FITSelect: An Invention to Select Microbial Strains Maximizing Product Formation from a Single Culture Without High-Throughput Screening

Rui Zhou

Thesis submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Biological Systems Engineering

Ryan Senger, Chair
Chenming Zhang, Committee Member
Justin Barone, Committee Member
Eva Collakova, Committee Member

August 8, 2010

Blacksburg, VA

Keywords: synthetic biology, growth competition assay, metabolic engineering,

combinatorial methods

FITSelect: An Invention to Select Microbial Strains Maximizing Product Formation from a Single Culture Without High-Throughput Screening

Rui Zhou

ABSTRACT

In metabolic engineering of prokaryotes, combinatorial approaches have developed recently that induce random genetic perturbations to achieve a desired cell phenotype. A screening strategy follows the randomized genetic manipulations to select strain(s) with the more optimal phenotype of interest. This screening strategy is often divided into two categories: (i) a growth competition assay and (ii) selection by high-throughput screening. The growth competition assay involves culturing strains together. The strain with the highest growth rate will ultimately dominate the culture. This strategy is ideal for selecting strain with cellular fitness (e.g., solvent tolerance), but it does not work for selecting a strain that can over-produce a product (e.g., an amino acid). For the case of selecting highly productive phenotypes, high-throughput screening is used. This method analyzes strains individually and is costly and time-consuming. In this research, a synthetic genetic circuit was developed to select highly productive phenotypes using a growth competition assay rather than high-throughput screening.

This novel system is called Feed-back Inhibition of Transcription for Growth Selection (FITSelect), and it uses a natural feedback inhibition mechanism in the L-

arginine production pathway to select strains (transformed with a random genomic library) that can over-produce L-arginine in *E. coli* DH10B. With FITSelect, the cell can thrive in the growth competition assay when L-arginine is over-produced (i.e., growth is tied to L-arginine production). Cell death or reduced growth results if L-arginine is not over-produced by the cell. This system was created by including an L-arginine concentration responsive *argF* promoter to control a *ccdB* cell death gene in the FITSelect system. The effects of *ccdB* were modulated by the antidote *ccdA* gene under control of an L-tryptophan responsive *trp* promoter. Several insights and construction strategies were required to build a system that ties the growth rate of the cell to L-arginine concentrations.

ACKNOWLEDGEMENTS

I would like to take this opportunity to be grateful to my advisor Dr. Ryan Senger. In the past two and half years, his patience and guidance leads me enter and explore this scientific ocean. His scientific attitude and passion for research influences and motives me.

Also I would like say thank you to my lab mates. It is my honor to work with you all, and I learned a lot from you.

I would also like to thank my committee numbers: Dr. Chenming Zhang, Dr. Justin Barone, and Dr. Eva Collakova for your kind advice and help with my thesis.

Last but not least, I would like to thank my parents. Without your support I could not be able to step this far. Your love will be with me forever.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

1.1. Engineering polygenic traits in cells

With advances in metabolic engineering over the last 20 years, it is now understood that simple cellular processes are controlled by or impacted by several genes (polygenic) in the cell. The approach of adding or modifying a single gene to achieve a desired cellular fitness (e.g., solvent tolerance) or optimally produce a desired product (e.g., a biofuel, amino acid, monomer, etc.) is no longer optimal. Current approaches in metabolic engineering focus on coordinated efforts to modify the expression of multiple genes simultaneously to achieve a desired outcome (*1-4*). To do this, two strategies of metabolic engineering exist: (i) rational design and (ii) the combinatorial approach. While successes have been seen with both strategies, significant differences exist. For the rational approach, a "rational" decision is first made by a researcher regarding how to engineer metabolism to achieve a desired outcome. These approaches may include: (i) over-expression of a specific gene, (ii) gene knock-outs, or (iii) modifying a specific regulatory circuit of the cell. Rational design is indeed hypothesis driven and often makes use of metabolic maps and models to design experiments. The combinatorial approach, on the other hand, is "phenotype driven" (*1*). Often, the combinatorial approach begins with admitting it is not known what genes are involved with a particular trait of the cell (e.g. solvent tolerance). There is no single "solvent tolerance" gene that can be over-expressed. This trait is acquired by the action of several genes and regulators working together. The same has been seen for over-production of amino acids (*5*). Simply over-expressing genes related to the production of L-phenylalanine does not result in an abundance of this amino acid because multiple mechanisms are in place within the cell to tightly control the L-phenylalanine concentration of the cytoplasm.

So, the combinatorial approach begins with random genetic modifications. These may be in the form of: (i) random mutations of the genome, (ii) random insertions of DNA into the genome (random knock-out), and/or (iii) over-expression of random fragments of genomic DNA (a genomic library). The genetic modifications are carried-out so that each cell within the culture contains a different random genetic perturbation. Then, each cell is analyzed for its ability to convey the desired trait. Once an optimal cell is identified, the genetic modifications are examined. Hypotheses are then formulated to explain how these particular modifications resulted in the cellular phenotype of interest.

1.2. The growth competition assay and its alternatives

The combinatorial approach to metabolic engineering has been extremely successful. However, major challenges exist with this approach. For example, the combinatorial approach requires testing each cell of a culture for a desired trait and effectively selecting that cell. For bacterial cultures, this number of cells is very large. Thus, there is a significant need to engineer "selection technologies" for the combinatorial metabolic engineering approach. The simplest method of selecting an optimum cell is by using a "growth competition assay" (*6*, *7*). For the case of solvent tolerance, a growth competition assay works very well. Random genetic mutations are made to a culture, and the culture is exposed to increasing levels of a solvent (e.g., ethanol or butanol). Cells with genetic manipulations that confer solvent tolerance will be able to grow at a faster rate than cells that are inhibited by the solvent. After several rounds of growing the culture in fresh aliquots of media containing solvent, the "solvent tolerant" cells will dominate the culture.

The growth competition assay is an effective selection procedure to engineer cellular fitness in response to a stress that inhibits the growth rate of the culture. However, what if the desired outcome is to over-produce an amino acid? The cell in the culture capable of this trait will not possess the highest growth rate. Thus, a growth competition assay cannot be used in this case. Several selection strategies have been developed for this case, but they all involve high-throughput screening of individual colonies of cells (*5, 8*). This high-throughput approach to selection is costly and time-consuming. It also limits the use of combinatorial metabolic engineering to labs with the necessary high-throughput screening equipment.

Thus, the goal of this research is to use synthetic biology to create a genetic circuit that ties the production of an amino acid (i.e., L-arginine in this case) to the growth rate of the cell. With this system, a cell producing more L-arginine has a higher growth rate than a cell producing less L-arginine. This synthetic circuit is called Feedback Inhibition of Transcription for Growth Selection (FITSelect). The building and testing of the FITSelect system for over-producing L-arginine by *E. coli* K12 is described in this thesis.

1.3. The FITSelect synthetic circuit

The FITSelect system was designed around the transcriptional feedback system for L-arginine production in bacteria. When the L-arginine concentration in an *E. coli* K12 cell is high, L-arginine binds to an ArgR protein. When this happens, ArgR becomes "activated" and binds DNA at an ARG box site. This binding halts transcription of the L-arginine producing genes. The idea driving FITSelect is to use ARG box regulatory sequence in front of a gene that inhibits growth in a synthetic construct. So, if L-arginine is over-produced, a growth inhibitory gene is shut-down in the cell. If L-arginine is scarce, the growth inhibitory gene is expressed highly.

With this approach, FITSelect enables over-producing L-arginine cells to be selected by a growth competition assay. In this research, FITSelect is built and demonstrated for the case of L-arginine over-production, but this system can be tailored easily for any product demonstrating feedback regulation. For other products, such as ethanol or butanol, future protein engineering efforts will be required to produce protein regulators that can be activated by a desired chemical to bind DNA to inhibit transcription.

1.4. References

1. Santos, C. N., and Stephanopoulos, G. (2008) Combinatorial engineering of microbes for optimizing cellular phenotype, *Curr Opin Chem Biol 12*, 168-176.
2. Soh, K. C., and Hatzimanikatis, V. (2010) DREAMS of metabolism, *Trends in biotechnology 28*, 501-508.
3. Tyo, K. E., Kocharin, K., and Nielsen, J. (2010) Toward design-based engineering of industrial microbes, *Curr Opin Microbiol 13*, 255-262.
4. Wackett, L. P. (2011) Engineering microbes to produce biofuels, *Curr Opin Biotechnol 22*, 388-393.
5. Santos, C. N., and Stephanopoulos, G. (2008) Melanin-based high-throughput screen for L-tyrosine production in Escherichia coli, *Appl Environ Microbiol 74*, 1190-1197.
6. Gill, R. T., Wildt, S., Yang, Y. T., Ziesman, S., and Stephanopoulos, G. (2002) Genome-wide screening for trait conferring genes using DNA microarrays, *P Natl Acad Sci USA 99*, 7033-7038.

7. Borden, J. R., and Papoutsakis, E. T. (2007) Dynamics of genomic-library enrichment and identification of solvent tolerance genes for Clostridium acetobutylicum, *Applied and environmental microbiology 73*, 3061-3068.

8. Yu, H., Tyo, K., Alper, H., Klein-Marcuschamer, D., and Stephanopoulos, G. (2008) A high-throughput screen for hyaluronic acid accumulation in recombinant Escherichia coli transformed by libraries of engineered sigma factors, *Biotechnology and Bioengineering 101*, 788-796.

2. Literature Review

2.1. An introduction to metabolic engineering

Thousands of metabolites and enzymes collaborate to be functional as a living cell, and the goal of living cells is to achieve optimal growth. For metabolic engineers, optimal growth is not always the desired goal. Other aspects such as maximizing productivity of a target metabolite or changing phenotype of cells are of interest to biotechnology. In response, metabolic engineering appeared as an independent subject of research more than 20 years ago following the advent of recombinant DNA technology, which allowed for the genetic manipulation of microbes. It is defined as "the purposeful modification of intermediary metabolism using recombinant DNA techniques" (*9*). The main tool sets of metabolic engineering are similar to those of molecular biology. But, the major difference between metabolic engineering and traditional molecular biology is that metabolic engineering focuses on metabolic pathways and global metabolism instead of individual enzymatic reactions in a cell. The main strategies of genetic manipulation used in metabolic engineering are (i) the adjustment of known pathways and (ii) the introduction of foreign pathways to increase process productivity or extend metabolic capability (*1*). There are many applications of metabolic engineering (*10*), and several are given here with specific examples from the literature.

1. ***Improve yield and productivity of native products synthesized by microorganisms***. In this application, improvement of ethanol production was the primary goal. *Escherichia coli* and *Klebsiella* species were the major focus of research because of (i) their native ability to produce ethanol, (ii) their wide range of substrate usage and (iii) the various

molecular biology tools available for these organisms. But, the production of ethanol from these organisms could not meet the requirement of industry. After the introduction of pyruvate decarboxylase (PDC) from *Zymomonas mobilis* to *Klebsiella planticola*, the yield of ethanol production was 1.3 mol/mol of xylose, or 25.1 g/liter *(11)*. Another example of metabolic engineering for ethanol production was from Ohta et al. (*12*). In this example, PDC and alcohol dehydrogenase from *Zymomonas mobilis* was introduced into *E. coli*. They harvested ethanol at the concentrations of up to 54.4 and 41.6 g/liter with 10% glucose and 8% xylose were used in the medium, respectively.

2. ***Extend the range of substrates***. *Corynebacterium glutamicum* is used industrially for the production of amino acids, mainly L-glutamate and L-lysine. Sucrose or glucose are commonly used as the carbon sources. Other cheaper substrates are desired in order to reduce cost. In this example, the lactose operon from *E. coli* was inserted into an *E.coli/C. glutamicum* shuttle vector and transformed into *C. glutamicum*. This provided *C. glutamicum* with the ability to grow in minimal media supplied with the cheap substrate lactose as the sole carbon source (*13*).
3. ***Produce products that are new to the host cell.*** Poly[(R)-(-)-3-hydroxybutyrate] (PHB) attracts great interest because of its thermoplastic properties. *Alcaligenes eutrophus* is a bacterial strain that can produce PHB as a carbon storage polymer. Three genes have been identified that catalyze the polymerization of acetyl-CoA to PHB in *A. eutrophus.* These three genes were modified and transferred into the plant *Arabidopsis thaliana.* A large amount of PHB granule was observed in *A. thaliana* plastid. This accounted for up to 14% of plant dry weight. This method demonstrated a useful procedure for making biodegradable plastics directly from sunlight and CO_2 through metabolic engineering *(14)*.

4. ***Improve general cellular properties.*** Acetate is one of the acidic metabolic byproducts of *E. coli*. In fermentation, the production of acetate is harmful not only to cell growth but also to recombinant protein yields. Researchers cloned acetolactate synthase (ALS) from *Bacillus subtilis* to redirect carbon flux in *E. coli* from acetate production to acetoin. Acetate secretion was minimized below the levels of toxicity, and the new product acetoin was demonstrated to be 50 times less harmful than acetate (*15*).
5. ***In the medical field for (i) analysis of the metabolism of whole organs and tissues and (ii) the identification of targets for disease control by gene therapy or nutritional strategies.*** With the development of diagnostic and therapeutic technologies, the observation of patient status has been transformed from macroscopic measurements like muscle mass and temperature to more microscopic level measurements such as metabolic status. At the microscopic level, metabolite distribution in the body can be examined to learn about the general health of a patient. Metabolic flux can be calculated and used to identify the presence of disease (*16*).

Some of the earliest applications of metabolic engineering involved genetic manipulations using a "trial and error" approach. This is based on detailed biochemical maps and mathematical analyses of particular metabolic reaction networks in an effort to identify each important step in the network (*17*). A shift was made from "trial and error" methods to more systematic and rational approaches involving recombinant DNA technology by using stoichiometric, kinetic, and regulatory knowledge to modify existing metabolic pathways by means of (i) gene deletion, (ii) gene over-expression, and (iii) de-regulation of genetic programs (*18*). This approach requires clear knowledge of metabolic pathways and the choice of a specific

gene to modify to achieve a desired goal. Although these two methods are conceptually different, they are complementary. The rational design approach is often seen as a step forward compared to the trial and error method, but a significant amount of information is required. Often, rational methods require information about the system that is not yet known. To effectively use the knowledge of metabolism we do possess, metabolic networks and models are often used with the rational design approach. Also, borrowing natural features from other organisms has been shown to be helpful for the rational approach. The following is a simple example of the rational approach. The final precursor for commercial ascorbic acid (Vitamin C) in 1990 was 2-keto-L-gulonic acid (2-KLG). The production of this precursor involved two separate fermentation steps in two organisms. First, in *Erwinia herbicola*, glucose was converted to 2,5-diketo-D-gluconic acid(2,5-DKG). Then, 2,5-DKG was converted into 2-KLG in *Corynebacterium*. Researchers were successful in cloning 2,5-DKG reductase from *Corynebacterium* into *E. herbicola* to simplify this full conversion (*19*). But, it was soon discovered that changing a single target gene caused many unexpected outputs. It was then realized that often several genes must cooperate to perform specific functions to achieve a desired phenotype.

In order to elucidate interactions between pools of genes and their resulting phenotypes, combinatorial methods appeared as a more advanced approach method. With the combinatorial method, previous knowledge about pathways or genetic regulation mechanisms is not necessary. This turns out to be the limiting factor of the rational method. In the combinatorial approach, libraries of genomic DNA, promoters, or random gene knock-outs are constructed and cultures are enriched to look for a phenotype of interest.

2.2. Synthetic biology

The term "synthetic biology" has been redefined multiple times and now has three meanings. This term first appeared in 1980 and was used to describe engineering the genetic structure in bacteria to accommodate recombinant DNA technology (*20*). The second use of "synthetic biology" was introduced about ten years ago (*21*) and referred to the biosynthesis of unnatural organic molecules in living organisms. The most recent definition refers to re-assembling natural biological parts in unnatural ways to perform new biological functions (*22*). This application of synthetic biology has a close relationship with metabolic engineering where the goal is to introduce and tune expression of foreign pathways by assembling different biological components from various organisms. The "emergent properties" of synthetic systems are important to their design. This means the combination of all components is greater than the sum of performances of the individual parts (*23*). This phenomenon refers to bio-complexity and is studied by systems biology, which looks at interactions between genes and regulatory mechanisms. Simply adding genetic elements together is a far from effective construction of functional foreign pathways. Regulation of synthetic pathways becomes an important issue in metabolic engineering. This topic will be discussed in detail later.

Control is also a vital topic for synthetic pathways. Controlling a pathway first requires observing gene expression levels. Optimization can then be used to enhance/limit gene expression levels and, ultimately, full pathway usage. This means bottlenecks in pathway expression and usage can be detected and eliminated with a number of metabolic engineering strategies. Static and dynamic control are two specialized methods of pathway control (*24*). The basic elements of the static method are genetically encoded components which play key roles in determining level of gene expression. Various elements of static control include: (i) the strength

of a constitutive promoter and (ii) the type of constitutive promoters as well as (iii) the copy number of the vector (if applicable). Promoters are important "parts" of synthetic biology that exert static control. Promoter strength tuning can be achieved by manipulating the DNA sequence between -35 and -10 basepairs preceding the transcriptional start site. In addition, the ribosome-binding site (RBS) sequence is another tunable static element that contributes greatly to the translation rate of mRNA. Recently, Salis et al. (*25*) developed an algorithm to calculate protein expression level for synthetic RBS sequences. Dynamic control, on the other hand, involves a more complex strategy and exists in native metabolic pathways. The dynamic control mechanisms involve all feedback regulation. An application of dynamic control was shown in the improvement of lycopene production in *E. coli* K12. Farmer et al. (*26*) constructed an intracellular feedback loop to control expression of limiting enzymes in transformation steps from glucose to lycopene. In this loop, glucose was converted to lycopene and the side-product acetyl phosphate (ACP). Through regulation by the *Ntr* regulon (NRI) in *E.coli*, the expression level of ACP in cell can be sensed and this signal is then used to modulate the *glnAp2* promoter. This adjusts expression levels of the limiting enzymes of this pathway. Through this dynamic control strategy, the expression level of lycopene was increased 18-fold.

2.3. Rational methods of metabolic engineering

Rational metabolic engineering appeared as a more advanced approach when compared to the trial and error method. Typically, the goal of metabolic engineering is to (i) increase production of a desired chemical, (ii) reduce undesired byproduct synthesis, (iii) engineer utilization of new substrates, or (iv) increase cellular fitness to environmental conditions. Several steps are need to achieve a metabolic engineering goal: (i) identify critical branch points

in metabolic pathways which have great influence on product formation and (ii) determine how to modify these enzymes using gene inhibition, amplification, or deregulation (*27*). First, critical branch points in the metabolic network must be located. Three types of critical branch points exist: (i) flexible, (ii) weakly rigid, and (iii) strongly rigid. (*28*). For flexible branch points, flux through this branch is controlled by the terminal product through feedback inhibition. A weakly rigid branch point is one in which one branch dominates the flux distribution because of high enzymatic activity or lack of feedback inhibition. Strongly rigid branch points are the most difficult to manipulate. End-products in each branch of the pathway provide feedback inhibition. In this case, the flux distribution is relatively stable compared to the flexible and weakly rigid systems. This also means it is harder to redistribute flux flow.

Enzyme modification may change the metabolic flux distributions through these three kinds of branch points. For a flexible branch, this goal is achieved by (i) inhibiting the enzyme catalyzing the undesired pathway or by (ii) enhancing desired pathways enzymes. Attenuation of enzymatic activity in dominating pathways can redirect flux flow in weakly rigid branches. It is often more difficult to engineer strongly rigid branches. Each case needs specific consideration such as the alternation of substrate affinities (*28*). Next, the limiting step in metabolism must be found and strategies should be taken to overcome these limitations. This is an important strategy for linear pathways consisting of many enzymes because multiple enzymes can coordinate to limit the overall reaction rate. Many published examples have shown the success with this method. Ethanol production in *E. coli* and *Klebsiella* species was explored, and an example about a flexible branch is given here. Expression of pyruvate decarboxylase (PDC) from *Z. Mobilis* (*11*) can highly increase ethanol productivity. Enteric bacteria such as *Klebsiella* ferment D-xylose, a cheap and plentiful pentose, to mixed acids and butanediol in addition to

ethanol. But, the productivity of ethanol was observed to be very low. *Z. Mobilis* can produce ethanol with high yield by consuming hexose instead of xylose. Researchers transformed plasmids harboring PDC gene from *Z. mobilis* to the *Klebsiella* strain and were successful in increasing the ethanol production, from 9.4g/L to 25.4g/L.

2.4. Combinatorial approach to metabolic engineering

Certain cellular phenotypes can be determined by single gene, as mentioned previously with the heterogeneous expression of acetolactate synthase (ALS) in *E.coli* to reduce acetate excretion (*15*). However, in several cases, multiple genes are co-functional in determining a cellular response and resulting phenotype. The mechanism of solvent tolerance is a good example of a co-functional (or polygenic) trait. Through genomic-library enrichment of *Clostridium acetobutylicum,* Borden et al. (*29*) found that at least four transcriptional regulators are responsible for conveying butanol tolerance. After transformation of two regulators into strains, butanol tolerance increased by 81%. As shown in this case, balanced expression of co-functional genes is critical. To be successful with rational design, many rounds of modification of related genes need to be performed to achieve balance. As such, a phenotype-driven method has evolved. Thus, the combinatorial method of metabolic engineering is defined as the random perturbation of genes or other related targets followed by a coordinated screening of the resulting cells in a way that allows for selection of the desired trait. Once the desired phenotype is observed, further analysis is done to elucidate the genetic changes leading to the phenotype of interest. In the combinatorial approach, previous knowledge about the metabolic pathways or co-functional gene interactions is not necessary. Several tools for combinatorial engineering at the transcriptional level have been reviewed by Stephanopoulos and Keasling (*1, 30*). Tools

targeting the post-transcriptional level have also been designed by (i) adjusting regulation of mRNA or RNAse cleavage sites, (ii) adjusting RBS sequences to alter translation level, (iii) engineering mRNA stability, and (iv) altering translation initiation (*31*). Global perturbation is another widely used method of combinatorial metabolic engineering. In this method, random gene knockout and over-expression libraries are constructed using antisense RNA (asRNA) inhibition or over-expression genomic libraries harbored by plasmids, respectively. An example of using an asRNA leading to global perturbation is identifying important genes for growth of the fungal pathogen *Candida albicans*. Identification of its critical growth genes can help to locate new targets for drug-design. In this research, Backer et al. (*32*) constructed antisense complementary cDNA fragments under control of inducible promoter GAL1 into plasmids and transformed them into *C. albicans* to get strains with slower growth rate phenotype. An analysis revealed 86 genes important for growth and an additional 45 needing further exploration.

The cellular processes of transcription and translation are important targets of combinatorial metabolic engineering. The RNA polymerase component-like sigma factor was perturbed by Stephanopoulos' group to achieve new regulatory functions (*33*). The *rpoD* gene encoding for main sigma factor, sigma 70, was mutated randomly. Thus, the altered preference of the RNA polymerase to native promoters would affect the global transcription level of the entire genome. This random mutagenesis library was screened for desired phenotypes. Ethanol tolerant strains were selected and analyzed further. To engineer global translation, the ribosome is an alternative engineering target. Also, genome shuffling is another form of combinatorial metabolic engineering. When this method was applied to *Streptomyces fradiae*, the effect of two rounds genome shuffling yielded strains with production equivalent to strains that had undergone up to 20 rounds of sequential random mutagenesis (*34*). Tylosin, a complex polyketide antibiotic,

was the valuable commercial product being produced by this strain. When genome shuttling was applied, three rounds of screening were done to obtain an even higher tylosin producing strain. The first step in this project was random mutagenesis. After screening over 20,000 individual strains, 11 strains were selected for further experiments. Then they were prepared for protoplast and mixed in equal fractions. This step allowed genome shuttling to occur. Then screening was applied based on tylosin production in individual strains. Nine-fold increased tylosin production was observed compared to the wild-type strain and over 30% higher than strains from previous work. Although still relatively new, significant achievements have been observed through combinatorial metabolic engineering. After genetic perturbations are made, the choice of a screening method is the next important thing to consider. High-throughput screening and the growth competition assay are the two current methods. The high-throughput screen strategy involves screening of all possible individual strains while the growth competition assay utilizes growth pressure to select certain strains by using a "survival of the fittest" competition approach. Details about these selection strategies are discussed in a later section.

2.5. Inverse metabolic engineering (IEM)

The concept of inverse metabolic engineering (IEM) was first described by Jay Bailey (*35*). It is the opposite of constructive metabolic engineering, which refers to finding a metabolic flux-limiting step or bottleneck in a pathway and employing the rational method. In the early years of metabolic engineering, many approaches with the constructive method did not reach a desired result. This caused widespread thinking about a possible reason, and several limitations in the biological knowledge at the time were considered. Secondary responses of an engineered cell were considered first. It was hypothesized that changing native or introducing foreign

pathways could result in large effects on sub-networks of metabolism that were not considered by the rational design. It was also considered that some protein or enzyme activities were regulated in cells by post-translational modifications such as phospholrylation, leading to activation or inhibition. The altered activities were also not considered in constructive metabolic engineering. Different algorithms for choosing potential targets for engineering and IEM became an effective alternative method. The IEM consists of three steps: (i) construction or selection of strains that fit a desired phenotype, (ii) elucidate the mechanism resulting in desired phenotype in selected strains, (iii) apply certain genes or regulatory elements to other strains achieve the same phenotype. This approach turned out to be quite successful. Alleviation of oxygen limitation in *E. coli* was achieved by IEM. This improved cell growth and is an example of an early application utilizing IEM (*36*). *Vitreoscilla* synthesized more heme cofactor for a simple hemoglobin when under oxygen limited environment. So, research was done to find genes responsible for this phenomenon. After identifying the gene, it was transformed into *E. coli* and grown in a large-scale fermentor, where an oxygen limited environment occurred quite often. Recently, xylose uptake rate was enhanced by IEM to produce more ethanol in *S. cerevisiae* (*37*). Most work to enhance growth rate in yeast is to borrow genes from *Pichia stipitis.* In this research, a genomic library from *P. stipitis* was constructed and inserted into *S. cerevisiae*. Then, screening was applied to obtain strains grow faster supplied xylose as only carbon source. Two genes were identified to enhance growth rate and resulted better ethanol fermentation rate.

2.6. Genomic DNA libraries

A genomic library consists of random genomic DNA (gDNA) fragments from one or multiple bacterial species. These fragments are then cloned into an expression vector and expressed in a host species. The collection of cloned gDNA molecules represents over 99.9% genome of the source organism. The process of constructing a genomic DNA library is shown in Fig. 2.1. Endonuclease restriction enzymes (*37*), sonication (*38*), and vibration are some approaches of introducing genomic DNA into fragments of appropriate lengths. These fragments are then blunted using standard molecular biology techniques. Cloning fragments into the expression vector is a critical step in library construction. The most common and straightforward method involves blunt-end cloning. But, low cloning efficiency is recognized as a major problem for this method. Improved cloning methods have been developed in response. TOPO cloning method is alternative way developed from Invitrogen. Basically, genes of interest are first inserted into entry-vector that is TOPO vector, shown in Fig. 2.2. This commercial system is convenient, easy to execute, and can save much time compared to blunt end cloning. However, these commercial cloning kits act as a "black box" only produces and end result. Full details of the molecular biology details are not known and troubleshooting is difficult. The method of TA cloning is another choice for library cloning. The basic method of TA cloning is a ligation between inserts with 3' A overhangs and vectors with 5' T overhangs (*39*). The A or T base overhangs of inserts or plasmid are added using a *Taq* polymerase which has the ability to add a single nucleotide to 3' ends of blunt-end DNA fragments. With the A and T overhangs on the insert and vector, cloning efficiency can reach up to 90%.

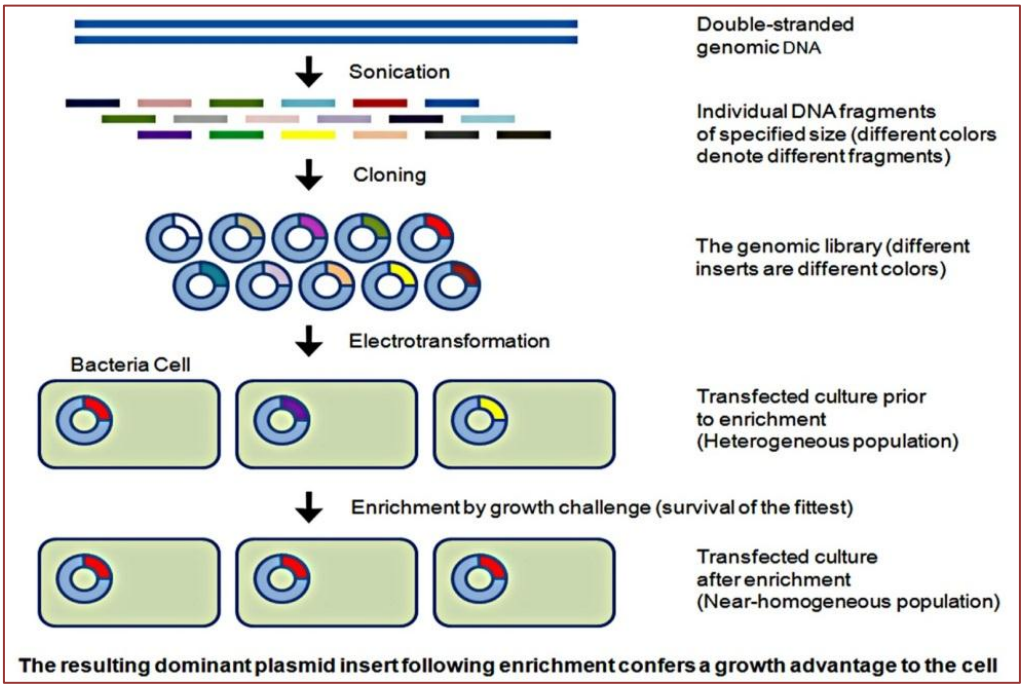

Fig 2.1. The process of genomic library construction.

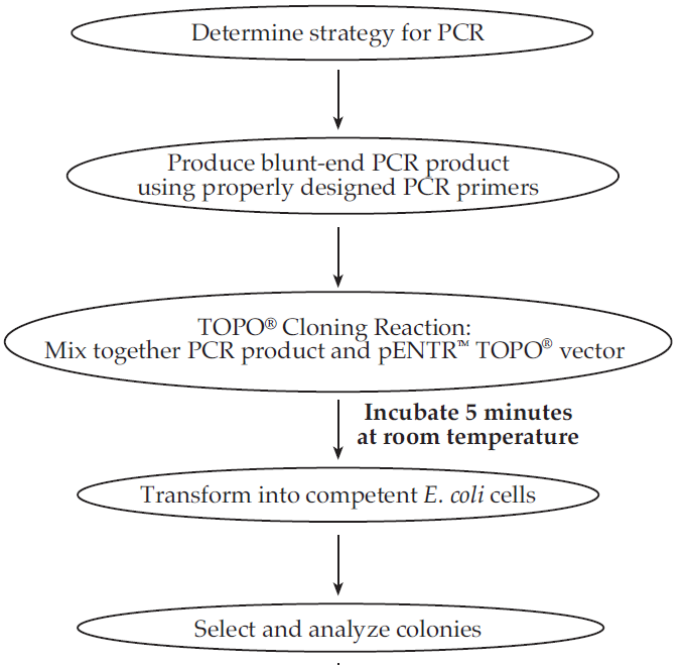

Fig 2.2. The process of TOPO cloning (Invitrogen).

2.7. Screening and library enrichment strategies

Once a genomic library is constructed and cloned into the host strain, executing a proper enrichment strategy is the next step. The proper enrichment approach is related to the screening strategy. Two categories of enrichment can be applied (i) with growth pressure or (ii) without growth pressure. A growth competition assay is the major tool used to apply enrichment with growth pressure. This tool requires an environmental pressure imposing on the strains. Only strains containing a gene plasmid with a gDNA insert that can resist this environment pressure can survive and become dominant in the culture. No extra screening operations are needed. Environment pressure can include: (i) solvent, (ii) pH, (iii) antibiotics, (iv) UV radiation, (v) temperature, and many others. Pine-Sol, an antibacterial compound, was used as selection pressure in identifying tolerance genes in *E. coli* using microarray technology (*38*). As mentioned previously, butanol tolerance can also act as cell growth pressure. Sixteen genes and 4 transcriptional regulators were identified in *Clostridium acetobutylicum* which affect butanol tolerance (*29*). For the condition of no growth pressure, vectors with library inserts cannot be enriched by competition. After the initial growth phase, a different screening strategy is needed to search for the desired phenotype. For these cases, high-throughput screening is applied to numerous strains grown as colonies on solid media based on some phenotype of colonies. Color is an easy feature to follow and is often applied to high-throughput screening. A high-throughput screen for hyaluronic acid (HA) production in *E. coli* is a good example of this method (*8*). HA is a mucopolysaccharide which can be stained by alcian blue, a water soluble copper-phthalocyanine dye to form blue color. The different color of an *E. coli* colony indicates the HA excretion amount. Sigma D and sigma S are subunits of RNA polymerase in *E. coli* encoded by *rpoD and rpoS.* These two genes were mutagenized randomly and a library was of these

randomly mutagenized genes was generated. These two genes of the library can affect HA transcription level. According to the different color of colonies, the one with high HA production can be harvested for further analysis. Another example refers to L-tyrosine production in *E. coli* strain (*5*). Tyrosine can form easily observable black, diffusible pigment through a series of reactions. So, the darkest colonies in plates were picked and analyzed. About 200 darkest colonies were analyzed and 30 needed second round analysis to determine genes in the darkest colonies. Basically, if the goal is to use a genomic library to increase expression of a desired product, a growth competition assay will not work because cells usually express product at the expense of the cellular growth rate. Thus, a high-producer will most likely not be enriched by a growth competition assay. As these two strategies are compared, it is clear that the growth competition assay offers significant advantages to enrichment. These include the absence of high-throughput equipment and they take less time, labor, and money. The limitation of the growth competition assay is also obvious. Here, growth pressure is needed for selection. This means only superior cellular fitness phenotypes can be selected in response to environmental variables. For over-producing amino acids, high-throughput screening becomes the only choice for selection.

Based on these observations, a new strategy to tie amino acids production to the specific growth rate of the cell has been developed in this research. Thus, a synthetic circuit has been created to ensure cells with maximum production of an amino acid will also show the maximum growth rate. The amino acid chosen for over-production in this research is L-arginine. This synthetic system is named Feedback Inhibition of Transcription for Growth Selection (FITSelect) and is discussed in detail in the following section.

2.8. FITSelect system Introduction

The FITSelect system is a novel synthetic biological system that enables the use of a growth competition assay to screen for cultures of a genomic library maximizing L-arginine production. The basic idea is to utilize transcriptional feedback to regulate gene expression of genes that control the growth rate of the cell. As shown in Fig. 2.3, the system is composed of several basic elements. In addition to those described below, a genomic library fragment is also included. The following major components, or "parts," are explained in greater detail.

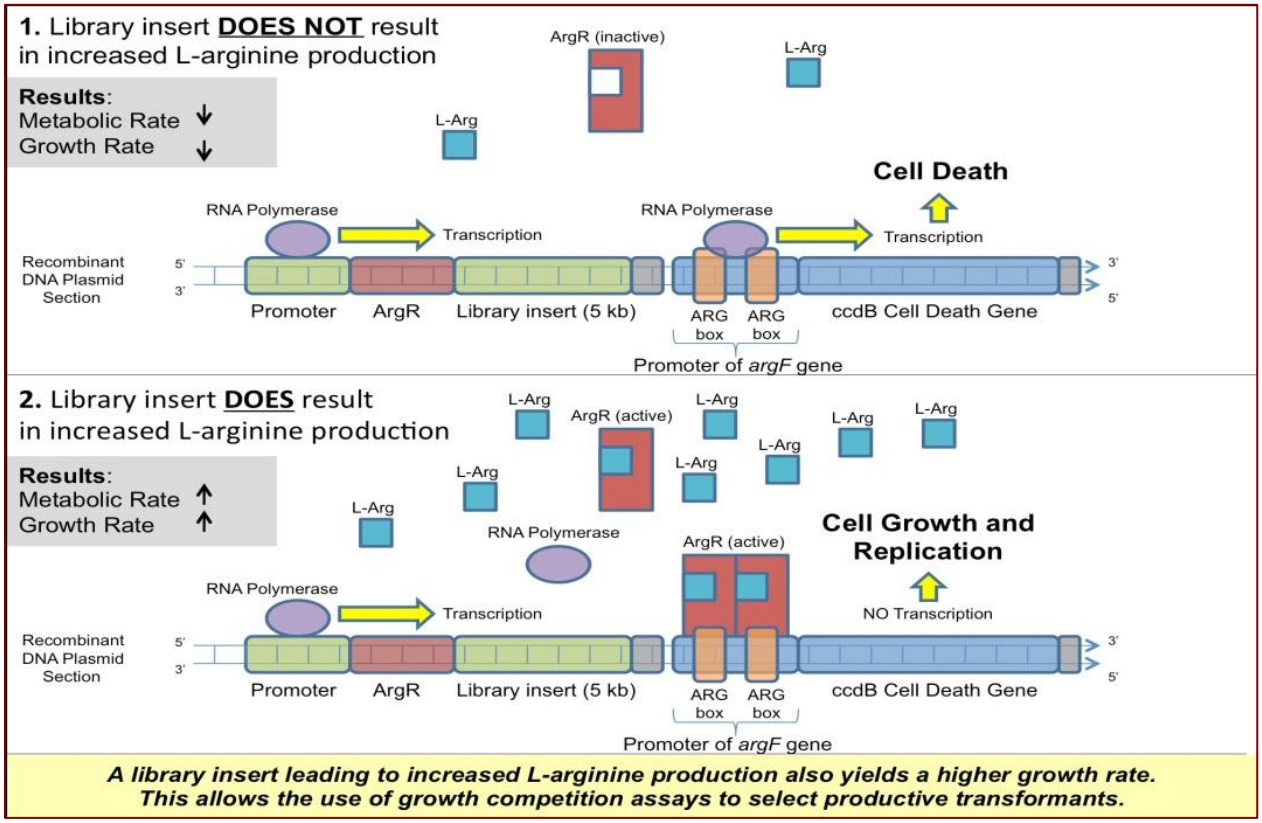

Fig 2.3. Mechanism of FITSelect system.

1. **The *ccdB* gene (or any gene that can kill the cell or decrease its growth rate).**

The function of this gene is to inhibit normal cell growth or result in cell death. The *ccdB* and *ccdA* genes are pair genes known as post-segregational killing systems.

This system consists of two components; poison (*ccdB*) and antidote (*ccdA*) (*40*). Its natural function relies on different stabilities of poison and antidote (*41*). The *ccdA* and *ccdB* pair of genes resides in the F plasmid which contains 3 different post-segregational systems, *flm*, *srn*, and *ccd* in pathogenic *E. coli* (*42-44*). The F plasmid is a bacterial DNA sequence determining sex of a bacterium and is responsible for conjugation. The *ccdB* gene targets gyrase in cells as its toxic mechanism. Gyrase belongs to topoisomerase II class of enzymes which is found in all cells and is essential to cell growth. The function of gyrase is to unwind supercoiled DNA at expense of ATP during duplication. The mechanism of this system is that when cells harbor the F plasmid, both genes are expressed. These two proteins bind as (CcdA)2-(CcdB)2 to form a complex *in vivo* to inactivate the CcdB toxin. When cells do not harbor the F plasmid, both proteins still exist in the cytoplasm but synthesis is stopped. In this case, the CcdA protein would be degraded by an ATP-dependent protease while the CcdB toxin protein is present and stable. Once CcdA does not bind to the CcdB protein, the CcdB binds to gyrase causing reduced DNA synthesis and an activated SOS system, ultimately leading to cell death (*41*, *45*).

2. *rrnB* terminators

The *rrnB* operon is known as one of seven transcription units for ribosome RNA (rRNA) and was the first one to have its genome sequence determined (*46*). The *rrnB* terminators are genetic structures that terminate transcription of this operon. The *rrnB* terminators contain two separate terminator sequences named t1 and t2. They are both rho-independent terminators. Two classes of terminators exist. Class I termination

requires formation of an RNA hairpin followed by U-rich sequence, depending on DNA template strand. Class II termination depends on a conserved sequence (CS) followed by a T-rich sequence. This causes the RNA polymerase to pause and change its conformation (*47, 48*). A T1 terminator contains these two signals and has a higher terminator efficiency than other terminators (~89%) (*49*).

3. ArgF promoter

The *argF* regulon was first raised by Maas and Clark (*50*) to describe "a system which the production of all enzymes can be controlled by a single repressor substance". Synthesis of L-arginine starts from glutamate and involves eight enzymatic steps and a total 10 enzymes (*51*). The regulon, however, involves a total of 11 genes that form 10 enzymes and 1 regulator protein. This regulator is the arginine repressor (ArgR) protein (*52*). The mechanism of ArgR regulation is related to the concentration of L-arginine *in vivo*. The ArgR repressor (aporepressor) and L-arginine (co-repressor) form a complex with DNA by binding to two conserved 18 bp elements named as ARG boxes. This acts to repress transcription (*53, 54*). The number of copies of the ArgR repressor protein has been measured in the *E. coli* cytoplasm and is about 330-510 copies (*52*).

4. *Trp* promoter

The *Trp* promoter is the promoter region of *trp* operon, which is responsible for L-tryptophan biosynthesis, and was first discovered by Jacques Monod in 1953 (*55*). Regulation involving the *trp* promoter is determined by the presence of L-tryptophan. Unlike the well-characterized *lac* promoter, which is activated by lactose present in cells,

the *trp* promoter is inactivated by high levels of L-tryptophan. Two L-tryptophan molecules bind to inactive a dimeric *trp* aporepressor protein (TrpR). This activates the repressor and results in ceased transcription in high tryptophan environments. Thus, active repressor binds to *trp* promoter to block the transcription of L-tryptophan producing genes (*56, 57*).

2.9. Mechanism of FITSelect system

The goal of FITSelect system is to explore an alternative way to replace complex high-throughput screening methods for enriching a genomic library to select for cells with maximized amino acids production. In this research, over-production of L-arginine is the goal. The approach is to construct a genomic library from *E. coli* gDNA then enrich and screen for strains with maximized L-arginine production. In this combinatorial strategy, the growth competition assay will be applied for selection instead of an expensive high-throughput screening method. Here, it is described how the FITSelect synthetic system can be used to tie the production of L-arginine to the specific growth rate, enabling the use of a growth competition assay. The FITSelect system can be separated into three parts, as shown in Fig. 2.3. The first part is responsible for expression of genomic library. In this part, the *pLac* promoter, native to the pUC19 expression plasmid, is responsible for controlling transcription of the cloned genomic library sequence. This transcription step is terminated by *rrnB* terminator. The second part is responsible for expression of *ccdB* gene to result in slowed cellular division and even cell death. This gene is controlled by the *argF* promoter, which contains two ARG boxes. The third part contains *ccdA* which uses the *trp* promoter to control its expression. It is also terminated by *rrnb* terminator. The reason to add this part is in the early phase of cell growth not enough L-arginine

is present to bind the ArgR protein and block transcription of the *ccdB* gene, resulting cell death of the entire culture. So, *ccdA* is expressed to inactivate CcdB gene in the early phases of the culture. After the initial lag phase, L-tryptophan will be added into media to inhibit the transcription the *ccdA* gene.

Two conditions exist with expression of the library sequence. The first (and least probable) situation is that the particular library sequence enhances the production of L-arginine (by some unknown mechanism). In this condition, over-expressed L-arginine will bind to the ArgR protein and activate the repression mechanism (increase its affinity to bind the ARG boxes). This binding will repress the expression of the *ccdB* gene and halt the cell death mechanism. Once the transcription of *ccdB* gene is inhibited, cells will reproduce normally and thrive in the culture. The second condition is when the library sequence does not result in an increase in L-arginine expression. This scenario will occur for over 99.9% of library constructs. In this case, because of a lack of intracellular L-arginine, the ArgR repressor protein will stay in inactive (unbound to L-arginine). Thus, the L-arginine promoter will remain functional and transcribe the *ccdB* gene. The active CcdB toxin will then bind to gyrase to interfere with duplication of DNA and result in cell death. So, when grown together in liquid culture, cells that do not harbor a plasmid with a sequence for improving production of L-arginine (by any mechanism) will die. Cells containing a beneficial sequence will grow normally and eventually dominate the culture. At this point, the dominant cells are further analyzed to identify the gDNA insert resulting in enhanced production of L-arginine.

2.10. References

1. Cameron, D. C., and Tong, I. T. (1993) Cellular and metabolic engineering. An overview, *Appl Biochem Biotechnol 38*, 105-140.
2. Santos, C. N., and Stephanopoulos, G. (2008) Combinatorial engineering of microbes for optimizing cellular phenotype, *Curr Opin Chem Biol 12*, 168-176.
3. Stephanopoulos, G. (1999) Metabolic fluxes and metabolic engineering, *Metab Eng 1*, 1-11.
4. Tolan, J. S., and Finn, R. K. (1987) Fermentation of d-Xylose to Ethanol by Genetically Modified Klebsiella planticola, *Appl Environ Microbiol 53*, 2039-2044.
5. Ohta, K., Beall, D. S., Mejia, J. P., Shanmugam, K. T., and Ingram, L. O. (1991) Genetic improvement of Escherichia coli for ethanol production: chromosomal integration of Zymomonas mobilis genes encoding pyruvate decarboxylase and alcohol dehydrogenase II, *Appl Environ Microbiol 57*, 893-900.
6. Brabetz, W., Liebl, W., and Schleifer, K. H. (1991) Studies on the utilization of lactose by Corynebacterium glutamicum, bearing the lactose operon of Escherichia coli, *Arch Microbiol 155*, 607-612.
7. Nawrath, C., Poirier, Y., and Somerville, C. (1994) Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of Arabidopsis thaliana results in high levels of polymer accumulation, *Proc Natl Acad Sci U S A 91*, 12760-12764.
8. Aristidou, A. A., San, K. Y., and Bennett, G. N. (1994) Modification of central metabolic pathway in escherichia coli to reduce acetate accumulation by heterologous expression of the bacillus subtilis acetolactate synthase gene, *Biotechnol Bioeng 44*, 944-951.

9. Yarmush, M. L., and Berthiaume, F. (1997) Metabolic engineering and human disease, *Nat Biotechnol 15*, 525-528.

10. Bailey, J. E., Birnbaum, S., Galazzo, J. L., Khosla, C., and Shanks, J. V. (1990) Strategies and challenges in metabolic engineering, *Ann N Y Acad Sci 589*, 1-15.

11. Bailey, J. E. (1991) Toward a science of metabolic engineering, *Science 252*, 1668-1675.

12. Anderson, S., Marks, C. B., Lazarus, R., Miller, J., Stafford, K., Seymour, J., Light, D., Rastetter, W., and Estell, D. (1985) Production of 2-Keto-L-Gulonate, an Intermediate in L-Ascorbate Synthesis, by a Genetically Modffied Erwinia herbicola, *Science 230*, 144-149.

13. Hobom, B. (1980) [Gene surgery: on the threshold of synthetic biology], *Med Klin 75*, 834-841.

14. Rawis, R. L. (2000) 'Synthetic Biology' Makes Its Debut, *Chemical & Engineering News 78*, 49-53.

15. Ball, P. (2004) Synthetic biology: starting from scratch, *Nature 431*, 624-626.

16. Woese, C. R. (2004) A new biology for a new century, *Microbiol Mol Biol Rev 68*, 173-186.

17. Holtz, W. J., and Keasling, J. D. (2010) Engineering static and dynamic control of synthetic pathways, *Cell 140*, 19-23.

18. Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression, *Nat Biotechnol 27*, 946-950.

19. Farmer, W. R., and Liao, J. C. (2000) Improving lycopene production in Escherichia coli by engineering metabolic control, *Nat Biotechnol 18*, 533-537.

20. Stephanopoulos, G., and Sinskey, A. J. (1993) Metabolic engineering--methodologies and future prospects, *Trends Biotechnol 11*, 392-396.

21. Stephanopoulos, G., and Vallino, J. J. (1991) Network rigidity and metabolic engineering in metabolite overproduction, *Science 252*, 1675-1681.

22. Borden, J. R., and Papoutsakis, E. T. (2007) Dynamics of genomic-library enrichment and identification of solvent tolerance genes for Clostridium acetobutylicum, *Appl Environ Microbiol 73*, 3061-3068.

23. Keasling, J. D. (2008) Synthetic biology for synthetic chemistry, *ACS Chem Biol 3*, 64-76.

24. Pfleger, B. F., Pitera, D. J., Smolke, C. D., and Keasling, J. D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes, *Nat Biotechnol 24*, 1027-1032.

25. De Backer, M. D., Nelissen, B., Logghe, M., Viaene, J., Loonen, I., Vandoninck, S., de Hoogt, R., Dewaele, S., Simons, F. A., Verhasselt, P., Vanhoof, G., Contreras, R., and Luyten, W. H. M. L. (2001) An antisense-based functional genomics approach for identification of genes critical for growth of Candida albicans, *Nat Biotech 19*, 235-241.

26. Alper, H., and Stephanopoulos, G. (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype, *Metabolic Engineering 9*, 258-267.

27. Zhang, Y. X., Perry, K., Vinci, V. A., Powell, K., Stemmer, W. P., and del Cardayre, S. B. (2002) Genome shuffling leads to rapid phenotypic improvement in bacteria, *Nature 415*, 644-646.

28. Bailey, J. E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W. A., and Tsai, P. S. (1996) Inverse metabolic engineering: A strategy for directed genetic engineering of useful phenotypes, *Biotechnol Bioeng 52*, 109-121.

29. Khosla, C., and Bailey, J. E. (1988) Heterologous expression of a bacterial haemoglobin improves the growth properties of recombinant Escherichia coli, *Nature 331*, 633-635.

30. Jin, Y. S., Alper, H., Yang, Y. T., and Stephanopoulos, G. (2005) Improvement of xylose uptake and ethanol production in recombinant Saccharomyces cerevisiae through an inverse metabolic engineering approach, *Appl Environ Microbiol 71*, 8249-8256.

31. Gill, R. T., Wildt, S., Yang, Y. T., Ziesman, S., and Stephanopoulos, G. (2002) Genome-wide screening for trait conferring genes using DNA microarrays, *P Natl Acad Sci USA 99*, 7033-7038.

32. Zhou, M. Y., and Gomez-Sanchez, C. E. (2000) Universal TA cloning, *Curr Issues Mol Biol 2*, 1-7.

33. Yu, H., Tyo, K., Alper, H., Klein-Marcuschamer, D., and Stephanopoulos, G. (2008) A high-throughput screen for hyaluronic acid accumulation in recombinant Escherichia coli transformed by libraries of engineered sigma factors, *Biotechnology and Bioengineering 101*, 788-796.

34. Santos, C. N., and Stephanopoulos, G. (2008) Melanin-based high-throughput screen for L-tyrosine production in Escherichia coli, *Appl Environ Microbiol 74*, 1190-1197.

35. Bex, F., Karoui, H., Rokeach, L., Dreze, P., Garcia, L., and Couturier, M. (1983) Mini-F encoded proteins: identification of a new 10.5 kilodalton species, *EMBO J 2*, 1853-1861.

36. Van Melderen, L. (2002) Molecular interactions of the CcdB poison with its bacterial target, the DNA gyrase, *Int J Med Microbiol 291*, 537-544.

37. Loh, S. M., Cram, D. S., and Skurray, R. A. (1988) Nucleotide sequence and transcriptional analysis of a third function (Flm) involved in F-plasmid maintenance, *Gene 66*, 259-268.

38. Golub, E. I., and Panzer, H. A. (1988) The F factor of Escherichia coli carries a locus of stable plasmid inheritance stm, similar to the parB locus of plasmid RI, *Mol Gen Genet 214*, 353-357.

39. Ogura, T., and Hiraga, S. (1983) Mini-F Plasmid Genes That Couple Host-Cell Division to Plasmid Proliferation, *P Natl Acad Sci-Biol 80*, 4784-4788.

40. Miki, T., Yoshioka, K., and Horiuchi, T. (1984) Control of cell division by sex factor F in Escherichia coli. I. The 42.84-43.6 F segment couples cell division of the host bacteria with replication of plasmid DNA, *J Mol Biol 174*, 605-625.

41. Brosius, J., Dull, T. J., Sleeter, D. D., and Noller, H. F. (1981) Gene organization and primary structure of a ribosomal RNA operon from Escherichia coli, *J Mol Biol 148*, 107-127.

42. Hartvig, L., and Christiansen, J. (1996) Intrinsic termination of T7 RNA polymerase mediated by either RNA or DNA, *EMBO J 15*, 4767-4774.

43. Sohn, Y., and Kang, C. (2005) Sequential multiple functions of the conserved sequence in sequence-specific termination by T7 RNA polymerase, *Proc Natl Acad Sci U S A 102*, 75-80.

44. Nojima, T., Lin, A. C., Fujii, T., and Endo, I. (2005) Determination of the termination efficiency of the transcription terminator using different fluorescent profiles in green fluorescent protein mutants, *Anal Sci 21*, 1479-1481.

45. Maas, W. K. (1964) Studies on the Mechanism of Repression of Arginine Biosynthesis in Escherichia Coli. Ii. Dominance of Repressibility in Diploids, *J Mol Biol 8*, 365-370.

46. Caldara, M., Charlier, D., and Cunin, R. (2006) The arginine regulon of Escherichia coli: whole-system transcriptome analysis discovers new genes and provides an integrated view of arginine regulation, *Microbiology 152*, 3343-3354.

47. Maas, W. K. (1994) The arginine repressor of Escherichia coli, *Microbiol Rev 58*, 631-640.

48. Cunin, R., Eckhardt, T., Piette, J., Boyen, A., Pierard, A., and Glansdorff, N. (1983) Molecular basis for modulated regulation of gene expression in the arginine regulon of Escherichia coli K-12, *Nucleic Acids Res 11*, 5007-5019.

49. Charlier, D., Roovers, M., Van Vliet, F., Boyen, A., Cunin, R., Nakamura, Y., Glansdorff, N., and Pierard, A. (1992) Arginine regulon of Escherichia coli K-12. A study of repressor-operator interactions and of in vitro binding affinities versus in vivo repression, *J Mol Biol 226*, 367-386.

50. Marmorstein, R. Q., Joachimiak, A., Sprinzl, M., and Sigler, P. B. (1987) The Structural Basis for the Interaction between L-Tryptophan and the Escherichia-Coli Trp Aporepressor, *Journal of Biological Chemistry 262*, 4922-4927.

51. Bass, S. H., and Yansura, D. G. (2000) Application of the E. coli trp promoter, *Mol Biotechnol 16*, 253-260.

3. Materials and Methods

3.1. Strains, plasmids and culture media

Escherichia coli Top 10 strain (DH10B, F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G λ-) (Invitrogen) was used for construction, cloning, and expression of the FITSelect synthetic circuit. *E. coli ccdB* survival strain (F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*Δ139Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1*nup*G *fhuA::IS2*) (Invitrogen) is resistant to the *ccdB* cell death gene and was used to clone and express plasmids that had an active *ccdB* gene. The high-copy number pUC19 plasmid (Invitrogen) and low-copy number pACYC177 plasmid (New England Biolabs) were used for expression of the FITSelect constructs. Approximately 500~700 copies of the pUC19 plasmid exist in the cell because of its lack of the *rop* gene and a single point mutation in *rep* of pMB1 origin, while the *rep* of p15A origin of replication of the pACYC177 plasmid ensures approximately 10 copies are maintained in the cell. Lysogeny broth (LB) media and M9 media was used for cell growth. LB media was formulated as follows.

1. 10 g tryptone, 5g yeast extract and 10g NaCl (15g agar if prepare solid plates) dissolved in 800 mL Type 1 purity water.
2. Adjust pH to 7.0 and fill to 1 L with Type 1 water.
3. Autoclave and cool to 55 °C.
4. Add antibiotics in proper concentration. 100ug/ml of Ampicillin or 25ug/ml of chloramphenicol is added into media.

5. For solid plates, plate the liquid media into each plate with about 25ml media and wait liquid media to solid then seal with parafilm.
6. Store liquid media and plates at +4 ℃.

M9 minimal media was formulated as follows.

1. 64g Na_2HPO_4-$7H_2O$, 15g KH_2PO_4, 2.5g NaCl, 5.0g NH_4Cl dissolved in 800ml Type 1 purity water.
2. Adjust pH to 7.0 and fill to 1 L with Type 1 water to form M9 salt solution.
3. Autoclave and cool down
4. 200ml M9 salt solution, 2ml of 1M $MgSO_4$, 20 ml of 20% glucose, 100ul of 1M $CaCl_2$ dissolved in 700ml Type 1 purity water (all sterile ingredients)
5. Add antibiotics in proper concentration. Usually, 100ug/ml of Ampicillin or 25ug/ml of chloramphenicol is added into media.
6. For solid plates, 200ml M9 salt solution and 15g agar dissolved in 800ml Type 1 purity water
7. Autoclave and cool to 55 °C and add antibiotics in proper concentration as described before.
8. Plate the liquid media into each plate with about 25ml media and wait liquid media to solid then seal with parafilms.
9. Store liquid media and plates at +4 ℃.

3.2. Molecular biology protocols

The polymerase chain reaction (PCR) was used to amplify target DNA sequences from template DNA using an Eppendorf mastercycler gradient (Thermo) and matercycler personal

thermal cycler (Thermo). *E.coli* DH10B genomic DNA (gDNA) was extracted using a Generation Capture Kit (Promega) according to the manufactuer's protocol.The proof-reading VentR polymerase (New England Biolabs) was used for gene amplification for cloning, and the Taq polymerase (New England Biolabs) was used for all other PCR applications (e.g., checking cloning results). All PCR reagents were obtained from New England Biolabs, and all primers were synthesized by Integrated DNA Technologies. PCR primers were designed using Primer3Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). The basic procedure of traditional PCR includes (i) initialization, (ii) denaturation, (iii) annealing, (iv) extension, and (v) a final elongation step. Touch-down PCR was used to amplify target genes in this research. The main difference between traditional and touch-down PCR is the annealing temperature varies every 5 cycles. This procedure helps avoid non-specific hybridization and allows for slight inaccuracies in primer-DNA melting temperature calculations. Specific PCR protocols for individual genes and genetic parts are given in Section 3.4. PCR products were purified using a PCR purification kit (QIAGEN) (Cat. No 28104) according to manufacturer's protocol. In selected cases, gel-band purification was performed using SYBR-safe dye (Invitrogen) and a kit (QIAGEN) (Cat. No 28704) according to the manufacturer's protocol.

Following the PCR reaction, products were analyzed by gel electrophoresis using a 0.8% gel containing 0.28g agarose and 35ml TAE buffer. Gels were stained with ethidium bromide and imaged under UV light using a gel imager (Biorad). Traditional directional cloning was used to construct cassettes and clone individual genes. Primers were designed containing restriction cut sites. Proper restriction enzymes (New England Biolabs) and corresponding buffer were added to digest PCR products for cloning. Most digestions were performed at 37°C for 3 hours. Ligation to construct the FITselect cassettes and clone into plasmids was performed

with T4 DNA ligase (New England Biolabs). In most cases, a 1:1 molar ratio of DNA/plasmid (or of DNA products) was used in the ligation reaction, and incubation was performed at 16°C overnight. In some cases, the PCR product was ligated directly into a plasmid using a PCR product to plasmid ratio 2:1.

Constructed plasmids were then transformed into chemically competent *E. coli* DH10B (Top 10) cells (see above). The following general steps were followed for the transformation. First, 10pg to 10ng of plasmid was added to one vial of competent cells. The mixture was incubated on ice for 30 minutes. Then, the culture was heat shocked at 42 ℃ for 30 seconds. Then, 250 mL of SOC medium (1 liter containing 20g tryptone, 5g yeast extract, 2ml of 5M NaCl, 2.5ml of 1M KCl, 10ml of 1M $MgCl_2$, 10ml of 1M $MgSO_4$, 20ml of 1M glucose) was added to the culture, and the mixture was incubated for 1 hour at 37 ℃ under 200 rpm agitation in an incubator. The transformation procedure was completed by spreading 50 to 200 μL of the cell mixture on solid media plates and incubating the culture overnight at 37 ℃. Plasmids were replicated in *E. coli* DH10B (Top 10) cells in LB media and were extracted using a Miniprep kit (Gerard Biotech) according to the manufacturer's protocol.

3.3. Description of plasmids constructed

The FITSelect constructs cloned into the pUC19 and the pACYC177 plasmids are shown in Fig. 3.1. Two separate cassettes are required of the FITSelect circuit, which is designed to select genomic library inserts enhancing L-arginine production. The components of each cassette and the details of enzymes cutting sites are show in Figs. 3.2 and 3.3. The cassettes were constructed in a piecewise manner so the individual functions of the components could be tested individually. These are given in Table 3.1.

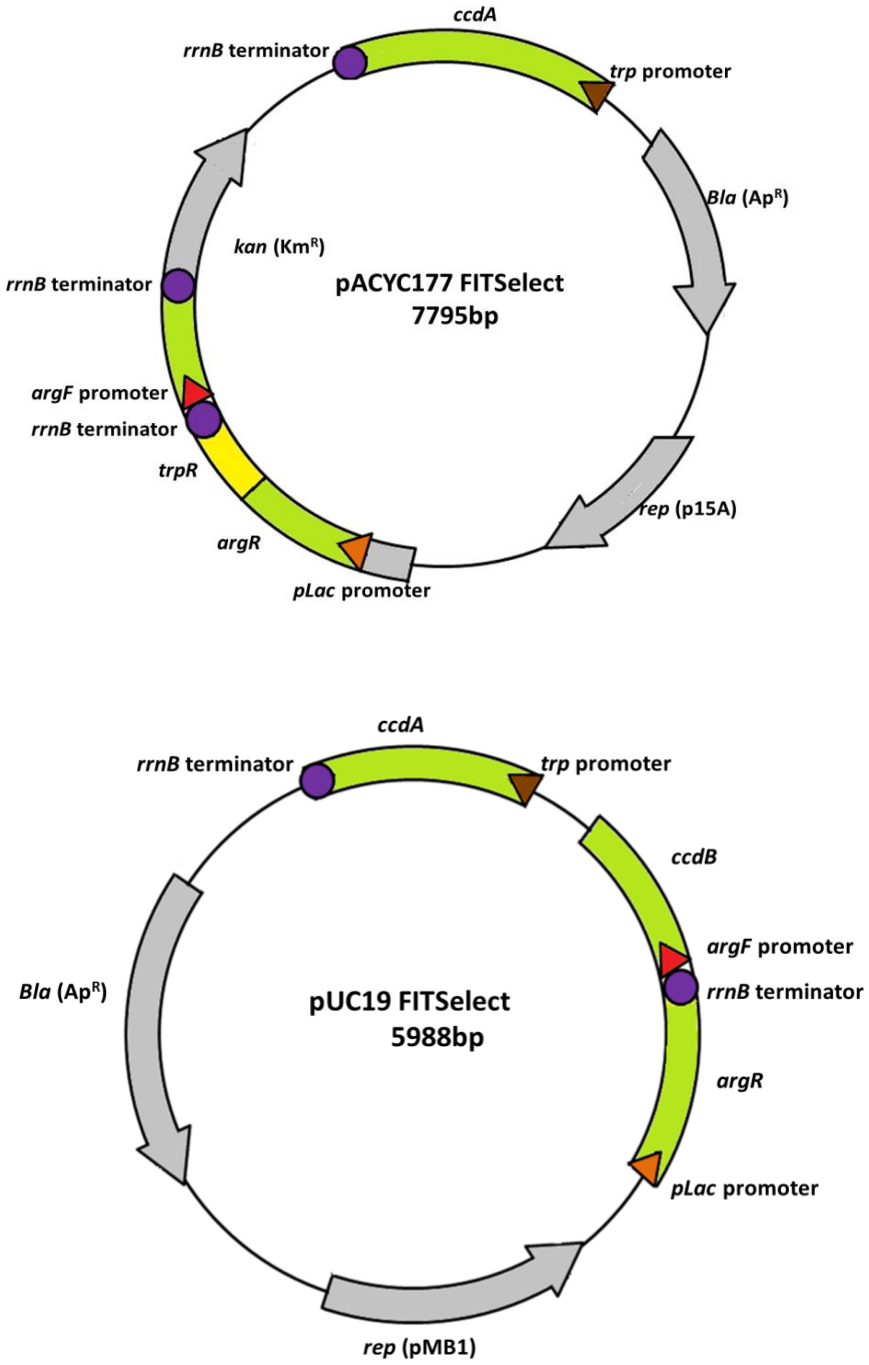

Fig. 3.1. The FITSelect constructs cloned into the pUC19 and pACYC177 plasmids.

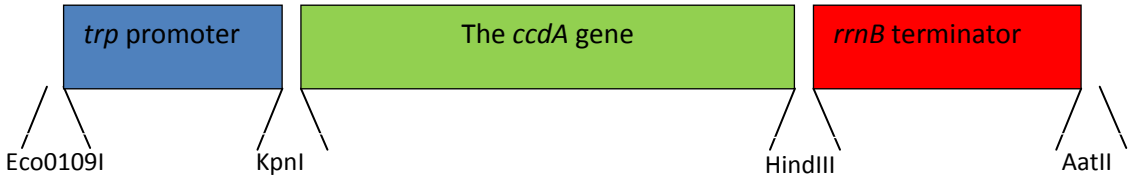

Fig. 3.2. Components of the *ccdA* cassette with enzyme cut-sites.

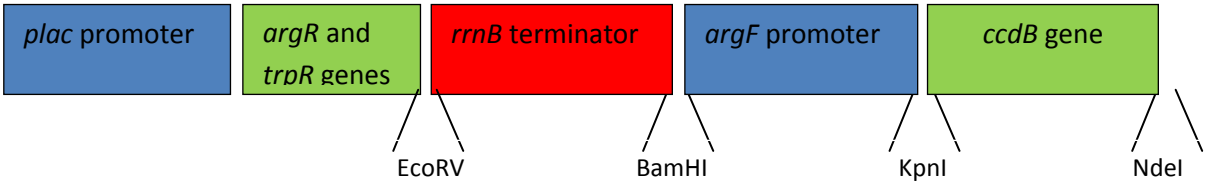

Fig. 3.3. Components of the *ccdB* cassette with enzyme cut-sites.

Table 3.1. List of all plasmids constructed for this research.

Name	**Components of FITSelect contained in the plasmid**
plcB	*ccdB*, ArgF promoter, rrnB terminator (pUC19)
plcBaR	*ccdB*, ArgF promoter, rrnB terminator, *argR* (pUC19)
plcBaRrBaP-	*ccdB*, *argR* (pUC19)
plcBaRaP-	*ccdB*, rrnB terminator, *argR* (pUC19)
plcBaRcB-	ArgF promoter, rrnB terminator, *argR* (pUC19)
plrbscBaR	rbs *ccdB*, ArgF promoter, rrnB terminator, *argR* (pUC19)
plcBaRcA	*ccdB*, ArgF promoter, rrnB terminator, *argR*, trp promoter, *ccdA* (pUC19)
pACYCcBaR	*ccdB*, ArgF promoter, rrnB terminator, *argR* (pACYC177)
pACYCcBaRcA	*ccdB*, ArgF promoter, rrnB terminator, *argR*, trp promoter, *ccdA* (pACYC177)
pACYCcBaRcAtR	*ccdB*, ArgF promoter, rrnB terminator, *argR*, *trpR*, trp promoter, *ccdA* (pACYC177)

3.4. PCR amplification of: (i) *argF* promoter region, (ii) *rrnB* terminator, (iii) *ccdB* gene, (iv) *argR* repressor, (v) *trp* promoter region, and (vi) the *trpR* repressor

All PCR primers designed and used in this research are shown in Table 3.2. The gDNA PCR templates for (i) the *rrnB* transcriptional terminator, (ii) the *argF* promoter region, (iii) the

argR repressor, (iv) the *trp* promoter region, and (v) the *trpR* repressor were *E. coli* DH10B Top10 (Invitrogen). The *ccdB* gene, with the chromophenocal resistance gene, was amplified from the pDEST14 plasmid of the TOPO Isomerase cloning system (Invitrogen). The *ccdA* gene was synthesized by Integrated DNA Technologies.

Table 3.2. PCR primers used to amplify the rrnB terminator, argF promoter, argR repressor, ccdB cell death gene, rbs ccdB gene, trp promoter, and trpR repressor. The enzyme cut sites in the primer sequences are highlighted (green). Underscores are used to separate enzyme cut sites, linkers, and primer sequences.

Primer Name	**Gene**	**Enzyme Cutting Sites**	**Sequence**
rrnB_left_HindIII EcoRV (256):	*rrnB* terminator	HindIII; EcoRV	AATCT_AAGCTT_AATCAGT CAA_GATATC_ATGCGAGAG TAGGGAACT
rrnB_right_bamHI	*rrnB* terminator	BamHI	TCTAA_GGATCC_AGTTTGTA GAAACGCAAAA
ArgFpr4_left_bam HI (128)	*argF* promoter	BamHI	TCTAA_GGATCC_GCAGATA CGATTATTTTCAC
ArgFpr4_right_ kpnI	*argF* promoter	KpnI	ATTAT_GGTACC_GCGAACG CCACTTAT
ccdB2_left_KpnI (1866)	*ccdB*	KpnI	ATTAT_GGTACC_TCCCTCTA GATCACAAGTTT
ccdB3_right_NdeI	*ccdB*	NdeI	TATTA_CATATG_GGATCATC ACCACTTTGTA
ccdB rbs_left_kpnI (411)	rbs *ccdB*	KpnI	ATTAT_GGTACC_TGCCCGG AGTATCCGTTCAGTTCAAAG AGAGGC_ATGCAGTTTAAGG TTTACA
ccdB rbs_right_ndeI		NdeI	TATTA_CATATG_AACACAA CATATCCAGTCAC
Trppr2_left Eco0109I (166)	*trp* promoter	Eco0109I	ATTAT_AGGACCT_GTAAAT CACTGCATAATTCG
Trppr2_right_KpnI	*trp* promoter	KpnI	ACTCA_GGTACC_TGTCGAT ACCCTTTTTACG
ArgR_left (594)	*argR* repressor		GTTGTATCAACCACCATATC
ArgR_right	*argR* repressor		AACGCACTATTCGTTAAAT
TrpR left (352)	*trpR* repressor		GACATATTATGGCCCAAC
TrpR right	*trpR* repressor		TTATCAGGCCTACAAAATC

A touch-down PCR method was used for amplification. Denaturing and annealing times of 1 min were applied to all PCR amplifications. Specific conditions used for *rrnB* terminator amplification were: (i) 5 cycles at 60°C, (ii) 5 cycles at 58°C, (iii) 5 cycles at 56°C, and (iv) 15 cycles at 54°C. The extension time was 20 seconds to reduce non-specific bands with higher numbers of base pairs. PCR conditions used for the *argF* and *trp* promoter region amplification were: (i) 5 cycles at 54°C, (ii) 5 cycles at 52°C, (iii) 5 cycles at 50°C and (iv) 15 cycles at 48°C. The extension time was also 20 seconds to reduce non-specific bands. PCR conditions for the *ccdB* gene amplification were: (i) 5 cycles at 60°C, (ii) 5 cycles at 58°C, (iii) 5 cycles at 56°C and (iv) 15 cycles at 54°C. An extension time of 2 min was used. PCR conditions for *argR* and *trpR* repressor genes were: (i) 5 cycles at 60°C, (ii) 5 cycles at 58°C, (iii) 5 cycles at 56°C and (iv) 15 cycles at 54°C. An extension time of 1 min was used. PCR conditions for rbs *ccdB* genes were: (i) 5 cycles at 54°C, (ii) 5 cycles at 52°C, (iii) 5 cycles at 50°C and (iv)15 cycles at 48°C. An extension time of 1 min was used.

3.5. Digestion and ligation of multiple genes into a cassette.

The PCR products of the (i) *rrnB* transcriptional terminator, (ii) the *argF* promoter, and (iii) the *ccdB* cell death gene (see Fig. 3.3) were ligated together to form a single product before cloning into the pUC19 plasmid. The following enzymes (New England Biolabs) were used: Antarctic phosphatase (M0289S), HindIII (R0104S), BamHI (R0136S), KpnI (R0142S), NdeI (R0111S), EcoRV (R0195S) and T4 DNA ligase (M0202S). PCR products were digested with the proper restriction enzymes according to the manufacturer's protocol (3 hours incubation at 37°C). The *rrnB* terminator and *ccdB* gene product were digested with BamHI and KpnI, respectively, and treated with antarctic phosphatase to avoid blunt-end self-ligation at the "non-

sticky" end. The *argF* promoter was also digested with BamHI and KpnI (3 hours incubation at 37°C). Digested products were then purified using the QIAGEN PCR purification kit. The three purified products were ligated using the T4 DNA ligase with a molar ratio of PCR products of 1:1:1 overnight at 16°C. The ligation products were identified by gel electrophoresis (using SYBR Safe (Invitrogen) dye) and the "correctly sized" ligation product was extracted using gel band purification (QIAGEN). Following ligation of three PCR products into a single product, the restriction enzymes HindIII and NdeI were used to digest this product (3 hours incubation at 37°C). This product was then cloned into the pUC19 plasmid as described in Section 3.6.

The PCR products of *trp* promoter, *rrnB* terminator (see Fig. 3.2) were ligated with the *ccdA* cell death antidote gene to form a single product. The *ccdA* gene was synthesized by Integrated DNA Technologies, provided in a pIDTSmart plasmid with ampicillin resistance, and was excised using KpnI and HindIII restriction enzymes. To build the cassette, the *trp* promoter was first digested with KpnI, and the *rrnB* terminator was cut with the HindIII restriction enzyme (3 hours incubation at 37°C). Then, the *trp* promoter and *rrnB* terminator were treated with Antarctic phosphatase (New England Biolabs) to block ligation at their blunt ends. The three cut PCR products were purified using a PCR purification kit (QIAGEN) and ligated using T4 DNA ligase (New England Biolabs) with a 1:1:1 molar ratio (overnight incubation at 16°C), according to the manufacturer's protocol. Following creation of the cassette, it was treated with Eco0109I and AatII restriction enzymes (3 hours incubation at 37°C), purified (QIAGEN PCR purification kit), and cloned into the pUC19 plasmid as described below.

3.6. Ligation of constructs into the pUC19 plasmid and transformation into ccdB resistant E. coli K12 (Top10) cells

Initially, 10pg~10ng of the common pUC19 plasmid was replicated by transforming into Top10 cells using the manufacturer's protocol. Cultures were initially grown on solid media, and colonies were transferred to liquid media (containing ampicillin) and grown at 37°C. Cultures were harvested when the optical density at 600 nm (OD600) value reached 1.2 (during mid-exponential growth). Plasmid extraction was performed using a Miniprep kit. This yielded a high concentration of pure pUC19 plasmid. To insert the *ccdB* cassette (Fig. 3.2), about 6 μg of pUC19 plasmid was digested with HincIII and NdeI (3 hours incubation at 37°C). Digested pUC19 plasmid and the *ccdB cassette* were ligated using T4 DNA ligase at 16°C overnight using a plasmid to cassette molar ratio 1:2. This plasmid was labeled the plcB plasmid (see Table 3.1).

After the ligation, the new plasmid plcB harboring the 3 inserts was transformed into Top10 *ccdB* resistant chemically competent *E. coli* DH10B. Cells were spread on LB plates containing ampicillin antibiotic and IPTG with X-gal (BP4200-10) (Fisher Scientific). The ITPG is required to induce the pLac promoter of the pUC19 plasmid. X-gal is used to detect the presence of β-galactosidase and turns blue in color in its presence. The *ccdB* cassette was inserted into the middle of a *lacZ* gene (encoding β unit of β-galactosidase). With successful cloning, *lacZ* is not functional and colonies remain white in color. Unsuccessful cloning yields blue colonies. Upon cloning the *ccdB* cassette, white colonies were found to contain all 3 inserts. Colony PCR using the Taq polymerase was performed to check the insert size. Primers for this "checking" procedure are listed in Table 3.3. To perform the colony PCR procedure, a single colony was dissolved in 20 μL Type 1 purity water. Then, 2 μL was used as the DNA template for a 50 μL PCR reaction. A touch-down method was used at temperatures of (i) 54°C, (ii) 52°C,

and (iii) 50°C for 5 cycles each and 48°C for 15 cycles. A 3 min extension time was used, and the remaining PCR parameters were the same as described above. The target PCR product size for the plcB plasmid was about 2700 bp.

The method of TA cloning was applied to clone the *argR* gene into plcB plasmid to form the plcBaR plasmid (see Table 3.1). The plcB plasmid was digested with HindIII (3 hours incubation at 37°C). Digested plasmid was purified with the using a kit as described previously. Purified plasmid with "sticky ends" was blunt ended using a T4 polymerase (M0203S, New England Biolabs). Then blunt ended plasmid was treated with terminal transferase (Roche) (1.5 hours incubation at 37°C) to add T-tails to the plasmid. Then the mixture was purified with the PCR purification kit. The Taq polymerase and dNTP mix was added into purified *argR* PCR product to add A-tails to this product (25 min incubation at 72°C). A 1:2 molar ratio of T-tail plcB plasmid to *argR* gene was ligated via T4 ligase at 16°C overnight.

After the ligation, new plasmid plcBaRwas transformed into Top 10 chemical competent cells. The transformed cells were spread on LB plates containing ampicillin antibiotic. Colonies were picked out to be tested by colony PCR as described above. The primers used were the same for checking *ccdB* cassette while the extension time was 1 min instead of 3min. The target PCR product was about 3000bp.

The plcBaRcA plasmid was constructed by inserted *ccdA* cassette (see Fig. 3.3) into plcBaR plasmid (see Table 3.1). Both the *ccdA* cassette and plcBaR plasmid were digested with restriction enzymes Eco0109I and AatII (3 hours incubation at 37°C). Digested products were purified using the PCR purification kit. Ligation with T4 DNA ligase was performed to ligate *ccdA* cassette into the cut plcBaR plasmid. The molar ratio between cassette and plasmid was 2:1. The ligation mixture was incubated at 16°C overnight.

After the ligation, the new plasmid, plcBaRcA, was transformed into Top10 *ccdB* resistant chemically competent *E. coli* DH10B. Cells were spread on LB plates containing ampicillin antibiotic. Colonies were located containing the *ccdA* cassette using colony PCR. Primers used for these PCR "checking" purposes are listed in Table 3.3. To perform colony PCR, a single colony was dissolved in 20 μL Type 1 water. Then, 2 μL was used as template in a 50 μL PCR reaction. A touch-down method was used at temperatures of (i) 54°C, (ii) 52°C, and (iii) 50°C for 5 cycles each and 48°C for 15 cycles. An extension time of 1 min was used and other parameters were identical to the PCR reactions described above. For the plcBaRcA plasmid, the target PCR product size was about 800 bp.

Table 3.3. PCR primers used to check insertions in pUC19 plasmid and RBS ccdB gene.

Primer Name	**Gene**	**Sequence**
pUC19_check2_left	*ccdB* cassette	GTATGTTGTGTGGAATTGT
pUC19_check2_right		GTCACAGCTTGTCTGTAAG
pUC19_as_check_left	*ccdA* cassette	AACGTCGTGACTGGGAAAAC
pUC19_as_check_right		ACGAGGCCCTTTCGTCTC

3.7. Testing the function of the different FITSelect components

While assembling the FITSelect system, the individual components were tested, during its construction, to ensure functionality of the individual parts. After the plcB plasmid (see Table 3.1) was constructed, the function of the different parts of the *ccdB* cassette (*rrnB* terminator, *argF* promoter, and *ccdB* gene) was tested. The plasmid was replicated in *ccdB* resistant *E. coli*

DH10B and a miniprep was used to extract the plasmid. Three combinations of restriction enzymes, (i) HindIII and KpnI, (ii) BamHI and KpnI and (iii) KpnI and NdeI were used to digest plcB plasmid separately. The three digestion products were purified through gel purification kit. The three plasmids were blunt ended by T4 polymerase (New England Biolabs) and ligated back together with the individual parts using T4 ligase. This procedure generated three new plasmids, plcBaRrBaP-, plcBaRaP-, and plcBaRcB- (see Table 3.1). Following their construction, the three new plasmids were transformed into normal Top10 cells and cultured on LB plates containing 100 μg/mL ampicillin. The growth of colonies was used to assess functionality of each individual part.

3.8. Testing the RBS modification of the *ccdB* gene (rbs*ccdB*)

In response to excessive cell death observed from the FITSelect system during initial trials, the ribosome binding site (RBS) of the *ccdB* cell death gene was altered to reduce the translation efficiency of *ccdB* mRNA. For this procedure, LB agar plates and liquid media containing 100 μg/mL ampicillin was used for culturing cells. The LB media was also supplemented with L-arginine (Sigma) to specified amounts. PCR primers (see Table 3.2) were designed to amplify the RBS region of the *ccdB* gene. The following PCR conditions were required for this amplification: (i) 5 cycles at 54°C, (ii) 5 cycles at 52°C, (iii) 5 cycles at 50°C, and (iv) 15 cycles at 48°C. The extension time was 1 min. PCR product was purified using the PCR purification kit. Purified PCR rbs*ccdB* was digested with KpnI and NdeI (3 hours incubation at 37°C). In the meantime, plcBaR plasmid was digested with KpnI and NdeI as well (3 hours incubation at 37°C) to form the plcBaRcB- plasmid. Then, digested plasmid was purified by a gel purification kit. Ligation between plcBaRcB- and *rbsccdB was* done to re-join

the plasmid back as plrbscBaR plasmid. After that, plrbscBaR plasmid was transformed into *ccdB* survival Top10 cells and culture in LB plates with 100μg/mL ampicillin. Colonies were harvested and tested by colony PCR using the same conditions as for the *ccdB* cassette checking (see above). PCR product was 1600bp. Then, cells containing plrbscBaR were grown in LB liquid culture with 100μg/mL ampicillin to replicate the plasmid. A miniprep was performed to extract pure plrbscBaR plasmid and transform into normal Top 10 cells. These were cultured on LB plates containing 100μg/mL ampicillin and 500ug/ml L-arginine.

3.9. Transfer the *ccdB* and *ccdA* cassettes into the low-copy number plasmid pACYC177 and test the function of FITSelect.

After unsuccessful trials with the FITSelect circuit using the pUC19 plasmid, both cassettes (see Figs. 2 and 3) were transferred to the low-copy number pACYC177 plasmid. LB plates and liquid media with 100ul/ml Ampicillin (Amp) and 25ul/ml Chloramphenicol (CM) were used for cell culture. The *ccdB* survival and normal Top10 cells were used in transformation. Primers used for check *ccdB* and *ccdA* cassettes in pUC19 list in Table 3.4 were used for amplification of the *ccdB* and *ccdA* cassettes. The *ccdB* cassette was ligated into pACYC177 by TA cloning with 2:1 ratio. The ligation product was transformed into *ccdB* survival Top10 cells and cultured on LB plates with 100μg/mL ampicillin and 25μg/mL chloramphenicol. Colonies were harvested and tested by colony PCR. Primers used are listed in Table 3.4. Special PCR conditions for the *ccdB* cassette amplification were 5 cycles at 60°C, 5 cycles at 58°C, 5 cycles at 56°C and 15 cycles at 54°C. Extension time was 3 minutes. PCR product was about 3000bp. Colonies containing plasmid with right size insert were cultured in liquid LB media and extracted by miniprep. This plasmid was named as pACYCcBaR plasmid.

The *ccdA* cassette PCR product was purified by the PCR purification kit and then digested with Eco0109I and AatII (3 hours incubation at 37°C). Then, the pACYCcBaR plasmid was digested with Eco0109I and AatII (3 hours incubation at 37°C) and was ligated with the *ccdA* cassette using a 1:2 molar ratio. The ligation product was transformed into normal Top10 cells. Several colonies appeared on the plate were harvested and tested by colony PCR. The "pACYC check 2" primers were used (see Table 3.4). Special PCR conditions for the *ccdA* cassette check were 5 cycles at 54°C, 5 cycles at 52°C, 5 cycles at 50°C and 15 cycles at 48°C. Extension time was 1 minute. PCR product was about 800bp. Then, colonies containing plasmid were harvested and the plasmid was purified as described before. The new plasmid was named pACYCcBaRcA. TA cloning was applied to insert *trpR* gene downstream of *argR* gene as described above. The new plasmid was named pACYCcBaRcAtR. Primers for *trpR* check were list in Table 3.4. Special PCR conditions for the *trpR* check were 5 cycles at 54°C, 5 cycles at 52°C, 5 cycles at 50°C and 15 cycles at 48°C. Extension time was 1 minute. PCR product was about 750bp.

Table 3.4. PCR check ccd, ccdA and trpR in pACYC plasmid.

Primer Name	**Gene**	**Sequence**
pACYC check left primer	Region contain *ccdB* cassette	GACTCAAGACGATAGTTACC
pACYC check right primer	Region contain *ccdB* cassette	CTGTAATGAAGGAGAAAACT
pACYC check left 2 primer	Region contain *ccdA* cassette	ATAAAATTCTTGAAGACGAA
pACYC check right 2 primer	Region contain *ccdA* cassette	AGAAAAATAAACAAATAGGG
TrpR check left primer	Region contain part of *argR* and part of *trpR*	ATTCATCTGTATGCACAATA
TrpR check right primer	Region contain part of *argR* and part of *trpR*	AGGTTTAACAACGGTAAAT

3.10. TOPO cloning

The following procedures were used for the TOPO cloning procedure. The TOPO cloning kit, pDEST14 vector, and Gateway LR Clonase II enzyme mix were from Invitrogen. LB plates containing 50ug/ml kanamycin (Kan) and 100ug/ml Ampicillin (Amp) respectively were used for strain selection. LB liquid media with 100ug/ml Ampicillin was used for insertion expression. L-arabinose was used to induce insertion expression.

The PrsA gene sequence was located from *E. coli* DH10B genomic DNA. Antisense RNA (asRNA) constructs were designed and genes encoding these constructs were synthesized by Integrated DNA Technologies. For TOPO cloning, the molar ratio between the gene product and TOPO vector was 1:1. Cloning steps were peformed as described in the manufacturer's protocol. The, 2 μL of cloning mixture was transformed into Top10 chemically competent cells and cultured on LB plates with Kan. Several colonies were harvested and checked by colony PCR. Primers are listed in Table 3.5. Special PCR conditions for the TOPO check amplification were: 5 cycles at 54°C, 5 cycles at 52°C, 5 cycles at 50°C and 15 cycles at 48°C. Extension time was 1 minute. The correctly sized PCR product was about 480bp. Sequencing was performed to determine direction of the insert. Then, the plasmid containing the insert was replicated and extracted using miniprep kit. Purified plasmid was used to perform the LR recombination reaction with the Gateway destination vector (pDEST14) according to the manufacturer's protocol. This was done to to transfer the asRNA PrsA gene from TOPO vector to pDEST14 vector. The pDESR14 plasmid was then transformed into DH5α cells and replicated. This plasmid was then extracted and checked using PCR ("pDEST14 primers, see Table 3.5). Special PCR conditions for the pDEST14 check amplification were 5 cycles at 54°C, 5 cycles at 52°C, 5 cycles at 50°C and 15 cycles at 48°C. Extension time was 1 minutes. Also, sequencing was

performed to ensure the correct insert direction. Purified pDEST14 plasmid harboring asRNA PrsA was transformed into BL21 competent cells to determine function of the asRNA PrsA. Expression steps were performed according to the manufacturer's protocol. Cells without adding the L-arabinose inducer were cultured as a control. Time-course OD600 measurements were obtained to determine the growth rates of each induced (and the control) culture.

Table 3.5. PrsA PCR and TOPO check primers.

Primer Name	**Gene**	**Sequence**
TOPO_check_forward_Left primer	TOPO vector region with insertion	TTTTCCCAGTCACGA
Topo_check_forward_Right primer		CTCATAACACCCCTTGT
pDEST14_check_left	pDEST14 vector region with insertion	CAACGGTTTCCCTCT
pDEST1417_check_right		ATGCGTTGATGCAAT

4. Results and Discussion

4.1. Overview of the FITSelect system.

To produce a strain of *E. coli* DH10B that can over-produce L-arginine, a combinatorial metabolic engineering strategy was conceived. The goal was to increase the growth rate of a cell as a result of increased L-arginine production. This would enable selection of productive strains using a growth competition assay. To do this, the Feed-back Inhibition of Transcription for Growth Selection (FITSelect) system was developed to tie the production of L-arginine to the growth rate of the cell. The first version of FITSelect is shown in Fig. 4.1. This chapter describes the implementation and testing of all the genetic "parts" used to construct FITSelect. This chapter also details all of the updates and new designs that were required of this initial FITSelect design. In Fig. 4.1, a genomic library insert is cloned into the FITSelect genetic circuit. In Scenario 1 (top pane of Fig. 4.1), the library insert does not impact L-arginine production. This will be the case for over 99.9% of inserts. In this case, the T7 polymerase binds to the T7 promoter and amplifies the insert. With no impact on L-arginine production, the intracellular concentration of L-arginine remains low. In this case, the native ArgR repressor protein remains inactive. This allows RNA polymerase to bind to the ARG boxes of the *argF* promoter and amplify antisense RNA (asRNA) to the PrsA mRNA. When these bind, translation of PrsA is halted and the growth rate of the cell is decreased. This mechanism is explained further below and is based on the fact that PrsA is essential for culture growth. Without this protein, the cell cannot grow. In Scenario 2 (bottom pane of Fig. 4.1), the library insert results in over-production of L-arginine. This leads to a higher concentration of L-arginine inside the cell and binding of L-arginine to the ArgR repressor protein is facilitated. Once this binding occurs,

ArgR becomes "activated" and binds the ARG boxes of the *argF* promoter. When this happens, the RNA polymerase cannot bind the *argF* promoter region, and the PrsA asRNA is not transcribed. This leaves normal PrsA activity and a normal growth rate of the cell. Thus, FITSelect results in a cell over-producing L-arginine with a higher growth rate than a cell with normal L-arginine production. When grown in a growth competition assay, the cell with the highest growth rate will become dominant. This makes FITSelect ideal for selecting strains that can over-produce L-arginine using combinatorial metabolic engineering.

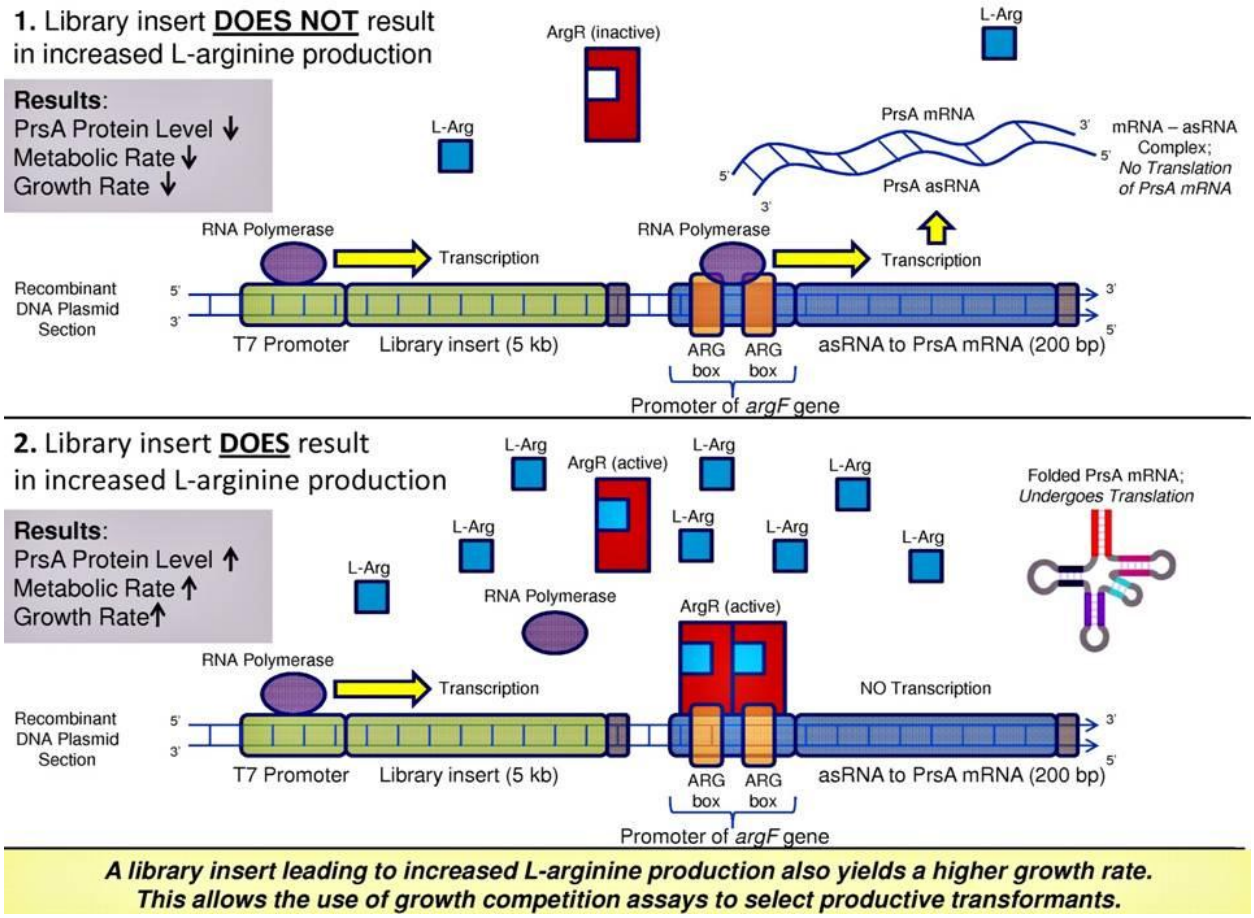

Fig. 4.1. Original conception of the FITSelect genetic circuit.

4.2. Down-reguation of *prsA* to control the growth rate of *E. coli* DH10B.

The *prsA* gene in *E. coli* DH10B encodes the phosphoribosylpyrophosphate synthase. While the details of the reaction catalyzed by this enzyme are not important for this research, the *prsA* gene is an "essential" gene in *E. coli*. This means that without this gene, the cell cannot grow. Also, with a down-regulated *prsA* gene, the cell will grow slowly, compared to the wild-type. The approach in this research was to create an antisense RNA (asRNA) to the *prsA* gene to down-regulate its translation. An asRNA construct was made by assessing the secondary structure of the *prsA* mRNA using mFOLD (*58*). The asRNA was complementary to a major loop region of this structure and overlapped the ribosome binding site (RBS). This was designed based on a published method (*59*). The asRNA construct was synthesized (IDT) and cloned using TOPO cloning (Invitrogen). Fig. 4.2 (left) shows the TOPO vector PCR check for the *prsA* asRNA sequence to verify the first successful cloning step of adding the asRNA construct to the TOPO vector. The second step of the cloning procedure was to transfer this gene to a destination vector (pDEST14) using homologous recombination. The PCR check of the pDEST14 vector to ensure it received the asRNA construct is shown in Fig. 4.2 (right). Gel results show the cloning was successful, and the asRNA construct was transferred to the destination vector. Sequencing results verified the correct direction of *prsA* asRNA sequence in both vectors. The destination vector was transformed into *E. coli* BL-21 cells (Invitrogen), and the growth curves of the culture containing the asRNA and a culture containing a control sequence were measured. Results are shown in Fig. 4.3. The asRNA and control sequences are given in Table 4.1. Almost no difference was apparent between these two growth curves; although a slight decrease in growth was noted of the *prsA* asRNA culture. No differences in growth were observed for additional control sequences as well (results not shown). These results

suggested the antisense design did not work as planned. Although, it is not known whether (i) the asRNA did not effectively bind the PrsA mRNA or (ii) the asRNA-mRNA complex of PrsA did not reduce the growth rate of the cell. From these results, it was concluded that expression of this particular *prsA* asRNA sequence could not decrease the growth rate of the cell. The next logical approach involves designing new asRNA constructs or choosing a new essential gene. However, at this point it was decided to abandon the asRNA approach for a "cell death" gene. The *ccdB* gene was chosen for this purpose, and it is commonly used in molecular biology cloning protocols.

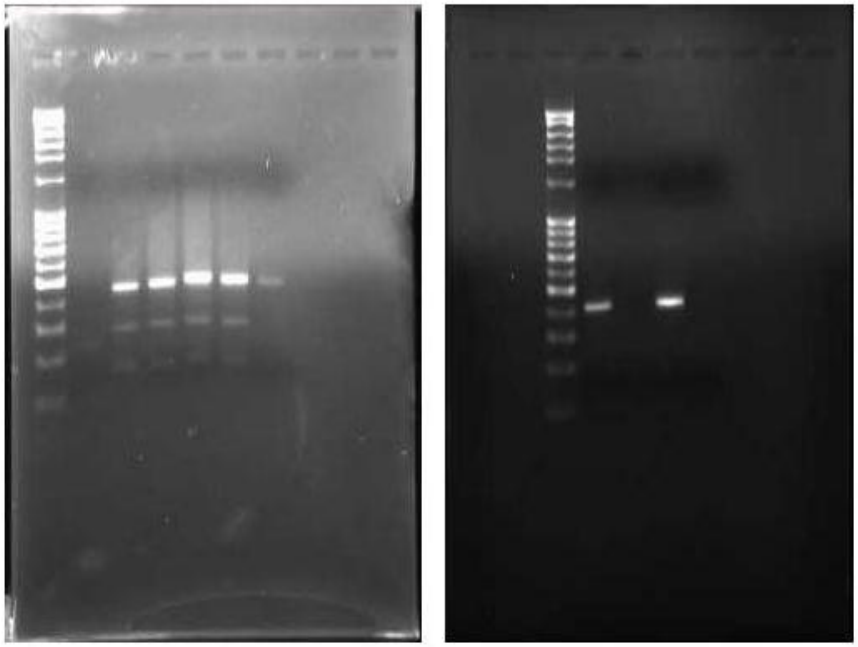

Fig. 4.2. Successful TOPO cloning of the *prsA* asRNA sequence in the TOPO vector (left) and the pDEST14 destination vector (right).

Table 4.1. asRNA and control sequences. The 5' CACC sequence is required for TOPO cloning.

Name	Sequence
prsA asRNA	CACC_CACGGACTATACTTCGAAAA
Control 1	CACC_AAAAAAAAAATTTTTTTTTT
Control 2	CACC_AAAAAAAAAAAAAAATTTTTTTTTTTTTTT
Control 3	CACC_AAAAAAAAAAAAAAAAAAAATTTTTTTTTTTTTTTTTTT

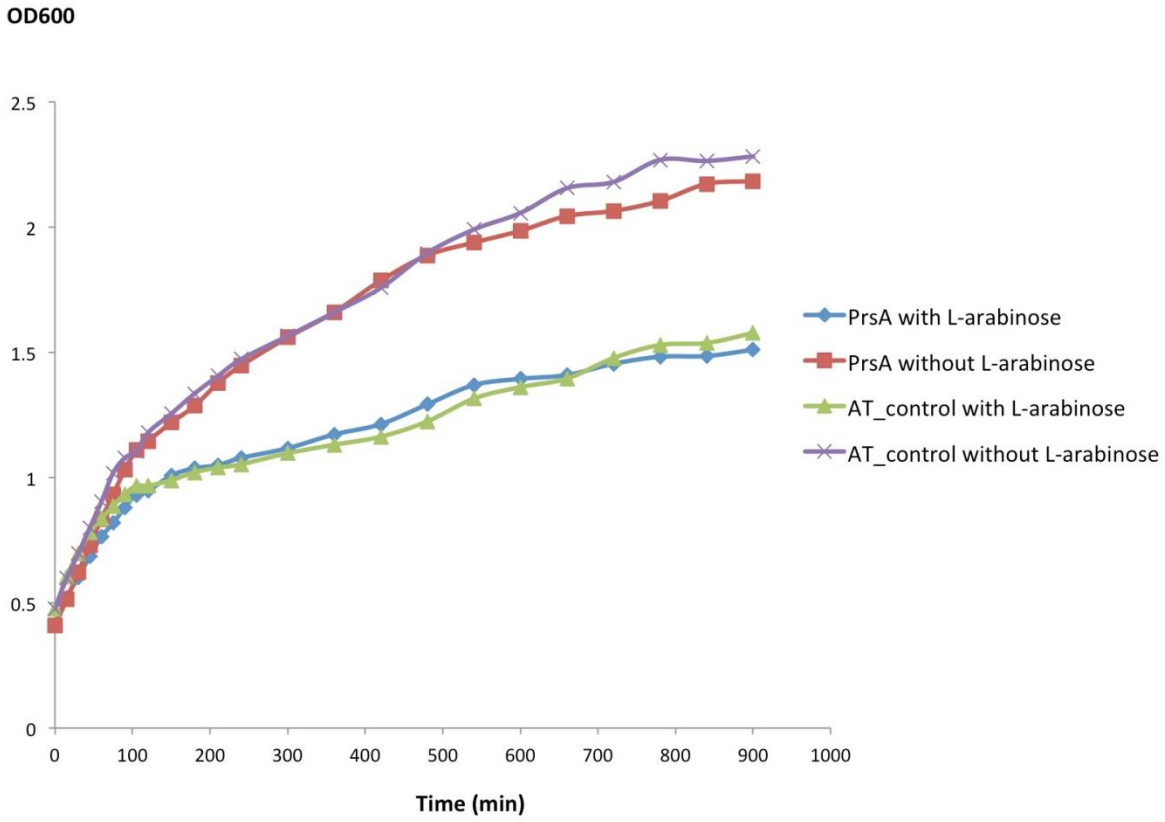

Fig. 4.3. Growth curves of BL-21 cells with pDEST14 plasmid harboring the prsA asRNA sequence and the AT Control 1 sequence (see Table 4.1). The pDEST14 plasmid was induced with L-arabinose.

4.3. The *ccdB* gene to replace the *prsA* asRNA, the inclusion of *argR*, and the initial construction of FITSelect.

The *argF* promoter region, containing the ARG regulatory boxes, and the *rrnB* terminator region were PCR amplified from *E. coli* DH10B gDNA as described in the Materials and

Methods chapter. The *ccdB* gene was obtained from pDEST14 plasmid DNA. These genetic "parts" are the major components of the FITSelect construct. The PCR amplification of *argF* promoter region, *ccdB* gene, and the *rrnB* terminator are shown in Fig. 4.4. The parts were ligated and cloned into the pUC19 plasmid. This plasmid provided the inducible *plac* promoter. Due to blue-white selection, and the presence of a multi-cloning site, this plasmid was preferred. Thus, the T7 promoter of Fig. 4.1 was replaced by the *plac* inducible promoter. The successful amplification of these parts from gDNA is shown in Fig. 4.4. These parts were ligated and successfully cloned into the pUC19 plasmid as shown in Fig. 4.5. This plasmid was named plcB.

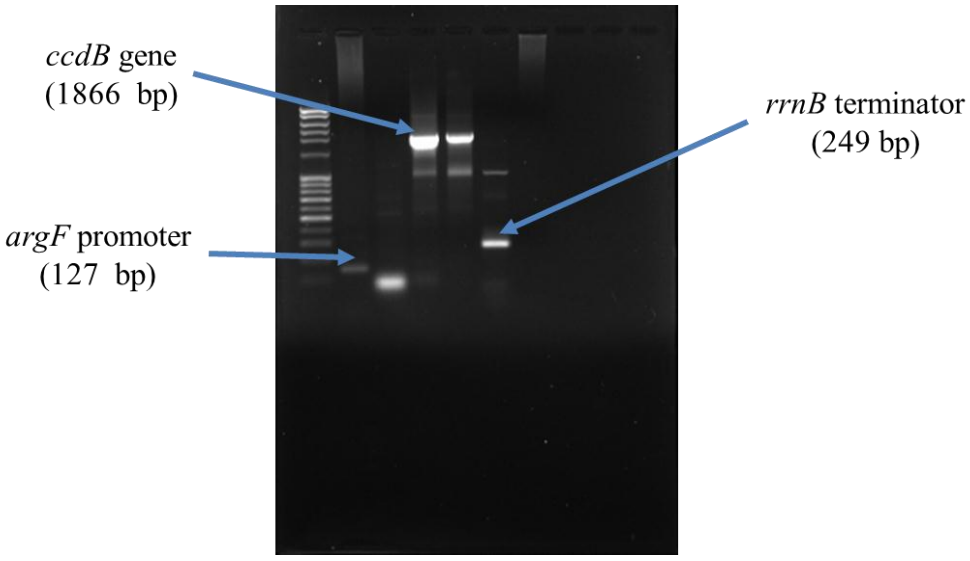

Fig. 4.4. Successful PCR amplification of the *argF* promoter, the *ccdB* cell death gene and the *rrnB* terminator from genomic and plasmid DNA.

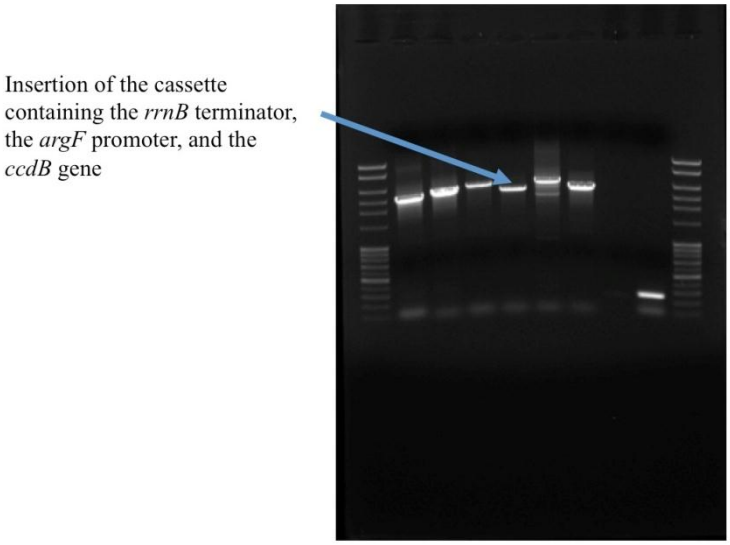

Fig. 4.5. Verification of successful cloning of the cassette containing the *rrnB* terminator, *argF* promoter, and *ccdB* gene.

The plcB plasmid was transformed into normal *E. coli* Top10 cells. When cultured on LB solid media, all cells died due to expression of the *ccdB* cell death gene. However, this cloning was successful in *ccdB* resistant Top 10 cells and yielded many colonies on LB plates. The LB media was supplemented with 20 g/L L-arginine to repress expression of the *ccdB* gene by

blocking the *argF* promoter, and the cloning was again attempted in normal Top10 cells. However, cell death still resulted and zero colonies were observed on LB plates. Because L-arginine must interact with the arginine repressor (ArgR) protein to block expression of the *argF* promoter (to silence *ccdB*), it was reasoned that cell death may have been due to an insufficient quantity of ArgR protein in the cells. Thus, the gene for ArgR (*argR*) was PCR amplified from Top10 gDNA and cloned into plcB to form plcBaR plasmid. The *argR* PCR amplification is shown in Fig. 4.6 and the PCR check showing its successful cloning is shown in Fig. 4.7. Following cloning, the plasmid harboring the *argR* gene was sent for sequencing to determine its direction in the plasmid. The plcBaR plasmid with the correct *argR* insert was transformed into normal Top10 cells and *ccdB* resistant Top10 cells. The media combinations in Table 4.2 were used to test for survivability of the normal Top10 cells. Ultimately, all media combinations resulted in death of the normal Top10 cells. However, the *ccdB* resistant Top10 cells again survived in LB media. This confirmed that even with an over-expressed *argR* gene, high expression of the *ccdB* gene was still leading to cell death in normal Top10 cells.

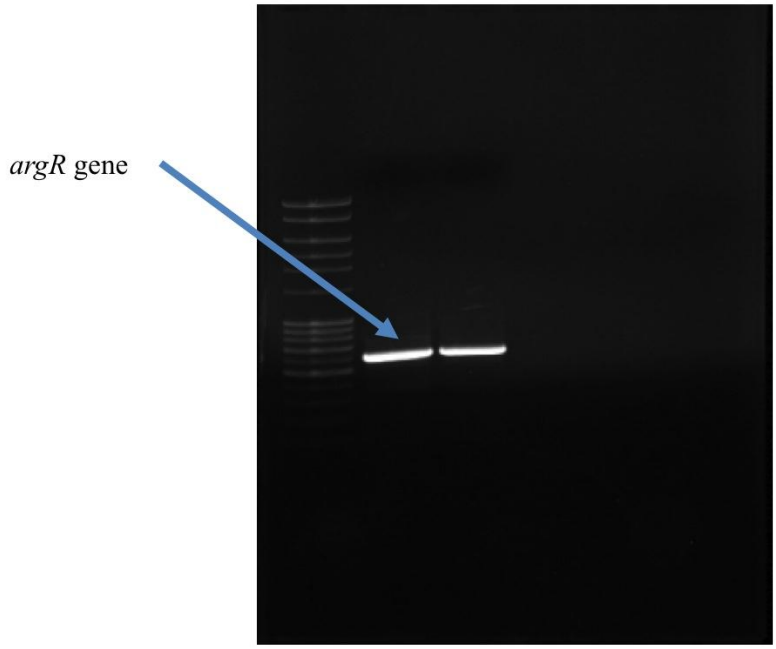

Fig. 4.6. PCR amplification of the *argR* gene from gDNA.

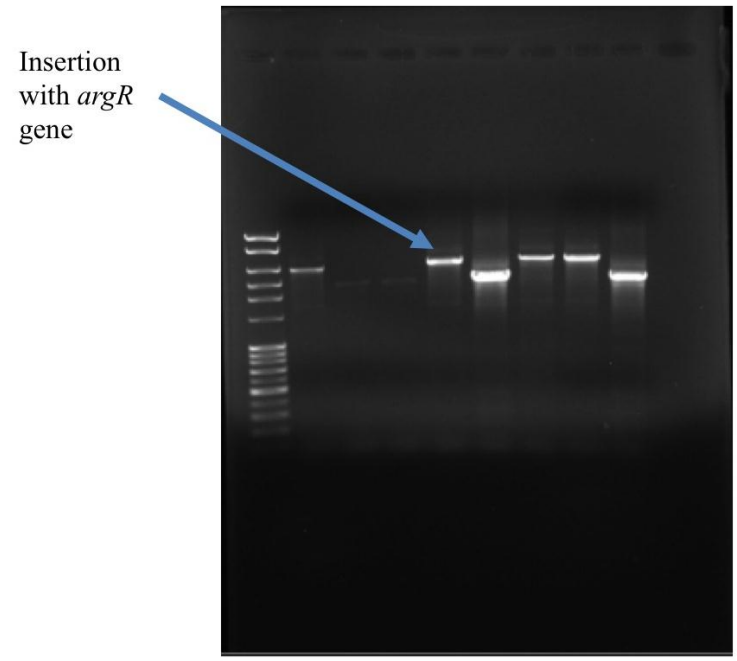

Fig. 4.7. Verification of successful *argR* cloning.

Table 4.2. Different media combinations for testing normal Top10 cells containing plcBaR.

Media	**Normal Top10 Cells**	***ccdB* Resistant Top10 Cells**
LB	No survival	Survival
M9	No survival	Survival
LB + L-arginine	No survival	Survival
M9 + L-arginine	No survival	Survival
M9 + L-arginine (no NH_4Cl)	No survival	Survival

4.4. Testing the function of the different "parts" of the plcBaR plasmid

Following cloning of the *ccdB* cassette (containing the *rrnB* terminator, the *argF* promoter, and the *ccdB* gene) and the *argR* gene, survivability was not obtained in normal Top10 cells. So, the function of each of the different synthetic "parts" was tested for its individual function. This was also done to see if cell death was due to any unexplained function of the cassette. Three separate parts (i) the *rrnB* terminator, (ii) the *argF* promoter, and (iii) the *ccdB* gene were knocked-out of the plcBaR plasmid. The following plasmids were created: (i) plcBaRrBaP- (without *rrnB* terminator and *argF* promoter), (ii) plcBaRcB- (without the *ccdB* gene), and (iii) plcBaRaR- (without the *argF* promoter). These plasmids were transformed into normal Top10 cells separately and cultured on solid LB media. Results of these plates are shown in Fig. 4.8 and approximate colony counts are listed in Table 4.3.

Table 4.3. Colony counts for plcBaR plasmid "part" knockouts when transformed into normal Top10 cells.

Plasmid name	Gene knockout	Media	Colony numbers
plcBaRaP-	*argF* promoter	LB	~100
plcBaRrBaP-	*rrnB* terminator, *argF* promoter	LB	~0
plcBaRcB-	*ccdB* gene	LB	~300

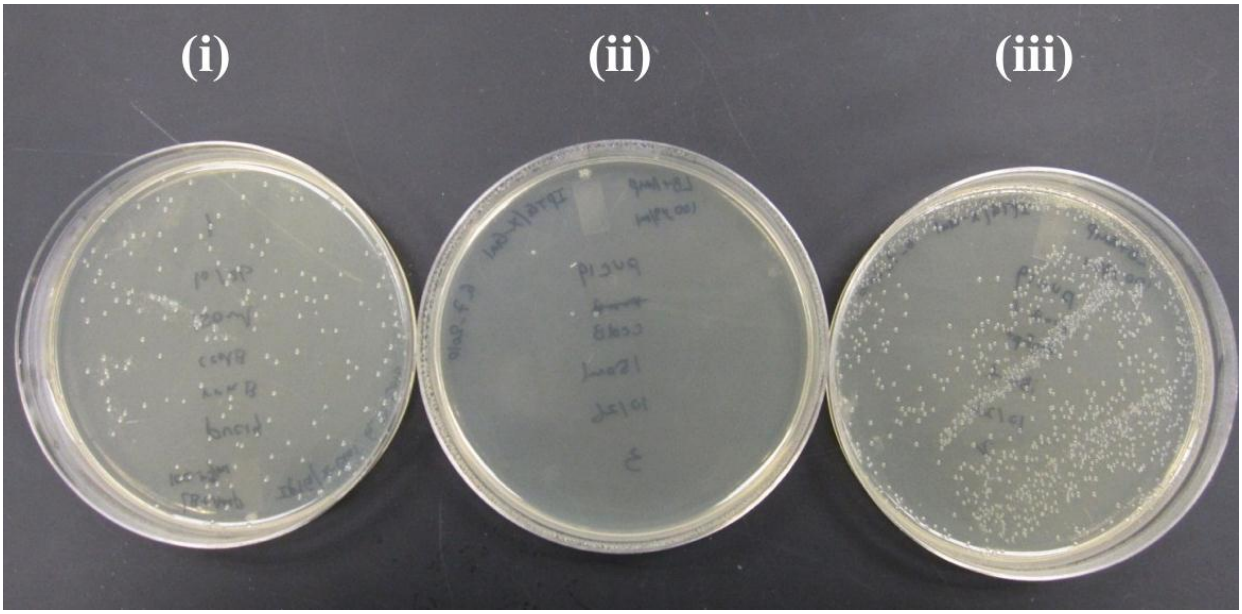

Fig. 4.8. Normal Top10 cells harboring (from left to right): (i) plcBaRaP-, (ii) plcBaRrBaP-, and (iii) plcBaRcB- plasmids.

From these results, it was concluded that all three "parts" of the *ccdB* cassette (the *rrnB* terminator, *argF* promoter, and the *ccdB* gene) are functional in the plcBaR plasmid. For the plcBaRaP- plasmid (*argF* promoter knock-out), a healthy colony count was observed (Fig. 4.8). In this case, the *rrnB* terminator terminated transcription from *plac* promoter. This left no promoter for expression of the *ccdB* gene. The survivability of these cells concludes that the *rrnB* terminator is functional and that the *argF* promoter adequately expresses the *ccdB* gene. In the plcBaRrBaP- plasmid (*rrnB* terminator and *argF* promoter knock-out) no cell growth was observed (Fig. 4.8). In this case, the transcription originating from the *plac* promoter is not terminated by the *rrnB* terminator and leads to the transcription of the *ccdB* cell death gene. This leads to cell death. This provides further evidence that the *rrnB* terminator is functional. If the *rrnB* terminator is functional, then the *argF* promoter must also be functional. These conclusions are further supported by the *argF* promoter knockout (plcBaRaP-). The last

knockout performed was the *ccdB* gene itself (plcBaRcB- plasmid). When transformed into normal Top10 cells and grown on LB solid media, many colonies showed up. This verifies that the *ccdB* gene of the cassette causes cell death, instead of some unknown action by the *rrnB* terminator or *argF* promoter. As shown in Table 4.3, the number of colonies present for the pclBaRcB- plasmid (*ccdB* gene knockout) (~300) was greater than that for the plcBaRcP- plasmid (*argF* promoter knockout) (~100). This could mean that the *rrnB* terminator is not completely effective in this cassette; however, this conclusion cannot be made based on this colony count data. These experiments did confirm that the *rrnB* terminator, the *argF* promoter, and the *ccdB* gene "parts" of the cassette are independently active. However, these "parts" still resulted in complete cell death, even with over-expression of *argR*. Next, a "less-effective" *ccdB* was engineered by modifying its ribosome binding site (RBS) in attempt to "fine-tune" the system.

4.5. Modifying the RBS of the *ccdB* gene

It has been shown recently that the RBS sequence is a key to determining the translation efficiency of an mRNA. A predictive tool called the "RBS Calculator" has also been published and uses a thermodynamic model to calculate this efficiency (*25*). In trials of the FITSelect system to this point, cell death from the *ccdB* gene resulted. The approach here was to "tune" the system by down-regulating the effectiveness of the cell death gene. Since the presence of L-arginine is responsible for turning "on/off" *ccdB* expression through the *argF* promoter, attenuation of *ccdB* translation was attempted. Using the RBS Calculator software (*25*), a new RBS for the *ccdB* gene was designed that was 5-times less effectively translated than the wild-type RBS. The wild-type and the new RBS sequences are given in Table 4.4. Fig. 4.9 shows the newly constructed PCR product of the *ccdB* gene with the modified RBS. This product was

named rbs*ccdB*. Then, the rbs*ccdB* was inserted into the plcBaR plasmid to replace original *ccdB* gene. This plasmid was named the plrbscBaR plasmid. This new plasmid was cloned into normal and *ccdB* resistant Top10 cells. The colony PCR result for the rbs*ccdB* gene, showing successful cloning is shown in Fig. 4.10.

Table 4.4. Wild-type and modified RBS sequences of the *ccdB* gene.

RBS	**Sequence**
Wild-type *ccdB*	TCTTTTGCTGACGAGAACAGGGACTGGTGAA
Modified (rbs*ccdB*)	TGCCCGGAGTATCCGTTCAGTTCAAAGAGAGGC

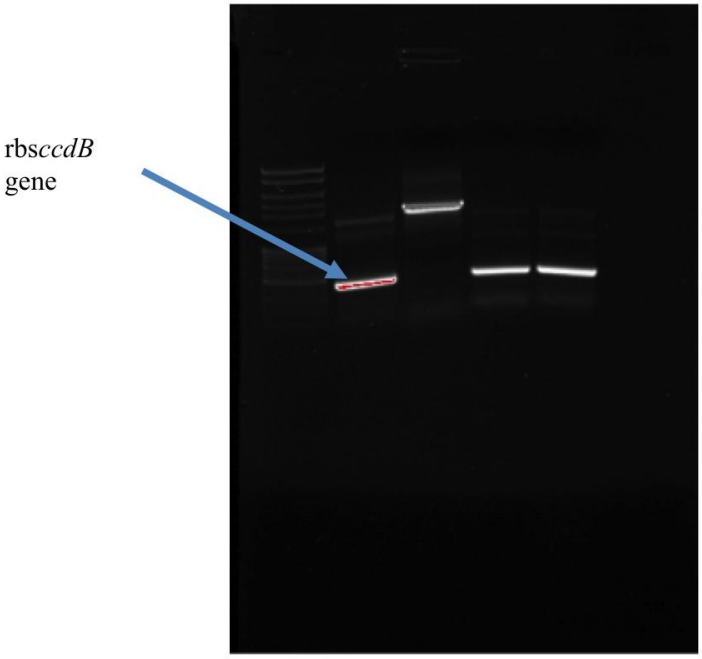

Fig. 4.9. PCR amplification of the rbs*ccdB* gene.

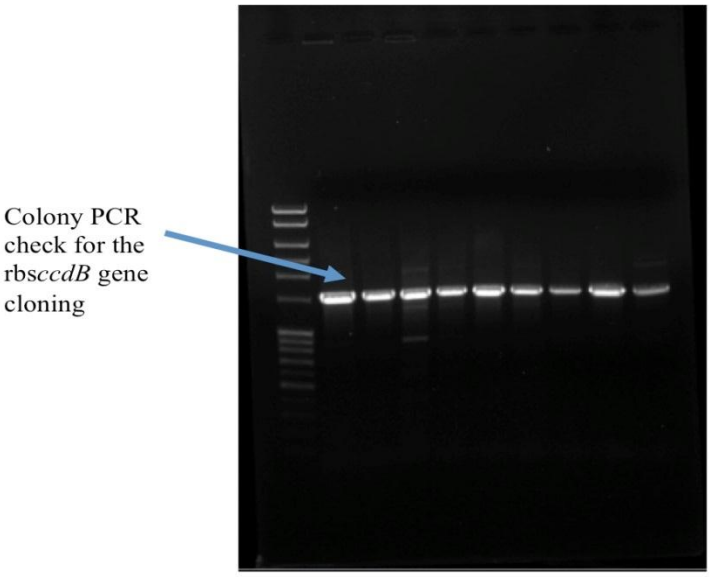

Fig. 4.10. Colony PCR check for successful cloning of the rbs*ccdB* gene in *ccdB* resistant Top10 cells.

When the plrbscBaR plasmid was transformed into *ccdB* resistant Top10 cells, normal growth was observed. However, when cloned into normal Top10 cells and cultured on LB plates (with additional L-arginine), no cell growth was observed. This meant the attenuated RBS of the *ccdB* still did not provide enough "tuning" of the system to avoid cell death by the *ccdB* gene. The next approach was to provide additional "tuning" of the *ccdB* gene by adding the *ccdA* "antidote" to the cell death gene.

4.6. Addition of the *ccdA* gene to modulate the *ccdB* gene

Since cell death by *ccdB* expression could not be modulated by simply adding L-arginine to repress the *argF* promoter, or by turning its RBS sequence, another method was developed. The *ccdA* gene is the "antidote" to the *ccdB* cell death gene (*41*). The *ccdA* gene under control of *trp* promoter which can be induced by L-tryptophan was added to the system. The *trp* promoter is regulated by feedback inhibition by L-tryptophan. In the absence of L-tryptophan, the

promoter is “on.” When L-tryptophan is added to the culture, the promoter turns “off.” This halts transcription. Thus, in this approach, the *ccdA* gene was cloned into the same plasmid harboring the *ccdB* gene. Under normal conditions (without L-tryptophan), the *ccdA* gene will be expressed and keep the cell alive. When a genomic library is used to over-produce L-arginine, sufficient quantities will accumulate inside the cell (of selected clones) to block *argF* promoter transcription. At this point, the absence of excess L-tryptophan will ensure all cells survive. Then, L-tryptophan will be added into media to repress the *trp* promoter. In this case, only the cells that can over-produce large amounts of L-arginine can survive. The *ccdA* gene used in this research was synthesized by Integrated DNA Technologies (Coralville, IA). Fig. 4.11 shows the *trp* promoter and *rrnB* terminator PCR amplification.

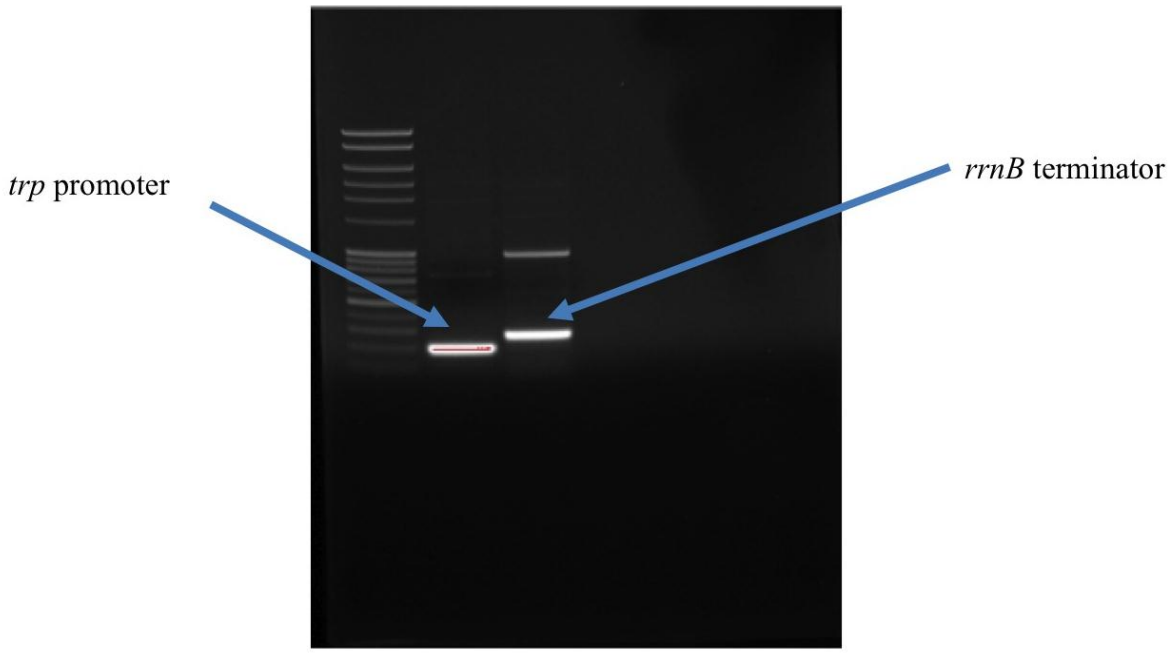

Fig. 4.11. PCR amplification of the *trp* promoter and *rrnB* terminator

After PCR amplification, *ccdA* cassette (consisting of the *trp* promoter, the *ccdA* gene, and the *rrnB* terminator) was ligated into the plcBaR plasmid to form the plcBaRcA plasmid.

The plcBaRcA plasmid was transformed into *ccdB* resistant and normal Top10 cells. The *ccdA* gene was functional as many colonies appeared on both sets of plates. This confirms the *ccdA* "antidote" counteracts the *ccdB* cell death gene. Next, repression of the *trp* promoter and the *ccdA* gene was examined by the addition of L-tryptophan to the culture media. Different media combinations were examined at this step and are listed in Table 4.5. Normal Top10 cells were able to survive in all media formulations and L-tryptophan concentrations. This meant the *trp* promoter was not effectively turned "off," and the CcdA antidote easily counteracted the cell death effect of CcdB. This led to the additional challenge of attenuating the *trp* promoter so that the optimal balance of CcdA and CcdB could be achieved.

Table 4.5 Combination of L-tryptophan concentrations to test the *trp* promoter.

Culture Media	L-Tryptophan Concentration (μg/mL)		
LB	0	100	300
M9	0	100	300

4.7. Transfer the full *ccdB* and *ccdA* cassettes into the low copy plasmid

The pUC19 is high copy number plasmid that has ~500-700 copies in single cell. It is hypothesized that problems related to turning "off" the *trp* promoter is a result of the high copy number of the plasmid. Because of the high number of copies of the plasmid existing in the cell, too much L-tryptophan may be required to effectively turn "off" all copies of the *trp* promoter present in the cell. In response, the *ccdB* and *ccdA* cassettes, originally transformed into pUC19, were transferred to the low copy plasmid pACYC117 as a possible solution. The pACYC117 plasmid has a low copy number of about 10 (copies) in a single cell. The *ccdB* and *ccdA*

cassettes were amplified separately and ligated into pACYC117 plasmid to form the pACYCcBaRcA plasmid. Fig. 4.12 shows the linearized pACYC177, the pACYCcBaR (containing the *ccdB* cassette only) and the pACYCcBaRcA (containing both *ccdB* and *ccdA* cassettes) plasmids following construction.

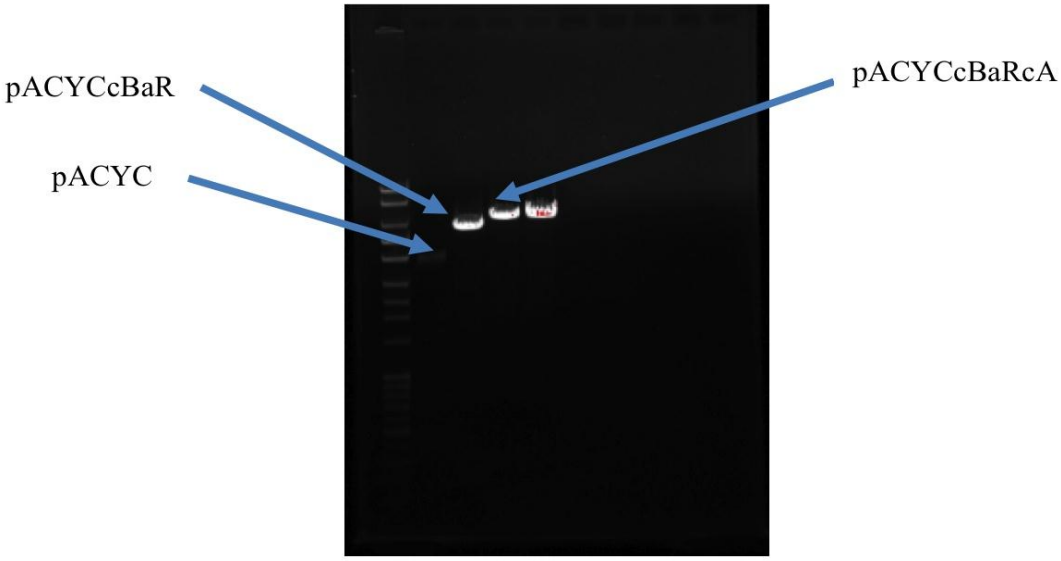

Fig. 4.12. The linearized pACYC, pACYCcBaR, pACYCcBaRcA plasmids.

The pACYCcBaRcA plasmid was then transformed into normal Top10 cells. The culture survived on LB plates. Then, the experiment was performed to test if added L-tryptophan was able to turn "off" the *trp* promoter and *ccdA* expression. The different combinations of culture media used in this experiment are listed in Table 4.6. Growth curves of the cultures in liquid media are shown in 4.13. No changes in the growth rate of the cultures were observed with increasing L-tryptophan concentrations in LB or M9 minimal media. Some deviations were observed with M9 media, but these changes did not correlate with L-tryptophan concentrations. Thus, these experiments show the *trp* promoter was still not being repressed effectively. Similar

to the *argF* promoter being repressed by the activated ArgR protein, the *trp* promoter is repressed by activated TrpR. Thus, the *trpR* gene was next added to the pACYCcBaRcA plasmid to effectively regulate the *trp* promoter.

Table 4.6. Combination of L-tryptophan concentrations to test the trp promoter in pACYCcBaRcA.

Media	L-Tryptophan Concentration (μg/mL)							
LB	0	100	200	300	500	1000	1500	3000
M9								

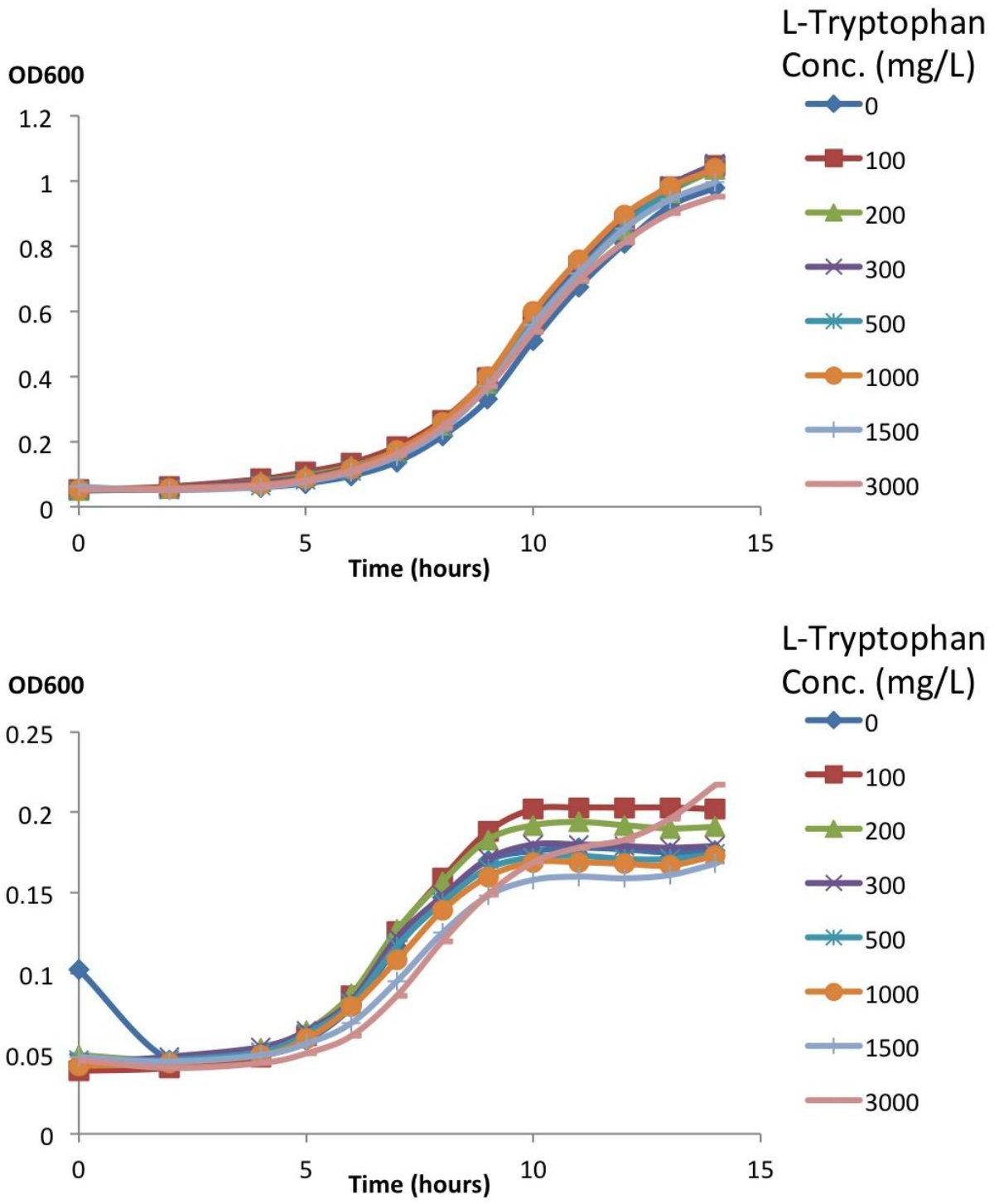

Fig. 4.13. Growth curve of normal Top10 cells harboring the pACYCcBaRcA plasmid in (i) LB media (top) containing L-tryptophan concentrations (given in mg/L) and (ii) M9 minimal media (bottom) containing L-tryptophan concentrations (given in mg/L).

4.8. Addition of *trpR* to the *ccdB* cassette

The *trpR* gene was PCR amplified from *E. coli* Top10 gDNA and was inserted into the pACYCcBaRcA plasmid downstream of the *argR* gene to form the pACYCcBaRcAtR plasmid. The *trpR* PCR amplification and PCR check for successful cloning is shown in Fig. 4.14.

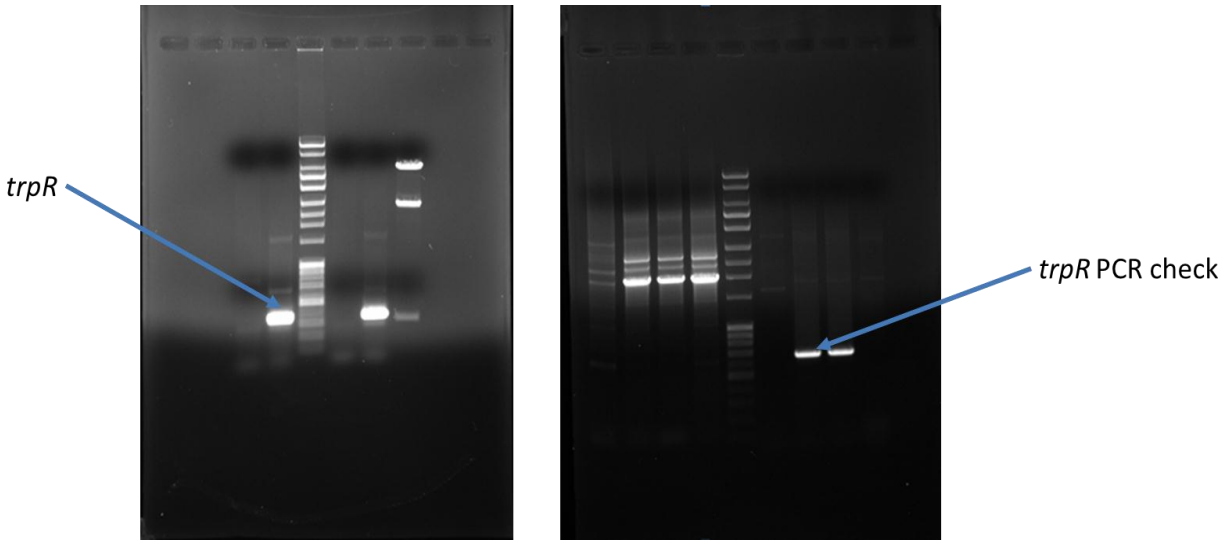

Fig.4.14. PCR amplification of *trpR* (left) and PCR check for successful cloning of *trpR* to form the pACYCcBaRcAtR plasmid (right).

The pACYCcBaRcAtR plasmid was transformed into normal Top10 cells, and the culture was grown on solid LB plates. Again, the culture survived, and several colonies were observed. Liquid media experiments were designed to test the regulation of the *argF* and *trp* promoters given the presence of over-expressed ArgR and TrpR regulator proteins. Different concentrations of L-tryptophan and L-arginine were added to liquid media as listed in Table 4.7. Growth curves of the cultures were generated and are shown in Fig. 4.15. As seen in Fig. 4.15 (top), the addition of L-tryptophan in varying amounts decreases growth of the culture accordingly (i.e., growth decreases as L-tryptophan increases). Very little difference was seen

between the growth rates observed when 1500 mg/L and 3000 mg/L were added. Thus, the amount of 3000 mg/L was seen as the "upper limit" of L-tryptophan leading to the maximum decrease in the growth rate of the culture. Thus, this amount of L-tryptophan was added to new cultures along with a varying amount of L-arginine. The results of culture growth for these conditions are also shown in Fig. 4.15 (bottom). Although somewhat less evident than the case for L-tryptophan addition (Fig 4.15, top), the addition of L-arginine led to a concentration-dependent increase in the growth rate of the culture. The specific growth rate during the exponential growth phase (between 4-9 hours) was calculated for each case. As a control, the growth curve of normal Top10 cells harboring the pACYC117 plasmid with 3000 mg/L L-tryptophan and varying amounts of L-arginine is shown in Fig. 4.16. No apparent growth rate difference is observed with increasing L-arginine concentrations. This control shows that the cells are not simply using L-arginine as a substrate to attain higher growth rates when using the FITSelect system. Results are shown as a function of L-arginine concentration for the FITSelect system in Fig. 4.17 and for the control in Fig. 4.18. For the FITSelect system (Fig. 4.17), a linear relationship with a positive slope is observed between the average specific growth rate of the culture in the exponential growth phase and the amount of L-arginine added to the culture. While significant tuning is still needed, these results show that the FITSelect synthetic circuit is functional in its response to L-arginine. This relationship between the growth rate and L-arginine concentration was not observed for the control culture (Fig. 4.18).

Table 4.7. Combination of (i) L-tryptophan concentrations to test the *trp* promoter in pACYCcBaRcAtR (ii) L-arginine concentration to test the *argF* promoter in pACYCcBaRcAtR.

Scenario [(i) or (ii)]	L-Tryptophan Conc. (mg/L)	L-Arginine Conc. (mg/L)
(i) Fig. 4.15 (top)	0	0
(i) Fig. 4.15 (top)	100	0
(i) Fig. 4.15 (top)	200	0
(i) Fig. 4.15 (top)	300	0
(i) Fig. 4.15 (top)	500	0
(i) Fig. 4.15 (top)	1000	0
(i) Fig. 4.15 (top)	3000	0
(ii) Fig. 4.15 (bottom)	3000	0
(ii) Fig. 4.15 (bottom)	3000	100
(ii) Fig. 4.15 (bottom)	3000	200
(ii) Fig. 4.15 (bottom)	3000	300
(ii) Fig. 4.15 (bottom)	3000	500
(ii) Fig. 4.15 (bottom)	3000	1000
(ii) Fig. 4.15 (bottom)	3000	3000

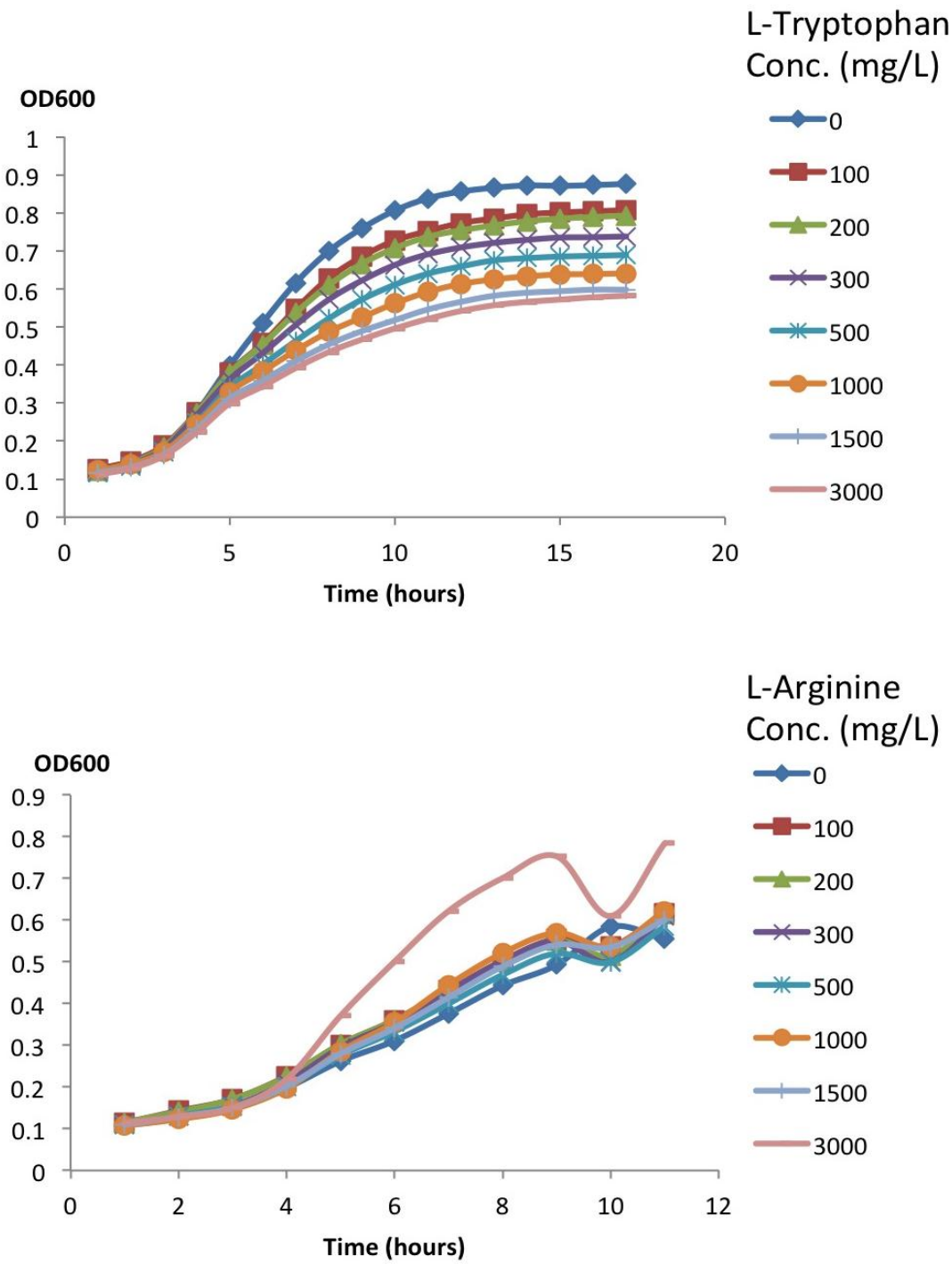

Fig 4.15. Growth curve of normal Top10 cells harboring the pACYCcBaRcAtR plasmid in (i) LB media (top) containing L-tryptophan concentrations (given in mg/L) and (ii) LB media (bottom) containing 3000 mg/L L-tryptophan and L-arginine concentrations (given in mg/L).

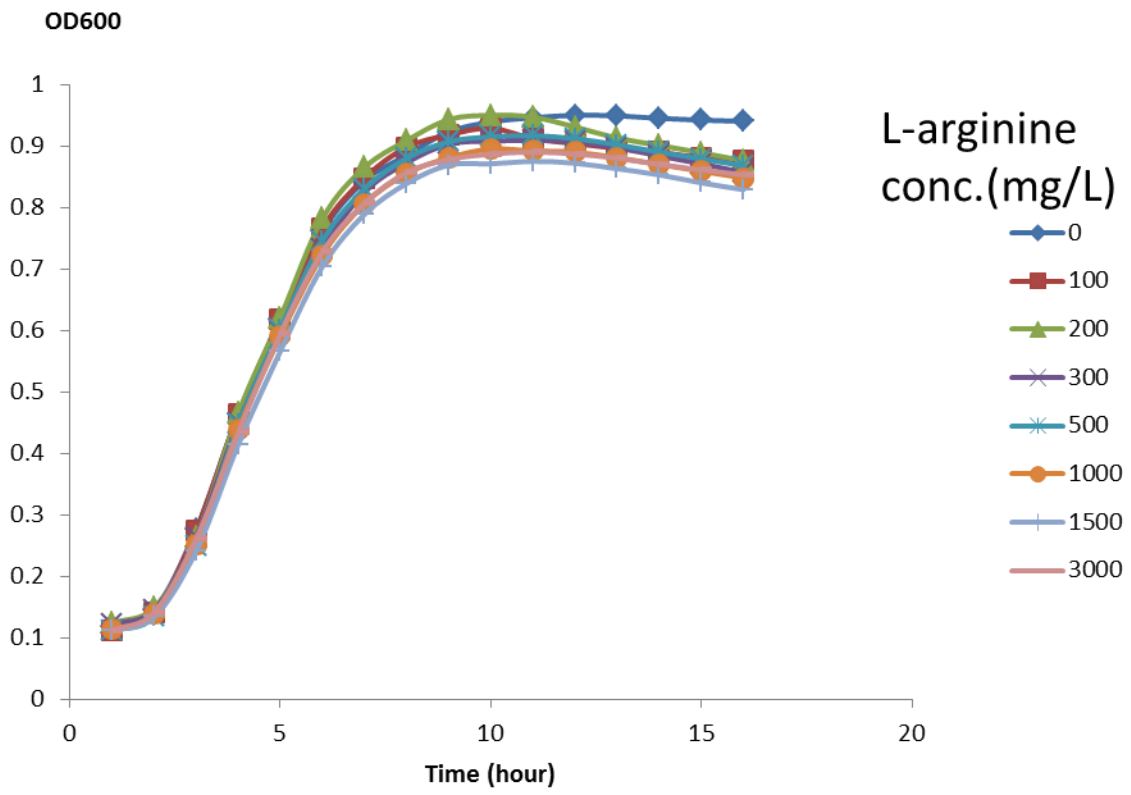

Fig. 4.16. Growth curve of normal Top10 cells with control plasmid (pACYC117) in LB media containing 3000 mg/L L-tryptophan and various amounts of added L-arginine (mg/L).

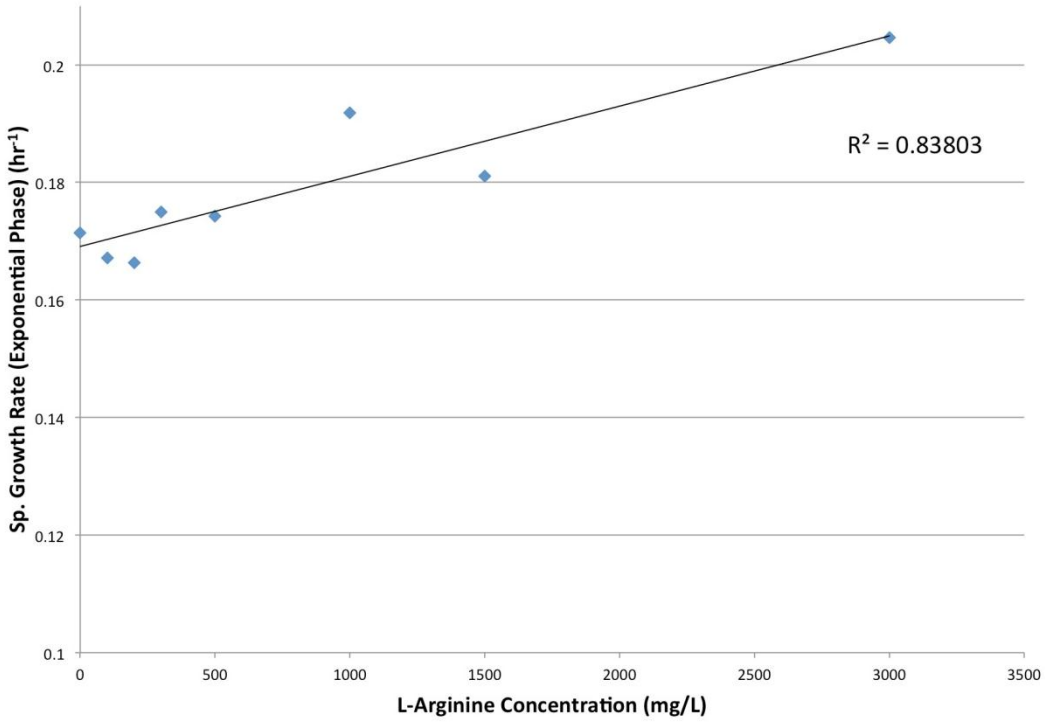

Fig 4.17. Growth curve of normal Top10 cells harboring the pACYCcBaRcAtR plasmid in LB media containing 3000 mg/L L-tryptophan as a function of added L-arginine concentration (mg/L).

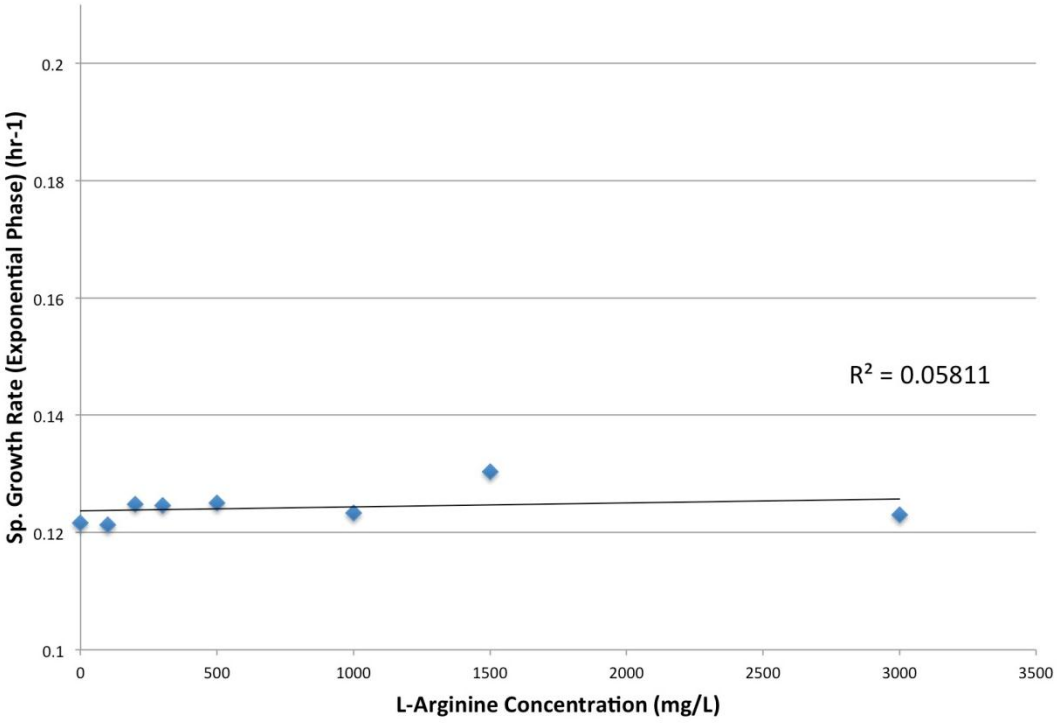

Fig 4.18. Growth curve of normal Top10 cells harboring the control pACYC177 plasmid in LB media containing 3000 mg/L L-tryptophan as a function of added L-arginine concentration (mg/L).

4.9. Addressing library cloning

The ultimate goal of the FITSelect system is to select genomic library inserts that result in L-arginine over-production. The research presented in this thesis resulted in the construction of a functional FITSelect genetic circuit, but library enrichment has not yet been performed. This is largely due to the time required to iteratively build this system and the complexity that resulted in comparison to the initial design (Fig. 4.1). Also, it is recognized that even though the FITSelect system is functional, additional fine-tuning is necessary. Ideally, the slope of the line relating the specific growth rate of the culture to L-arginine concentration (Fig. 4.17) must be much steeper for effective library enrichment. Ultimately, fine-tuning can eliminate the need for the modulation of CcdB with CcdA. One major issue that is recognized

with the FITSelect system in its current form is that a library insert may act to increase the growth rate of the cell not by causing over-expression of L-arginine but by reducing the amount of L-tryptophan in the cell. Currently, this is addressed by adding the "upper limit" of L-tryptophan to the culture so that growth is not increased if L-tryptophan is somehow reduced by a minor amount. However, this remains an area of attention for future designs and is discussed further in the Conclusions and Recommendations chapter.

4.10. References

1. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res 31*, 3406-3415.
2. Desai, R. P., and Papoutsakis, E. T. (1999) Antisense RNA strategies for metabolic engineering of Clostridium acetobutylicum, *Appl Environ Microbiol 65*, 936-945.
3. Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression, *Nat Biotechnol 27*, 946-950.
4. Van Melderen, L. (2002) Molecular interactions of the CcdB poison with its bacterial target, the DNA gyrase, *Int J Med Microbiol 291*, 537-544.

5. Conclusions, Recommendations and Possibilities

5.1. Conclusions of this research

In this project, a cheaper and less time-consuming selection method was developed to replace high-throughput screening methods for cells that over-produce L-arginine in a combinatorial metabolic engineering approach. This novel system was called Feed-back Inhibition of Transcription for Growth Selection (FITSelect), and it contained several synthetic "parts." In this system, "targets" are genes or sequences that can improve L-arginine production. In order to select for these targets in a growth competition assay, several cassettes were required including: (i) a *ccdB* cassette, (ii) a *ccdA* cassette, (iii) repressor genes and (iv) a genomic expression library (not performed in this research). In the *ccdB* cassette, the *ccdB* (cell death gene) is under control of *argF* promoter. The *argF* promoter can be inhibited by the presence of L-arginine. Similarly, the *ccdA* (antidote gene to *ccdB*) is transcribed under control of the *trp* promoter, which is inhibited by L-tryptophan. The repressor genes of FITSelect express the ArgR and TrpR repressor proteins that are activated by L-arginine and L-tryptophan respectively and bind the *argF* and *trp* promoters respectively to regulate transcription. The following is a simple description of the mechanism of the FITSelect system. The *ccdA* antidote gene is utilized to neutralize *ccdB* cell death gene during early culture growth, after transforming a genomic library. When the genomic library is expressed, *ccdA* transcription is stopped by adding a large amount of L-tryptophan to the culture media. Then, the genomic sequence which is able to enhance the over-production of L-arginine leads to cell survival by blocking *ccdB* transcription. On the other hand, cells harboring a genomic library insert that does not improve L-arginine production cannot fully repress the *ccdB* cassette, leading to cell death. Thus, selection of

productive strains can be achieved by simply culturing all cells together in a single flask. Research in this thesis resulted in construction of a FITSelect system that was responsive to concentrations of L-tryptophan and L-arginine. Individual testing of genetic "parts" showed that each one was functional, but considerable tuning of the system was needed to correlate the growth rate of the cell with L-arginine concentration. Ultimately, this required the following modifications to the initial conception of the FITSelect sytem: (i) inclusion of the *ccdB* cell death gene and modulation by the *ccdA* antidode gene (under control of the *trp* promoter), (ii) over-expression of the ArgR and TrpR repressors, (iii) down-regulating *ccdB* translation through modification of the RBS, and (iv) expression in a low copy number plasmid.

5.2. Recommendations

In this research, a functional FITSelect system was constructed. As shown in chapter 4, L-tryptophan was able to slow growth rate of the culture while adding L-arginine was also able to increase the culture growth rate. Fine-tuning is most likely necessary to achieve the anticipated target. To address fine-tuning in future research, a mathematical model of the FITSelect circuit may be built to quantity transcription and translation levels of *ccdB* and *ccdA* since these two proteins bind as (CcdA)2-(CcdB)2 to form a complex. Controlled studies with a green fluorescent protein (GFP) reporter gene can help learn the relative strengths of the *argF* and *trp* promoters in the presence and absence of their activated repressors. Modifying these promoter sequences and the RBS of the *ccdB* and *ccdA* genes can achieve more effective interaction between these two. This procedure can also help in the presence of low concentrations of L-tryptophan or L-arginine to regulate these promoters effectively. The intracellular concentrations of L-arginine and L-tryptophan were not measured in this research.

Additional “parts” to be added to the FITSelect sytem could include membrane transporters for each of these amino acids. This increase response time to L-tryptophan and ensure over-produced L-arginine is transported out of the cell (as long as enough is present to activate ArgR). This will prevent a situation where a large amount of *ccdA* and *ccdB* genes are already expressed before repressor proteins are functional. After these improvements, better system performance can be expected. Ultimately, through increasing the sensitivity of the *argF* promoter, the goal is to eliminate the *ccdA* cassette altogether.

5.3. Possibilities

Although the goal of this project is to select library inserts lead to higher production of L-arginine, exception still exists that higher cell growth rate is not caused by single library insert resulting in higher amount of L-arginine. Several possibilities are list here. First, library inserts act as antisense gene to *ccdB* gene. Thus *ccdB* gene is not functional to kill cells. Second, products from several library inserts collaborate to enhance production of L-arginine in cells. Isolation of single library insert does not result in higher production of L-arginine in cells. Third, false positive can be caused by certain proteins which have similar structure to L-arginine can binding to *argF* promoter. In this situation, survival cells are due to deficient of *ccdB* protein in cells.