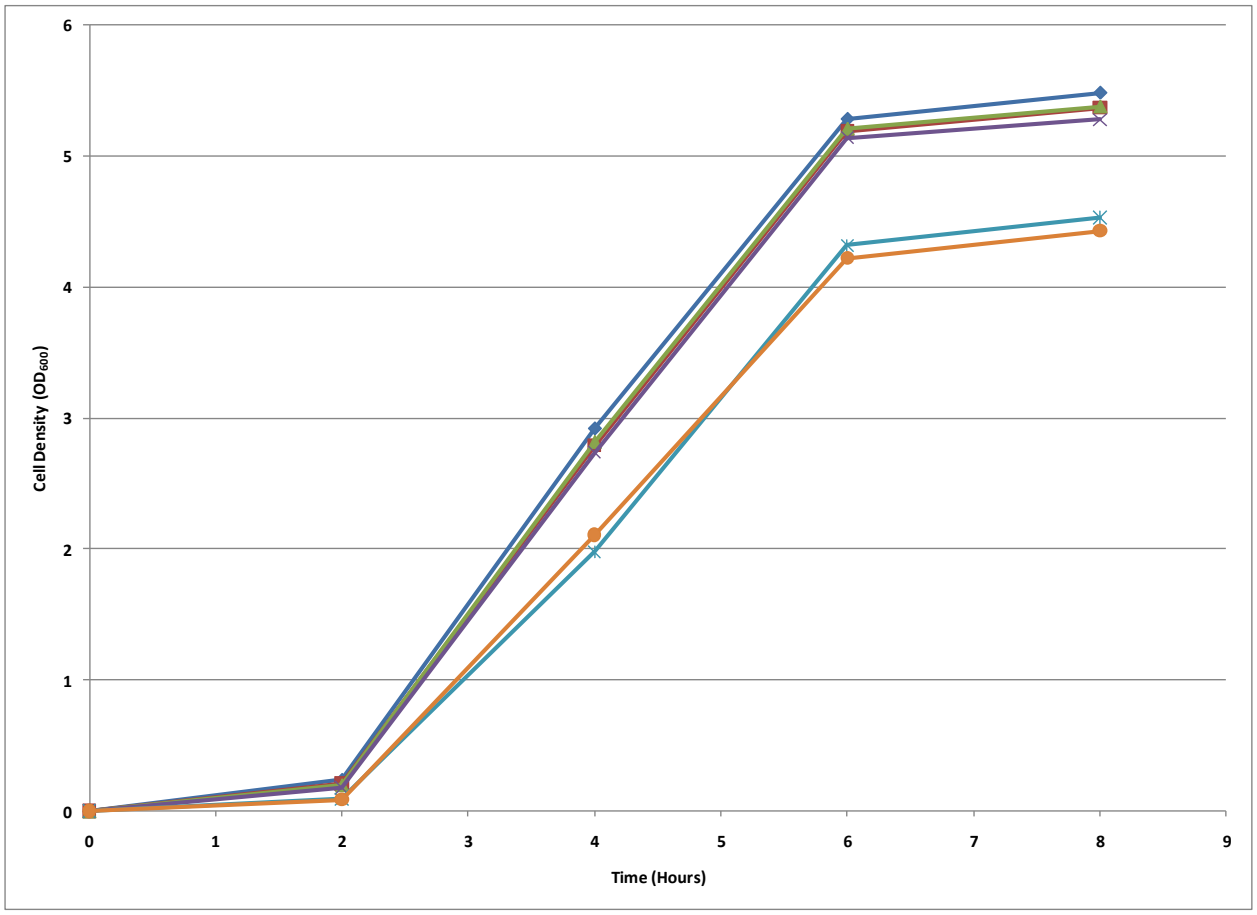


Figure 4.7 Growth curves of various HB101 and HB101-derivative strains used in this study. Wild-type HB101 (blue diamonds), the recombineered strains HB101PAC (brown squares) , HB101PalB (green triangles) , HB101PACPalB (purple crosses) and the plasmid-bearing strains HB101 (pTrc100catPAC) (blue crossed diamonds) and HB101(pAr4KnPalB) (orange circles) were grown in 25 ml LB broth in shake flasks incubated at 37ºC. All cell density values are average of triplicate analysis samples.

Chapter 5
Discussion

The recombineering method, based on λ Red and FLP-FRT recombinations, is a powerful tool commonly used to delete genes from the *E. coli* genome (*16, 18*). In this study, we have proposed a systematic approach based on the recombineering method for more extensive genomic engineering of *E. coli*. We have successfully inserted and replaced multiple genes within the bacterial genome.

This knock-in approach has four major steps. First, a template vector should be derived (Figure 3.1). Virtually, any expression vector can be used as a starting construct. The gene(s) of interest fused with various regulatory elements and the antibiotic marker gene flanked with two FRT sites are respectively cloned to derive the template plasmid. Second, the DNA region containing the gene(s) of interest to be knocked-in, various regulatory elements, and the antibiotic marker is amplified by PCR using a primer pair with overhang termini homologous to the targeted site of the genome. Third, the amplified PCR product is introduced into *E. coli* host cells for mediating λ Red recombination and recombinant cells are selected against an appropriate antibiotic. Finally, the antibiotic resistance gene in the genome of recombinant cell is deleted with FLP-FRT recombination to form an antibiotic marker-less derivative. The approach retains several technical advantages associated with the original recombineering protocols, such as the availability of various antibiotic markers and the derivation of antibiotic marker-free recombinants. Since the heterologous genes are integrated into the genome instead of being carried by a multicopy plasmid, technical issues, such as genetic stability and metabolic burden, no longer exist. Since the antibiotic marker is removed after the heterologous gene is introduced, the approach can be repeated to knock-in multiple genes in different genomic loci. Alternatively, multiple genes can be cloned into the same template vector for simultaneous knock-in at the same genomic locus.

Multiple insertions of genes using this method introduce multiple FRT sequences into the *E. coli* genome. This may raise concerns regarding undesired homologous recombinatons between these FRT sequences that may delete part of the genome, including essential genes. However, we have not found any literature reporting undesired DNA excision due to homologous recombination between the FRT sites in *E. coli (13, 76)*. Moreover, multiple regenerations of the HB101PACPalB2, which has two FRT sites, in liquid culture and solid agar media always yielded colonies similar to the parent strain. To further verify the stability of the insert, the HB101PACins strain (containing two FRT sites

flanking the Cm^R gene) was regenerated eight times. Colony PCR of the regenerated strains shows that the inserted genes are stable in each generation. This further proves that undesired excision of DNA by homologous recombination between the FRT sites is unlikely.

Given the flexible and extensive natures associated with the proposed approach, several technical aspects should be considered. First, the primers used to amplify the knock-in DNA fragment containing at least a heterologous gene and an antibiotic marker are large in size (up to 56 base pairs in this study). Large primers tend to bind to the template DNA non-specifically during PCR, causing amplification of erroneous DNA fragments (*68*). Although the correct DNA insert can be isolated using gel extraction, the low DNA recovery and poor DNA quality significantly affect the subsequent cloning. In this study, touchdown PCR (*68*) was used to resolve this issue. Second, the strain used in this approach should be carefully selected since the genetic background could affect certain applications (*77*). For example, BL21 appears to be an unsuitable host for this approach due to a poor transformation efficiency of either electroporation or P1-phage transduction . Also, HB101 was a poor host for expressing the *palB2* gene in this study. Although the reason is unknown, this observation is consistent with previous literature studying the expression from the *palB2* gene in HB101 (*75*). Third, while the heterologous gene can be knocked-in either via gene insertion or gene replacement, it appears that the efficiency of gene insertion is significantly lower than that of gene replacement possible due to the lack of a sizable DNA fragment between the two homologous arms. In this study, we observed that the recombination efficiency for inserting *pac* between *lacI* and *mhpR* was at least 90% lower than that for replacing *lacZYA* with *pac*. Finally, targeted sites for genomic engineering, either knock-in or knockout, should be carefully determined. These chosen target sites should not be part of an essential gene. Also, presumably, DNA regions that are not well exposed for recombination should be avoided though it might be hard to systematically identify these regions.

In addition to the construction of various knock-in *E. coli* strains, we also showed the capacity of the proposed approach for regulating the expression of heterologous genes that are integrated into the genome. The demonstration represents a potential advantage for several biotechnological applications, such as metabolic engineering which often requires expressional fine-tuning of a selection of genes for optimal production of target metabolites.

Recombineered strains carrying a chromosomal copy of the protein-encoding genes can reach slightly better cell densities than plasmid-bearing strains. In this study, the recombineered HB101 strains carrying a chromosomal copy of the gene(s) of interest showed up to 1 OD_{600} unit better cell

density than plasmid-bearing HB101 strains. This difference is growth rate is presumably due to the lack of plasmid-related metabolic stress on the cells of the recombineered strains.

Chapter 6
Conclusions

Genomic engineering has become a popular tool for manipulation of microbial strains. The availability of recombineering protocols and genomic sequences significantly facilitates the manipulation. However, most of the protocols for genomic engineering focus on gene knockout. To complement the technical deficiency, we have proposed a systematic approach based on the λ Red and FLP-FRT recombinations to introduce foreign genes into the *E. coli* genome.

Using the two heterologous genes of *pac* and *palB2*, various scenarios for gene knock-in (including both gene insertion and gene replacement) were demonstrated in this study. The final recombinant strains are marker-less strains, which allows researchers to repeat this strategy to insert any number of genes into the genome. Whereas plasmid-based expression of genes requires antibiotic pressure to maintain gene copy during cultivation, antibiotics are not necessary with recombinant derivatives engineered using the proposed strategy. Furthermore, expression of these genes can be regulated using inducible promoter systems, as also demonstrated in this study.

By proposing a systematic approach for knocking-in heterologous genes to the *E. coli* genome, this study has paved a way towards developing a more comprehensive genomic engineering strategy for "strain tailoring". Foreign pathways can be introduced into model bacterial genomes for the safe and controlled overproduction of economically valuable metabolites. Currently, we are applying this approach to graft novel biofuel-producing genes and pathways in *E. coli*.

References

1. S. Makrides, *Microbiology Reviews* **60**, 512 (September 1, 1996, 1996).
2. K. Zhang, H. Li, K. M. Cho, J. C. Liao, *Proceedings of the National Academy of Sciences* **105**, 20653 (2010).
3. K. L. J. Prather, C. H. Martin, *Curr. Opin. Biotechnol.* **19**, 468 (2008).
4. H. Minami *et al.*, *Proc. Natl. Acad. Sci.* **105**, 7393 (2008).
5. B. Wróbel, G. Wegrzyn, *Plasmid* **39**, 48 (1998).
6. P. Balbás, G. Gosset, *Mol. Biotechnol.* **19**, 1 (2001).
7. L. A. Palomares, S. Estrada-Mondaca, O. T. Ramírez, *Methods Mol. Biol.* **267**, 15 (2004).
8. N. I. Gutterson, D. E. K. Jr., *Proc. Natl. Acad. Sci.* **80**, 4894 (August 1983, 1983).
9. M. Y. Peredelchuk, B. G.N., *Gene* **187**, 231 (1997).
10. M. El Karoui, P. Dabert, A. Gruss, S. K. Amundsen, *Nucl. Acids Res.* **27**, 1296 (March 1, 1999, 1999).
11. C. M. Hamilton, M. Aldea, B. K. Washburn, P. Babitzke, S. Kushner, *J. Bacteriol.* **177**, 4617 (1989).
12. D. E. Shevell, A. M. Abou-Zamzam, B. Demple, G. C. Walker, *J. Bacteriol.* **170**, 3294 (1988).
13. L. C. Huang, E. A. Wood, M. M. Cox, *J. Bacteriol.* **179**, 6076 (1997).
14. G. Pósfai, V. Kolisnychenko, Z. Bereczki, F. R. Blattner, *Nucl. Acids Res.* **27**, 4409 (1999).
15. Y. Zhang, F. Buchholz, J. P. Muyrers, A. F. Stewart, *Nat. Genet.* **20**, 123 (1998).
16. K. A. Datsenko, B. L. Wanner, *Proc. Natl. Acad. Sci.* **97**, 6640 (June 6, 2000, 2000).
17. K. C. Murphy, *J. Bacteriol.* **180**, 2063 (April 15, 1998, 1998).
18. T. Baba *et al.*, *Mol. Syst. Biol.*, 2006 (2006).
19. O. W. Merten, European Federation of Biotechnology. Section on Microbial Physiology. *EFB Section on Microbial Physiology, Semmering, Austria, 5th-8th October 2000.* (Kluwer Academic Publishers, Dordrecht ; Boston, 2001), pp. x, 400 p.
20. B. R. Glick, C. L. Patten, J. J. Pasternak, *Molecular biotechnology : principles and applications of recombinant DNA.* (ASM Press, Washington, DC, ed. 4th, 2010), pp. xvii, 1000 p.
21. R. G. Buckholz, M. A. Gleeson, *Biotechnology (N Y)* **9**, 1067 (Nov, 1991).
22. T. U. Gerngross, *Nat. Biotechnol.* **22**, 1409 (Nov, 2004).
23. K. Terpe, *Appl. Microbiol. Biotechnol.* **72**, 211 (Sep, 2006).
24. C. R. Sweet, *Methods Mol. Biol.* **235**, 277 (2003).
25. T. Fowler, R. M. Berka, *Curr. Opin. Biotechnol.* **2**, 691 (Oct, 1991).

26. R. M. Berka, C. C. Barnett, *Biotechnol. Adv.* **7**, 127 (1989).

27. D. L. Jarvis, *Virology* **310**, 1 (May 25, 2003).

28. Z. Zou, P. D. Sun, *Protein Expr. Purif.* **50**, 9 (Nov, 2006).

29. A. Kriz *et al.*, *Nat. Commun.* **1**, 120 (2010).

30. S. N. Cohen, A. C. Chang, H. W. Boyer, R. B. Helling, *Proc. Natl. Acad. Sci.* **70**, 3240 (Nov, 1973).

31. B. H. Frank, R. E. Chance, *MMW Munch. Med. Wochenschr.* **Suppl 1**, S14 (May 4, 1983).

32. I. S. Johnson, *Science* **219**, 632 (Feb 11, 1983).

33. D. R. Moser, J. L. Campbell, *J. Bacteriol.* **154**, 809 (May, 1983).

34. S. Lin-Chao, W. T. Chen, T. T. Wong, *Mol. Microbiol.* **6**, 3385 (Nov, 1992).

35. J. Perez-Perez, J. Gutierrez, *Gene* **158**, 141 (1995).

36. H. A. de Boer, L. J. Comstock, M. Vasser, *Proc. Natl. Acad. Sci.* **80**, 21 (Jan, 1983).

37. M. W. Qoronfleh, C. Debouck, J. Keller, *J. Bacteriol.* **174**, 7902 (Dec, 1992).

38. S. H. Bass, D. G. Yansura, *Mol. Biotechnol.* **16**, 253 (Nov, 2000).

39. J. Brosius, M. Erfle, J. Storella, *J. Biol. Chem.* **260**, 3539 (Mar 25, 1985).

40. J. H. Miller, K. Ippen, J. G. Scaife, J. R. Beckwith, *J. Mol. Biol.* **38**, 413 (1968/12/28/, 1968).

41. F. W. Studier, B. A. Moffatt, *J. Mol. Biol.* **189**, 113 (May 5, 1986).

42. J. Furlong, M. Meighan, J. Conner, J. Murray, J. B. Clements, *Nucl. Acids Res.* **20**, 4668 (Sep 11, 1992).

43. R. Caspi, P. D. Karp, *Curr. Protoc. Bioinformatics* **Chapter 1**, Unit1 17 (Dec, 2007).

44. D. G. Yansura, S. H. Bass, *Methods Mol. Biol.* **62**, 55 (1997).

45. E. Amann, B. Ochs, K. J. Abel, *Gene* **69**, 301 (1988).

46. J. Sambrook, D. W. Russell, *Molecular cloning : a laboratory manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., ed. 3rd, 2001).

47. N. L. Lee, W. O. Gielow, R. G. Wallace, *Proc. Natl. Acad. Sci.* **78**, 752 (Feb, 1981).

48. C. G. Miyada, L. Stoltzfus, G. Wilcox, *Proc. Natl. Acad. Sci.* **81**, 4120 (Jul, 1984).

49. B. Csorgo, G. Posfai, *Acta. Biol. Hung.* **58 Suppl**, 1 (2007).

50. M. V. Francia, J. M. Garcia Lobo, *J. Bacteriol.* **178**, 894 (Feb, 1996).

51. M. Jasin, P. Schimmel, *J. Bacteriol.* **159**, 783 (August 1984, 1984).

52. J. Law *et al.*, *J. Bacteriol.* **177**, 7011 (1995).

53. D. J. Lee *et al.*, *BMC Microbiol.* **9**, 252 (2009).

54. F. Martinez-Morales, A. C. Borges, A. Martinez, K. T. Shanmugam, L. O. Ingram, *J. Bacteriol.* **181**, 7143 (Nov, 1999).

55. H. M. Ellis, D. Yu, T. DiTizio, D. L. Court, *Proc Natl Acad Sci U S A* **98**, 6742 (Jun 5, 2001).

56. P. G. Mitsis, J. G. Kwagh, *Nucl. Acids Res.* **27**, 3057 (Aug 1, 1999).

57. K. C. Murphy, *J. Mol. Biol.* **371**, 19 (Aug 3, 2007).

58. N. Marsic, E. Salaj-Smic, I. Stojiljkovic, Z. Trgovcevic, *Biochimie* **73**, 501 (Apr, 1991).

59. K. Muniyappa, E. Mythili, *Biochem. Mol. Biol. Int.* **31**, 1 (Sep, 1993).

60. L. A. Malureanu, *Methods Mol. Biol.* **693**, 181 (2011).

61. S. C. Lee, P. Liu, *Cold Spring Harb Protoc* **2009**, pdb prot5291 (Sep, 2009).

62. C. D. Herring, J. D. Glasner, F. R. Blattner, *Gene* **311**, 153 (Jun 5, 2003).

63. C. P. Chou, C.-C. Yu, J.-H. Tseng, M.-I. Lin, H.-K. Lin, *Biotechnol. Bioeng.* **63**, 263 (1999).

64. N. Narayanan, C. P. Chou, *App. and Environ. Microbiol.* **75**, 5424 (August 15, 2009, 2009).

65. L. A. Fernandez, I. Sola, L. Enjuanes, V. d. Lorenzo, *App. and Environ. Microbiol.* **66**, 5024 (2000).

66. D. M. Woodstock *et al.*, *Nucl. Acids Res.* **17**, 3469 (1989).

67. H. W. Boyer, D. Roulland-Dussoix, *J. Mol. Biol.* **41**, 459 (1969).

68. R. Don, P. Cox, B. Wainwright, K. Baker, J. Mattick, *Nucl. Acids Res.* **19**, 4008 (July 25, 1991, 1991).

69. V. Meevootisom, P. Somsuk, R. Prachaktam, T. W. Flegel, *App. and Environ. Microbiol.* **46**, 1227 (1983).

70. N. Narayanan, M.-Y. Hsieh, Y. Xu, C. P. Chou, *Biotechnol. Prog.* **22**, 617 (2006).

71. T. Vorderwulbecke, K. Kieslich, H. Erdmann, *Comparison of lipases by different assays*. (Elsevier, Amsterdam, PAYS-BAS, 1992), vol. 14.

72. C. M. Diges, O. C. Uhlenbeck, **20**, 5503 (2001/10/01/print, 2001).

73. J. M. Hudson, M. G. Fried, *J. Mol. Biol.* **214**, 381 (1990/7/20/, 1990).

74. B. Torres, G. Porras, J. L. Garcia, E. Diaz, *J. Biol. Chem.* **278**, 27575 (2003/07/25, 2003).

75. N. Narayanan, M. Khan, C. P. Chou, *J. Ind. Microbiol. Biot.* **37**, 349 (2010).

76. M. R. Snaith, J. A. H. Murray, C. A. Boulter, *Gene* **166**, 173 (Dec 1, 1995).

77. H. Tegel, J. Ottosson, S. Hober, *Febs J* **278**, 729 (Mar, 2011).

Appendix A

Molecular Cloning Protocols

A.1 Isolation of Genomic DNA of *E. coli*

(DNeasy® Blood and Tissue Kit, Qiagen)

1. Harvest *E. coli* BW25141 strain from a 10 mL overnight culture (grown at 37ºC at 225 rpm) by centrifuging for 10 minutes at 10,000 rpm at room temperature. Discard supernatant.
2. Resuspend pellet in 180 µL of Buffer ATL.
3. Add 4 µl RNase A provided, vortex and incubate at room temperature for 2 minutes.
4. Add 20 µl proteinase K. Mix by vortexing, and incubate at 56°C until the cells are completely lysed.
5. Add 200 µl Buffer AL to the sample, and vortex for 1 minute. Add 200 µl of 96% ethanol, and vortex for another minute.
6. Pipette the mixture from step 3 into the DNeasy Mini spin column placed in a 2 mL collection tube. Centrifuge at 8000 rpm for 1 minute. Discard flow-through and collection tube.
7. Place the DNeasy Mini spin column in a new 2 mL collection tube, add
8. 500 µl Buffer AW1, and centrifuge for 1 minute at 8000 rpm. Discard flow-through and collection tube.
9. Place the DNeasy Mini spin column in a new 2 mL collection tube, add 500 µl Buffer AW2, and centrifuge for 3 minutes at 14,000 rpm to dry the DNeasy membrane. Discard flow-through and collection tube.
10. Place the DNeasy Mini spin column in a clean 1.5 mL microcentrifuge tube, and pipette 100 µl Buffer AE directly onto the DNeasy membrane.
11. Incubate at room temperature for 1 minute, and then centrifuge for 1 minute at 8000 rpm to elute.
12. Repeat elution once as described in step 10.

13. Quantify DNA concentration using a Nanodrop® spectrophotometer, and store at -20ºC.

A.2 Plasmid DNA extraction from *E. coli*

(Qiaprep Spin Miniprep Kit, Qiagen)

1. Add the provided RNase A solution to Buffer 1 and mix.
2. Harvest the plasmid-bearing *E. coli* strain from a 10 mL overnight culture (grown at 37ºC at 225 rpm with appropriate antibiotic) by centrifuging for 10 minutes at 10,000 rpm at room temperature. Discard supernatant.
3. Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifuge tube.
4. Add 250 µl Buffer P2 and mix thoroughly by inverting the tube 4–6 times.
5. Add 350 µl Buffer N3 and mix immediately and thoroughly by inverting the tube 5 times.
6. Centrifuge for 10 minutes at 13,000 rpm in a microcentrifuge. A compact white pellet will form.
7. Apply the supernatants to the QIAprep spin column by pipetting.
8. Centrifuge for 1 minute. Discard the flow-through.
9. Wash the QIAprep spin column by adding 0.5 mL Buffer PB and centrifuging for 1 minute. Discard the flow-through.
10. Wash QIAprep spin column by adding 0.75 mL Buffer PE and centrifuging for 1 minute.
11. Discard the flow-through, and centrifuge for an additional 1 min to remove residual wash buffer.
12. Place the QIAprep column in a clean 1.5 mL microcentrifuge tube. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) to the center of each QIAprep spin column, let stand for 1 minute, and centrifuge for 1 minute.
13. Quantify DNA concentration using a Nanodrop® spectrophotometer, and store at -20ºC.

A.3 Restriction Digest of Plasmid or Insert DNA

1. Set up restriction digest as in a micro centrifuge tube, on ice, as follows:

Component	Stock Concentration	Required Concentration	Volume
Reaction Buffer*	10X	1X	3 µL
DNA	50-200 µg/ml	25-50 ng	0.25-2 µL
Restriction Enzyme(s)	10,000-20,000 U/mL	10 U	0.5-1 µL
Deionised Water			Up to 30µL
Total			30 µL

* Supplied reaction buffers for restriction enzymes were used.

2. Mix the reaction by tapping the microcentrifuge tubes gently.
3. Incubate the mixture at supplier's recommended temperature (37ºC for most restriction enzymes used in this study) for at least 1 hour.
4. Inactivate enzyme by heating at supplier's recommended temperature (65ºC for most restriction enzymes used in this study) for 20 minutes.
5. Purify digested DNA fragments using a PCR purification kit. Alternatively, specific digested fragments can be isolated by gel electrophoresing the digestion mixture through a 0.75% low melt agarose gel, excising a piece of the gel containing the fragment of interest, and extracting the fragment through a gel extraction kit.

A.4 DNA Ligation

1. Digest the plasmid or insert DNA with appropriate restriction enzymes (see section A.3).
2. Isolate and extract the desired plasmid DNA fragment by gel electrophoresis and gel extraction kit, and quantify using a Nanodrop® spectrophotometer.
3. PCR purify the insert digestion, and quantify using a Nanodrop® spectrophotometer.
4. Set up the ligation reaction as follows, on ice:

Component	Amount	Volume
Insert DNA	0.1-1 µg	
DNA	0.05-1 µg	
Ligation buffer		1 µL
T4 DNA Ligase	0.1 Weiss Units	0.2 µL
Deionised Water		Up to 10 µL
Total		10 µL

5. Mix reaction mixture by tapping the tube gently.
6. Incubate the reaction mixture 1-4 hours at room temperature.
7. Purify DNA using PCR purification kit, elute with 10 µL deionized, sterile water.
8. Transform entire elution in *E. coli* DH5α strain via electroporation.

A.5 Gel Extraction of DNA fragments

(QiaexII Gel Extraction Kit, Qiagen)

1. Excise the DNA band from the agarose gel with a clean, sharp scalpel.
2. Weigh the gel slice in a colorless tube. Add 3 volumes of Buffer QX1 to 1 volume of gel.
3. Resuspend QIAEX II by vortexing for 30 seconds. Add 30 µL of QIAEX II to the sample according and mix.
4. Incubate at 50°C for 10 minutes to solubilize the agarose and bind the DNA. Mix by vortexing every 2 minutes to keep QIAEX II in suspension. Check that the color of the mixture is yellow.
5. Centrifuge the sample for 30 seconds and carefully remove supernatant with a pipette.
6. Wash the pellet with 500 µl of Buffer QX1.
7. Wash the pellet twice with 500 µl of Buffer PE.
8. Air-dry the pellet for 15 minutes or until the pellet becomes white.

9. To elute DNA, add 20 µl of sterile deionised water and resuspend the pellet by vortexing. Incubate at 50ºC for 10 minutes.
10. Centrifuge for 1 minute, and carefully pipette the supernatant into a clean microcentrifuge tube.
11. Quantify DNA using a Nanodrop® spectrophotometer, and store at -20ºC.

.

A.6 Purification of PCR-amplified DNA

(QIAquick PCR Purification Kit, Qiagen)

1. Add 96% ethanol to Buffer PE.
2. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix.
3. Place a QIAquick spin column in a provided 2 ml collection tube.
4. To bind DNA, apply the sample to the QIAquick column and centrifuge for 1 minute.
5. Discard flow-through. Place the QIAquick column back into the same tube.
6. To wash, add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 1 minute.
7. Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 minute at maximum speed.
8. Place QIAquick column in a clean 1.5 ml microcentrifuge tube.
9. To elute DNA, add 30 µl sterile deionised water to the center of the QIAquick membrane and centrifuge the column for 1 minute.
10. Let the column stand for 1 minute, and then centrifuge.
11. Quantify DNA using a Nanodrop® spectrophotometer, and store at -20ºC.

A.7 DNA Transformation using Electroporation

(Modified from Electroporation Manual, BioRad)

Preparation of Electrocompetent Cells

1. Inoculate 100 ml of L-broth with 1/100 volume of a fresh overnight *E. coli* culture.

2. Grow the cells at 37 °C shaking at 200 rpm to an OD600 of approximately 0.5–0.7.

3. Chill cells on ice for ~20 min. For all subsequent steps, keep the cells as close to 0 °C as possible (in an ice/water bath) and chill all containers in ice before adding cells. To harvest, transfer the cells to two cold 60-ml centrifuge bottles and spin at 6000 x g for 15 minutes at 4 °C.

4. Carefully pour off and discard the supernatant. It is better to sacrifice the yield by pouring off a few cells than to leave any supernatant behind.

5. Gently resuspend the pellet in 50 ml of ice-cold 10% glycerol each. Centrifuge at 6000 x g for 15 minutes at 4 °C; carefully pour off and discard the supernatant.

6. Resuspend the pellet in 50 ml of ice-cold 10% glycerol, transfer to one 60-ml bottle. Centrifuge at 6000 x g for 15 minutes at 4 °C; carefully pour off and discard the supernatant.

7. Resuspend the pellet in ~4 ml of ice-cold 10% glycerol. Transfer to two 2-ml sterile microfuge tubes. Centrifuge at 6000 x g for 15 minutes at 4 °C; carefully pour off and discard the supernatant.

8. Resuspend the cell pellet in a final volume of 400µl of ice-cold 10% glycerol. The cell concentration should be about 1–3 x 10^{10} cells/ml. This suspension may be frozen in aliquots (40µl each in a sterilized 1.5ml microfuge tube) on dry ice and stored at -80 °C. The cells are stable for at least 6 months under thes

conditions.

Electroporation

1. Thaw the cells on ice. For each sample to be electroporated, place a 1.5 ml microfuge tube and either a 0.1 or 0.2 cm electroporation cuvette on ice.

2. In a cold, 1.5 ml polypropylene microfuge tube, mix 40 µl of the cell suspension with 1 to 2 µl of DNA (DNA should be in deioned water). Mix well and incubate on ice for approximately 1 minute. (Note: it is best to mix the plasmids and cells in a microfuge tube since the narrow gap of the cuvettes prevents uniform mixing.)

3. Set the MicroPulser to “Ec1” when using the 0.1 cm cuvettes. Set it to "Ec2" or "Ec3" when using the 0.2 cm cuvettes. See Section 4 for operating instructions.

4. Transfer the mixture of cells and DNA to a cold electroporation cuvette and tap the suspension to the bottom. Place the cuvette in the chamber slide. Push the slide into the chamber until the cuvette is seated between the contacts in the base of the chamber. Pulse once.

5. Remove the cuvette from the chamber and immediately add 1 ml of SOC medium to the cuvette. Quickly but gently resuspend the cells with a Pasteur pipette. (The period between applying the pulse and transferring the cells to outgrowth medium is crucial for recovering *E. coli* transformants (Dower *et al.*, 1988). Delaying this transfer by even 1 minute causes a 3-fold drop in transformation. This decline continues to a 20-fold drop by 10 minutes.

6. Transfer the cell suspension to a 17 x 100 mm polypropylene tube and incubate at 37 °C for 1 hour, shaking at 225 rpm.

7. Check and record the pulse parameters. The time constant should be close to 5 milliseconds. The field strength can be calculated as actual volts (kV) / cuvette gap (cm).

8. Plate on selective medium.

Solutions and Reagents for Electroporation

1. L-Broth: 10 g Bacto tryptone, 5 g Bacto yeast extract, 5 g NaCl; dissolve in 1.0 L water. Autoclave.

2. 10% (v/v) Glycerol: 12.6 g glycerol (density = 1.26 g/cc) in 90 ml of water. Autoclave or filter sterilize.

3. SOC: seperately autoclave,

(1) 2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl,

(2) 10 mM, MgCl2,

(3) 10 mM MgSO4,

(4) 20 mM glucose.

Appendix B

Restriction Mapping and Analysis of the Plasmids

The following plasmid maps were constructed in Vector NTI software by Invitrogen. Restriction enzymes *EcoRI* and *HindIII* were used for restriction analysis of plasmids. DNA gels were obtained by gel electrophoresis method.

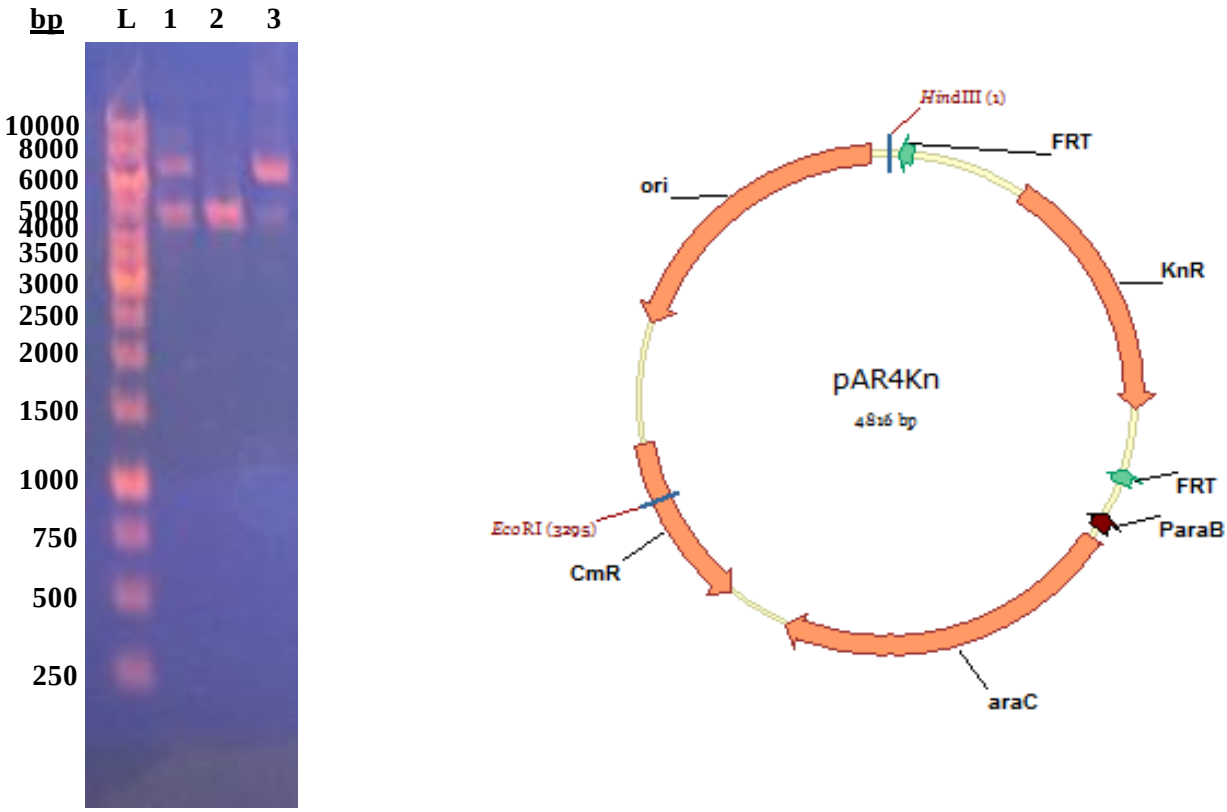


Figure B.0.1 Restriction Analysis of pAR4Kn vector. Restriction digest by *EcoRI* restriction enzyme gave the expected band at 4.8 kb (Lane 1). The remaining bands in lane 1 correspond to the uncut vector, as they are also present in the negative control lane (lane 3, uncut plasmid). Restriction digest by *HindIII* also shows the expected band at 4.8 kb (lane 2).

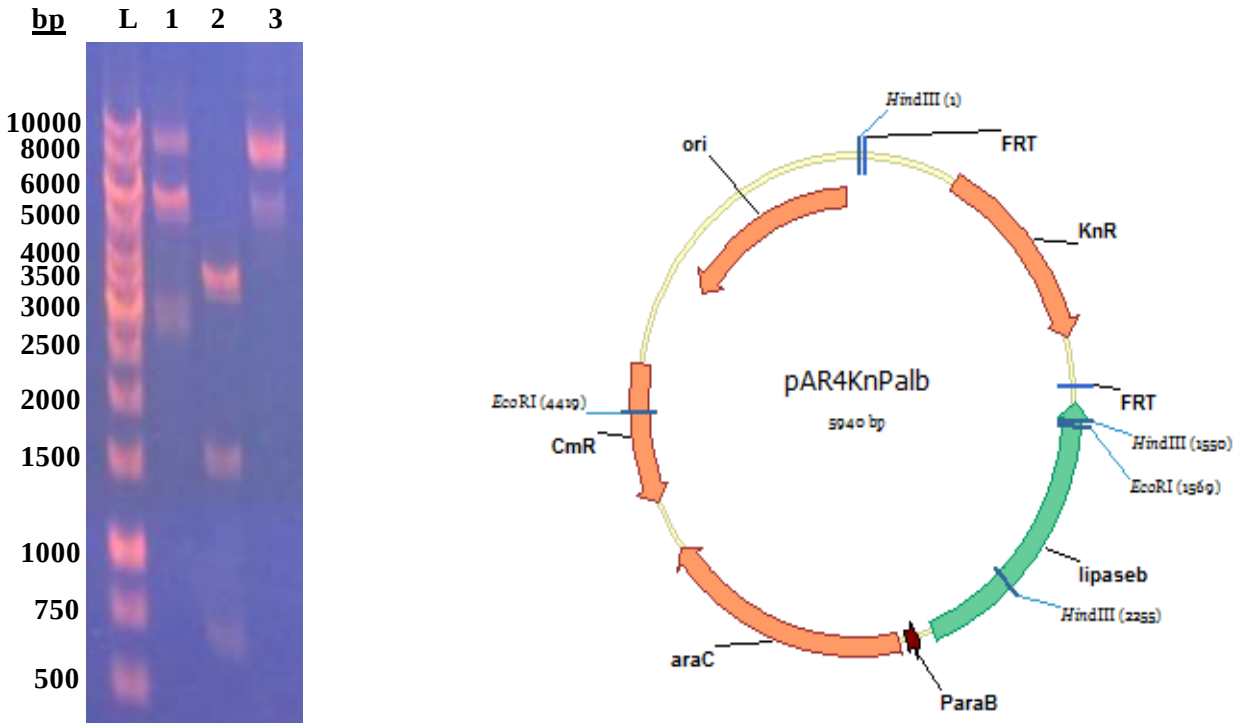
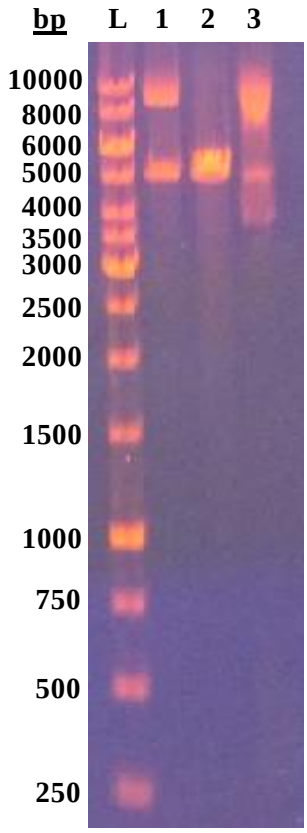


Figure B.0.2 Restriction Analysis of template vector pAR4KnPalB2. Restriction digest by *EcoRI* restriction enzyme gave the expected bands at approximately 3.1 kb and 2.9 kb (Lane 1). The remaining bands in lane 1 correspond to the uncut vector, as they are also present in the negative control lane (lane 3, uncut plasmid). Restriction digest by *HindIII* also shows the expected bands at approximately 3.7 kb, 1.5 kb, and 0.7 kb (lane 2).



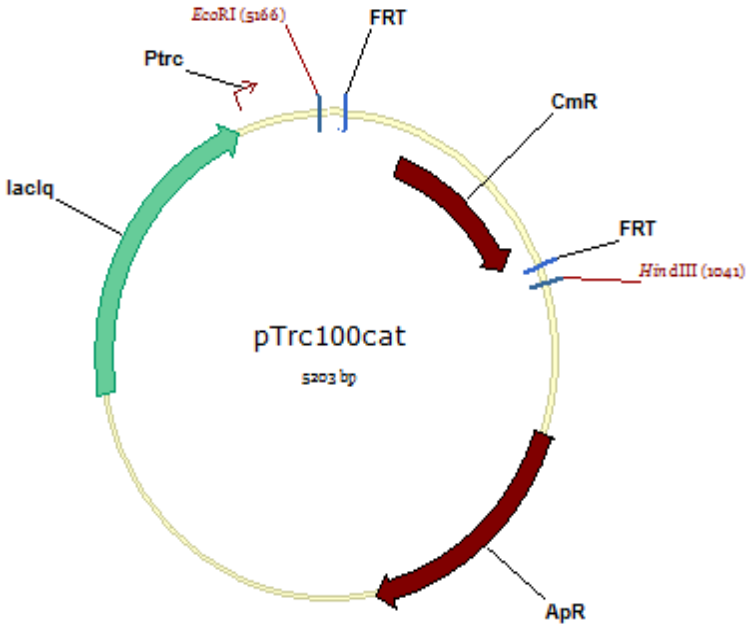


Figure B.0.3 Restriction Analysis of pTrc100cat. Separate restriction digests by *EcoRI* (lane 1) and *HindIII* (lane 2) restriction enzymes linearized the vector, yielding the expected bands at 5.2 kb. The remaining bands in lane 1 correspond to the uncut vector, as they are also present in the negative control lane (lane 3, uncut plasmid).

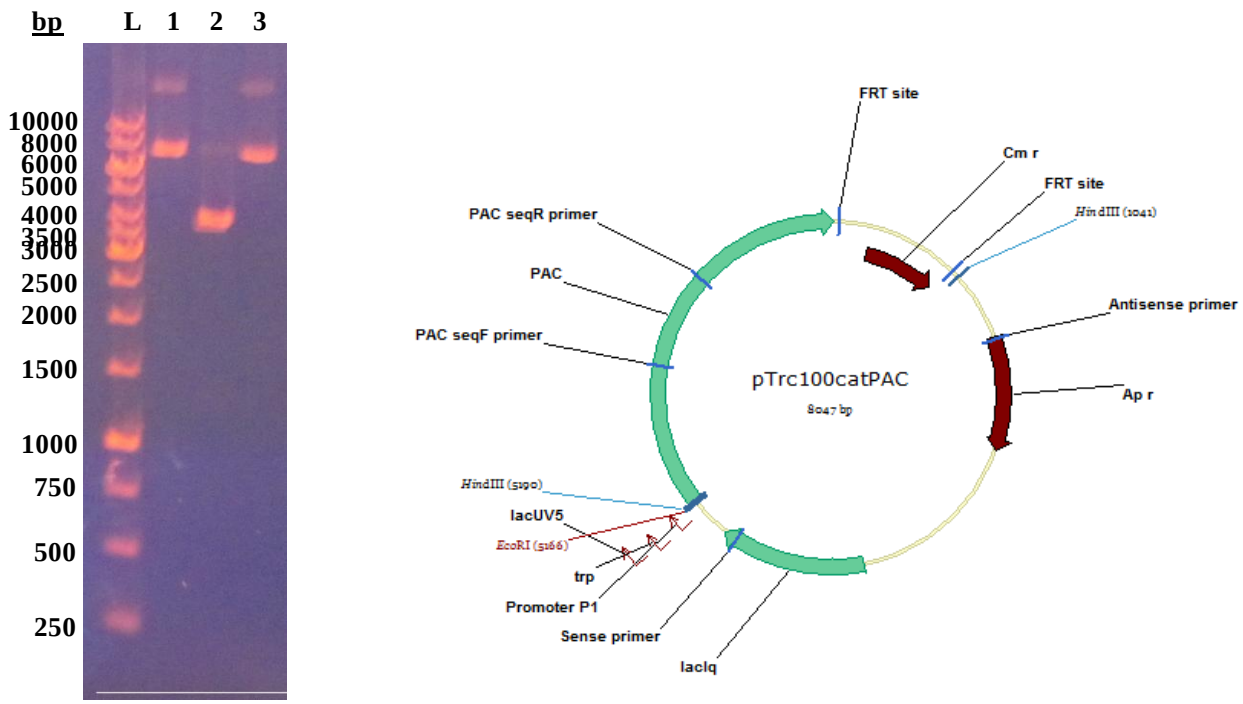


Figure B.0.4 Restriction Analysis of template vector pTrc100catPAC. Restriction digest by *EcoRI* restriction enzyme linearized the vector, yielding the expected band at ~ 8 kb (Lane 1). The remaining bands in lane 1 correspond to the uncut vector, as they are also present in the negative control lane (lane 3, uncut plasmid). Restriction digest by *HindIII* also shows the expected bands at approximately 4.1 kb and 3.9 kb (lane 2).