Combinatorial Search Strategies for the Metabolic Engineering of Microorganisms

by

Christine Nicole S. Santos

B.S. Chemical Engineering, Stanford University, 2004
M.S. Chemical Engineering, Stanford University, 2004

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Signature of Author: _____
Department of Chemical Engineering
May 7, 2010

Certified by: _____
Gregory Stephanopoulos
Willard Henry Dow Professor of Chemical Engineering and Biotechnology
Thesis Supervisor

Accepted by: _____
William Deen
Carbon P. Dubbs Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Combinatorial Search Strategies for the Metabolic Engineering of Microorganisms

by

Christine Nicole S. Santos

Submitted to the Department of Chemical Engineering on May 7, 2010 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Chemical Engineering

ABSTRACT

Although the field of microbial metabolic engineering has traditionally been dominated by rational and knowledge-driven approaches, recent advances in genetic engineering have led to the emergence of a new methodology based on phenotypic diversification and screening. Unlike "classical strain improvement," which requires the use of general mutagens to introduce nonspecific chromosomal substitutions, these novel combinatorial methods enable sampling of a wider range of phenotypic space and, additionally, offer the important feature of genetic traceability. As an example, the use of transposon mutagenesis allows for the random integration of a genetic cassette within the chromosome for the generation of gene knockout libraries. More recently, the mutagenesis of cellular transcriptional components (global transcription machinery engineering, gTME) has enabled a complete reprogramming of the transcriptome, a useful feature for eliciting a broad array of phenotypes. Despite these advances in library generation, however, the application of these combinatorial approaches has surprisingly been limited to the engineering of only a handful of cellular properties. Thus, there remains a pressing need for a full evaluation of these techniques and, more specifically, an objective comparison of their relative strengths and weaknesses when applied towards strain improvement endeavors.

We decided to explore these specific issues using the metabolic engineering framework of L-tyrosine overproduction in *Escherichia coli*. Although this particular strain optimization problem merely represents a "model system" for these studies, such endeavors do have important industrial implications, as L-tyrosine serves both as a dietary supplement and a valuable precursor for a myriad of polymers, adhesives and coatings, pharmaceuticals, biocosmetics, and flavonoid products. To establish the early foundations for a combinatorial approach, we began with the construction of a "parental" or starting strain for the generation of these genetic libraries. This was achieved by utilizing several common rational engineering strategies to both deregulate and increase the flux through the aromatic amino acid biosynthetic pathway. The resulting strains, P1 and P2, exhibited L-tyrosine production levels of 358 mg/l and 418 mg/l, respectively, thus establishing an already high base line for this study. In a parallel investigation, we also worked on developing a simple high-throughput screen for L-tyrosine production in *E. coli*, another prerequisite for the use of these combinatorial approaches. This was accomplished through the heterologous expression of a bacterial

tyrosinase which provided a visual link between L-tyrosine production and the synthesis of the colored pigment, melanin. When implemented on a solid agar format, this assay allowed for the identification and isolation of high L-tyrosine producers from combinatorial libraries of more than 10^6 mutants.

Having established the basis for a combinatorial study, these strains and tools were subsequently applied for the generation and screening of three separate libraries - a random knockout library constructed through transposon integration and two plasmid-encoded gTME libraries based on the mutagenesis of the α subunit and the σ^{70} sigma factor of RNA polymerase (*rpoA* and *rpoD*, respectively). Several strains were isolated, with some gTME mutants exhibiting impressive titers of up to ~900 mg/l L-tyrosine, a 114% increase over the parental. Upon further examination, however, we discovered that phenotypic transferability was somewhat hampered in these strains due to an unusual requirement for both the plasmid-encoded *rpoA*/*rpoD and* a mutated chromosomal background to achieve the desired phenotype. Furthermore, the biochemical mechanisms triggered by these factors appeared to be nonspecific, as several plasmid-background combinations were found to lead to the same cellular behaviors.

To elucidate the biochemical underpinnings for these phenomena, we decided to conduct a full characterization of three isolated gTME strains through both microarray analysis and whole genome sequencing. Interestingly, whole genome sequencing revealed the presence of a separate unique mutation within each strain in two biochemically-related loci (*hisH*, *purF*). Although microarray experiments generally yield intractable results, we were also fortunate to find patterns of expression linking this phenotype to two different cellular responses – the acid stress resistance pathway and the stringent response. Indeed, the overexpression of two transcriptional regulators for these pathways (*evgA*, *relA*) was able to supplant the need for the mutant *rpoA* or *rpoD* plasmids, thus validating the contributions of these specific mechanisms towards determining cellular phenotype.

The successful identification of these critical genetic factors led us to the construction of a novel, genetically-defined strain (rpoA14^R) exhibiting a titer of 902 mg/l L-tyrosine and a yield of 0.18 g L-tyrosine/g glucose in 50 ml cultures. To put these numbers into perspective, this yield on glucose is more than 150% greater than a classically-improved strain (DPD4195) currently used for the industrial production of L-tyrosine and, when excluding biomass-related glucose utilization, represents 85% of the maximum theoretical yield. As an added feature, further engineering of this strain has established its capacity to produce the flavonoid precursor naringenin at competitive levels, thus providing a route for the synthesis of other important L-tyrosine derivatives.

During this study, we have successfully applied a combinatorial engineering approach for both eliciting a complex phenotype and identifying novel biochemical and genetic avenues by which to engineer future strains. As such, these combinatorial techniques have certainly proven to be valuable tools within the metabolic engineer's ever-expanding arsenal.

Thesis Supervisor: Gregory Stephanopoulos
Title: Willard Henry Dow Professor of Chemical Engineering and Biotechnology

ACKNOWLEDGEMENTS

There are several people to whom I remain indebted for their help and support throughout these years. First, thank you to my thesis advisor, Greg Stephanopoulos, for his guidance and perpetual optimism, even during times when good results seemed impossible to come by. He granted me the freedom and financial support to pursue my own ideas, and I am grateful to have experienced this level of independence during my training as both a scientist and a scholar. My thesis committee members – Charles Cooney, Boris Magasanik, Kristala Jones Prather, and Tina van Dyk – also deserve my gratitude for the assistance they have given me throughout this entire process. Thank you for always being generous with your time, your advice, and your encouraging words.

During my time at MIT, I have had the pleasure of collaborating with several colleagues on various aspects of this project. Thank you to Tina Lütke-Eversloh for her enthusiasm for L-tyrosine and assisting with the rational engineering of several strains. Thanks also to Hal Alper and Daniel Klein-Marcuschamer for their expertise in transcriptional engineering and sharing their *rpoD* and *rpoA* plasmids (and knowledge) with me. Wenhai Xiao was an invaluable asset during the large-scale fermentation studies – thank you for the sleepless nights and for showing such genuine excitement for foam and crystals. Lastly, I would like to thank Charlie Whitaker of the Koch Institute Bioinformatics and Computing Core Facility for his eager willingness to help with both the alignment and analysis of the genome sequencing data. Although not direct contributors to this work, several other people in the Stephanopoulos lab have made graduate school a truly memorable experience. Curt Fischer, Brian Mickus, Ajikumar Parayil, Mitchell Tai, Keith Tyo, Jason Walther, Benjamin Wang, Matthew Wong, Vikram Yadav, Hang Zhou, and many others - thanks for all the insightful discussions (both science-related and not) and for always being willing to share the ups and downs of research.

I would not be where I am today without the help of my family – my parents, Bien and Res, and my two siblings, Kae and Mykee. Thank you for the unconditional love and support you have given me throughout my entire life, and thank you for teaching me to always strive for knowledge and excellence.

Last but certainly not least, I would like to express my gratitude to Albert Chen, my best friend and cohort in life. Thank you for accompanying me during late-night time points, for lending a hand for tedious assays, and for spending your Saturdays in lab just to keep me company. Your love, devotion, and kindness never cease to amaze me, and it is to you that I dedicate this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Engineering microbes for chemical production

Harnessing the power of microbes for the synthesis of food products, fuels, and other chemicals is certainly not a novel concept. Although an understanding of the underlying biochemical processes was initially limited, humans have in fact been exploiting this microbial potential for millennia, whether it be for the fermentation of sugars to yield alcoholic beverages, the conversion of milk to cheese, or the cultivation of yeast to make bread. Despite this apparent long history of microbial use, the field of industrial microbiology did not formally emerge until the latter half of the 19th century. It was during this time that Louis Pasteur proved the existence of microorganisms that could act as tiny living reactors for the production of a variety of chemicals. Set off by this finding along with a growing understanding of both microbial physiology and bioprocess engineering, the next century saw the development of fermentation-based processes for the production of a myriad of compounds, including solvents, antibiotics, enzymes, vitamins, amino acids, and polymers. Early successes in this field often relied on the satisfaction of two important criteria: 1) that an organism capable of producing the desired compound either exists in nature or can be readily engineered, and 2) that environmental conditions, such as pH, temperature, and nutrient availability, can be manipulated to further control and enhance microbial conversion of the specified product (Demain 1981; Demain 2001).

Satisfying the first criterion was not a trivial task at the time, given that organisms in nature evolve with a single goal: to maximize their growth and survival under (often) nutrient-deprived conditions. Natural variants that overproduce industrially-relevant primary or secondary metabolites are at an evolutionary disadvantage, as they waste precious resources that could be otherwise diverted towards growth. Thus, they are rarely found in nature, and it often falls on the shoulders of the industrial microbiologist to "engineer" such properties into a strain. Furthermore, because these microbes have evolved in environments far different from the process conditions of an industrial fermentation, special consideration must also be made in developing robust organisms that can thrive in these new settings.

1.2 Classical strain improvement and its shortcomings

Until the development of recombinant DNA technology, scientists had no direct way of manipulating microbial genetics and instead had to rely on environmental and chemical factors to engineer strains through a "forced" evolutionary process. This often involves growing strains in a specific medium composition to ensure that variants exhibiting the fastest growth (designed to be linked to the phenotype of interest) can overtake the population and be isolated for further study. Although there is a low frequency of spontaneous mutations introduced during normal replication and growth, these selection-based methods are oftentimes coupled with the use of general mutagens, such as N-methyl-N'-nitro-N-nitrosoguanidine (NTG) or UV irradiation, to introduce genetic variability and diversity at the outset (Parekh, Vinci et al. 2000; Adrio and Demain 2006).

As demonstrated by the number of examples still being published in the literature (Parekh, Vinci et al. 2000; Selifonova, Valle et al. 2001; Gong, Sun et al. 2007; Wisselink, Toirkens et al. 2007), random mutagenesis followed by screening, or "classical strain improvement," remains a popular method for optimizing the performance of industrial strains. However, there are several disadvantages associated with such strategies. First, because the introduced mutations are non-specific and non-targeted, a large number of strains must be screened in order to isolate an improved mutant from a mixed population. In the absence of a simple selection method such as competitive growth, picking out one good mutant amongst millions literally becomes the equivalent of finding a needle in a haystack. Second, although random mutagenesis may introduce trait-conferring mutations, strains also have a tendency to accumulate deleterious mutations that negatively affect growth, performance, or survivability. Due to difficulties associated with discerning each individual mutation and its specific effect, it becomes nearly impossible to separate desirable changes from the undesirable ones, thus making this route for strain optimization quite inefficient. Lastly, because the strain is largely treated as a black box, these methods offer no additional insight into the underlying biochemical processes required for a specific phenotype. As a result, advantageous modifications become restricted to the strain in which they are found and cannot be transferred to new production hosts or strains.

1.2.1 Recombinant DNA technology paves the way for metabolic engineering

The development of advanced recombinant DNA technology in the early 1980s paved the way for a new field capable of circumventing many of the problems associated with classical strain

improvement. With an arsenal of sophisticated techniques for manipulating DNA, a more complete mapping of cellular biochemical networks, and the growing availability of genome sequence data, metabolic engineering, as it came to be known, ushered in a new approach for engineering cellular properties in targeted and tractable ways. Early successes in the field demonstrated the use of gene deletions, gene overexpressions, and pathway deregulation for altering nutrient uptake, redirecting metabolite flow, and revising metabolic regulation. Such tools also permitted flexibility in the selection of hosts, as it allowed for the introduction of completely novel properties through the transfer of whole heterologous pathways (Bailey 1991; Stephanopoulos 1999).

Although this rational or systematic approach is truly seen as the hallmark of traditional metabolic engineering, the field has also evolved in recent years to accommodate phenotypes that, for one reason or another, remain inaccessible through these methods. Most of these failures stem from the sheer complexity and interconnectivity of biological networks, which often preclude the recognition of simple genotype-phenotype relationships to guide cellular modifications. These difficulties in accurately predicting cellular responses to genetic perturbations have made combinatorial methods increasingly attractive approaches in metabolic engineering. With combinatorial search strategies, random yet easily traceable genetic-level modifications are introduced into a cell to yield a new population of strains exhibiting a diverse range of phenotypes. A high-throughput screen or alternate selection method is then required to isolate mutants exhibiting desirable characteristics. Due to the random nature of these approaches, no significant *a priori* knowledge is required of the system, making it quite versatile for a myriad of applications. The current repertoire of tools available

for enhancing genetic diversity includes transposon mutagenesis as a way of introducing random gene knockouts and genomic library complementation for the identification of overexpression targets (Alper, Jin et al. 2005; Alper, Miyaoku et al. 2005; Jin and Stephanopoulos 2007). More recent studies stemming from our laboratory have also shown that modulating the cell's transcriptional machinery (global transcriptional machinery engineering, or gTME) can be an effective way of reprogramming the transcriptome to elicit unique phenotypes (Alper and Stephanopoulos 2007; Klein-Marcuschamer and Stephanopoulos 2008).

1.2.2 The advantages of traceability in metabolic engineering techniques

Due to the nature of the metabolic engineering techniques describe above, all introduced mutations (by both direct and random means) can be easily traced through DNA sequencing. Amazingly, this one property by itself is able to overcome almost all of the disadvantages associated with classical strain improvement. Because only a single mutation is introduced during each round of engineering, improvements in phenotype can oftentimes be attributed to a change in either the expression and/or sequence of a single gene or protein product. Having this information at hand can greatly facilitate the transfer of a desired property to other hosts or strains of interest. In addition, individual mutations known to confer advantages can now be introduced and studied in combination in order to assess the potential for positive additive or synergistic effects. Finally, and perhaps most significantly, the ability to pinpoint the location and nature of combinatorially-introduced mutations lends significant credence to the concept of "inverse metabolic engineering" (Bailey, Sburlati et al. 1996). Indeed, knowing the identities

of the most relevant mutations can help elucidate the biochemical mechanisms in play and should ultimately prove useful in formulating directed strategies for future rounds of engineering.

1.3 Evaluation of combinatorial strategies using L-tyrosine as a model system

The plethora of promising techniques and approaches now available for engineering cellular properties makes strain improvement endeavors simultaneously hopeful and daunting. Given constraints with both time and funding, how does one go about choosing the techniques with the highest probabilities of a successful outcome? In particular, because combinatorial methods are a relative newcomer to the field and have only been applied to a handful of cases, it becomes worthwhile to evaluate the efficacy of such library-based strategies in engineering novel phenotypes.

There are several related questions and themes that this thesis hopes to address during the course of this project.

1) *Rational versus combinatorial*: Are there cases when a rational approach to metabolic engineering is preferred over a combinatorial strategy (and vice versa)? Given that rational and combinatorial techniques often act by orthogonal means, do they have the potential to act as complementary approaches?
2) *Versatility of gTME*: Studies demonstrating the successful use of advanced combinatorial techniques, such as gTME, have, to date, been limited to a small number of cases. Is gTME a technique that is generally applicable to a wide range of

phenotypes? How does this method of library generation compare to more established methods, such as transposon mutagenesis?

3) *Inverse metabolic engineering in practice*: Global approaches for cellular engineering, such as gTME, often lead to transcriptional changes across hundreds of genes, some of which may not be directly related to the property of interest. Can sophisticated 'omics tools, such as DNA microarrays, be used to identify a smaller subset of genes/proteins responsible for the cell's altered characteristics? In an inverse metabolic engineering approach, can this information be used in turn to formulate targeted, rational approaches for further rounds of engineering?

To explore these themes, we have chosen to apply and further develop combinatorial strategies for engineering strains using L-tyrosine production in *Escherichia coli* as a model system. *E. coli* is a natural biological platform of choice, not only because of its heavy use in industry but also because techniques for manipulating its genetics are perhaps the most advanced of any organism. L-tyrosine production was selected because it is a compound of significant industrial interest (Lütke-Eversloh, Santos et al. 2007) and because this phenotype has previously been studied using systematic metabolic engineering approaches (Lütke-Eversloh and Stephanopoulos 2005; Lütke-Eversloh and Stephanopoulos 2007; Olson, Templeton et al. 2007; Chavez-Bejar, Lara et al. 2008). Thus, this model system taken altogether allows for an easy and direct comparison of the merits, drawbacks, and potential of these techniques.

1.4 Thesis Objectives and Specific Aims

The main objective of this thesis is to evaluate the use of combinatorial engineering approaches for optimizing L-tyrosine production in *E. coli*. Along with the use of more established methods, such as transposon mutagenesis, we aim to utilize and expand on newer techniques such as gTME to achieve this goal. Strains recovered during these searches will then be rigorously characterized with a variety of genomic tools to gain a more complete understanding of the cell's shifting metabolic network and to serve as a guide for future engineering strategies. At the end of this thesis, it is our hope to comment on the validity and potential of such combinatorial-based approaches for strain improvement endeavors.

In accordance with the objectives of this thesis, the following specific aims have been proposed:

Specific Aim #1: Library generation and screening

A. Utilize rational engineering approaches to construct a L-tyrosine-producing strain to be used as a starting point for these combinatorial engineering studies.
B. Develop a high-throughput screen for the identification and isolation of mutants with improved levels of L-tyrosine production from large libraries of strains.
C. Apply and develop various combinatorial engineering methods, including transposon mutagenesis and gTME, for generating phenotypically-diverse libraries and identify mutant strains from within these libraries that exhibit high L-tyrosine titers.

Specific Aim #2: Strain characterization and inverse metabolic engineering

A. Conduct a full transcriptional analysis of isolated strains to determine significant changes in gene expression and identify potential genetic targets for further engineering.
B. Study the performance of select strains in large-scale fermentations.

Specific Aim #3: Flavonoid production from glucose

A. Demonstrate additional applications for the engineered strains by expressing a four-step heterologous pathway for the production of flavonoids from glucose.

1.5 Thesis Organization

In Chapter 2, we will present an overview of the growing field of metabolic engineering with an emphasis on recent tools that have allowed for a) the creation of combinatorial libraries and b) the subsequent analysis of isolated strains. In Chapter 3, we will introduce L-tyrosine as an important model system for these studies and discuss previous attempts at overproducing this valuable compound through both chemical and biological-based methods. We will then describe the rational engineering of a parental L-tyrosine-producing strain in Chapter 4, followed by discussions on the development of a high-throughput screen for L-tyrosine production in Chapter 5. Chapter 6 and 7 will focus on the use of two different combinatorial techniques, transposon mutagenesis and gTME, respectively, for uncovering genetic changes that result in high-L-tyrosine-producing strains. Because of the complex genetic nature of strains isolated through gTME, Chapter 8 will focus on the use of genomic techniques to gain a better understanding of the physiological changes occurring within these strains. We will show

that such results enable the formulation of new, directed strategies for developing strains with an improved phenotype. In Chapter 9, we will talk about the potential of using L-tyrosine-overproducers for the synthesis of flavonoid compounds, such as naringenin. Finally, we will conclude with Chapter 10 to discuss the impact of these results, as well as offer recommendations for future work.

CHAPTER 2 COMBINATORIAL APPROACHES AND 'OMICS TOOLS FOR METABOLIC ENGINEERING

2.1 Introduction

Metabolic engineering has emerged in the past two decades as the discipline aiming at the construction of organisms for the production of fuels and chemicals using modern genetic tools. Defined as "the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell" (Bailey 1991), this field was formalized in part from the advent of more sophisticated recombinant DNA techniques that allowed for the targeted modification of existing biochemical reactions or the introduction of completely heterologous pathways. As such, the earliest examples in the field focus on engineering cellular phenotype using rational modifications (gene deletions/overexpressions, pathway deregulation) based on existing stoichiometric, kinetic, and regulatory knowledge of a system (Bailey 1991; Stephanopoulos 1999). Although this "rational design" approach has been successful in many applications, it soon became apparent that making accurate genotype-phenotype predictions to guide these changes was no easy task. Owing to the incredible complexity and interconnectivity of biological networks, a single genetic perturbation often has a variety of unpredictable secondary responses within the cell. In a similar vein, the performance of biosynthetic pathways often depends on distal genes through kinetic and regulatory interactions whose origins are poorly understood (Stephanopoulos, Alper et al. 2004; Jin, Alper et al. 2005). To add

an additional layer of intricacy, engineering complex phenotypes may also require the modulation of several of these factors *simultaneously* (Bailey 1999).

Such challenges led to the development of a new scheme called "inverse metabolic engineering" for cellular optimization. This methodology involves three main steps: 1) the construction or identification (by selection) of strains possessing a desired cellular phenotype, 2) the evaluation and determination of genetic and/or environmental factors that confer the phenotype, and 3) the transfer of this phenotype to another strain through direct modifications of the identified genetic and/or environmental factors (

Figure 2.1) (Bailey, Sburlati et al. 1996).

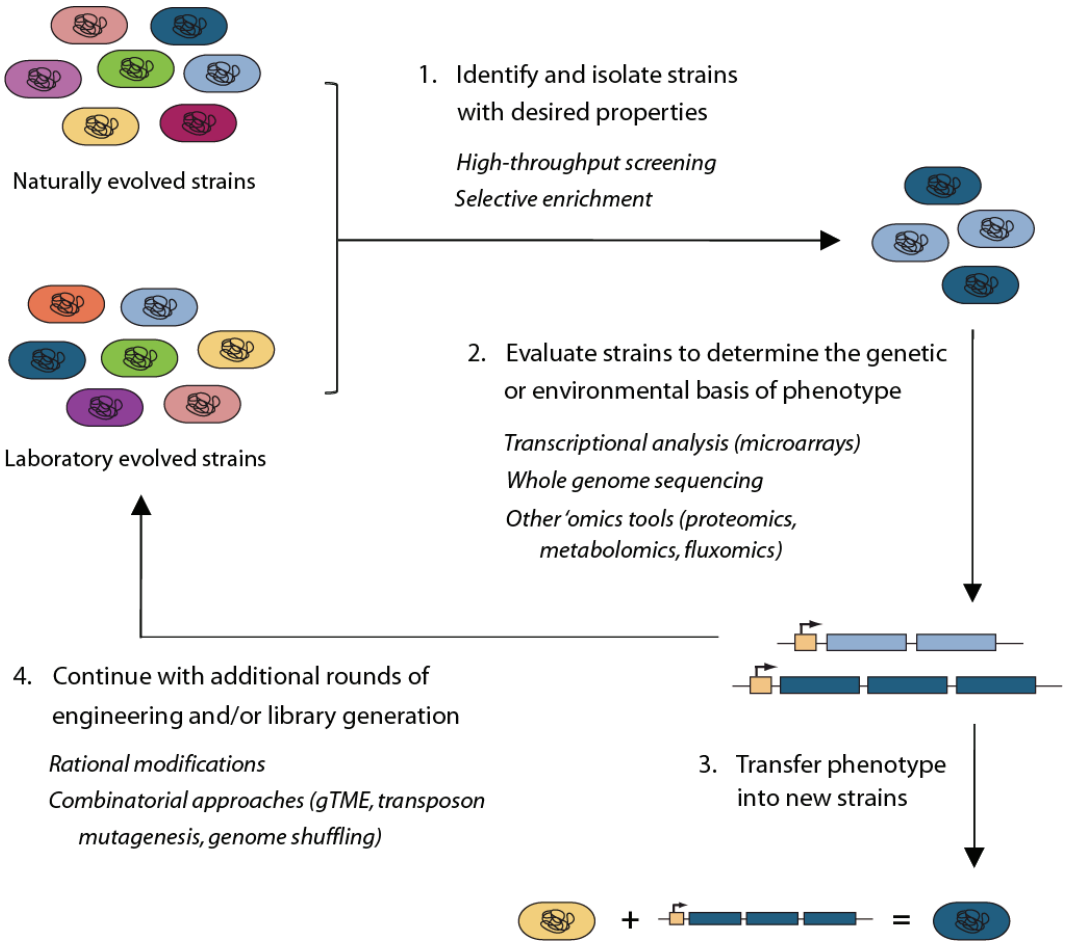


Figure 2.1 Inverse metabolic engineering approach

Combinatorial methods for generating library diversity have greatly facilitated the identification of strains for the first step, particularly for cases when no suitable organism can be found in nature. Of note, this process of library creation and screening can be repeated as many times as is required to isolate better performing strains. Once a functional strain has been isolated, a variety of 'omics tools are then utilized to characterize the cell's changing physiology and to extract the main biochemical determinants of phenotype. In this chapter, we will discuss the wide array of tools available for tackling both of these important challenges.

2.2 Combinatorial engineering tools for generating libraries

Although fine-tuning the relative levels of genes and gene expression in a specific production pathway may lead to improved phenotypes, the majority of applications require engineering techniques that can alter the entire cellular milieu in a less targeted and more global fashion. This concept of strain randomization and selection/screening is not a foreign one and is a methodology that has been employed extensively for the generation of industrial strains. For example, "classical strain improvement" relies on mutagenic agents such as NTG or UV irradiation to introduce genetic diversity (in the form of point mutations) in a population of strains (Parekh, Vinci et al. 2000; Adrio and Demain 2006). Other techniques, such as "evolutionary engineering," apply a selection pressure and allow the organism to naturally evolve in a chemostat, thus relying on the cell's inherent capacity to introduce adaptive mutations (Sauer 2001). Others have tried to combine aspects of these two approaches by temporarily inducing a mutator phenotype in a strain prior to selection and screening (Selifonova, Valle et al. 2001). Although these methods have been successful in engineering

several phenotypes, they suffer from being laborious and time-consuming, often requiring multiple rounds of screening and mutagenesis and, for the case of evolutionary engineering, extensive cultivation periods (Yomano, York et al. 1998; Wisselink, Toirkens et al. 2007). Furthermore, because silent mutations tend to accumulate, it is extremely difficult to identify which mutation(s) are necessary for conferring a particular phenotype.

Because of the inherent disadvantages with classical strain improvement and evolutionary engineering, several groups have focused their attentions on designing novel methods for generating phenotypic diversity in a population of strains. Here, we review several alternatives to these more established classical methods that offer the important features of traceability (thus making it amenable to phenotypic transfer) and an enhanced ability to sample a larger portion of the phenotypic space available to a cell.

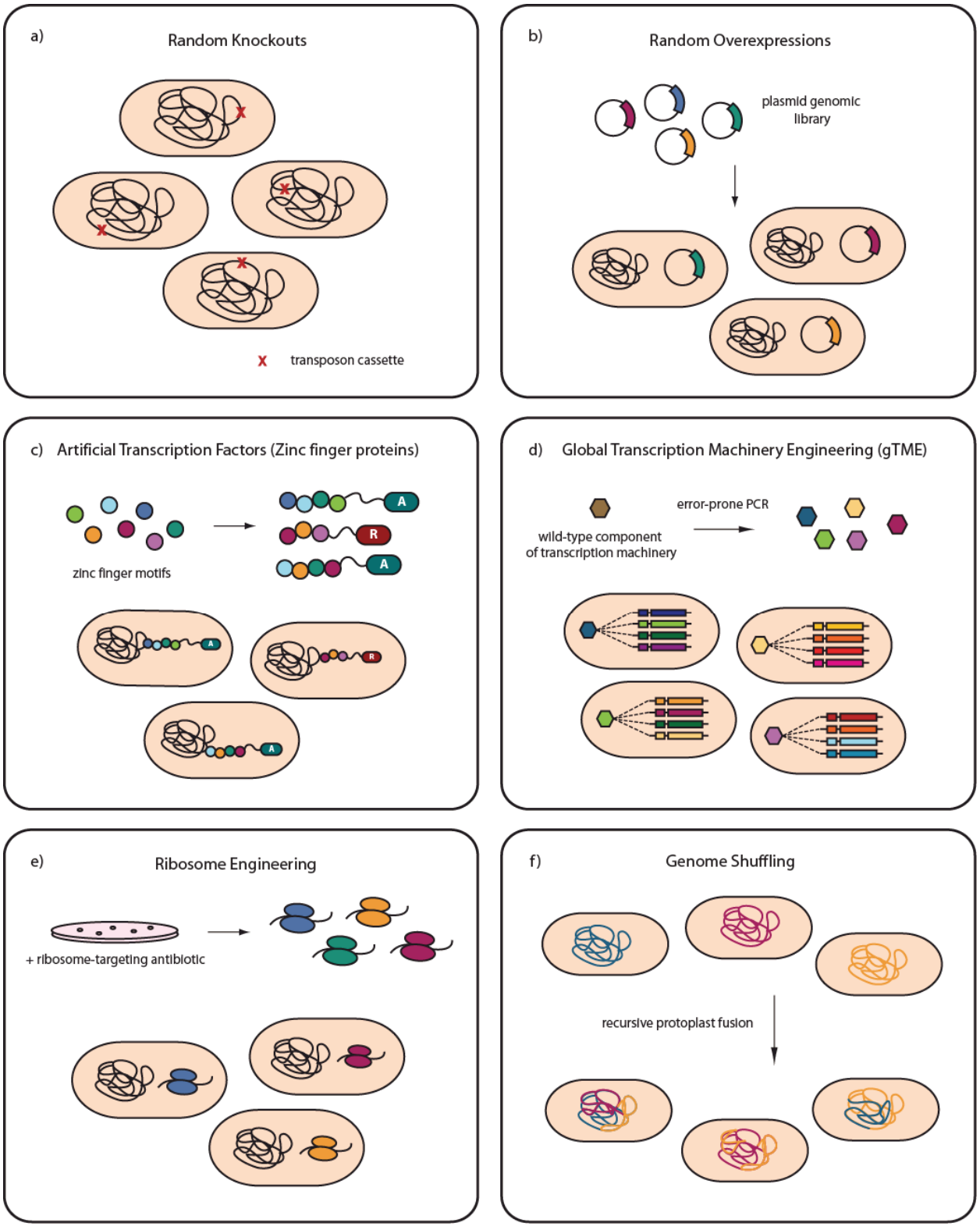
2.2.1 Random knockout and overexpression libraries

Methods for predicting knockout and overexpression targets rely heavily on *a priori* knowledge of the biochemical network and often take the form of stoichiometric models and algorithms, such as flux balance analysis (FBA) (Alper, Jin et al. 2005) and minimization of metabolic adjustment (MOMA) (Segre, Vitkup et al. 2002). However, a variety of recombinant DNA techniques are also available for introducing these genetic changes in a randomized fashion, which introduces the possibility of uncovering regulatory, kinetic, or unknown/poorly understood targets that are not encompassed in the model. For example, transposon mutagenesis, which involves the random integration of an antibiotic marker into the genome mediated, is an effective tool for creating random knockout libraries (Figure 2.2a). Although

previously prone to insertions into preferred "hot spot" regions, this system has now been optimized for effective and unbiased integration into the entire genome (Alexeyev and Shokolenko 1995; Badarinarayana, Estep et al. 2001). Other knockout approaches have also been demonstrated, including one based on antisense RNA inhibition and promoter interference in *Candida albicans*. For this system, a library of antisense complementary DNA (cDNA) fragments is placed under the control of an inducible GAL1 promoter which, when expressed, is able to inhibit gene expression (De Backer, Nelissen et al. 2001). Overexpression studies are often conducted simply by transforming the parental strain with a plasmid-encoded genomic library (Figure 2.2b) (Gill, Wildt et al. 2002; Jin and Stephanopoulos 2007; Lynch, Warnecke et al. 2007).

Figure 2.2 Phenotypic diversification through whole-cell engineering techniques

a) Random knockout libraries are created through the arbitrary insertion of a transposon cassette into the host genome, resulting in the disruption of the genetic element at the site of integration. b) Random overexpression libraries are generated by transforming the host strain with a plasmid-encoded genomic library. c) Artificial transcription factor (zinc finger protein) libraries are constructed by combinatorially assembling three or four zinc finger motifs, followed by attachment of an activator (A) or repressor (R) domain. Expression of these proteins in a strain leads to changes at the level of transcription. d) Global transcription machinery engineering (gTME) begins with the mutagenesis of a component of the transcription machinery (often in charge of DNA recognition and binding) by error-prone PCR. The introduction of this mutant constituent results in a complete alteration of the global transcriptome. e) In ribosome engineering, strains are first selected on media containing antibiotics that target ribosomal proteins and rRNA. The surviving strains, which contain mutations in the translational machinery of a cell, exhibit changes in protein translation and hence, cellular phenotype. f) Genome shuffling employs recursive protoplast fusion in order to create a library of cells containing shuffled genomic DNA from two or (often) more parental strains.


a)
Random Knockouts
transposon cassette
b)
Random Overexpressions
plasmid genomic library
c) Artificial Transcription Factors (Zinc finger proteins)
zinc finger motifs
A
R
A
d) Global Transcription Machinery Engineering (gTME)
error-prone PCR
wild-type component of transcription machinery
e)
Ribosome Engineering
+ ribosome-targeting antibiotic
f)
Genome Shuffling
recursive protoplast fusion

The advantage of these knockout/overexpression approaches is that the phenotype can easily be linked back to a single genetic perturbation. The presence of an integration cassette within a transposon library allows one to identify the region of insertion through a modified Thermal Asymmetric Interlaced PCR reaction (TAIL-PCR) (Jin, Alper et al. 2005), and the genomic pieces selected for overexpression can be identified by plasmid sequencing (Jin and Stephanopoulos 2007). In addition, many groups have developed microarray-based techniques to increase the throughput of target identification, which is particularly important if several unique genes are found to confer phenotypic improvements. The incorporation of a T7 promoter within the transposable cassette allows for the creation and analysis of cDNA libraries at the point of insertion and has been demonstrated to be successful in identifying gene clusters that are negatively selected during growth competition experiments of *E. coli* in rich and minimal media (Badarinarayana, Estep et al. 2001). More recently, a method for a multiscale analysis of library enrichment (SCALEs) has been developed which facilitates the identification of single gene or whole operon overexpressions which contribute to a phenotype of interest (Lynch, Warnecke et al. 2007).

Although these approaches are often studied in the context of growth phenotypes, they have also been shown to be remarkably successful for the case of lycopene production in *E. coli*. A random knockout search in the background of a previously constructed high-lycopene-producing parental strain led to the identification of three targets – *hnr*, *yjfP,* and Δ_p*yjiD* – that resulted in further increases in lycopene content. It is important to note that each of these genes encode for proteins with regulatory or unknown functions and thus would not have been predicted by a simple stoichiometric model (Jin, Alper et al. 2005). In a later study, the

introduction of a plasmid-encoded genomic library in the background of a mutant containing systematically-identified and transposon-based knockouts yielded additional overexpression targets (*rpoS, yjiD, ycgW*) conferring a lycopene-producing phenotype (Jin and Stephanopoulos 2007). This clearly demonstrates the potential in combining these two complementary approaches for phenotypic improvement.

2.2.2 Targeting transcription/translation for introducing global perturbations

Sequential approaches for cellular optimization are not always capable of finding the global optimum of a phenotype due to the often convoluted form of the metabolic landscape (Jin, Alper et al. 2005). As a result, accessing these complex phenotypes from a single round of strain diversification requires techniques for introducing genetic changes with known pleiotropic effects. Most of the approaches that have been developed target proteins that have widespread regulatory functions in important cellular processes such as transcription and translation. Thus, the alteration of these proteins holds enormous potential for generating widely diverse libraries for strain improvement studies.

Artificial transcription factor libraries based on zinc finger motifs

Zinc fingers are highly specific DNA-binding domains that recognize 3-base pair (bp) sequences and are found in a variety of transcriptional regulatory proteins. A single transcription factor may contain several of these motifs, which can be assembled in a highly modular fashion. This modularity, coupled with the ability to attach additional effector (activator/repressor) domains, make zinc finger proteins powerful tools for introducing transcriptional diversity (Figure 2.2c).

The utility of this approach was first demonstrated in *S. cerevisiae* through the creation of a randomized plasmid library of over 10^5 three-finger and four-finger proteins. Transformation of these plasmids and subsequent screening led to the isolation of zinc-finger proteins that were capable of conferring a variety of tolerance phenotypes in *S. cerevisiae*, including resistance to heat treatment, osmotic pressure, and the antifungal drug ketoconazole (Park, Lee et al. 2003). 11 different zinc finger proteins were found to grant varying degrees of ketoconazole resistance. Significantly, coexpression of two factors often led to enhanced resistance with the best performing double expression strain experiencing a 1,000-fold enhancement in the resistance phenotype. The most surprising result, however, was the discovery that swapping the effector domain of a zinc finger protein selected for in one phenotype (from activator to repressor or vice versa) could confer the *opposite* phenotype. Although the broad applicability of such a result needs to be examined, this introduces the possibility of utilizing this technique for situations in which the opposite phenotype is more amenable to screening.

In a later study conducted in *E. coli*, 40 different zinc fingers were used to construct libraries of 6.4×10^4 and 2.6×10^6 three-fingered and four-fingered proteins, respectively. Despite the absence of an activator or repressor domain in these proteins, the authors were successful in identifying a zinc finger protein (T9) conferring a thermotolerance phenotype (Park, Jang et al. 2005). After measuring the transcript levels of genes containing potential binding sites, the thermotolerance trait was shown to be a result of the downregulation (but not complete inhibition) of the ubiquinone biosynthesis gene *ubiX*. Thus, these zinc finger proteins are also capable of directing subtle changes in gene expression, a trait that is not accessible through the previously discussed knockout and overexpression approaches.

Global Transcription Machinery Engineering (gTME)

Global transcription machinery engineering (gTME) is a powerful new tool which enables a complete reprogramming of the cellular transcriptome. It does so through the targeted mutagenesis (via error prone PCR) of select components of the transcriptional machinery, particularly those which are involved in DNA sequence recognition and thus dictate the promoter preferences of RNA polymerase (Figure 2.2d). In prokaryotic systems, sigma factors bind to the promoter regions of genes with varying degrees of affinity and help to preferentially recruit the RNA polymerase holoenzyme to initiate transcription. Thus, slight variations in these proteins have the potential to greatly affect the subset of genes which are bound by RNA polymerase and expressed. Mutated and plasmid-encoded versions of the principal sigma factor in *E. coli* (σ^{70}) resulted in strains optimized for a variety of phenotypes, including ethanol tolerance, lycopene overproduction, and simultaneous tolerance to sodium dodecyl sulfate (SDS) and ethanol (Alper and Stephanopoulos 2007). These phenotypes, some of which were previously unattainable through traditional methods of strain improvement, were generated in a highly efficient manner and required only a few rounds of screening to promote substantial phenotypic improvements. This methodology has also been shown to be effective in *S. cerevisiae* through the mutation of the TATA-binding protein (encoded by *SPT15*) and one of the TATA-binding protein-associated factors (*TAF25*). With this approach, Alper et al. were successful in isolating a strain that acquired a high degree of tolerance to high glucose/high ethanol conditions and exhibited a 70% improvement in the volumetric productivity of ethanol. We note that the phenotype of cells harboring the mutant sigma (or transcription) factors were indistinguishable from that of the control in the absence of stress (Alper, Moxley et al. 2006).

This work has recently been extended to another prokaryotic organism, *L. plantarum*, and this study reveals several important findings about the gTME approach. By directly comparing gTME with the more established method of NTG mutagenesis, the authors were able to demonstrate that the former is significantly more effective in generating phenotypic diversity as measured by the complex phenotype of growth/colony size. This increased diversity was also found to be linked to a greater probability of finding a desirable mutant in a given screen as tested with the trait of malic acid tolerance. Finally, diversity was shown to be well-correlated with the frequency of error generation during the mutagenic PCR reaction with higher mutation frequencies leading to a wider range of phenotypes (Klein-Marcuschamer and Stephanopoulos 2008). This last point is particularly important, since it offers an additional knob for controlling this system and adapting it to various applications.

Ribosome Engineering

Modulating ribosomal proteins and rRNA is an alternative way for altering bacterial gene expression at the level of translation. This technique, termed "ribosome engineering," begins with the introduction of mutations that confer resistance to a variety of antibiotics that attack the ribosome, including streptomycin, gentamicin, kanamycin, and chloramphenicol. Selection on different concentrations of antibiotics facilitates the isolation of strains with a wide variety of mutations in known ribosomal components (Figure 2.2e). Although the mechanism leading to phenotypic variations is not well understood, the changes are hypothesized to be a result of the activation of a "stringent response" which leads to elevated levels of protein synthesis during the late growth phase and is beneficial for a select number of specialized phenotypes.

Such a technique was successful in conferring antibiotic-producing properties in several *Streptomyces* organisms, enzyme (α-amylase and protease) overproduction in *Bacillus subtilis*, and improved tolerance to aromatic compounds, such as 4-hydroxybenzoate, toluene, and *m*-xylene in *Pseudomonas putida* (Ochi, Okamoto et al. 2004; Ochi 2007). It has yet to be determined whether a more randomized mutagenesis approach will be capable of yielding strains optimized in other cellular phenotypes. Because mutations are introduced directly into the chromosomal copies of these translational proteins, the majority will likely lead to an inhibition of growth and may ultimately limit the broader utilization of this technique.

2.2.3 Genome Shuffling

Whole-genome shuffling is a technique for generating a combinatorial library of complex progeny from a few previously selected parental strains exhibiting subtle improvements in a desired property. The utilization of a method called recursive protoplast fusion ensures a thorough shuffling of the genome, which allows for the generation of strains that may accumulate mutations that have additive or even synergistic effects with respect to the phenotype of interest (Figure 2.2d). Compared to classical strain improvement strategies, this technique has been shown to be more effective in producing strains with enhanced phenotypes. Impressively, when applied towards tylosin production in *Streptomyces fradiae*, two rounds of genome shuffling were found to be successful in achieving the same titers generated from twenty rounds of classical strain improvement. As stated aptly, "this is the difference between 24,000 assays and approximately 1 year of effort, and roughly 1,000,000 assays and 20 years of effort" (Zhang, Perry et al. 2002). Since then, this tool has been used to

engineer a myriad of other complex phenotypes, including improved acid tolerance in *L. plantarum*, pesticide degradation by *Sphingobium chlorophenolicum*, hydroxycitric acid production in *Streptomyces* sp. U121, and epithilone production by S*organgium cellulosome* (Patnaik, Louie et al. 2002; Dai and Copley 2004; Gong, Sun et al. 2007; Hida, Yamada et al. 2007). Thus, genome shuffling is clearly an important metabolic engineering tool for strain improvement, particularly when strains with intermediate properties have already been isolated.

2.2.4 Summary

Although we now have access to a diverse collection of tools for the generation of libraries, these combinatorial approaches have thus far been limited to the engineering of only a few cellular properties. As evidenced by the body of literature in this area, the major thrusts have largely been on identifying genetic changes that alter the growth characteristics of a strain (Badarinarayana, Estep et al. 2001; Gill, Wildt et al. 2002; Patnaik, Louie et al. 2002; Park, Lee et al. 2003; Park, Jang et al. 2005; Alper, Moxley et al. 2006; Lynch, Warnecke et al. 2007), a phenotype that is easily selected by serial subculturing or chemostat cultivation. Because of the relative youth of these techniques, particularly newer methodologies such as gTME, the next major thrusts in this field must clearly focus on characterizing the potential of these techniques for accessing a wider array of phenotypes.

2.3 'Omics tools for strain characterization

Unfortunately, the global nature of several of the approaches discussed above often becomes a hindrance when it comes time to reconstitute the desired features in alternate hosts. Indeed, a full 'omics analysis of the strain may be needed to understand the basis for these phenotypic differences, and full genome sequencing/reconstructions may be necessary to initiate phenotypic transfer into a new strain. Techniques ranging from proteome analysis, metabolite profiling, and comparative flux analysis have been developed in recent years to access information about the metabolic states of a cell. However, the successful application of these tools towards an inverse metabolic engineering paradigm has been quite difficult due to the vast amount of data generated through these approaches (Bro and Nielsen 2004; Tyo, Alper et al. 2007; Kim, Kim et al. 2010).

Despite these challenges associated with data interpretation, genomic tools for strain characterization have surprisingly led to a handful of success stories. In this section, we will present a few case studies on the use of two genomic techniques – transcriptional profiling and whole genome sequencing – on the characterization and engineering of cellular phenotype.

2.3.1 Transcriptional profiling

Transcriptional profiling utilizes DNA microarrays to provide a descriptive snapshot of the genome-wide mRNA expression levels within a cell. Its ability to probe global differences in gene expression has, in many cases, proven quite useful in characterizing cellular phenotypes; however, successful applications of transcriptional profiling towards *engineering* cellular phenotype have so far been limited to a small number of cases. As mentioned earlier, this is

largely due to the copious amounts of data generated and the absence of methods available for translating this information into useful engineering strategies. Indeed, it is not unusual for DNA microarrays to uncover hundreds of differentially expressed genes between two cellular states. With so much strain to strain variability, distinguishing which changes in gene expression are *causal* to a phenotype as opposed to being a mere *result* of the phenotype certainly becomes a daunting task (Gill 2003; Bro and Nielsen 2004; Lynch, Gill et al. 2004).

This genome-wide transcriptional approach has been applied to a number of studies with only varying degrees of success. In 2003, Bro and coworkers tried to utilize this approach to characterize differences between wild type *S. cerevisiae* and two rationally engineered strains exhibiting 26% and 41% increases in their maximum specific galactose uptake rates. Their ultimate goal was to use a global analysis in order to identify novel genetic targets for improving the flux through the galactose utilization pathway. However, when no clear explanation or correlations were identified from a genome-wide search, the authors decided to focus their attention on the differential expression of genes previously known to be involved with galactose metabolism. Through this, they recovered *PGM2* (the major isoform of phosphoglucomutase) as a target and showed that subsequent overexpression of this gene could increase galactose uptake by 70% over the wild-type (Bro, Knudsen et al. 2005).

In a similar example, transcriptional profiling was used to characterize 11 engineered strains of *Aspergillus terreus* exhibiting a wide range of lovastatin titers (Askenazi, Driggers et al. 2003). Association analysis was then performed to find genes whose expression profiles correlated well with the production profile of this cholesterol drug. Although the authors' search led to the identification of several interesting biosynthetic clusters, this information was

ultimately abandoned in favor of a more simplistic approach for increasing lovastatin titers. Seeing a correlation between lovastatin concentrations and *lovF* (a lovastatin biosynthetic gene) expression, the authors decided to link the *lovF* promoter with the *ble* gene encoding for phleomycin resistance. They then demonstrated the ability to select for higher lovastatin producers through the addition of phleomycin to the media. The utility of such a technique, however, seems to have been overstated, as the recovered strains did not appear to perform better than those engineered by rational means (Askenazi, Driggers et al. 2003).

As a final case study, transcriptional profiling has also been used to characterize the growth of *S. cerevisiae* on xylose minimal medium. The expression profiles of a classically improved strain exhibiting enhanced xylose utilization and growth was compared to that of a wild-type strain. Again, rather than examining transcriptional differences in all 6000 yeast genes, the authors chose to limit their search to specific clusters and pathways known to be involved with xylose metabolism. Despite having uncovered a potential transcriptional target, direct modification of this regulator (either through overexpression or deletion) had either no significant effects or a detrimental effect on growth on xylose (Wahlbom, Cordero Otero et al. 2003).

Although these three specific examples are certainly not meant to be taken as a comprehensive overview of this topic, they do serve to illustrate a couple important points and themes. First, despite early attempts to approach these problems from a *global* perspective, each of these groups eventually chose to focus on a smaller subset of genes known to be involved with the desired phenotypes. Although this is an understandable “second pass” given the complexity of these datasets, this unfortunately eliminates all possibilities of uncovering

novel and unexpected genetic targets for further engineering. This is a particularly important point when dealing with more complex phenotypes, as many of the relevant biochemical pathways, reactions, or mechanisms may still be uncharacterized or very poorly understood.

Second, these examples show that despite having supporting transcriptional data along with seemingly sound and plausible mechanisms, genetic modifications, once introduced into a cell, often have unexpected physiological effects. Indeed, even when a potential target is identified (such as PET18 for the case of xylose utilization), there is no guarantee that its modification by overexpression, deletion, or sequence revision will result in any real benefits on cellular phenotype. It was particularly disappointing to see that although several pathways and biosynthetic clusters were shown to correlate with lovastatin production (and despite having formulated detailed hypotheses on how these may be influencing phenotype), the authors were not able to use this information to their advantage (Askenazi, Driggers et al. 2003). Rather than making attempts to elucidate these interactions, their follow-up studies focused on the use of a *lovF-ble* gene cassette as a selection mechanism, which did not make use of any novel transcriptional data and also failed to isolate strains that performed better than their rationally-engineered counterparts. Clearly, more efficient ways for both filtering through the data and testing potential genetic targets must be developed in order for this global approach to be useful.

2.3.2 Whole genome sequencing

In these past few years, we have seen the development of several next-generation sequencing platforms which have drastically reduced both the costs and investment in time normally

associated with large-scale sequencing projects (Smith, Quinlan et al. 2008). As a result, whole genome sequencing has quickly emerged as an attractive and financially viable option for comparing the full genetic make-up of interesting strains. As one might imagine, these tools are particularly useful for characterizing strains which have been subjected to chemical mutagenesis and as a result, have accumulated random insertions, deletions, and point mutations within its genome. Although these techniques are not as crucial for studying strains in which traceable mutations have been introduced (as is the case with random overexpression/knockout or gTME studies), they could still ultimately prove to be useful for the identification of natural mutational events that may occur during the course of cultivation and selection. When combined with experimental verification, these comprehensive sequencing studies allow one to distinguish the effects of each mutation to facilitate the maintenance of beneficial changes and the removal of detrimental modifications.

To this end, Herring and coworkers used a comparative genome sequencing approach to track the evolution of *E.coli* cultures cultivated on glycerol minimal medium. Over the course of 44 days (~660 generations), a total of 13 *de novo* mutations were identified and verified to be present within five representative members of the population (2-3 mutations/strain). Interestingly, each strain had acquired a mutation in *glpK* (glycerol kinase), the enzyme responsible for the first step in glycerol metabolism. In addition, 3 of the 5 strains contained deletions or point mutations within the β subunits of the RNA polymerase, suggesting that widespread transcriptional alterations may also be at play. When single, double, and triple mutants were constructed, the growth rate of four of these strains matched those of the evolved clones when grown on glycerol minimal medium. Clearly, this comparative genome

sequencing approach was successful in identifying all phenotypically-relevant mutations and as such, the authors were able to reconstruct several fully functional strains (Herring, Raghunathan et al. 2006).

In the most impressive example to-date, genome sequencing was recently used to characterize mutations in an industrial L-lysine producer that had previously undergone decades of classical strain improvement. Due to the large number (>1000) of mutations that had accumulated during this period, the authors focused their initial characterization on genes known to be relevant to L-lysine biosynthesis. Mutations were reintroduced in a clean genetic background in a step-wise fashion (starting with L-lysine terminal pathways and moving out to central carbon metabolism) in an attempt to reconstitute the L-lysine production phenotype. Mutations that had a beneficial effect on L-lysine titers were retained in the engineered host while those that did not exhibit positive effects were eliminated. When this "minimally mutated" strain was characterized in large-scale fermentations, it was not able to achieve the highest titers seen with the original L-lysine producer (90 g/L versus over 100 g/L L-lysine); however, the strain exhibited a much more robust constitution, with a significant increase in productivity (thus requiring a shorter overall fermentation time) and the added capability of growing at higher temperatures. Although additional mutational analysis must be performed to increase titers, this study demonstrates the utility of employing whole genome sequencing for the reconstruction of strains exhibiting improved growth characteristics (Ikeda, Ohnishi et al. 2006).

2.3.3 Summary

As demonstrated by these examples, genomic tools such as transcriptional profiling and whole genome sequencing have been applied towards inverse metabolic engineering with varying degrees of success. Whole genome sequencing of classically mutated strains can oftentimes lead to the reconstruction of robust performers, particularly when the number of identified mutations remains tractable. In these cases, the effects of individual mutations can be easily tested, and advantageous changes can be combined in order to achieve additive or possibly synergistic improvements on phenotype.

Unfortunately, moving just one step away from the raw sequence data to a gene *expression* analysis results in more challenging obstacles in both data analysis and target identification. To overcome such difficulties, groups will oftentimes focus on a smaller subset of genes known to be involved in the phenotype of interest, which unfortunately limits the exploration of *genome*-wide differences. And as we saw in many cases, once plausible targets are isolated and tested, there is absolutely no guarantee that the modified strains will behave in the way that one would have predicted. Altogether, these results are really not that surprising given the complexity of a cell's biochemical makeup and the multiple layers of regulation governing this large network of reactions. When you also take into consideration that of the 4460 *E. coli* genes, functions have been assigned to only 76% (66% of which have been verified experimentally), it really is no wonder that our current biochemical knowledge is not always sufficient for accurately guiding strain optimization endeavors (Karp, Keseler et al. 2007). Clearly, although genomic tools and other 'omics strategies (proteomics, metabolomics, fluxomics) hold quite a bit of potential for aiding inverse metabolic engineering studies,

additional advances in data analysis and interpretation must be made in order to increase the success rate of these techniques being reduced to practice.

CHAPTER 3 L-TYROSINE: APPLICATIONS AND PRODUCTION

3.1 Introduction

As mentioned in Chapter 1, we have chosen to explore the potential of combinatorial engineering techniques using L-tyrosine production as a model system. L-tyrosine has been a compound of significant industrial interest, and its production has been explored for several decades. In this chapter, we will discuss some of the industrial uses of this compound and then summarize both chemical and biologically-based processes that have been developed for its production. A special emphasis will be placed on the rational engineering of microbes such as *E. coli*, as some of these studies will form the basis of the construction of a "parental" strain for future library generation.

3.2 Applications and industrial importance of L-tyrosine

The demand for the three aromatic amino acids (L-tyrosine, L-phenylalanine, and L-tryptophan) has been high due to several uses, primarily in the food, feed, and pharmaceutical industries (Breuer, Ditrich et al. 2004). L-Phenylalanine, which is the precursor for the artificial sweetener aspartame, comprises the largest market volume by far with an estimated production scale of 18,000 tons in 2006 (Sprenger 2007). In contrast, L-tyrosine's market size of less than 200 tons per year presumably explains why the fermentative production of L-tyrosine has garnered less attention than microbial L-phenylalanine and L-tryptophan synthesis.

Despite its lower market volume, L-tyrosine remains a valuable compound with a myriad of applications. It itself is used as a common dietary supplement, primarily due to anecdotal reports of its ability to stimulate brain activity (for improved memory and mental alertness), to act as an appetite suppressant, to control depression and anxiety, and to enhance physical performance (Deijen and Orlebeke 1994; Neri, Wiegmann et al. 1995; Gerth, Mann et al. 1999; Williams 2005; O'Brien, Mahoney et al. 2007). In addition, L-tyrosine is often taken by people with phenylketonuria (PKU), as these patients have a deficiency in the enzyme that converts L-phenylalanine to L-tyrosine. Along with the requirement for low-L-phenylalanine, high-L-tyrosine protein substitutes, various groups have pointed to the potential advantages of consuming free L-tyrosine, particularly for the treatment of maternal PKU (Rohr, Lobbregt et al. 1998; van Spronsen, van Rijn et al. 2001; Giovannini, Verduci et al. 2007).

L-tyrosine also serves as an important precursor for a variety of high value compounds (Figure 3.1). For example, it is used to produce 3,4-dihydroxy-L-phenylalanine (L-DOPA or levodopa), which is currently the most powerful symptomatic drug for the treatment of Parkinson's disease (a condition caused by a deficiency in the action and formation of dopamine by the dopaminergic neurons of the brain). In newer drug formulations, levodopa is combined with the structurally similar chemical carbidopa to help prevent its metabolism in the bloodstream before it is able to reach the brain (Bonuccelli and Del Dotto 2006; Rajput and Rajput 2006). L-tyrosine can also be used for the generation of melanins through the action of a single enzyme. These compounds possess physiochemical properties that make them particularly suitable for use as UV absorbers, cation exchangers, drug carriers, and amorphous semiconductors (Bell and Wheeler 1986; della-Cioppa, Garger et al. 1990;

Cabrera-Valladares, Martinez et al. 2006). Finally, L-tyrosine can be easily converted to *p*-coumaric acid and *p*-hydroxystyrene, both of which are important components for a variety of novel polymers, adhesives and coatings, pharmaceuticals, biocosmetics, and flavonoid products (Leonard, Lim et al. 2007; Qi, Vannelli et al. 2007; Sariaslani 2007; Vannelli, Wei Qi et al. 2007; Leonard, Yan et al. 2008).

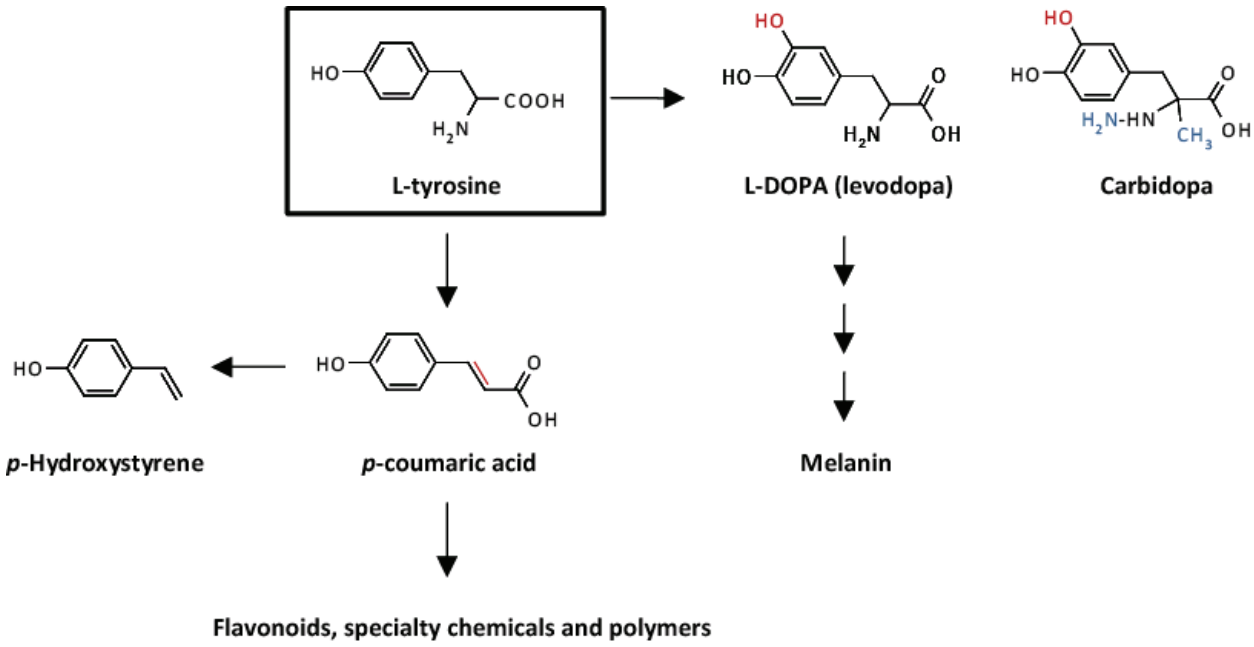


Figure 3.1 Industrial and pharmaceutical compounds derived from L-tyrosine

3.3 Chemical L-tyrosine production from protein hydrolysates

Before the advent of microbial production processes, amino acids were traditionally obtained by first breaking down proteinous materials, such as casein or wheat and corn gluten, by acid, alkali, or enzymatic hydrolysis. Most methods made use of acid hydrolysis with hydrochloric

acid, followed by a partial neutralization step with sodium hydroxide or another suitable alkali solution. Amino acids were then individually extracted/precipitated from these protein hydrolysates and, if needed, subjected to downstream purification steps (Barnett 1935).

Interestingly, the first methods for producing L-tyrosine focused on its recovery from byproducts of the monosodium glutamate production process. Following the hydrolysis of wheat and corn glutens, a crude tyrosine-leucine cake with small amounts of cystine was separated and often discarded as waste (Vassel 1956). Thus, most efforts concentrated on the separation of these three amino acids. Unfortunately, this was not always an easy task, as these three compounds exhibit very similar solubility characteristics in water, alkali, and acid solutions. Thus, several techniques were developed for this separation and included the use of pH, temperature, and density adjustments to alter solubility profiles (Mark 1939; Steinmetzer 1983), the use of alkali metal sulfides to convert cystine to the more soluble cysteine (Cardinal 1953), and the use of ammonium hydroxide to preferentially solubilize cystine (Vassel 1956). These pretreatment steps often allowed for the precipitation and separation of L-tyrosine by filtration or centrifugation.

3.4 Biotechnological L-tyrosine production using tyrosine phenol lyases

Owing to the harsh and extreme process conditions associated with chemical hydrolysis and separation, several groups also explored the use of the enzyme tyrosine phenol lyase to develop a biological process for the production of L-tyrosine. Tyrosine phenol lyase (EC 4.1.99.2), also known as β-tyrosinase, is a pyridoxal phosphate-requiring enzyme that possesses the ability to catalyze a series of α,β-elimination, β-replacement, and racemization

reactions. It has been of particular interest due to its capacity for synthesizing L-tyrosine (or L-DOPA) from the substrates phenol (or pyrocatechol), pyruvate, and ammonia. Although this enzyme is expressed in a number of phylogenetic groups, most studies have focused on the use of three bacterial species – *Erwinia herbicola*, *Citrobacter intermedius* (formerly *Escherichia intermedia*), and *Citrobacter freundii* – due to their high tyrosine phenol lyase activities (Enei, Matsui et al. 1972; Tysyachnaya, Yakovleva et al. 1979). Because of the industrial relevance of such a process, the reaction has been tested in various implementations, from the use of whole cell biocatalysts (either in suspension, immobilized, or encapsulated) to the direct application of purified and/or immobilized enzymes (Yamada, Kumagai et al. 1972; Enei, Nakazawa et al. 1973; Fukui, Ikeda et al. 1975; Fukui, Ikeda et al. 1975; Tysyachnaya, Yakovleva et al. 1979; Nagasawa, Utagawa et al. 1981; Para, Lucciardi et al. 1985; Lloyd-George and Chang 1993; Lloyd-George and Chang 1995).

3.4.1 Manipulating ammonium, pyruvate, and phenol concentrations

Regardless of the form in which the enzyme is presented, experiments examining the synthesis of L-tyrosine from phenol, ammonia, and pyruvate consistently revealed the same findings. In all cases, the α,β-elimination reaction could be easily driven in the synthetic direction, in part aided by the low solubility of L-tyrosine but more significantly by the ability to supply an excess of ammonia and pyruvate (thus allowing one to alter the thermodynamics of the reaction). As it turned out, providing this excess was also needed to replenish substrate pools, as they were found to be easily depleted by other cellular mechanisms. For example, pyruvate could participate in other metabolic processes when whole cell

biocatalysts were employed and undergo decomposition through a non-enzymatic route when enzyme preparations were provided as a catalyst. With intact cells, transport of reagents through the cell membrane was also considered to be a potential limiting factor, even in cases when substrates were present at high concentrations in the external solution. Despite the need for plentiful substrates, however, it was discovered that pyruvate concentrations above a certain threshold had inhibitory effects on the enzymatic activity of tyrosine phenol lyase; hence, lower concentrations of pyruvate were maintained by adding multiple substrate doses throughout the production process. In addition, phenol, which can partially destroy cell walls and denature proteins, was often infused at a minimum concentration to avoid inhibition and inactivation of the catalyst (Enei, Nakazawa et al. 1972; Enei, Nakazawa et al. 1973; Fukui, Ikeda et al. 1975; Tysyachnaya, Yakovleva et al. 1979; Nagasawa, Utagawa et al. 1981; Para, Lucciardi et al. 1985).

3.4.2 Production with whole cell biocatalysts

Production processes using solubilized, intact cells of *E. herbicola* achieved a maximum concentration of 60.5 g/l L-tyrosine from 20 g/l sodium pyruvate added twice, 50 g/l ammonium acetate, and a phenol concentration maintained at 10 g/l throughout the reaction (Enei, Nakazawa et al. 1973). For the case of *C. freundii*, a titer of about 0.11 M (20 g/l) L-tyrosine was achieved with the addition of 0.30 M (33 g/l) sodium pyruvate, 0.65 M (50 g/l) ammonium acetate, and 0.132 M (12.4 g/l) phenol after 2 hr (Tysyachnaya, Yakovleva et al. 1979) (Table 3.1).

Table 3.1 Biotechnological production methods for L-tyrosine

Biocatalyst	Characteristics	L-Tyrosine (g/L)	Reference
E. herbicola ATCC 21434, whole cells	Conversion of phenol, pyruvate and ammonia by TPL_{Eh}	60.5	(Enei, Nakazawa et al. 1973)
C. freundii, strain 62	Conversion of phenol, pyruvate, and ammonia by TPL_{Cf}	20	(Tysyachnaya, Yakovleva et al. 1979)
C. intermedius ATCC 21073, immobilized	Conversion of phenol, pyruvate and ammonia by TPL_{Ci} (batch)	10	(Para, Lucciardi et al. 1985)
C. intermedius ATCC 21073, immobilized	Same as above (continuous)	0.45	(Para, Lucciardi et al. 1985)
TPL from *S. toebii*	Conversion of phenol, pyruvate and ammonia by thermostable TPL_{St} from *E. coli* lysates	131	(Kim, Rha et al. 2007)
E. coli W3110 Δ*tyrR*, *tyrA* Δ*pheA*::Km/pMGL/pSTV-t*yrA*^{mut}	Δ*tyrR* Δ*pheA,* fbr-DAHP synthase, fbr-CM/PDH, shikimate kinase, kanamycin resistant	6.3	(Takai, Nishi et al. 2009)
E. coli T1	Δ*tyrR*, fbr-DAHP synthase, fbr-CM/PDH (3-l fed-batch reactor)	3.8	(Lütke-Eversloh and Stephanopoulos 2007)
E. coli T2	Δ*tyrR*, fbr-DAHP synthase, fbr-CM/PDH, transketolase, PEP synthase (3-l fed-batch reactor)	9.7	(Lütke-Eversloh and Stephanopoulos 2007)
E. coli DPD4193	fbr-DAHP synthase, *tyrR*^{mut}, *trpE*^{mut}, Δ*pheLA*, P_{trc}-*tyrA*, kanamycin resistant (shake flasks)	0.18	(Olson, Templeton et al. 2007)
E. coli DPD4195	fbr-DAHP synthase, *tyrR*^{mut}, Δ*pheLA*, P_{trc}-*tyrA*, kanamycin resistant (shake flasks)	0.147	(Olson, Templeton et al. 2007)
E. coli DPD4195	Same as above (10-l reactor)	10.5	(Sariaslani 2007)
E. coli DPD4195	Same as above (200-l reactor)	50	(Patnaik, Zolandz et al. 2008)
E. coli PB12CP	Δ*ptsHI*, fbr-DAHP synthase, $TyrC_{Zm}$, $PheA_{CM}$ (1-l reactor)	3	(Chavez-Bejar, Lara et al. 2008)

CM - chorismate mutase; DAHP - 3-deoxy-D-arabinoheptulosonate-7-phosphate; fbr - feedback inhibition resistant; mut - mutated gene; PDH - prephenate dehydrogenase; PEP - phosphoenolpyruvate; P_{trc} - trc promoter; TPL - tyrosine phenol lyase; Zm - *Z. mobilis*; Δ - deleted gene.

3.4.3 Enzyme and cell immobilization for added stability

The immobilization of enzymes and cells is an attractive method for retaining catalysts in continuous flow processes or repeated fill-and-draw (semi-batch) operations. In many cases, immobilization confers additional stability to the enzyme and also facilitates the separation of catalyst and reaction products at the completion of the production process. As an example, the direct coupling of tyrosine phenol lyase from *C. intermedius* to Sepharose beads led to a catalyst with a higher thermostability and higher resistance to denaturing agents than its free counterpart. When tested in a continuous reaction, an immobilized tyrosine phenol lyase column could perform at 25°C and a pH of 8.0 for five days with only a 30% decrease in enzyme activity (Fukui, Ikeda et al. 1975; Fukui, Ikeda et al. 1975). For the case of whole cell biocatalysts, entrapment of *C. intermedius* cells in a polyacrylamide gel resulted in a 60% decrease in specific activity but also led to a significant increase in enzyme stability. A batch operation with 75 mM (8.25 g/l) sodium pyruvate, 45 mM (5.95 g/l) ammonium sulfate, 45 mM (2.4 g/l) ammonium chloride, and 55 mM (5.5 g/l) phenol led to a final L-tyrosine concentration of 10 g/l after 5 h (productivity 2 g/l/hr). When tested in a continuous column reactor with 2.75 mM (0.26 g/l) phenol, L-tyrosine was produced at a concentration of 0.45 g/l, constituting a 90% conversion of phenol. Speaking to the stability of this setup, L-tyrosine production remained constant for more than 1300 hr (54 days) (Para, Lucciardi et al. 1985) (Table 3.1).

3.4.4 Discovery of a thermostable and chemostable tyrosine phenol lyase

More recently, a fed-batch L-tyrosine production process was reported employing an

engineered thermo- and chemostable tyrosine phenol lyase from *Symbiobacterium toebii* (recovered from recombinant *E. coli* cell lysates). Along with the option of using a wider range of operating temperatures, enzymes from thermophilic bacteria are also known for their stability, even in the presence of high concentrations of phenolic substances. With continuous feeding of 207 g/l phenol and 264 g/l sodium pyruvate at a rate of 20 ml/hr and 214 g/l ammonium chloride at a rate of 12 ml/hr, a total of 131 g/l L-tyrosine was synthesized in a 2.5 liter fed-batch reactor (Table 3.1). This constitutes a 94% conversion of phenol to L-tyrosine with a productivity of 4.3 g/l/hr. The stability of the process was in part aided by the addition of 0.2% polyethyleneglycol (PEG 3350) and alcohol to help with the refolding and solubilization of insoluble proteins and to suppress protein aggregation and nonspecific adsorption (Kim, Rha et al. 2007).

3.5 Biotechnological L-tyrosine production using engineered microbes

In light of growing environmental concerns and a rapidly declining supply of fossil fuel reserves, there has been a resurgence of interest in the development of non-petrochemical-based methods for the synthesis of chemicals and fuels. For the case of amino acid production, these green processes often take the form of microbial fermentations using engineered strains and renewable feedstocks. In this section, we will summarize several rational strategies that have been applied towards engineering strains of *E. coli* for the overproduction of L-tyrosine and other aromatic compounds.

3.5.1 The aromatic amino acid biosynthetic pathway and its regulation

The aromatic amino acid (AAA) biosynthetic pathway in *E. coli* is responsible for the synthesis of the three aromatic amino acids - L-tyrosine, L-phenylalanine, and L-tryptophan (Figure 3.2). The first committed step of the pathway is the condensation of phosphoenolpyruvate (PEP, from the glycolytic pathway) and erythrose-4-phosphate (E4P, from the pentose phosphate pathway) to form DAHP, a reaction catalyzed by the enzyme DAHP synthase. Subsequent steps then convert DAHP to chorismate through the compound shikimate. It is at this branch point that the pathway diverges into three separate terminal routes for the synthesis of L-tyrosine, L-phenylalanine, and L-tryptophan.

Despite decades of classical strain improvement by mutagenesis and selection, economically viable titers of the aromatic amino acids in *E. coli* were not attained, presumably due to the complex regulatory scheme of this branched pathway. Indeed, extensive study reveals that the AAA pathway contains three layers of regulation at the transcriptional and allosteric levels and, for the case of L-phenylalanine production, through attenuation at the transcriptional/translational interface. Perhaps the most highly regulated step of the pathway is the first one catalyzed by DAHP synthase. *E. coli* possesses three isoforms of this enzyme (encoded by *aroH*, *aroF* and *aroG*) which are feedback-inhibited by the end products L-tryptophan, L-tyrosine, and L-phenylalanine, respectively (Pittard, Camakaris et al. 2005; Sprenger 2007). A second site of regulation is found at the chorismate branchpoint with the enzymes chorismate mutase/prephenate dehydratase (CM/PDT, *pheA*) and chorismate mutase/prephenate dehydrogenase (CM/PDH, *tyrA*). Each of these bifunctional enzymes is feedback-regulated at their C-terminal PDT and PDH domains by L-phenylalanine and

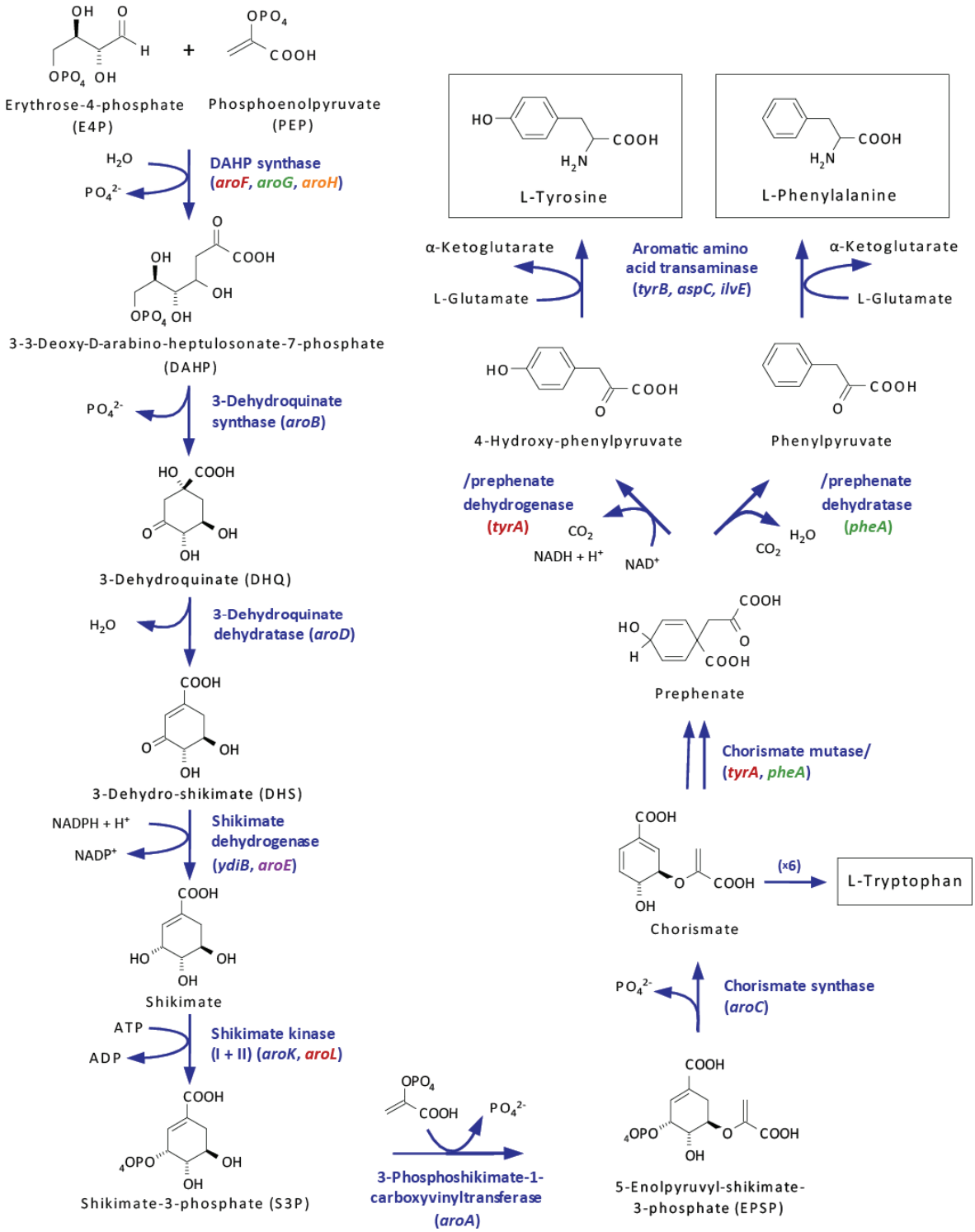


Figure 3.2 Aromatic amino acid biosynthetic pathway

Colors indicate allosteric feedback inhibition by L-tyrosine (red), L-phenylalanine (green), L-tryptophan (orange), or shikimate (purple).

L-tyrosine, respectively (Pohnert, Zhang et al. 1999; Chen, Vincent et al. 2003; Lütke-Eversloh and Stephanopoulos 2005).

In addition to allosteric control, transcriptional control is mediated by two effector proteins - TrpR (repression only) and TyrR (activation and repression). The TyrR regulon is comprised of eight transcriptional units that are each regulated in a distinct manner by the TyrR protein. In the context of AAA overproduction, TyrR can have several potentially phenotypically detrimental effects, including the repression of the biosynthetic genes *aroF*, *aroG*, *aroL*, *tyrA*, and *tyrB* in both a cofactor-independent and L-tyrosine-mediated manner (Pittard, Camakaris et al. 2005).

3.5.2 Overcoming regulatory hurdles for aromatic amino acid production

Because the synthesis of the aromatic amino acids is an energetically expensive process, it is not surprising that cells rely on efficient regulation of the AAA pathway to maintain sparing levels of these compounds. As a result, engineering efforts must first focus on deregulating this pathway in order to achieve the maximum possible flux for aromatic amino acid production.

Due to the ability of TyrR to repress transcription of several key AAA pathway genes, elimination of TyrR-mediated control was an important target for the production of aromatic amino acids in *E. coli*. This idea was further supported by a comparative transcriptome analysis of Δ*tyrR* and wild-type *E. coli* strains grown on inhibitory concentrations of L-phenylalanine. Differences in gene expression profiles suggested that inhibition was related to the TyrR-mediated control of the AAA pathway; inactivation of TyrR was therefore thought

to have the potential to improve cell viability and as a result, L-phenylalanine production (Polen, Kramer et al. 2005). Alleviation of transcriptional control was achieved through various means, including direct *tyrR* gene inactivation/deletion and through selection of antimetabolite-resistant mutants (Garner and Herrmann 1985; Tribe 1987; Berry 1996; Sprenger 2007).

In order to relieve allosteric control of rate-limiting metabolic reactions, several researchers also worked on the generation and overexpression of feedback-resistant versions of DAHP synthase and CM/PDT. As with some TyrR inactivation studies, these altered enzymes were oftentimes recovered through cultivation and selection with aromatic amino acid analogues. Such efforts were successful in generating overproducers and details of the work can be found in several excellent reviews (Bongaerts, Kramer et al. 2001; Kramer, Bongaerts et al. 2003; Ikeda 2006; Sprenger 2007).

3.5.3 Engineering central carbon metabolism to increase precursor supplies

Following deregulation of the AAA pathway, attentions then shifted towards engineering central carbon metabolism to increase the supply of the two main precursors, PEP and E4P. Several strategies were attempted for the production of DAHP, L-phenylalanine, L-tryptophan, and other aromatics and are summarized as follows.

To increase PEP availability:

(1) Deletion or inactivation of enzymes that consume PEP, including PEP carboxylase (*ppc*) *(Backman 1992)* and pyruvate kinases (*pykA*, *pykF*) (Gosset, Yong-Xiao et al. 1996).

(2) Overexpression of enzymes that convert pyruvate back to PEP such as PEP synthase (*pps*) (Liao, Hou et al. 1996).

(3) Inactivation of the phosphotransferase (PTS) system which consumes PEP for the phosphorylation and transport of glucose into the cell (Berry 1996).

(4) Engineering of the carbon storage regulator (Csr) system to increase flux through gluconeogenesis either by disruption of *csrA* and *csrD* or overexpression of *csrB* (Tatarko and Romeo 2001; Yakandawala, Romeo et al. 2008).

To increase E4P availability:

(1) Overexpression of enzymes in the pentose phosphate pathway to produce E4P, including transketolase (*tktA*) and transaldolase (*talB*) (Berry 1996; Lu and Liao 1997).

(2) Inactivation of phosphoglucose isomerase (*pgi*) to divert carbon from glycolysis and into the hexose monophosphate shunt for accelerated E4P production (Mascarenhas, Ashworth et al. 1991).

Although incorporation of these techniques individually often led to measurable increases in aromatic amino acid production, significant changes were observed only when a proper balance of *both* precursors was achieved. Such accomplishments often required the use of two or more of these strategies in combination. As a illustrative example, DAHP was produced at near theoretical yield (86% molar yield on glucose) through the simultaneous expression of transketolase, PEP synthase, and DAHP synthase (Patnaik and Liao 1994). For later comparison, it is important to note, however, that such strains were grown to stationary phase in rich media prior to resuspension in minimal media.

3.5.4 Engineering L-tyrosine production in *E. coli*

Surprisingly, the engineering of *E. coli* for L-tyrosine overproduction has been described only very recently by a few independent groups (Table 3.1). In 2005, the overexpression of feedback inhibition-resistant DAHP synthase and CM/PDH in *E. coli* Δ*tyrR* mutants was demonstrated in a patent filed by Ajinomoto Co., Inc. When all three changes were combined in a L-phenylalanine auxotrophic host strain (Δ*pheA*) with shikimate kinase (*aroL*) overexpression, the resulting strain yielded up to 6.3 g/l L-tyrosine (Takai, Nishi et al. 2009). Similarly, a L-tyrosine-producing prototrophic *E. coli* strain was generated in our laboratory by altering only regulatory circuits of aromatic amino acid biosynthesis. In wild-type *E. coli* K12, TyrR-mediated control was eliminated by the deletion of the *tyrR* gene, and regulatory hurdles were overcome through overexpression of feedback inhibition-resistant derivatives of the *aroG* and *tyrA* genes (strain *E. coli* T1). Such modifications resulted in the production of up to 3.8 g/l L-tyrosine in high-cell density fermentations. Further strain improvement was achieved through the overexpression of *ppsA* and *tktA* to boost supplies of the precursors PEP and E4P, and the final strain, *E. coli* T2, produced up to 9.7 g/l L-tyrosine (a yield of 0.1 g/g glucose) (Lütke-Eversloh and Stephanopoulos 2005; Lütke-Eversloh and Stephanopoulos 2007). More recently, our lab has also explored the coordinated overexpression of AAA pathway genes in order to identify additional bottlenecks in L-tyrosine production. By examining the effects of *aroB*, *aroD*, *aroE*, *ydiB*, *aroK*, *aroL*, *aroA*, *aroC* and *tyrB* overexpression (either singly or in combination), the *ydiB*-encoded shikimate dehydrogenase and the *aroK*-encoded shikimate kinase were identified as the next major impediments in the pathway. Indeed, the simultaneous overexpression of these

two genes was successful in increasing L-tyrosine titers by 26% over the parental *E. coli* T2 strain (Lütke-Eversloh and Stephanopoulos 2008).

Scientists at DuPont's Central Research and Development station took an alternative approach by making use of preexisting L-phenylalanine-producing strains (NST37 and NST74) and converting them into L-tyrosine overproducers. These strains had previously been engineered for L-phenylalanine production through multiple rounds of mutagenesis and selection on L-phenylalanine analogs and thus exhibited a high flux through the AAA pathway (Sariaslani 2007). By replacing the *pheLA* genes with a strong trc promoter to drive downstream *tyrA* gene expression, researchers were able to divert the preexisting flux towards L-tyrosine production (Tribe 1987; Olson, Templeton et al. 2007). In 50 ml shake flask experiments, these engineered strains, named *E. coli* DPD4193 (auxotrophic for L-phenylalanine and L-tryptophan) and DPD4195 (auxotrophic for L-phenylalanine), produced 180 mg/l and 147 mg/l L-tyrosine, respectively (yields of 0.09 and 0.07 g/g glucose). More recently, strain DPD4195 was described to produce 10.5 g/l L-tyrosine in 10 liter fed-batch fermentations (Sariaslani 2007). Remarkably, the deletion of the *lacI* repressor, which presumably led to an increase in *tyrA* expression, resulted in a five-fold increase in titer to more than 50 g/l L-tyrosine at a 200 liter scale (yield of ~0.3 g/g glucose) (Sariaslani 2007; Patnaik, Zolandz et al. 2008).

Although the generation of feedback-resistant CM/PDH ($tyrA^{fbr}$) has been successfully used to relieve allosteric control by L-tyrosine (Lütke-Eversloh and Stephanopoulos 2005), Chávez Béjar et al. also explored the use of a different method for circumventing feedback regulation. Strains were engineered to overexpress two enzymes with a natural insensitivity

to L-tyrosine inhibition – cyclohexadienyl dehydrogenase from *Zymomonas mobilis* (TyrC) and the CM domain ($PheA_{CM}$) of *E. coli* CM/PDT. In this alternate reaction scheme, $PheA_{CM}$ generates prephenate from chorismate, and TyrC subsequently catalyzes its conversion to *p*-hydroxyphenylpyruvate. When simultaneously expressed with a feedback-resistant DAHP synthase ($aroG^{fbr}$) in a PTS^- background, these strains produced 3 g/l of L-tyrosine in 1 liter fed-batch fermentations (yield of 0.066 g/g glucose) (Chavez-Bejar, Lara et al. 2008).

3.6 Conclusions

As demonstrated by these examples, the biotechnological production of L-tyrosine can be achieved through various strategies employing both whole microbial cells and isolated enzymes (Table 3.1). Although early work in this area was initially dominated by the use of tyrosine phenol lyases, we have seen a shift in recent years to the development of fermentation-based processes for L-tyrosine overproduction. Employing whole cells as biocatalysts for L-tyrosine production offers clear advantages over *in vitro* enzymatic conversion, since microbial synthesis proceeds under environmentally-friendly conditions and uses biomass-derived feedstocks as substrates. Fortunately, advancements in recombinant DNA technology have greatly facilitated the engineering of strains for this purpose. Various techniques have been applied to generate and characterize feedback insensitive mutations and to overexpress deregulated enzymes of the AAA pathway. As seen with L-tryptophan and L-phenylalanine production, combining such deregulation with strategies for increasing PEP and E4P precursor supplies can have a significant impact on final titers and yields. Although basic principles for increasing precursor availability have been explored in the context of L-

tyrosine production (Lütke-Eversloh and Stephanopoulos 2007), it is clear that lessons learned from L-phenylalanine and L-tryptophan overproduction have not been completely exploited and may lead to further improvements in microbial L-tyrosine synthesis.

Due to our fairly comprehensive understanding of the AAA pathway, rational metabolic engineering approaches have been quite successful in altering regulation and increasing pathway flux for the production of L-tyrosine. Compared to wild-type *E. coli* levels, gains in production have been quite significant with fed-batch fermentation titers reaching the g/l scale. However, it is interesting to note that the best reported strain (DPD4195) was actually constructed through a combination of both rational techniques and classical strain improvement (through extensive rounds of mutagenesis and selection on aromatic amino acid analogs) (Sariaslani 2007; Patnaik, Zolandz et al. 2008). While it is reasonable to assume that several beneficial mutations may have acted by altering transcriptional and allosteric control of key AAA enzymes, it is also likely that selected strains may have accumulated additional mutations with more distal effects. In this latter scenario, whole genome sequencing would be required to identify mutations occurring at loci beyond those of the AAA pathway genes and PEP/E4P consuming or utilizing enzymes.

Although we can only speculate on the nature of the introduced mutations, strain DPD4195 presents a clear impetus for the development and application of combinatorial metabolic engineering techniques for microbial L-tyrosine production. The ability of this strain to synthesize L-tyrosine at levels that exceed exclusively rational strain constructions provides early evidence for a possibly complementary nature between rational techniques and more random genetic approaches. Based on this early success, future research on the

biotechnological synthesis of L-tyrosine should clearly transition to encompass traditional methods based on rational design as well more global and integrated engineering approaches. When adopted within an inverse metabolic engineering paradigm, these combinatorial approaches offer unique opportunities for identifying novel targets for engineering phenotypic improvement. Through this work, we hope to explore the as-yet untapped potential of these approaches for optimizing L-tyrosine overproduction in *E. coli*.

CHAPTER 4 RATIONAL ENGINEERING OF A PARENTAL STRAIN

4.1 Introduction

Although the broader objective of this thesis is to investigate the utility of combinatorial techniques for improving a L-tyrosine production phenotype, our main goal is certainly not limited to recovering similar titers as those achieved by rational approaches. All too often, publications touting the utility of novel engineering techniques base their comparisons on wild-type strains which naturally possess very little capacity for the evaluated phenotype (Dueber, Wu et al. 2009; Wang, Isaacs et al. 2009). While gains through these methods may initially seem impressive, more often than not, they fall short when compared to strains already engineered through more traditional approaches. Thus, their actual utility for practical strain engineering applications comes into question. Can these techniques really be considered useful tools in a metabolic engineer's arsenal if they simply replace the function of a preexisting method without the delivery of improved results? Clearly, the important metric for evaluation is not whether similar gains can be made but rather, whether these newer methods can access phenotypes that have *not* been achieved through more traditional methods.

With this in mind, our objective in evaluating these combinatorial metabolic engineering techniques is to determine whether such methods can lead to L-tyrosine yields and titers even beyond those realized by exclusively rational constructions. In doing so, we hope to investigate the possibly complementary nature of these orthogonal approaches. To achieve this goal, a

clear first step must then be the rational assembly of strains that possess an already elevated capacity for L-tyrosine synthesis. Though already quite competitive in terms of L-tyrosine yields and titers, these strains will then be used as the starting point for evaluating various combinatorial techniques for library generation. In Chapter 3, we reviewed several successful strategies for boosting L-tyrosine synthesis involving both deregulation of the AAA pathway and enhancement of precursor (PEP, E4P) supplies. In this chapter, we apply results from these previous works towards the construction and evaluation of two parental (or starting) strains to be used in this study.

4.2 Materials and Methods

4.2.1 Construction of plasmid-based parental strain

E. coli K12 Δ*pheA* Δ*tyrR* (T. Lütke-Eversloh and G. Stephanopoulos, unpublished data) was transformed with pCL1920::*tyrA*^{fbr}*aroG*^{fbr} (Lütke-Eversloh and Stephanopoulos 2007). This strain, which will henceforth be referred to as parental 1 (P1), was used as the starting strain for the creation of transposon mutagenesis libraries.

4.2.2 Cloning of pZE-*kan*^{FRT}-*tyrA*^{fbr}*aroG*^{fbr} plasmids

Plasmid pCL1920::*tyrA*^{fbr}*aroG*^{fbr} was used as a template for the amplification of *tyrA*^{fbr}-*aroG*^{fbr} with *Pfu* Turbo DNA polymerase (Stratagene) and the following primers: CS114 tyrAfbr sense KpnI (5'-GCT CGG TAC CAT GGT TGC TGA ATT GAC CGC ATT ACG-3') and CS278 aroGfbr anti MluI (5'-CGA CGC GTT TAC CCG CGA CGC GCT TTT ACT G-3'). This PCR product was then

digested with KpnI and MluI and ligated into five KpnI/MluI digested pZE-*gfp* plasmids taken from a synthetic constitutive promoter library (Alper, Fischer et al. 2005). Each of the five versions of pZE-*tyrA*^{fbr}*aroG*^{fbr} corresponds to a different strength promoter chosen for these studies (R, Y, W, B, and P_L). To construct pZE-*kan*^{FRT}-*tyrA*^{fbr}*aroG*^{fbr}, primers CS279 pKD13 sense SacI (5'-TCC GAG CTC TTG TGT AGG CTG GAG CTG CTT CGA-3') and CS280 pKD13 anti AatII (5'-TCT TAG ACG TCG GAA TTG ATC CGT CGA CCT GCA GTT CGA A-3') were used to amplify an FRT-flanked kanamycin resistance gene (*kan)* on the plasmid pKD13 (Datsenko and Wanner 2000). After digestion with SacI and AatII, this product was ligated to SacI/AatII digested pZE-*tyrA*^{fbr}*aroG*^{fbr} and transformed into chemically competent *E. coli* DH5α cells as described in the protocol. Following transformation, all plasmid constructs were isolated, verified by sequencing, and transformed into *E. coli* K12 Δ*pheA* Δ*tyrR*. All enzymes used in the cloning procedure were purchased from New England Biolabs.

4.2.3 Chromosomal integration of *tyrA*^{fbr}-*aroG*^{fbr} cassette

The *kan*^{FRT}-*tyrA*^{fbr}-*aroG*^{fbr} cassette was integrated into the *lacZ* locus of *E. coli* K12 Δ*pheA* Δ*tyrR* using a lambda-red recombination based method (Datsenko and Wanner 2000). Briefly, *kan*^{FRT}-*tyrA*^{fbr}-*aroG*^{fbr} was amplified from pZE-*kan*^{FRT}-*tyrA*^{fbr}*aroG*^{fbr} with primers CS173 aroG-lacZ anti (5'-TTC CGG CAC CAG AAG CGG TGC CGG AAA GCT GGC TGG AGT GCG ATC TTC CTG AGG CCG ATA CTG TCG TCG TCC CCT TTA CCC GCG ACG C-3') and CS281 pKD13-lacZ sense (5'-CGC GTG CAG CAG ATG GCG ATG GCT GGT TTC CAT CAG TTG CTG TTG ACT GTA GCG GCT GAT GTT GAA CTG GAA GTC GTG TAG GCT GGA GCT GCT TCG A-3') and Platinum *Pfx* DNA polymerase (Invitrogen). Both primers incorporated 75-77 bp of homology with the ends of the *lacZ* gene

to facilitate integration into the proper locus. Following transformation of the cassette into *E. coli* K12 Δ*pheA* Δ*tyrR* pJM12 (a pKD46 derivative), colonies were verified by colony PCR and sequencing. Excision of FRT-flanked *kan* from the resulting strains *E. coli* K12 Δ*pheA* Δ*tyrR* *lacZ*::*kan*^{FRT}-*tyrA*^{fbr}*aroG*^{fbr} was mediated by transformation with FLP recombinase-expressing pCP20 as described in the literature (Datsenko and Wanner 2000). Five versions of *E. coli* K12 Δ*pheA* Δ*tyrR* *lacZ*::*tyrA*^{fbr}*aroG*^{fbr} were constructed by this process with each utilizing a different strength promoter (R, Y, W, B, P_L).

To integrate a *tyrA*^{fbr}-*aroG*^{fbr} cassette within the *tyrR* locus, *kan*^{FRT}-*tyrA*^{fbr}-*aroG*^{fbr} was amplified from pZE-*kan*^{FRT}-*tyrA*^{fbr}*aroG*^{fbr} (using promoters W, B, and P_L only) with primers CS286 pKD13-tyrR sense (5'- TGC AAT ATC GGG TGC TGA CCG GAT ATC TTT ACG CCG AAG TGC CCG TTT TTC CGT CTT TGT GTC AAT GAT TGT TGA CAG GTG TAG GCT GGA GCT GCT TCG A-3') and CS287 aroG-tyrR anti (5'- TAA TTT AAT ATG CCT GAT GGT GTT GCA CCA TCA GGC ATA TTC GCG CTT ACT CTT CGT TCT TCT TCT GAC TCA GAC CAT TAC CCG CGA CGC GCT TTT ACT G-3'). These primers incorporated 77-78 bp of homology with the ends of the *tyrR* gene to facilitate integration into the proper locus. Following transformation into *E. coli* K12 Δ*pheA* Δ*tyrR* pJM12, verification and excision of *kan* were performed as described earlier. Integration of a second *tyrA*^{fbr}-*aroG*^{fbr} cassette into these strains was mediated by P1 transduction (Miller 1992) of *lacZ*::*kan*^{FRT}-$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr} from the previously constructed strain *E. coli* K12 Δ*pheA* Δ*tyrR* *lacZ*::*kan*^{FRT}-$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr}. Selection on kanamycin, verification, and subsequent *kan* excision resulted in three separate versions of *E. coli* K12 Δ*pheA* Δ*tyrR* *lacZ*::*tyrA*^{fbr}*aroG*^{fbr} *tyrR*::*tyrA*^{fbr}*aroG*^{fbr}. All three strains contain a P_L promoter at the *lacZ* locus and the promoters W, B, or P_L at the *tyrR* site. The strain *E. coli* K12 Δ*pheA* Δ*tyrR* *lacZ*::$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr} *tyrR*::

$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr}, which makes use of the highest strength promoter, P_L, for both cassettes, will henceforth be referred to as parental strain 2 (P2).

4.2.4 Cultivation conditions

L-tyrosine production experiments were performed at 37°C with 225 rpm orbital shaking in 50 ml MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) cultures supplemented with 5 g/l glucose and an additional 4 g/l NH_4Cl. All liquid cultivations were conducted in at least triplicates. When appropriate, antibiotics were added in the following concentrations: 50 µg/ml spectinomycin for maintenance of pCL1920::*tyrA*^{fbr}*aroG*^{fbr} and 20 µg/ml kanamycin for maintenance of pZE-derived plasmids. Isopropyl-β-D-thiogalactopyranoside (IPTG) (EMD Chemicals) was added at a concentration of 2 mM for the induction of pCL1920::*tyrA*^{fbr}*aroG*^{fbr}. For L-phenylalanine auxotrophs (Δ*pheA*), L-phenylalanine (Sigma) was supplied at a concentration of 0.35 mM.

4.2.5 Analytical methods

For the quantification of L-tyrosine, cell-free culture supernatants were filtered through 0.2 µm PTFE membrane syringe filters (VWR International) and used for HPLC analysis with a Waters 2690 Separations module connected with a Waters 996 Photodiode Array detector set to a wavelength of 278 nm. The samples were separated on a Waters Resolve C18 column with 0.1 % (vol/vol) trifluoroacetic acid (TFA) in water (solvent A) and 0.1 % (vol/vol) TFA in acetonitrile (solvent B) as the mobile phase. The following gradient was used at a flow rate of

1 ml/min: 0 min, 95 % solvent A + 5 % solvent B; 8 min, 20 % solvent A + 80 % solvent B; 10 min, 80 % solvent A + 20 % solvent B; 11 min, 95 % solvent A + 5 % solvent B.

4.3 Results

Previous studies in our lab have demonstrated that significant gains in titer can be achieved from simple AAA pathway deregulation (Lütke-Eversloh and Stephanopoulos 2007). We therefore applied these principles to guide the genetic construction of two parental L-tyrosine producers. Specifically, we chose to incorporate three important genetic modifications: 1) deletion of *tyrR* to circumvent transcriptional regulation, 2) deletion of *pheA* to eliminate the loss of AAA pathway intermediates to competing reactions, and 3) overexpression of feedback resistant derivatives of DAHP synthase (*aroG*) and CM/PDH (*tyrA*) to increase flux through the AAA pathway.

4.3.1 Plasmid-based overexpression of *tyrA*^{fbr} and *aroG*^{fbr}

Our initial approach to strain construction was to supply feedback-resistant versions of *aroG* and *tyrA* on a low (~5) copy plasmid. Plasmid-based overexpression can be quite advantageous, not only because of its genetic simplicity, but also because it allows for the introduction of multiple copies of a gene within a cell. This is oftentimes an important consideration, particularly if the expressed gene represents a critical bottleneck in a process or pathway. To this end, we made use of a previously constructed plasmid (pCL1920::*tyrA*^{fbr}*aroG*^{fbr}) which contained *tyrA*^{fbr} and *aroG*^{fbr} under the control of an IPTG-inducible trc promoter. Although its functionality was recently demonstrated in an *E. coli* K12 Δ*tyrR* background (Lütke-Eversloh and

Stephanopoulos 2007), we tested its utility in a L-phenylalanine auxotroph (*E. coli* K12 Δ*pheA* Δ*tyrR*) to eliminate the loss of AAA pathway intermediates, such as prephenate, to L-phenylalanine production.

As expected, these simple genetic modifications had a significant impact on extracellular accumulation of L-tyrosine in the medium. Although deletion of *tyrR* and *pheA* together only brought nominal increases in L-tyrosine titer (from non-detectable levels in wild-type *E. coli* K12 to 6 mg/l in a Δ*pheA* Δ*tyrR* background), combining these deletions with the overexpression of feedback-resistant DAHP synthase and CM/PDH resulted in titers as high as 347 mg/l (Figure 4.1a). This engineered strain, named parental 1 (P1, *E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{fbr}*aroG*^{fbr}), was used as a plasmid-based parental for the generation of random knockout libraries.

4.3.2 Chromosomal Integration of *tyrA*^{fbr} and *aroG*^{fbr}

Because the traceability of many combinatorial engineering methods requires the use of antibiotic casettes and plasmid-based expression, marker and origin of replication incompatibilities can oftentimes arise between hosts and tools. To circumvent such problems, we constructed a second marker- and plasmid-free parental strain by chromosomal integration of a constitutively-expressed *tyrA*^{fbr}-*aroG*^{fbr} operon into *E. coli* K12 Δ*pheA* Δ*tyrR*. As an added advantage of this genetic scheme, this strain can be cultivated in media without antibiotics or inducers, thus transforming its fermentation into a simpler and more economical process.

We began this work by transferring transcriptional control of *tyrA*^{fbr} and *aroG*^{fbr} from an inducible trc promoter to one capable of constitutive overexpression. Rather than selecting a

single bacterial promoter, we utilized a recently constructed synthetic library of constitutive promoters capable of directing a wide range of expression levels (Alper, Fischer et al. 2005). This tunable parameter allowed us to explore whether different strengths of expression would have a measurable influence on cellular phenotype. The specific promoters tested and their relative strengths are listed in Table 4.1.

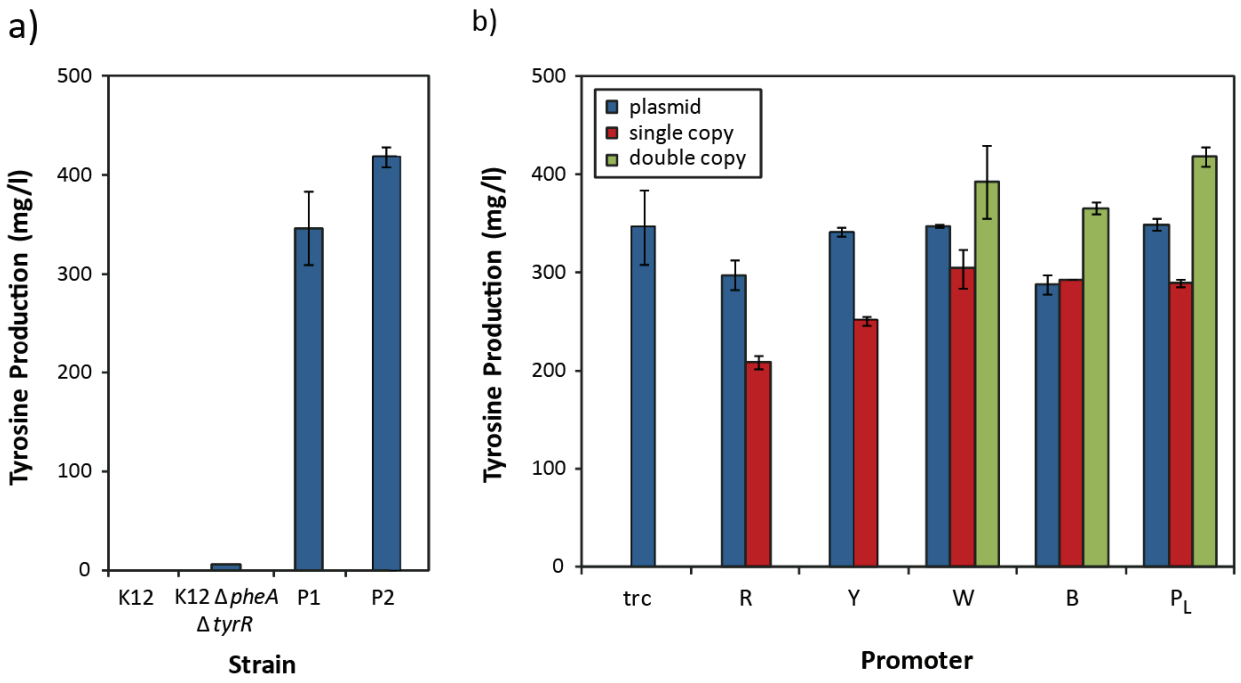


Figure 4.1 L-tyrosine production of rationally engineered strains after 24 hr

a) L-tyrosine production of wild-type *E. coli* K12, a double deletion mutant (*E. coli* K12 Δ*pheA* Δ*tyrR*), and two parental strains (P1, plasmid-based parental and P2, chromosome-based parental). b) L-tyrosine production of *E. coli* K12 Δ*pheA* Δ*tyrR* strains containing plasmid-based overexpression of the *tyrA*^{fbr}-*aroG*^{fbr} operon (blue), a single integrated copy (red), or two chromosomal copies (green). Plasmid-based copies of *tyrA*^{fbr}*aroG*^{fbr} under the control of the synthetic constitutive promoters (R, Y, W, B, P_L) were provided on pZE-*kan*^{FRT}-*tyrA*^{fbr}*aroG*^{fbr}. Plasmid-based trc expression was provided on pCL1920::*tyrA*^{fbr}*aroG*^{fbr}. Promoters used in the first integrated copy are indicated below each bar. Promoter P_L was utilized to drive expression from the second integration copy for all strains.

Table 4.1 Relative strengths of five synthetic constitutive promoters

Promoter	Average promoter strength metric (0 to 1)[a]
R	0.14
Y	0.31
W	0.54
B	0.82
P_L	0.87

[a] data provided by C. Fischer

Interestingly, we observed that integration of only a single copy of *tyrA*^{fbr} and *aroG*^{fbr} into the *lacZ* locus led to lower L-tyrosine titers when compared to our inducible, plasmid-based parental, P1 (Figure 4.1b). This trend held for all constitutive promoters tested with the weakest promoters, R and Y, exhibiting the lowest titers. To validate the functionality of our constitutively-expressed operons, we measured the L-tyrosine production capabilities of strains bearing the plasmid-based constructs (*E. coli* K12 Δ*pheA* Δ*tyrR* pZE-*kan*^{FRT}-*tyrA*^{fbr}*aroG*^{fbr}). Because appropriately high L-tyrosine levels were seen for all five versions of pZE-*kan*^{FRT}-*tyrA*^{fbr}*aroG*^{fbr} (corresponding to the five promoters), we hypothesized that the poor performance of the single integration strains may be a result of differences in gene dosage and copy number (Figure 4.1b).

4.3.3 A second copy of *tyrA*^{fbr}*aroG*^{fbr} is required for higher L-tyrosine titers

In an effort to boost L-tyrosine production, we integrated a second copy of *tyrA*^{fbr}-*aroG*^{fbr} into the *tyrR* locus of *E. coli* K12 Δ*pheA* Δ*tyrR lacZ*::$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr} using the three strongest promoters W, B, and P_L. As we suspected, the presence of a second copy improved final L-

tyrosine titers to levels even beyond those seen with P1 (Figure 4.1b). Because no clear trends were seen between L-tyrosine production and the promoter strength used to drive expression from the second operon, we assumed that DAHP synthase and CM/PDH no longer exist as the major bottlenecks of the AAA pathway. Nevertheless, we selected the strain with the strongest promoter (P_L) at the *tyrR* locus as the second parental strain (P2) for these studies. P2 (*E. coli* K12 Δ*pheA* Δ*tyrR* *lacZ*::$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr} *tyrR*::$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr}), which requires no antibiotic selection or induction, was used as the starting strain for the generation of all gTME-derived libraries.

4.4 Discussion

4.4.1 The advantages of L-phenylalanine auxotrophy

With the exception of an additional *pheA* deletion, our plasmid-based parental strain, P1, is a very similar construction to another strain assembled in our lab (*E. coli* K12 Δ*tyrR* pCL1920::*tyrA*^{fbr}*aroG*^{fbr} (Lütke-Eversloh and Stephanopoulos 2007). However, titers from P1 are almost three-fold higher than those previously reported for this strain (347 mg/l versus 127 mg/l), demonstrating the importance of eliminating the prephenate-utilizing reaction in this particular background. Surprisingly, we see that these gains are lost once DAHP synthase and CM/PDH levels are increased by promoter replacement to form T1, with L-tyrosine titers reaching 346 mg/l in a L-phenylalanine prototroph (Lütke-Eversloh and Stephanopoulos 2007). Thus, it seems that high expression of *tyrA*^{fbr} is sufficient for diverting flux of prephenate away from L-phenylalanine production, even with an intact CM/PDT in place.

Despite this apparent loss of benefit, we decided to continue our work in a L-phenylalanine auxotroph for a number of reasons. First, even small percentages of L-phenylalanine produced at a large industrial scale can be economically unfavorable, not only because of precursor consumption but also due to the added costs of downstream processing and separation. Although reports have shown that L-phenylalanine synthesis can still occur nonenzymatically in auxotrophic strains (Sariaslani 2007; Patnaik, Zolandz et al. 2008), levels would undoubtedly be higher in the presence of an intact *pheA* locus. Secondly, L-phenylalanine auxotrophy can be used quite advantageously to control and limit biomass formation. Although L-tyrosine production has long been known to be associated with exponential growth, a strict conflict still exists between biomass formation and product synthesis, as both processes compete for limited supplies of carbon substrates. The ability to control the amount of L-phenylalanine supplemented at the beginning of the culture allows one to select and predetermine a minimum amount of biomass needed to convert the carbon source to product efficiently.

4.4.2 Avoiding network rigidity

In designing our parental strains, we chose to focus only on genetic modifications that would deregulate the AAA pathway despite growing evidence of the importance of modulating precursor availability. This strategy was taken to avoid imposing too much rigidity into the metabolic network. Although maximizing the capacity of the main biosynthetic pathway is clearly a necessary first step, modulating precursor production and utilization may work to our disadvantage by forcing a cell into a particular metabolic state. Such stringency may ultimately

limit the capabilities of combinatorial approaches for modulating and accessing a wide range of cellular states and phenotypes. Despite these self-imposed limitations, our rationally constructed strains still exhibited competitive titers and yields. The development of these two parental strains therefore puts us in an excellent position to explore combinatorial engineering strategies for their potential to further enhance these desired properties.

CHAPTER 5 HIGH-THROUGHPUT SCREEN FOR L-TYROSINE PRODUCTION

5.1 Introduction

Although the utility of the combinatorial approach has been demonstrated for a number of genetic tools, including transposon mutagenesis, genomic complementation, and gTME (Alper, Miyaoku et al. 2005; Jin and Stephanopoulos 2007), each of these prior examples has dealt with easily accessible phenotypes. For the case of lycopene production in *E. coli*, desirable clones assumed a red pigmentation and could therefore be selected by visual inspection. For the case of SDS or ethanol tolerance in *E. coli* and *Saccharomyces cerevisiae*, simple growth competition assays were used to heavily enrich a population with the best performers. For most other systems of interest, however, the widespread use of these combinatorial approaches is hampered by the absence of a high-throughput method for selecting strains with the desired cellular properties. This is particularly so for the case of L-tyrosine production in *E. coli*. Although rational approaches have certainly led to significant increases in aromatic amino acid production, further gains in yield and productivity may require the modulation of factors that are not directly involved in the biosynthetic pathway or the related precursor forming/utilization reactions. Implementation of the combinatorial metabolic engineering approaches discussed earlier would allow for the identification of these more obscure targets, which may act through unknown or poorly understood mechanisms. A high-throughput screen capable of selecting L-tyrosine-producing mutants from a large, diverse population thus becomes an important tool for the future engineering of these production strains.

In this chapter, we present the development of a high-throughput screen for L-tyrosine production based on the synthesis of the black and diffusible pigment melanin. This is accomplished through the heterologous expression of a bacterial tyrosinase in the production strain(s) of interest. Tyrosinases, which contain a pair of cupric ions in their active site, use molecular oxygen to catalyze the *ortho*-hydroxylation of L-tyrosine to L-DOPA, followed by its oxidation to dopachrome. The reactive quinones that are generated then polymerize nonenzymatically to form melanin (Figure 5.1) (Claus and Decker 2006). Since melanin is a black pigment with a characteristic absorbance profile, the production of melanin can be easily detected by both visual and spectrophotometric means. Coupling L-tyrosine production and melanin synthesis in this way therefore allows for a simple method for identifying high L-tyrosine producers within a mixed population of cells. In this particular study, we have introduced the *melA* gene from *Rhizobium etli* (Cabrera-Valladares, Martinez et al. 2006; Lagunas-Munoz, Cabrera-Valladares et al. 2006) into a series of *E. coli* L-tyrosine production strains. Strains that either produced or were exposed to greater amounts of L-tyrosine could be distinguished by the unique pigmentation imparted by the synthesis of melanin.

5.2 Materials and Methods

5.2.1 Bacterial strains and cultivation conditions

R. etli CFN42 was kindly provided by G. Dávila (Gonzalez, Bustos et al. 2003) and cultured in Peptone-Yeast extract (PY) medium at 30°C (Bravo and Mora 1988). *E. coli* DH5α (Invitrogen)

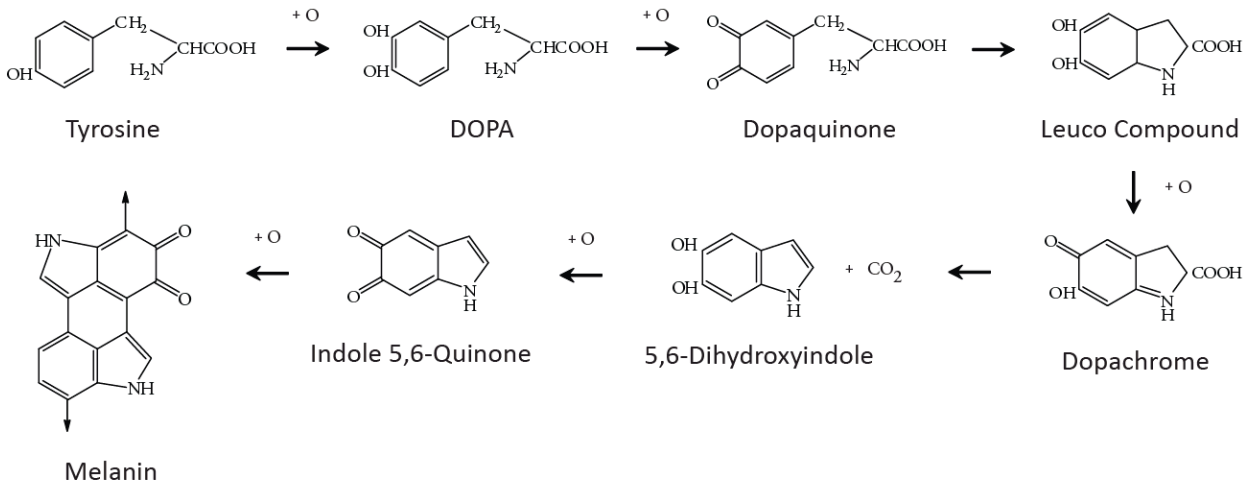


Figure 5.1 Synthesis of melanin from L-tyrosine

The enzyme tyrosinase catalyzes the *ortho*-hydroxylation of L-tyrosine to L-DOPA and the oxidation of L-DOPA to dopachrome. The polymer melanin is then generated from dopachrome through a series of nonenzymatic steps as shown above.

was used for routine transformations as described in the protocol and cultivated in Luria-Bertani (LB) medium. The following plasmids were transformed into *E. coli* K12 Δ*pheA* Δ*tyrR* and/or *E. coli* K12 Δ*pheA* Δ*tyrR* Δ*wecB* (C.N.S. Santos and G. Stephanopoulos, unpublished data) and used for L-tyrosine and melanin production experiments: pCL1920::*tyrA*^{WT}*aroG*^{WT}, pCL1920::*tyrA*^{fbr}*aroG*^{WT}, pCL1920::*tyrA*^{WT}*aroG*^{fbr}, pCL1920::*tyrA*^{fbr}*aroG*^{fbr}, pTrc*melA*, and pTrc*melA*^{mut1}(Lutke-Eversloh and Stephanopoulos 2007). L-tyrosine production experiments were performed at 37°C with 225 rpm orbital shaking in 50 ml MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) cultures supplemented with 5 g/l glucose and an additional 4 g/l NH_4Cl. Liquid melanin production experiments were performed at 30°C with 225 rpm orbital shaking in 50 ml M9 (Sambrook, Fritsch et al. 1989) or MOPS minimal medium cultures supplemented with 5 g/l glucose, an additional 4 g/l NH_4Cl, 40 µg/ml $CuSO_4$, and L-tyrosine at the indicated concentrations. All liquid cultivations were conducted in at least

triplicates. Solid melanin production experiments were performed at 30°C in MOPS minimal medium supplemented with 5 g/l glucose, an additional 4 g/l NH_4Cl, 0.4 µg/ml $CuSO_4$, 15 g/l Bacto Agar (BD Diagnostics), and L-tyrosine at the indicated concentrations. When appropriate, antibiotics were added in the following concentrations: 100 µg/ml carbenicillin for maintenance of pTrc*melA* and 50 µg/ml spectinomycin for maintenance of pCL1920-derived plasmids. Carbenicillin was chosen in place of ampicillin due to its improved stability during the longer cultivations (>48 hr) required for the synthesis of melanin. IPTG was added at a concentration of 1 mM for the induction of pTrc*melA* and 3 mM for the induction of both pTrc*melA* and pCL1920-derived plasmids. All chemicals, including those used in the supplementation experiments - $CaCl_2$, NaCl, Na_2HPO_4, NaH_2PO_4, and K_2HPO_4 - were purchased from Sigma, J.T. Baker, or Mallinckrodt Chemicals.

5.2.2 Construction of pTrc*melA*

R. etli CFN42 genomic DNA was extracted with the Wizard Genomic DNA Purification Kit (Promega) and used as a template for the amplification of *melA* with Pfu Turbo DNA polymerase (Stratagene) and the following primers: CS168 melA sense NcoI (5'- TAA ACC ATG GCG TGG CTG GTC GGC A - 3') and CS169 melA anti Hind III (5' – ACG AAG CTT TTA GGC GGA CAC TAT GGC TAT TTC TAG CTT – 3'). In order to introduce an *Nco*I restriction site for cloning, the start codon was changed from TTG to ATG, and the second codon was changed from CCG to GCG. This second alteration resulted in a proline to alanine substitution in the second amino acid. The *melA* PCR product was digested with *Nco*I and *Hind*III and ligated into the *Nco*I/*Hind*III-digested plasmid pTrcHis2B (Invitrogen) for one hour at room temperature. The

plasmid was then transformed into chemically competent *E. coli* DH5α cells (Invitrogen) and plated on LB-agar plates containing 100 µg/ml ampicillin, 1 mM IPTG, 500 mg/l L-tyrosine and 0.4 µg/ml $CuSO_4$. This latter step was designed to facilitate the selection of clones with correct plasmids, which should synthesize melanin and form dark colonies. Plasmid constructs were isolated and verified by sequencing. All enzymes used in the cloning procedure were purchased from New England Biolabs.

5.2.3 Analytical methods

For the quantification of L-tyrosine, cell-free culture supernatants were filtered through 0.2 µm PTFE membrane syringe filters (VWR International) and used for HPLC analysis with a Waters 2690 Separations module connected with a Waters 996 Photodiode Array detector set to a wavelength of 278 nm. The samples were separated on a Waters Resolve C18 column with 0.1 % (vol/vol) trifluoroacetic acid (TFA) in water (solvent A) and 0.1 % (vol/vol) TFA in acetonitrile (solvent B) as the mobile phase. The following gradient was used at a flow rate of 1 ml/min: 0 min, 95 % solvent A + 5 % solvent B; 8 min, 20 % solvent A + 80 % solvent B; 10 min, 80 % solvent A + 20 % solvent B; 11 min, 95 % solvent A + 5 % solvent B. For the quantification of melanin, the optical densities of cell-free culture supernatants at 400 nm were determined with an Ultrospec 2100 *pro* UV/Visible spectrophotometer (Amersham Biosciences) and compared to a synthetic melanin standard (Sigma). For cell density determinations, the optical densities of cultures and cell-free culture supernatants were measured at 600 nm. Since melanin affects the absorbance measurements at this wavelength, the cell density is calculated

by taking the difference between these two values. pH measurements were taken with a SympHony SP20 pH meter and electrode (VWR International).

5.2.4 Library construction and screening

Transposon mutagenesis (random knockout) libraries for P1 (*E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{fbr}*aroG*^{fbr}) pTrc*melA*^{mut1} were generated by transformation with 1000-1300 ng of the pJA1 vector (Badarinarayana, Estep et al. 2001). After an initial 1 hr outgrowth period at 37°C, cells were centrifuged at 2000 x g and resuspended in 1 ml MOPS minimal medium. Cells were then plated on 150 x 15 mm petridishes containing MOPS minimal medium with 5 g/l glucose, an additional 4 g/l NH_4Cl, 40 μg/ml $CuSO_4$, and 20 mM Na_2HPO_4. Additionally, the media was supplemented with 10 μg/ml kanamycin to select for strains with transposon-mediated chromosomal integrations. After an incubation period of 120-144 hr at 30°C, 165 of the darkest colonies (representing 7.9% of the total population) were chosen and restreaked on a fresh set of MOPS-agar plates. 30 colonies exhibiting the greatest melanin production after an additional 120-144 hr of incubation were used to inoculate 200 μl of LB medium containing 1 mM IPTG and 50 μg/ml spectinomycin. After four rounds of subculturing (with each round lasting at least 5 to 6 hr), individual colonies were isolated and tested for the loss of pTrc*melA*^{mut1} by streaking on Amp^{+} and Amp^{-} LB plates. Ampicillin-sensitive colonies were then analyzed for L-tyrosine production under the cultivation conditions described above. A modified Thermal Asymmetric Interlaced PCR (Tail-PCR) protocol was used to sequence and identity promising transposon targets (Alper, Miyaoku et al. 2005).

5.3 Results

5.3.1 Isolation of a *melA* variant with an enhanced capacity for melanin synthesis

The *melA* gene was amplified from *R. etli* genomic DNA and introduced into the pTrcHis2B vector under the control of the IPTG-inducible promoter, P_{trc}. During sequence verification of the plasmid construct, a variant was discovered containing a C→T base pair substitution at the 1000th nucleotide of the *melA* gene, a change which results in a proline to serine switch in the 334th amino acid. This single amino acid substitution led to a significant reduction in the lag time before the onset of melanin production, with the mutant showing signs of melanin synthesis 12 hr ahead of the wild-type (Figure 5.2). The mutated plasmid variant, named pTrc*melA*^{mut1}, was therefore selected for use in subsequent melanin production experiments.

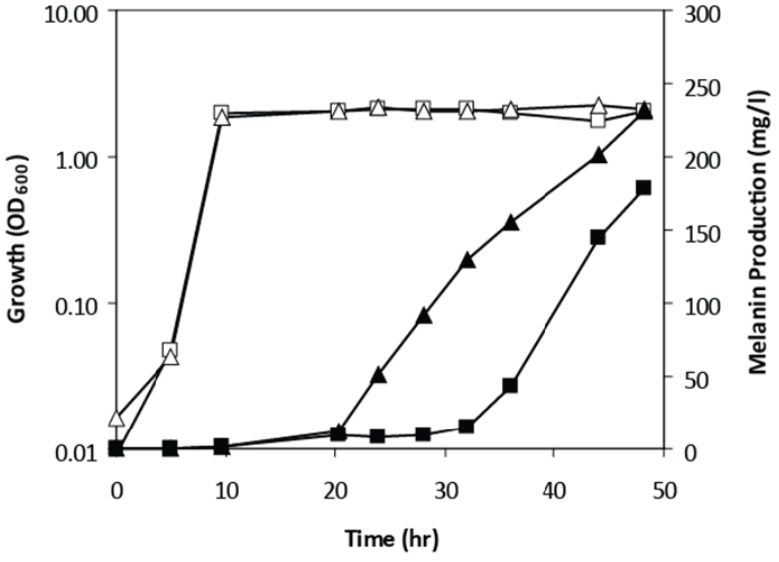


Figure 5.2 Growth and melanin production of K12 Δ*phe*A Δ*tyrR* expressing two versions of the *R. etli melA* gene.

Cultures were grown in M9 minimal medium with 500 mg/l L-tyrosine supplementation. Growth (OD_{600}): □, pTrc*melA*; △, pTrc*melA*^{mut1}. Melanin production (mg/l): ■, pTrc*melA*; ▲, pTrc*melA*^{mut1}.

5.3.2 Melanin Production in M9 and MOPS minimal media

In order to demonstrate the feasibility of probing L-tyrosine concentrations through melanin production, *E. coli* K12 Δ*phe*A Δ*tyrR* pTrc*melA*^{mut1} was cultured in liquid M9 minimal medium supplied with varying amounts of L-tyrosine (0 to 500 mg/l in 50 mg/l increments). As expected, a positive linear trend was observed between extracellular L-tyrosine supplementation and melanin production after 48 hr of cultivation, with a linear regression R^2 value of 0.922. Cultures grown with higher concentrations of L-tyrosine (> 250 mg/l) continued to produce melanin after this period, leading to even greater resolution and higher R^2 values after 72 and 96 hr (Figure 5.3a).

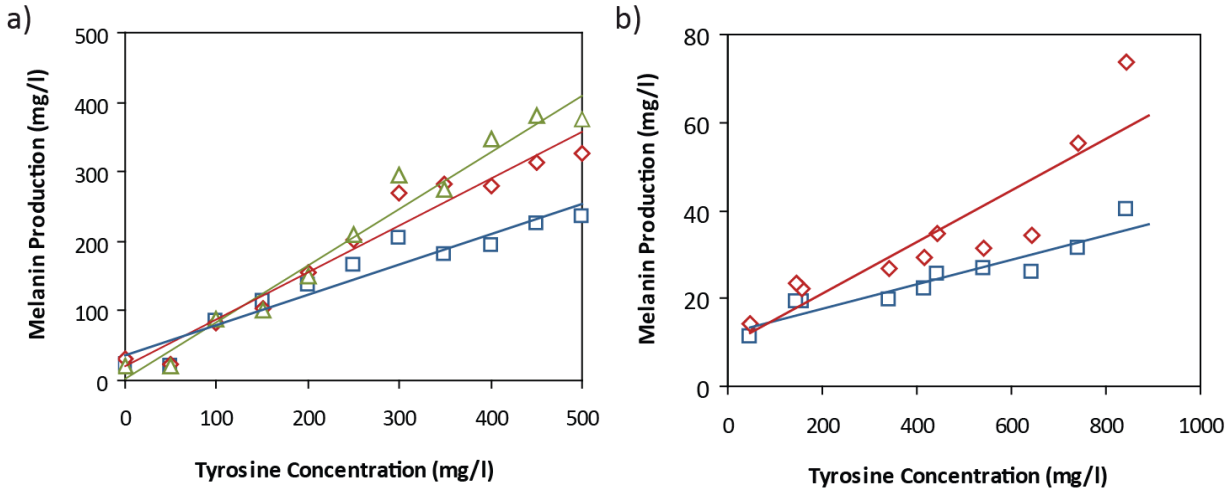


Figure 5.3 Correlation between melanin production and external L-tyrosine concentrations in different media formulations

a) M9 minimal medium – K12 Δ*pheA* Δ*tyrR* pTrc*melA*^{mut1} was cultivated in 0 to 500 mg/l L-tyrosine in 50 mg/l increments. Melanin measurements were taken after 48 hr (□), 72 hr (◇), and 96 hr (△) of cultivation. R^2 values for the linear regressions were 0.922, 0.956, and 0.968, respectively. b) MOPS minimal medium – Five L-tyrosine production strains (Table 5.1) were transformed with pTrc*melA*^{mut1} and assayed for melanin production in media without L-tyrosine supplementation. In order to probe a wider L-tyrosine concentration range, Strain D was also cultivated in media containing 100-500 mg/l L-tyrosine. Melanin measurements are shown after 313 hr (□) and 410 hr (◇) of growth. R^2 values for the linear regressions were 0.875 and 0.797, respectively.

Table 5.1 Production strains and L-tyrosine titers after 24 hr

Strain	Genotype	L-tyrosine production (mg/l)
A	*E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{WT}*aroG*^{WT}	7
B	*E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{WT}*aroG*^{fbr}	156
C	*E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{fbr}*aroG*^{WT}	175
D	*E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{fbr}*aroG*^{fbr}	347
E	*E. coli* K12 Δ*pheA* Δ*tyrR* Δ*wecB* pCL1920::*tyrA*^{fbr}*aroG*^{fbr}	433

Since L-tyrosine production is typically enhanced by as much as fifteen-fold in MOPS minimal medium as compared to M9 minimal medium (data not shown), steps were taken to determine whether these initial results could be reproduced in this alternate medium formulation. For these experiments, five L-tyrosine production strains (Table 5.1) were transformed with pTrc*melA*^{mut1} and cultivated in media without L-tyrosine supplementation. Additionally, *E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{fbr}*aroG*^{fbr} pTrc*melA*^{mut1} (Strain D) was cultured in media containing 100 to 500 mg/l L-tyrosine (in 100 mg/l increments) to extend the range of L-tyrosine concentrations tested. Under these conditions, two significant drawbacks with the use of MOPS minimal medium were encountered - 1) the poor resolving power of the assay due to the low levels of melanin produced and 2) the inordinate length of time needed for melanin synthesis to occur. Although a weaker linear correlation between L-tyrosine and melanin concentrations was still observed after 313 and 410 hr (Figure 5.3b), the highest melanin titers were five-fold lower than those produced in M9 minimal medium (74 mg/l versus 375 mg/l). Furthermore, a six-fold longer incubation period (313 hr versus 48 hr) was required for this trend to develop.

5.3.3 Optimizing melanin production in MOPS minimal medium

Supplementation experiments were conducted in order to determine which M9 component, if any, could improve tyrosinase enzyme activity in MOPS minimal medium. A comparison of media compositions revealed that M9 minimal medium contains 180-fold more calcium chloride ($CaCl_2$) and 32-fold more hydrogen phosphate (HPO_4^{2-}) than MOPS minimal medium (Neidhardt, Bloch et al. 1974); hence, the effects of $CaCl_2$ and sodium phosphate (dibasic, Na_2HPO_4) supplementation were examined. Although the addition of 0.09 mM $CaCl_2$ had a slightly detrimental effect on melanin synthesis, Na_2HPO_4 at both concentrations tested (40 and 60 mM) was sufficient for restoring melanin production in MOPS minimal medium (Table 5.2). Melanin concentrations measured after 72 and 96 hr were comparable to those previously observed for M9 minimal medium. In order to minimize the deviation from the standard recipe for MOPS minimal medium, lower concentrations of Na_2HPO_4 were also tested for their effect on melanin synthesis. Further optimization of the Na_2HPO_4 concentration revealed that only 20 mM was necessary to achieve adequate levels of melanin production (Figure 5.4).

Table 5.2 Melanin production of *E. coli* K12 Δ*phe*A Δ*tyrR* pTrc*melA*^{mut1} in MOPS minimal medium with supplementation

Supplementation[a]	Melanin Production (mg/l)	
	72 hr	**96 hr**
None	72	97
$CaCl_2$ (0.09 mM)	28	41
Na_2HPO_4		
40 mM	319	425
60 mM	338	349

[a] All cultures were supplemented with 500 mg/l L-tyrosine.

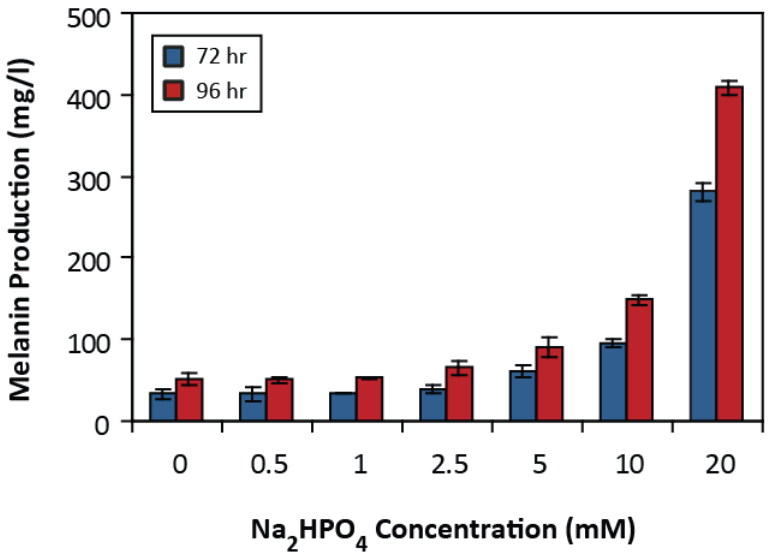


Figure 5.4 Melanin production by *E. coli* K12 Δ*phe*A Δ*tyrR* pTrc*melA*^{mut1} in MOPS minimal medium with different amounts of Na_2HPO_4 supplementation

All cultures were additionally supplemented with 500 mg/l L-tyrosine. Melanin measurements were taken after 72 hr (blue bars) and 96 hr (red bars).

To determine whether a correlation between melanin production and L-tyrosine concentration could be established with the optimized medium formulation, *E. coli* K12 Δ*phe*A Δ*tyrR* pTrc*melA*^{mut1} was once again cultured in varying concentrations of L-tyrosine. In stark contrast to the original MOPS minimal medium experiment, supplementation with 20 mM Na_2HPO_4 led to significant increases in the rates of melanin synthesis, as well as final titers of melanin. A linear trend was observed up to 300 mg/l L-tyrosine after 72 hr. After a cultivation period of 96 hr, a strong correlation was seen for the entire range of L-tyrosine concentrations tested, exhibiting a linear regression R^2 value of 0.992 (Figure 5.5).

Since the high-throughput implementation of this assay will likely require use in a solid medium format, a series of experiments was conducted to determine whether incremental differences in melanin production could also be distinguished by visual inspection on agar

plates. *E. coli* K12 Δ*phe*A Δ*tyrR* pTrc*melA*^{mut1} colonies were streaked on L-tyrosine-supplemented MOPS-agar plates and incubated at 30°C for the specified periods of time. After 72 hr, plates with L-tyrosine concentrations differing by as low as 50 mg/l were easily differentiated based on both the intensity of pigmentation and the radial diffusion of melanin (Figure 5.6a). The visual contrast between colonies became even more pronounced with increasing incubation times. These favorable trends were not just limited to externally supplemented L-tyrosine; a similar pigmentation pattern was observed among strains capable of different levels of L-tyrosine production (Table 5.1; A-E), with the highest L-tyrosine producer exhibiting the darkest coloration (Figure 5.6b).

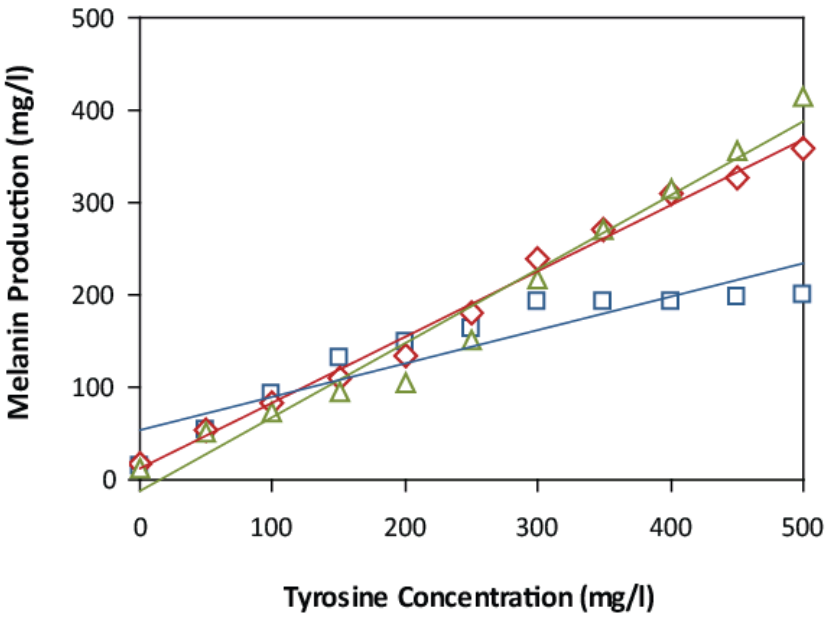


Figure 5.5 Correlation between melanin production and L-tyrosine supplementation in MOPS minimal medium with 20 mM Na_2HPO_4

K12 Δ*pheA* Δ*tyrR* pTrc*melA*^{mut1} was cultivated in 0 to 500 mg/l L-tyrosine in 50 mg/l increments. Melanin measurements were taken after 72 hr (□), 96 hr (◇), and 120 hr (△) of cultivation. R^2 values for the linear regressions were 0.853, 0.992, and 0.970, respectively.

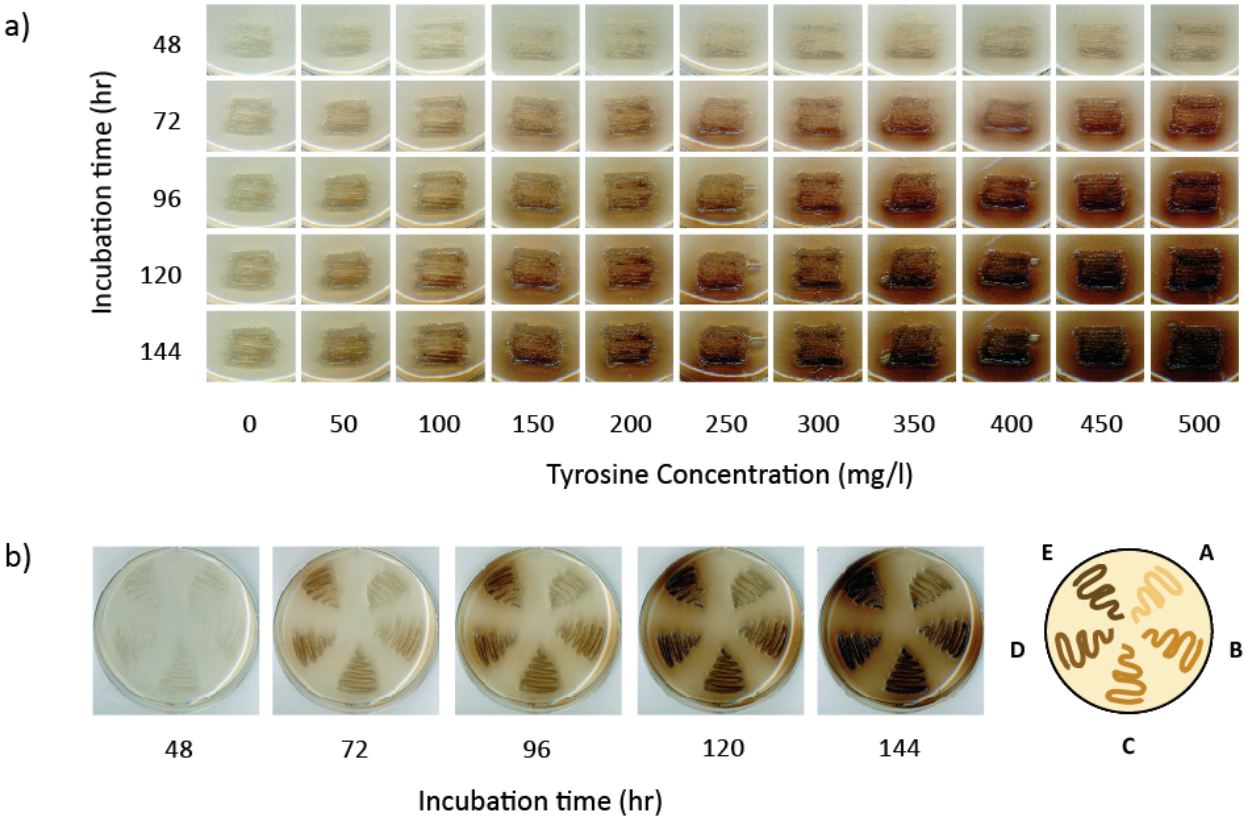


Figure 5.6 Melanin production on MOPS-agar plates with 20 mM Na_2HPO_4

a) Melanin production by K12 Δ*phe*A Δ*tyrR* pTrc*melA*^{mut1} with L-tyrosine supplementation. b) Melanin production by five L-tyrosine production strains, A-E, as listed in Table 5.1. Image brightness and contrast were adjusted with Adobe Photoshop CS2 (Brightness +25; Contrast +45).

5.3.4 Na_2HPO_4 supplementation increases the buffering capacity of MOPS minimal medium

To gain a better understanding of how Na_2HPO_4 exerts its effect on melanin synthesis, additional supplementation experiments were performed with the following chemicals: sodium chloride (NaCl), potassium phosphate (dibasic, K_2HPO_4), and sodium phosphate (monobasic, NaH_2PO_4). A proper basis for comparison was established by maintaining equivalent

concentrations of sodium (Na^+) or phosphate (HPO_4^{2-} or $H_2PO_4^-$) ions in all media formulations. The addition of NaCl to MOPS minimal medium had a slightly beneficial effect on melanin production but was only able to increase titers to 24-30% of the values observed with 20 mM Na_2HPO_4 (Table 5.3). Thus, the Na^+ concentration has only a marginal impact on melanin synthesis, and the bulk of the improvement in tyrosinase enzyme activity relies on the increase in HPO_4^{2-} availability. In accordance with this hypothesis, high melanin titers were once again attained when K_2HPO_4, another source of HPO_4^{2-}, was added to the medium. Interestingly, the addition of NaH_2PO_4 did not have a significant effect on melanin production, suggesting that Na_2HPO_4 supplementation is needed simply to impart the medium with extra buffering capacity. Further evidence arises from an apparent correlation between culture pH and melanin titers, with the highest pH values (6.55 - 6.67) resulting in the greatest melanin titers. When a 50-50 mixture of NaH_2PO_4 and Na_2HPO_4 was added to the medium, intermediate values for both pH and melanin concentrations were observed (Table 5.3). These findings are consistent with other recent reports on *R. etli* MelA activity. Using a partially purified MelA tyrosinase, Cabrera-Valladares et al. found that the pH optimum for the enzyme's L-dopa oxidase activity lies in the range of 6.5-7.5 (Cabrera-Valladares, Martinez et al. 2006). Subsequent optimization of melanin production in stationary-phase *E. coli* revealed a pH optimum of 7.5 (Lagunas-Munoz, Cabrera-Valladares et al. 2006).

Table 5.3 Melanin production and pH of *E. coli* K12 Δ*phe*A Δ*tyrR* pTrc*melA*^{mut1} in MOPS minimal medium with supplementation

Supplementation[a,b]	Melanin Production (mg/l)		Culture pH	
	72 hr	96 hr	72 hr	96 hr
None	25	37	4.60	4.57
Sodium phosphate, dibasic (Na_2HPO_4)[b]	305	401	6.57	6.55
Sodium chloride (NaCl)[c]	74	117	4.91	4.91
Potassium phosphate, dibasic (K_2HPO_4)[c]	303	392	6.65	6.67
Sodium phosphate, monobasic (NaH_2PO_4)[c]	22	34	4.74	4.76
Sodium phosphate, monobasic and dibasic (NaH_2PO_4 + Na_2HPO_4)[c,d]	85	136	5.04	5.04

[a] All cultures were supplemented with 500 mg/l L-tyrosine.

[b] Na_2HPO_4 was supplemented at a final concentration of 20 mM.

[c] For all other medium formulations, supplemented components were added such that the sodium (Na^+) or phosphate (HPO_4^{2-} or $H_2PO_4^-$) ion concentrations were equivalent to those found in 20 mM Na_2HPO_4.

[d] NaH_2PO_4 and Na_2HPO_4 contributed equal amounts of HPO_4^{2-} or H_2PO^{4-}.

5.3.5 Effect of Na_2HPO_4 supplementation and pTrc*melA*^{mut1} on L-tyrosine production

During the course of developing this assay, two major genotypic and environmental perturbations were introduced. Both the heterologous expression of the *R. etli melA* gene and the change in media formulation (as necessitated by the Na_2HPO_4 requirement for melanin synthesis) have the potential to interfere with L-tyrosine production in the strains of interest. Thus, studies were conducted to elucidate what effects, if any, such alterations have on the behavior of these strains. To ensure that Na_2HPO_4 supplementation does not have a negative effect on L-tyrosine synthesis, the production levels of four strains (Table 5.1; A, C-E) were

characterized in MOPS minimal medium containing 0 mM and 20 mM Na_2HPO_4. A plot of the concentrations achieved by these strains reveals that the overall trends in L-tyrosine production are maintained (Figure 5.7a). It is therefore reasonable to assume that the strains exhibiting the greatest capacity for L-tyrosine production in 20 mM Na_2HPO_4 will remain the top performers under standard cultivation conditions. To determine the impact of expressing a heterologous tyrosinase, the production levels of five strains (Table 5.1; A-E) were measured both in the presence and absence of pTrc*melA*^{mut1}. As seen in Figure 5.7b, the presence of the reporter plasmid had an unfavorable effect on the final L-tyrosine titers for the top two production strains, D and E, when cultivated at 37°C in liquid MOPS minimal medium. The L-tyrosine concentrations measured for these strains exhibited wide standard deviations, even for biological replicates within the same experiment. This demonstrates the importance of removing pTrc*melA*^{mut1} from strains identified through the melanin screen prior to a more rigorous quantification of L-tyrosine production through liquid cultivation and HPLC analysis.

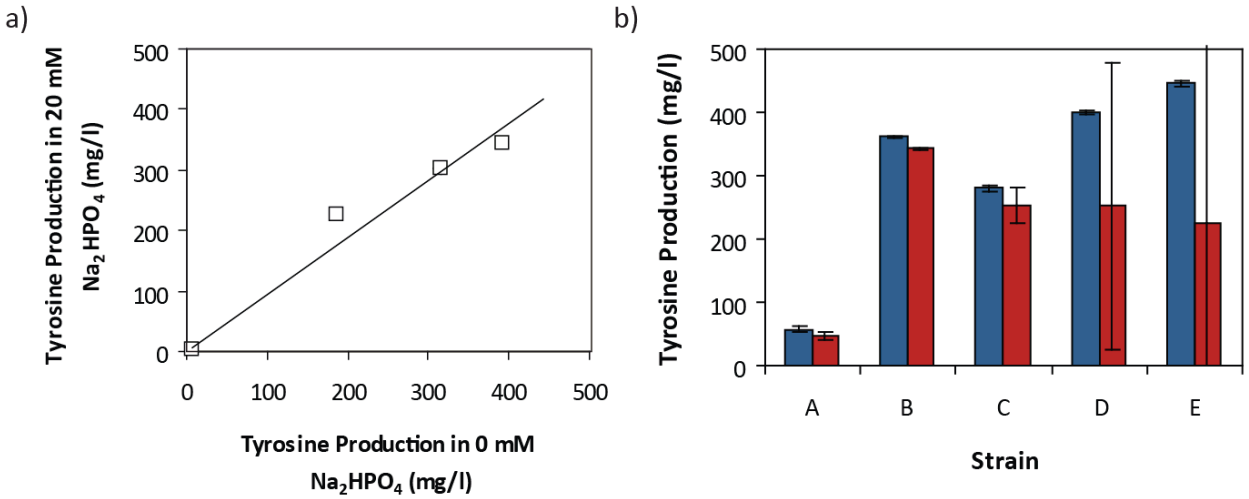


Figure 5.7 Effect of cellular and cultivation perturbations on L-tyrosine production

a) L-tyrosine production of Strains A, C, D, and E (Table 5.1) in 0 and 20 mM Na_2HPO_4 after 24 hr.
b) L-tyrosine production of Strains A-E (Table 5.1) with (red bars) and without (blue bars) pTrc*melA*^{mut1} after 24 hr.

5.3.6 Screening a random knockout library

Taking these results altogether, we have developed a selection strategy for utilizing this assay to screen large combinatorial libraries of mutant strains (Figure 5.8). In our initial application of the screen, a random knockout library was generated by the transposon-mediated mutagenesis of the parental strain P1 pTrc*melA*^{mut1} (Badarinarayana, Estep et al. 2001). Following transformation with the pJA1 vector, the library was plated directly onto MOPS-agar supplemented with 20 mM Na_2HPO_4 and 0.4 μg/ml $CuSO_4$, providing the optimum conditions for the synthesis of melanin. The plates were incubated at 30°C for a period of 120-144 hr during which colonies of noticeably different pigmentation intensities were detected. To achieve greater resolving power, the darkest colonies from this stage (165 colonies) were streaked out on a fresh set of MOPS-agar plates and incubated for an additional 120-144 hr. Subjecting strains through this second round of selection allowed us to more clearly differentiate between the levels of melanin produced by these isolates, as well as to limit further selection of false positives. Following this second incubation period, 30 strains exhibiting the most intense coloration underwent repeated rounds of subculturing in Amp$^-$ media to facilitate the loss of pTrc*melA*^{mut1}, which was shown earlier to have a detrimental effect on final L-tyrosine titers. In most cases, the plasmid was easily lost after four rounds of reinoculation, with each round lasting at least five to six hours (data not shown). Individual clones were isolated and tested for growth on both Amp$^+$ and Amp$^-$ media to verify the loss of the plasmid, and ampicillin-sensitive mutants were then cultivated under standard L-tyrosine production conditions and analyzed by HPLC.

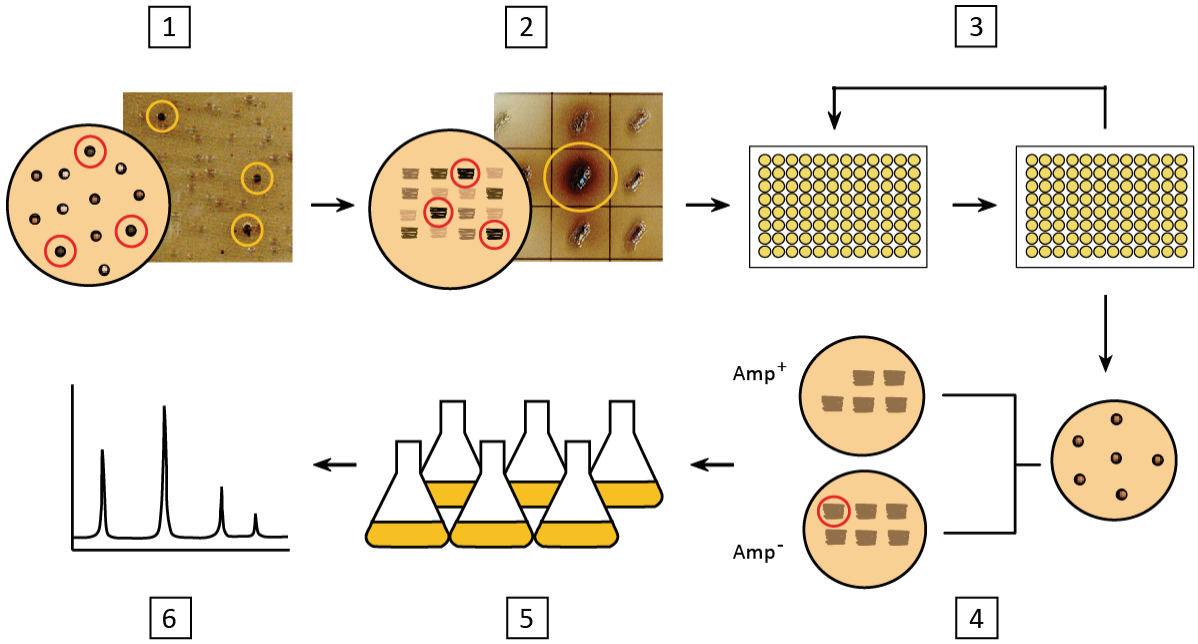


Figure 5.8 Strategy for screening libraries on solid media

(Step 1) Plate the library of mutants on MOPS-agar and incubate strains at 30°C until clear differences in melanin pigmentation are observed (120-144 hr). Select the darkest colonies from this first round of screening and (Step 2) streak them out on a fresh set of MOPS-agar plates. Incubate the plates for an additional 120-144 hr to allow for melanin synthesis. Select the darkest streaks from this round and (Step 3) proceed to the plasmid curing step. This is achieved by subculturing mutants in Amp^- medium at 37°C to facilitate the loss of pTrc*melA*^{mut1}. (Step 4) To verify loss of the plasmid, isolate single colonies and check for growth on Amp^- and Amp^+ plates. Strains that now exhibit ampicillin sensitivity are (Step 5) cultivated in the appropriate conditions for L-tyrosine production (MOPS minimal medium, 37°C). (Step 6) The cell-free culture supernatant is collected and analyzed via HPLC to quantify the L-tyrosine content of the medium. Image brightness and contrast were adjusted with Adobe Photoshop CS2 (Brightness +30; Contrast +30).

Out of an initial library size of 21,000 viable colonies, 30 mutants were chosen for a rigorous quantification of L-tyrosine production. Of these isolated strains, 2 mutants were found to possess L-tyrosine titers 57-71% above that of the parental strain (Table 5.4). The integrations were sequenced and verified to have occurred in the C-terminal region of the epsilon subunit of DNA polymerase III, encoded by *dnaQ* (Taft-Benz and Schaaper 1999), and the intergenic region between *rppH* (formerly *ygdP*) and *mutH*. For ease of notation, this latter region will be denoted as iHH.

Table 5.4 L-tyrosine production of strains isolated from a random knockout library (24 hr)

Strain	Genotype	L-tyrosine production (mg/l)	% increase above P1
P1	*E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{fbr}*aroG*^{fbr}	347	-
KO-*dnaQ*	*E. coli* K12 Δ*pheA* Δ*tyrR dnaQ*::Tn10-*kan* pCL1920::*tyrA*^{fbr}*aroG*^{fbr}	545	57
KO-iHH	*E. coli* K12 Δ*pheA* Δ*tyrR* iHH*::*Tn10-*kan* pCL1920::*tyrA*^{fbr}*aroG*^{fbr}	594	71

5.4 Discussion

5.4.1 Host versatility of melanin-based screen

Although the microbial production of aromatic amino acids, such as L-tyrosine, has been extensively studied and reviewed in recent years (Berry 1996; Flores, Xiao et al. 1996; Bongaerts, Kramer et al. 2001; Ikeda 2006; Lutke-Eversloh, Santos et al. 2007; Sprenger 2007), strategies for strain improvement have typically exhibited a narrow focus on well-characterized biochemical pathways and regulatory mechanisms. Combinatorial methods for metabolic

engineering were previously of limited use due to the absence of a high-throughput screen for assessing the phenotype of interest. In this study, we have described the development of a simple high-throughput screen for the microbial production of L-tyrosine based on the synthesis of the dark pigment melanin. Although proof-of-concept experiments were carried out in *E. coli*, such a screen can be readily applied to any microorganism capable of expressing a tyrosinase enzyme with high activity towards L-tyrosine. Many bacteria, including several species of *Rhizobia*, *Streptomyces*, *Pseudomonas*, and *Bacilli* naturally express these enzymes and produce melanin for protection against UV damage and increased virulence and pathogenesis (Wang, Aazaz et al. 2000; Nosanchuk and Casadevall 2003; Ruan, He et al. 2005; Cabrera-Valladares, Martinez et al. 2006; Claus and Decker 2006). Hence, such strains possess the innate ability to act as sensors for the production of L-tyrosine. Additionally, in cases where the host strain lacks an endogenous tyrosinase, the appropriate gene(s) can easily be introduced on a plasmid or integrated into the bacterial chromosome.

5.4.2 Screen implementation in liquid or solid formats

Screening for L-tyrosine production by monitoring melanin synthesis is a versatile technique that can be implemented in a variety of formats. Although the screening strategy presented here focuses on a solid medium implementation, liquid culture experiments can also be carried out in 96-well microtiter plates with individual strains from a combinatorial library inoculated into separate wells. After cultivation in media conducive for melanin production, the absorbance at 400 nm is measured, and desirable mutants are selected for further characterization. In addition, although the described method suggests the use of the same

strain for both L-tyrosine production and detection, it is possible to decouple these two steps by creating a strain exclusively for detection. In such a screening strategy, mutants from a combinatorial library are first individually cultured in 96-well microtiter plates. After a predetermined period of time, the culture supernatants, which contain different amounts of microbially-produced L-tyrosine, are then used as a growth medium component for a separate reporter strain expressing *melA*. Detection strains grown in the highest L-tyrosine concentrations will synthesize the most melanin and exhibit the highest absorbances at 400 nm, allowing for the identification of the best performing mutants. This latter strategy bears similarity to a recently published method for mevalonate detection with a GFP-expressing mevalonate auxotroph (Pfleger, Pitera et al. 2007). This "biosensor," as it was termed, allows one to measure the mevalonate content of a culture by monitoring the growth or fluorescence of the auxotrophic reporter strain. Although such a technique can also be used for L-tyrosine production through the construction of the appropriate auxotroph, coupling L-tyrosine production with melanin synthesis has the added convenience of requiring only one culturing step to simultaneously produce and detect the compound of interest. This important feature also allows for the simple execution of this screen in a solid medium format, which can be used to further enhance the high-throughput nature of the assay. As described earlier, with this approach, combinatorial libraries are plated directly on MOPS-agar medium, and colonies that exhibit the darkest pigmentation are selected for further analysis. Such a technique eliminates the need for expensive robotics to automate the selection and inoculation of colonies into microtiter plates and multi-plate scanners to increase the throughput of the absorbance measurements. Through this method, libraries on the order of 10^6 in size can be probed with

relative ease. Although the utility of the mevalonate biosensor strain was also demonstrated in a solid medium format by utilizing a plate spraying technique, this method was only shown to distinguish between mevalonate-producing and non-mevalonate-producing colonies (Pfleger, Pitera et al. 2007). This approach would be particularly difficult to implement on agar plates for the case of L-tyrosine production, since the parental strain that is used to generate these combinatorial libraries already produces an elevated level of L-tyrosine. It is therefore likely that the most severe growth-limiting factor for the auxotrophic strain will be the depletion of a carbon source rather than L-tyrosine.

5.4.3 Alternative high-throughput screen for L-tyrosine production

More recently, an alternative assay for L-tyrosine production has been described which utilizes a chemical reaction between 1-nitroso-2-naphthol and L-tyrosine to produce a yellow, fluorescent product. This method, which was originally developed for the determination of L-tyrosine levels in blood plasma (Waalkes and Udenfriend 1957), was adapted for the case of microbial L-tyrosine production in microtiter plates (Lütke-Eversloh and Stephanopoulos 2007). Again, however, the high-throughput implementation of this assay is heavily reliant on the availability of expensive robotics to automate sampling, reaction preparations, and fluorescence measurements.

5.4.4 Discovery of novel knockout targets

Screening a random knockout library with this melanin-based selection strategy has led to the discovery of two targets that were successful in eliciting significant increases in L-tyrosine

production. A *dnaQ*::*kan* mutation in the background of the parental strain *E. coli* K12 Δ*pheA* Δ*tyrR* pCL1920::*tyrA*^{fbr}*aroG*^{fbr} led to a 57% increase in L-tyrosine production; an insertion in the intergenic region between *rppH* and *mutH* resulted in even more substantial increases (71%). It is important to note that rational design approaches would not have been capable of predicting either of these genetic changes, particularly for the case of the *rppH-mutH* mutation, which occurred in a noncoding region. Certainly, for both mutant strains identified, further work must be conducted to elucidate the complex relationship between genotype and cellular phenotype. This example, however, serves to illustrate the great potential that can be unlocked by such a screening strategy. Indeed, the application of this simple assay for probing a variety of combinatorial libraries will likely lead to the discovery of additional targets that were previously unreachable through traditional methods of metabolic engineering.

CHAPTER 6 EVALUATION OF RANDOM KNOCKOUT TARGETS

6.1 Introduction

As described in Chapter 5, the use of a melanin-based screen to scan a random knockout library yielded two deletion mutants that led to higher L-tyrosine titers than the parental strain, P1. Interestingly, both transposon insertions occurred within or near loci not previously determined to be related to aromatic amino acid production, with one mutation occurring in a noncoding region. How then do these mutations contribute to enhanced L-tyrosine production in these recovered strains? This chapter delves into the characterization of the specific mutations isolated, as well as the subsequent development and testing of new rational strategies by which to engineer future strains.

6.2 Materials and Methods

6.2.1 Cultivation conditions

L-tyrosine production experiments were performed at 37°C with 225 rpm orbital shaking in 50 ml MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) cultures supplemented with 5 g/l glucose and an additional 4 g/l NH_4Cl. All liquid cultivations were conducted in at least triplicates. When appropriate, antibiotics were added in the following concentrations: 50 µg/ml spectinomycin for maintenance of pCL1920::*tyrA*^{fbr}*aroG*^{fbr} and 20 µg/ml kanamycin for

strains containing *kan* insertions. Isopropyl-β-D-thiogalactopyranoside (IPTG) (EMD Chemicals) was added at a concentration of 2 mM for the induction of pCL1920::*tyrA*^{fbr}*aroG*^{fbr}. For L-phenylalanine auxotrophs (Δ*pheA*), L-phenylalanine (Sigma) was supplied at a concentration of 0.35 mM.

6.2.2 P1*vir* phage transduction of transposon cassettes

Transfer of the transposon insertions into a clean parental (P1) genetic background was mediated by P1*vir* phage transduction (Miller 1992) of the isolated mutants, KO-*dnaQ* and KO-iHH. The presence of the appropriate cassettes, *dnaQ*::Tn10-*kan* and iHH::Tn10-*kan*, was verified by colony PCR and sequencing.

6.2.3 Construction of *dnaQ*^{mut1} and *dnaQ*^{mut2} strains

Strains *dnaQ*^{mut1} and *dnaQ*^{mut2} were constructed by amplification of *kan* from pKD13 (Datsenko and Wanner 2000) with two sets of primers. CS332 dnaQ trunc sense (5' AGT TAC GCG TTG TTT TTG CGA CAG ATG AAG AGA TTG CAG CTC ATG AAG CCT AAG TGT AGG CTG GAG CTG CTT C-3') and CS334 dnaQ trunc anti (5'- GCA AAA ATC GCC CAA GTC GCT ATT TTT AGC GCC TTT CAC AGG TAT TTA TGA TCC GTC GAC CTG CAG TTC GA-3') were used to amplify the *dnaQ*^{mut1} cassette for the insertion of a stop codon (underlined text) 51 bp from the end of the *dnaQ* gene. CS333 dnaQ-kan sense (5'-TTG TTT TTG CGA CAG ATG AAG AGA TTG CAG CTC ATG AAG CCC TGA TGA ATC CCC TAA TGA TTT TGG TAA AAA ATC ATT AAG TGT AGG CTG GAG CTG CTT C-3') and CS334 dnaQ trunc anti were used to amplify the *dnaQ*^{mut2} cassette for the replacement of the last 51 bp of *dnaQ* with 36 new bp derived from the Tn10 insertion cassette (underlined text) and a

stop codon. Chromosomal integration of these PCR cassettes into P1 was performed using a lambda-red recombination based method (Datsenko and Wanner 2000). Colonies were verified by colony PCR and sequencing, and excision of FRT-flanked *kan* was mediated by transformation with FLP recombinase-expressing pCP20 as described in the literature (Datsenko and Wanner 2000).

6.2.4 Construction of acetate-deficient strains

Deletion cassettes for *ackA-pta* were amplified from pKD13 with primers CS501 ackA-pta-kan sense (5'- CGT ATC AAT TAT AGG TAC TTC CAT GTC GAG TAA GTT AGT ACT GGT TCT GAA CTG CGT GTA GGC TGG AGC TGC TTC-3') and CS502 ackA-pta-kan anti (5'- GCA AAG CTG CGG ATG ATG ACG AGA TTA CTG CTG CTG TGC AGA CTG AAT CGC AAT CCG TCG ACC TGC AGT TCG A-3'). Deletion cassettes for *poxB* were amplified from pKD13 with primers CS514 poxB-kan sense (5'- AAT GCC ACC CTT TTT ACC TTA GCC AGT TTG TTT TCG CCA GTT CGA TCA CTA TCC GTC GAC CTG CAG TTC GA-3') and CS515 poxB-kan anti (5'- GAA ACA AAC GGT TGC AGC TTA TAT CGC CAA AAC ACT CGA ATC GGC AGG GGG TGT AGG CTG GAG CTG CTT C-3'). Integration of these PCR products into parental strain P2 and KO-*dnaQ* (either singly or in combination) was performed with a lambda-red recombination based method as described before (Datsenko and Wanner 2000).

6.2.5 Mutational analysis of strains

Mutation frequency analysis protocols were adapted from a previously published report on the use of *rpoB* to analyze the specificity of base substitutions in *E. coli* (Garibyan, Huang et al.

2003). Briefly, cells were inoculated at a starting OD_{600} of 0.01 and grown overnight in 5 ml LB cultures. After appropriate dilution, cultures were plated on LB-agar with and without 100 µg/ml rifampicin (Sigma). (Dilution is necessary to recover distinct colonies as opposed to lawn growth, with an approximate target of less than 200 colonies per plate). The number of colony forming units (CFUs) on each plate was quantified after 16-20 hr, and average mutation frequencies were determined by dividing the average number of rifampicin-resistant colonies by the average number of CFUs growing on unselective media (LB). At least ten culture and plate replicates were used in this calculation for each strain tested.

6.2.6 Analytical methods

For the quantification of L-tyrosine, cell-free culture supernatants were filtered through 0.2 µm PTFE membrane syringe filters (VWR International) and used for HPLC analysis with a Waters 2690 Separations module connected with a Waters 996 Photodiode Array detector set to a wavelength of 278 nm. The samples were separated on a Waters Resolve C18 column with 0.1 % (vol/vol) trifluoroacetic acid (TFA) in water (solvent A) and 0.1 % (vol/vol) TFA in acetonitrile (solvent B) as the mobile phase. The following gradient was used at a flow rate of 1 ml/min: 0 min, 95 % solvent A + 5 % solvent B; 8 min, 20 % solvent A + 80 % solvent B; 10 min, 80 % solvent A + 20 % solvent B; 11 min, 95 % solvent A + 5 % solvent B. Acetate concentrations were measured by HPLC using an Aminex HPX-87H ion-exclusion column (300 7.8 mm; Bio-Rad, Hercules, CA) and a mobile phase of 14 mM H_2SO_4 at 50 °C and 0.7 ml/min. Cell densities of cultures for growth rate calculations were determined by measuring their absorbance at 600 nm with an Ultrospec 2100 *pro* UV/Visible spectrophotometer (Amersham

Biosciences). pH measurements were taken with a SympHony SP20 pH meter and electrode (VWR International).

6.3 Results and Discussion

6.3.1 Phenotypic transfer is limited to *dnaQ*::Tn10-*kan*

In evaluating the two isolated knockout mutants, our first goal was to determine whether their enhanced phenotypes were an *exclusive* result of transposon integration. In this favorable scenario, identification of the site of insertion and subsequent replication of this deletion in another strain should result in the full recovery of phenotype. To test this property, we utilized P1*vir* phage transduction to transfer the integration cassette into a clean genetic background (parental P1). Surprisingly, we found that although a strain containing a reconstructed *dnaQ*::Tn10-*kan* led to comparable high titers as the recovered strain KO-*dnaQ*, the same trend was not seen for the iHH::Tn-10-*kan* mutation (Figure 6.1). Phage transduction of this altered locus into P1 led to a loss of almost all advantages previously seen in the original knockout mutant KO-iHH, with L-tyrosine titers just 11% above those of the parental strain (396 mg/l versus 356 mg/l for P1). In light of this result, it is possible that deletion of iHH works either additively or synergistically with an unidentified background mutation (presumably incurred during screening or cultivation) to impart a more striking L-tyrosine overproduction phenotype. Some may argue that this unknown chromosomal mutation may in fact be wholly responsible for the changes, but this theory seems less likely given that iHH deletion by itself still led to measurable and statistically significant increases in production ($p < 0.001$).

Although this unconventional result makes it quite difficult to initiate phenotypic transfer into a new strain, it was not altogether unexpected after considering the unusual location of the insertion within an intergenic/noncoding region. While it is true that more and more biologically-active noncoding RNAs (ncRNAs) are being discovered (Lease, Smith et al. 2004; Storz, Opdyke et al. 2004; Jin, Watt et al. 2009), most intergenic regions still play limited roles in dictating cellular function. Because of the underlying complexity associated with the iHH mutation, we decided to focus our efforts on characterizing the *dnaQ* mutation, which was transferable and did lead to sizeable gains in L-tyrosine production.

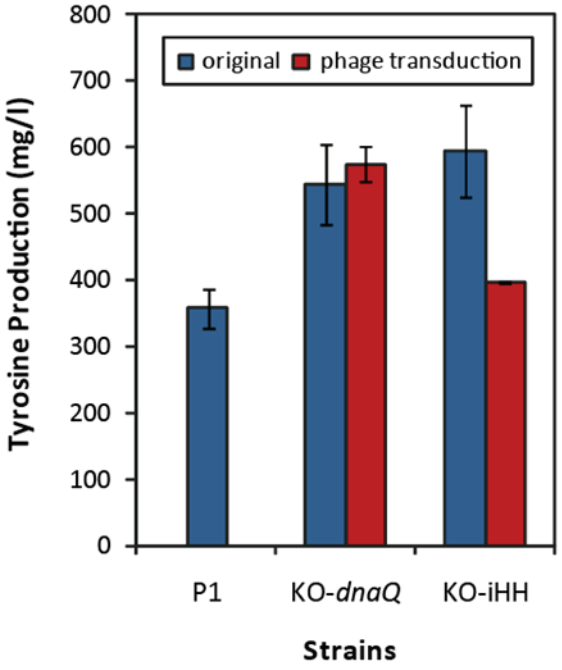


Figure 6.1 L-tyrosine production of knockout mutants and derivatives after 24 hr

L-tyrosine titers for parental strain P1 and original mutants isolated from a random knockout library are shown in blue. Titers for the corresponding strains reconstructed via P1*vir* phage transduction are depicted in red.

6.3.2 DNA polymerase III and the functional role of its ε subunit

Although *E. coli* possesses five different DNA polymerases, only one is capable of performing DNA elongation and proofreading during chromosome duplication. This function resides with

DNA polymerase III (pol III), an enzyme that can efficiently replicate the entire *E.* coli chromosome (4.6 million base pairs) with error frequencies of less than 1 in 10^{10} nucleotides. The pol III holoenyme exists as a complex of 18 polypeptides with 10 distinct subunits. Despite this large size, its polymerase and 3'→5' exonuclease activities have long been attributed to a smaller core assembly made up of the α, ε, and θ subunits (Kelman and O'Donnell 1995; O'Donnell 2006). Interestingly, our recovered knockout target *dnaQ* encodes for the ε subunit of this enzyme.

Functional studies on the ε subunit have elucidated its role in imparting two significant properties to pol III – replication fidelity and processivity. Because the ε subunit contains a 3'→5' exonuclease domain, it is largely responsible for ensuring accuracy of replication in pol III Surprisingly, ε also has an important structural role and, through its tight association with the polymerase subunit α, can stimulate activity, high processivity (the number of nucleotides added before dissociation from substrate), and rapid cycling of the enzyme (Studwell and O'Donnell 1990). This favorable interaction between α and ε seems to be mutually beneficial, as the presence of the α subunit not only leads to a 10-fold increase in ε exonuclease activity on single-stranded DNA but also allows it to extend its substrate specificity to double-stranded DNA as well (Maki and Kornberg 1987).

6.3.3 Domain structure and mutational analysis of ε subunit

Structural and mutational analyses have identified two specific domains in ε that allow it to mediate the two aforementioned functions. The N-terminal 186 amino acids of ε comprise the 3'→5' exonuclease domain, while a C-terminal α-association unit exists within the last 40 amino

acids of the protein. A Q-linker sequence tethers these two structurally distinct but interacting domains (Perrino, Harvey et al. 1999; Taft-Benz and Schaaper 1999). Within the exonuclease domain lie three conserved motifs ExoI, ExoII, and ExoIIIε, and, as one might expect, mutations isolated to these areas lead to severe deficiencies in replication fidelity (as exhibited by a 700-8000-fold increase in mutation frequency). These mutations tend to be dominant and can only be partially reversed by *dnaQ* complementation. This limited reversibility is hypothesized to be due to a fairly evenly-matched competition between both wild-type and exonuclease-deficient ε for α binding. Strains containing both versions of ε ultimately end up with about half the number of defective DNA polymerase holoenzymes, leading to mutation frequencies that are lower than those of a purely mutated strain but higher than the wild-type version (Taft-Benz and Schaaper 1998).

Interestingly, a second group of mutants exhibiting a lower fold increase in mutation frequency (6-20-fold above wild-type) has also been isolated. This subset consists mostly of recessive *dnaQ* mutators with phenotypes that can be reversed by wild-type *dnaQ* complementation. Such mutations are thought to lead to partial deficiencies in α subunit binding, either through the introduction of mutations in the ε subunit's α-association domain or by imposing other structural deformities that alter its binding properties. Because wild-type *dnaQ* products can compete quite effectively for binding with the pol III core subunits, one observes almost complete phenotypic reversal even in the presence of the mutant allele (Taft-Benz and Schaaper 1998).

Although a substantial amount of work has focused on the loss of replication fidelity in pol III, very few studies have commented on specific mutations that alter enzyme processivity.

This disparity most likely arises from difficulties in screening for these mutants; while selecting for a mutator phenotype has become a relatively straightforward experiment (Garibyan, Huang et al. 2003), no such technique exists for selecting less processive pol III variants. It has been shown, however, that decreasing *dnaQ* expression can lead to decreased processivity (as reflected by a slower growth rate) without having a concomitant effect on mutation frequency. This experiment, which makes use of antisense RNA for the conditional silencing of *dnaQ,* therefore proves that the two functions of ε can in fact be decoupled (Stefan, Reggiani et al. 2003).

6.3.4 Investigating functional loss in *dnaQ*:Tn10-*kan*

Because the ε subunit of pol III plays such a fundamental role in cell replication and growth, complete inactivation of *dnaQ* unsurprisingly leads to inviable strains (Stefan, Reggiani et al. 2003). From this data, it then becomes readily apparent that integration of the transposon within *dnaQ* did not completely abolish function, but rather, altered it in a very specific manner. In corroboration with this hypothesis, sequencing of the *dnaQ* locus revealed that transposon integration actually occurred within the C-terminal region of the ε subunit, a mere 16 amino acids from the end of the protein; thus, most of the protein remains intact and likely retains some functionality (Figure 6.2).

Several questions, however, remain unanswered. First, what specific effect does this insertion have on pol III function, and which properties of the ε subunit (replication fidelity or polymerization rate/processivity) are affected? Perhaps more importantly, how exactly does modulation of the cell's replication machinery lead to the seemingly unrelated phenotype of

amino acid overproduction? To elucidate these points further, we reconstructed and analyzed the effects of two mutant *dnaQ* alleles on cellular phenotype.

6.3.5 Construction of *dnaQ*^{mut1} and *dnaQ*^{mut2}

Although the transposon actually inserted 16 amino acids from the C-terminus of the ε subunit, a more detailed analysis of the resulting polypeptide sequence revealed much more than a mere truncation of this segment. Instead, cassette integration led to a completely modified sequence through the replacement of the last 16 amino acids with 12 completely new residues (Figure 6.2). Although previous studies have shown that truncation of ε by just three amino acids can lead to a recessive mutator phenotype (Taft-Benz and Schaaper 1998), the effect of this particular mutant allele unfortunately becomes much more difficult to predict given its completely altered C-terminal sequence. We therefore constructed two strains containing distinct but representative *dnaQ* alleles in order to probe some of the changes in pol III function imparted by these specific mutations. Strain *dnaQ*^{mut1} was built under the assumption that a modified C-terminus renders that portion of the protein nonfunctional; its *dnaQ* allele was therefore designed to be a complete truncation of the terminal 16 amino acids of the protein (Figure 6.2). Strain *dnaQ*^{mut2}, by contrast, was intended to more accurately mimic the recovered KO-*dnaQ* mutant through incorporation of the 12 novel amino acids. If this unique terminal sequence has a specific effect on ε subunit functionality, such changes should be reflected in this reconstructed mutant.

a)

	dnaQ^{WT}	GTT TTT GCG ACA GAT GAA GAG ATT GCA GCT CAT GAA GCC CGT CTC GAT CTG GTG
1210	*dnaQ*^{mut1}	GTT TTT GCG ACA GAT GAA GAG ATT GCA GCT CAT GAA GCC --- --- --- --- ---
	dnaQ^{mut2}	GTT TTT GCG ACA GAT GAA GAG ATT GCA GCT CAT GAA GCC CTG ATG AAT CCC CTA

	dnaQ^{WT}	CAG AAG AAA GGC GGA AGT TGC CTC TGG CGA GCA TAA
1276	*dnaQ*^{mut1}	--- --- --- --- --- --- --- --- --- --- --- TAA
	dnaQ^{mut2}	ATG ATT TTG GTA AAA AAT CAT --- --- --- --- TAA

b)

	dnaQ^{WT}	V F A T D E E I A A H E A R L D L V Q K K G G S C L W R A *
215	*dnaQ*^{mut1}	V F A T D E E I A A H E A - - - - - - - - - - - - - - - - - *
	dnaQ^{mut2}	V F A T D E E I A A H E A L M N P L M I L V K N H - - - - *

c)

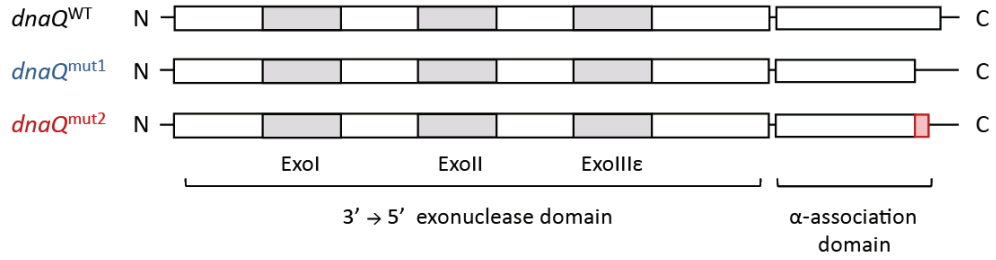


Figure 6.2 Comparison of WT *dnaQ* and mutant constructs

a) A comparison of the last 90 bp of the *dnaQ* sequence and b) the corresponding amino acids (using standard single-letter abbreviations; * indicates a stop codon). c) Protein schematic of ε subunit variants.

6.3.6 Mutational analysis reveals that replication fidelity was not affected

Our first goal in characterizing *dnaQ*^{mut1} and *dnaQ*^{mut2} was to determine whether the replication fidelity of pol III was altered in these strains. Specifically, we utilized a simple rifampicin-based assay to measure changes in background mutation frequency (Garibyan, Huang et al. 2003). In this test, mutagenesis of the *rpoB* gene is used as a metric for establishing the rates at which single base pair substitutions are introduced within the bacterial chromosome. *rpoB* functions as a fairly informative and accurate locus for this assay, since at least 69 single base pair substitutions within this gene (distributed among 24 coding positions) are able to grant rifampicin resistance to the cell.

As we might have predicted from previous studies, the truncation of *dnaQ* in *dnaQ*^{mut1} led to a mutation frequency that was about 9 times higher than that seen in the parental strain P1 (Figure 6.3a). This fold increase actually corroborates quite well with other publications on the effects of truncating or otherwise altering the α-association domain of the ε subunit (Taft-Benz and Schaaper 1998). Surprisingly, however, both *dnaQ*^{mut2} and the original knockout mutant KO-*dnaQ* exhibited very low mutation frequencies that were either comparable or even slightly lower than the parental. These unexpected results reveal two important points. First, the altered sequence introduced by transposon integration *does* have a specific effect on protein function and therefore cannot be modeled as a mere truncation as was assumed in *dnaQ*^{mut1}; clearly, strain *dnaQ*^{mut2} serves as a more accurate model for the mutational events that occurred in KO-*dnaQ*. While a mutational analysis reveals similarities in behavior between *dnaQ*^{mut2} and KO-*dnaQ*, this idea is also reinforced by the L-tyrosine production capabilities of these three strains. As seen in Figure 6.3b, the L-tyrosine titer recovered with *dnaQ*^{mut1}, while

still slightly higher than P1, very noticeably lags behind both *dnaQ*^{mut2} and KO-*dnaQ*. The second important conclusion to be drawn from these results is that, surprisingly, introduction of a unique C-terminal sequence in the ε subunit had no effect on replication fidelity, as no mutational deficiencies were observed in these strains.

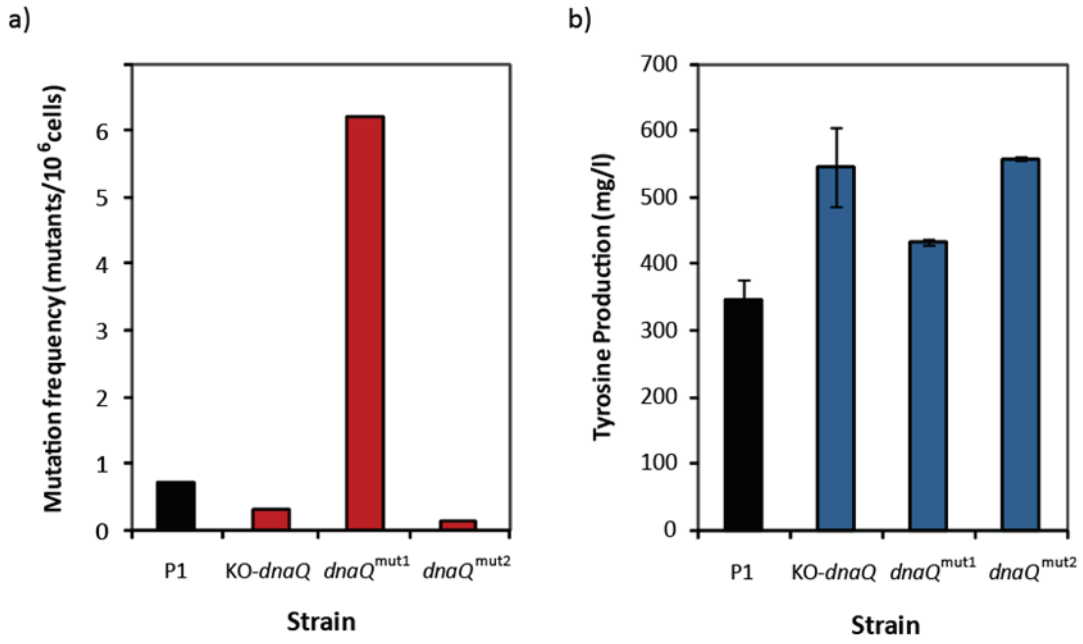


Figure 6.3 Cellular properties of *dnaQ*-mutated strains

a) Mutation frequency as determined by a rifampicin-based assay b) L-tyrosine production of strains after 24 hr

6.3.7 Growth and L-tyrosine production are negatively correlated

Since pol III replication fidelity was not affected in *dnaQ*^{mut2} and KO-*dnaQ*, we then set out to determine if DNA polymerization rate or processivity may have been altered by these mutant *dnaQ* alleles. Because no direct metric for *in vivo* pol III processivity exists, we were forced to access this phenotype by looking for more general, downstream effects on cellular growth rate

(Stefan, Reggiani et al. 2003). In doing so, clear differences between *dnaQ*^{mut2} and P1 finally emerged, with *dnaQ*^{mut2} exhibiting a 15% slower specific growth rate (Table 6.1). Given these results, it is therefore reasonable to assume that while the novel C-terminus had no effect on the replication fidelity of pol III, it did alter enzyme processivity, a property that became manifest as a reduction in growth rate.

Table 6.1 Specific growth rate of parental strain P1 and *dnaQ*^{mut2}

Strain	Specific growth rate, μ (hr^{-1})
P1	0.547
dnaQ^{mut2}	0.462

An examination of several rationally engineered strains and combinatorial mutants revealed an even clearer illustration of an apparent negative correlation between L-tyrosine production and specific growth rate (Figure 6.4). The linearity of this trend was actually quite striking, displaying an R^2 value of 0.91. This unusual result then begs the interesting question: how exactly does a reduction in cell growth lead to a higher capacity for L-tyrosine production in these strains?

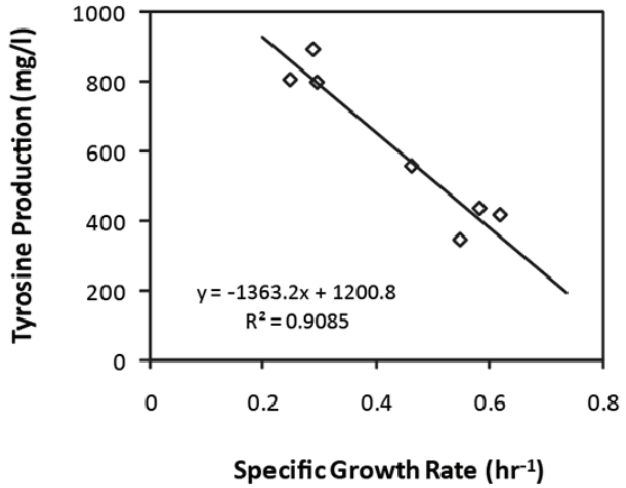


Figure 6.4 Negative correlation between L-tyrosine production and specific growth rate

Individual data points represent the following strains: P1, P2, P2 pHACm-*rpoD*^{WT}, *dnaQ*^{mut2}, rpoA14, rpoA27, and rpoD3. The last three strains are gTME-derived mutants that will be introduced in Chapter 7.

6.3.8 Strains with high L-tyrosine titers generate only low levels of acetate

Despite the strong negative correlation revealed in Figure 6.4, the interplay between growth and L-tyrosine production remains complex and multifaceted. Several studies including our own have shown that the vast majority of L-tyrosine is actually produced during exponential growth, a phenomenon that makes intuitive sense in light of a growing cell's requirement for amino acids to synthesize proteins and other cellular building blocks (Lütke-Eversloh and Stephanopoulos 2007; Patnaik, Zolandz et al. 2008). At the same time, however, *overproduction* of an amino acid often means that less carbon is available for biomass formation; thus, a slower growing cell which does not utilize carbon efficiently may ultimately divert flux of this carbon source towards other pathways of interest.

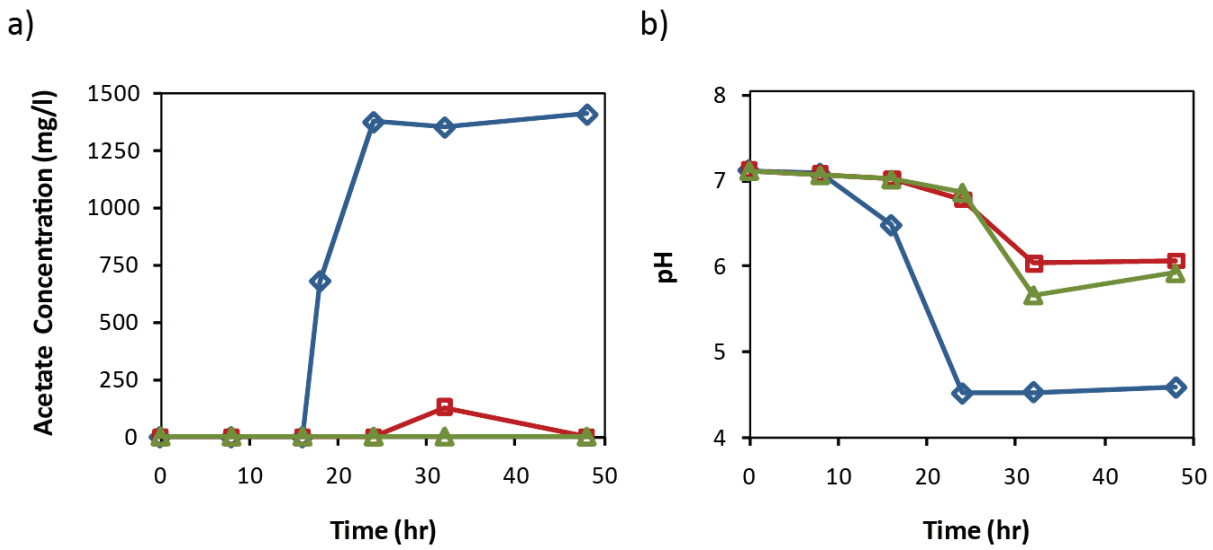


Figure 6.5 Acetate concentrations and pH of representative cultures

a) Acetate concentrations and b) culture pH for three strains: P2 pHACm-*rpoA*^{WT} (◇), rpoA14 (□), rpoA27 (△).

While carbon limitation remains a plausible hypothesis, organic acid concentrations in spent media led us to explore a more specific biochemical feature which could account for the negative correlation between growth and L-tyrosine production. This investigation arose from our observations that strains with elevated levels of L-tyrosine production secreted much lower amounts of acetate than their parental counterparts. In a representative comparison, two mutants isolated from an *rpoA*-derived gTME library (to be introduced in Chapter 7) produced close to undetectable levels of acetate, while the parental strain P2 carrying a wild-type *rpoA* secreted acetate at more than 1 g/l quantities (Figure 6.5a). In corroboration with these measurements, the level of acetate accumulation was also reflected by a much lower culture pH for the latter strain (Figure 6.5b).

6.3.9 Acetate production inhibits growth and acts as a carbon sink

Acetate is naturally produced in *E. coli* when growing cells surpass a threshold specific rate of glucose consumption. For cultures grown under glucose-limited conditions, this uptake rate is directly correlated to the specific growth rate of the cell and, for this reason, is often depicted as the determining variable for this phenomenon. Cells that replicate beyond this threshold rate are then faced with a unique biochemical dilemma. Because the rate of glycolysis far exceeds that of the TCA cycle, the intracellular concentrations of pyruvate and acetyl-coA begin to rise rapidly. As a mechanism for coping with this "overflow metabolism," acetate is generated from these accumulating metabolites and subsequently secreted into the culture medium. This biochemical conversion can be catalyzed by two separate pathways which function at different stages of the growth cycle. The enzymes phosphotransacetylase and

acetate kinase (encoded by *pta* and *ackA*) generate acetate from acetyl-coA and are both highly expressed during exponential growth. As a second route, pyruvate can also be converted to acetate by the action of pyruvate oxidase (encoded by *poxB*). This pathway has been shown to be active only during the early stationary phase (Figure 6.6) (Tomar, Eiteman et al. 2003; Dittrich, Bennett et al. 2005).

Although fast growers are often desired in biotechnological processes, acetate production can, unfortunately, inhibit a strain's inherent capacity to produce a desired compound. From a mass balance point of view, synthesis of acetate can act as a carbon sink and ultimately drains resources that could have otherwise been diverted towards product formation. Beyond competition for carbon, however, acetate accumulation also results in dramatic drops in culture pH and can therefore have a severe impact on cell viability. For both these reasons, it is not altogether surprising that our best performing strains exhibited reduced levels of acetate secretion. Furthermore, it seems likely that actively diminishing acetate production in these strains may be a simple metabolic engineering strategy for extending their L-tyrosine production capabilities.

6.3.10 Inhibiting acetate production through deletion of *ackA-pta* and *poxB*

Inhibiting acetate production in *E. coli* has been a popular objective for several years, particularly because of its negative effects on recombinant protein production (Jensen and Carlsen 1990; Nakano, Rischke et al. 1997; March, Eiteman et al. 2002). Although several environmental and genetic approaches have been previously adopted for this purpose, the

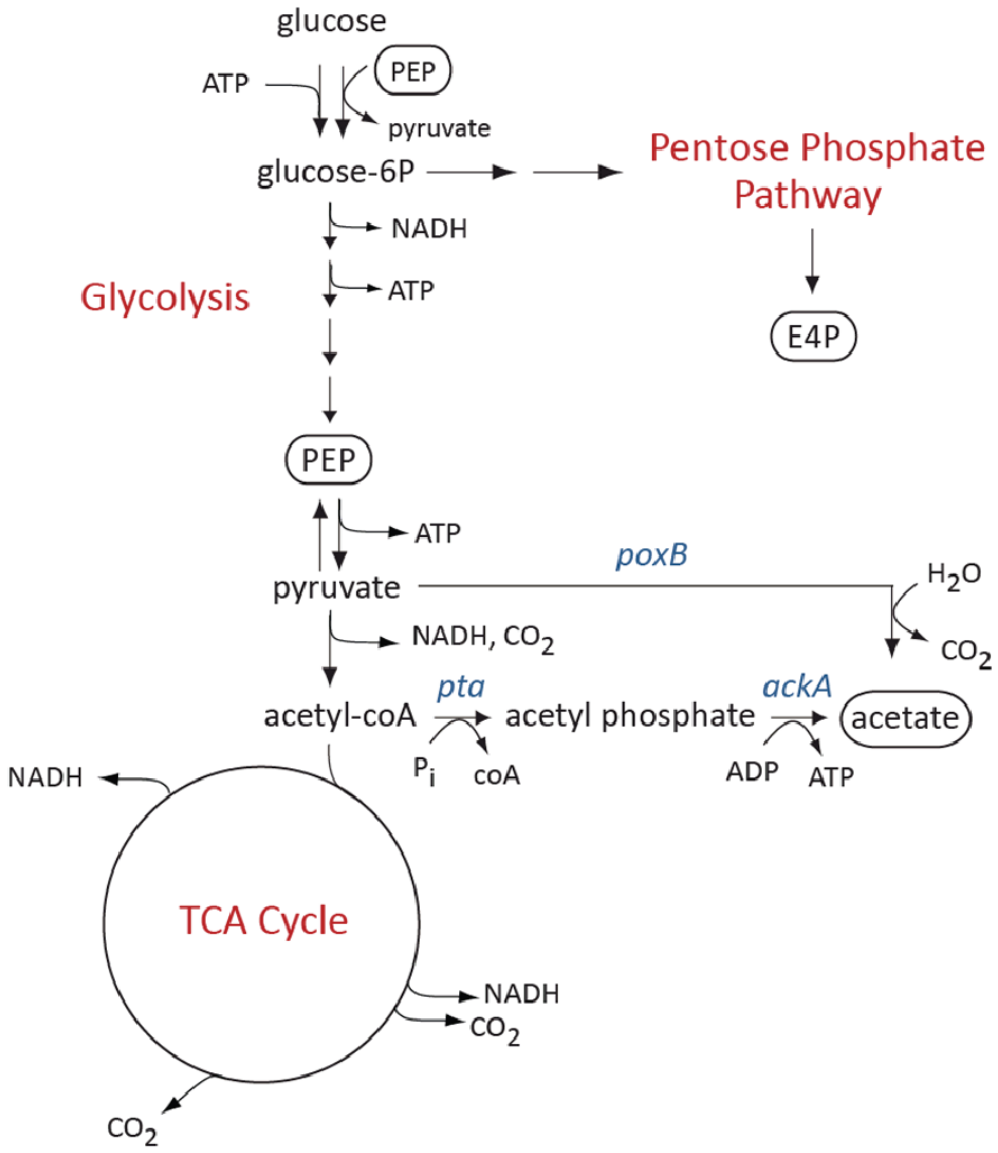


Figure 6.6 Key biochemical pathways in *E. coli* involved in the aerobic consumption of glucose and the synthesis of acetate, carbon dioxide, and biomass

Acetate is generated as a product of "overflow metabolism" when the rate of glycolysis exceeds the capacity of the TCA cycle. The build-up of pyruvate and acetyl-CoA is alleviated by the synthesis of acetate by phosphotransacetylase/acetate kinase (*pta-ackA*) and pyruvate oxidase (*poxB*). (Please note that not all pathways are indicated.)

most common and effective modifications have involved the deletion of the two main biosynthetic routes for acetate production. In this study, we sought to investigate the effects of *ackA-pta* and *poxB* deletion (either singly or in combination) on observed L-tyrosine titers. Based on our understanding of acetate biochemistry and its inhibitory impact on cell viability, we hypothesized that deletion of these genes should lead to increases in L-tyrosine production comparable to those seen in KO-*dnaQ*. Such a result would also provide a simple explanation for why slower growth rates, particularly those that lie below the threshold needed for acetate production, seem to be beneficial for this particular phenotype.

Unfortunately, deletion of these three genes led to unexpected results when tested in two different backgrounds -- parental strain P2 and $dnaQ^{mut2}$. As seen in Figure 6.7, elimination of *poxB* by itself resulted in very little change in L-tyrosine production, presumably because *poxB*-encoded pyruvate oxidase enzyme possesses little to no activity during exponential growth when the bulk of L-tyrosine production is known to occur. However, deletion of the *ackA-pta* operon either by itself or in combination with *poxB* was surprisingly detrimental for this phenotype, with strains exhibiting 45-60% decreases in final L-tyrosine titers. Thus, crude deletions in the acetate biosynthetic pathway are not sufficient for imparting this cellular property, and other metabolic changes must be imposed for the recovery of high L-tyrosine production.

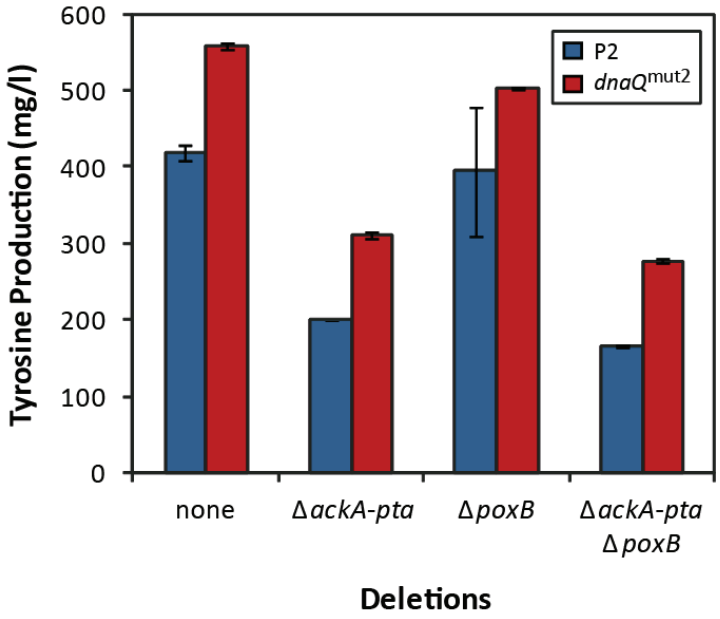


Figure 6.7 Effects of *ackA-pta* and *poxB* deletions on L-tyrosine titers after 48 hr

Single and double deletions of a*ckA-pta* and *poxB* were generated in two genetic backgrounds: 1) parental strain, P2 (blue bars) and 2) *dnaQ*^{mut2} (red bars).

6.4 Conclusions

6.4.1 Identified knockout targets were previously unreachable through rational engineering methods

In this chapter, we explored the use of transposon-mediated combinatorial engineering for the development of L-tyrosine-overproducing strains of *E. coli*. Although this particular technique is only capable of generating single deletion strains, screening of just one library led to the identification of two different mutants exhibiting significant gains in L-tyrosine production. While actual yields and titers did not exceed those of other rationally engineered constructions

(Lütke-Eversloh and Stephanopoulos 2007), this case study shows the utility of this simple combinatorial approach in identifying completely novel targets for metabolic engineering. As an example, the identification of *dnaQ* as an insertion target for engineering L-tyrosine overproduction was wholly unexpected, as its protein product has no apparent connection with any pathways involved with amino acid production. Furthermore, it plays a vital role in an essential global cellular process (DNA replication), making its recovery as a target for mutagenesis all the more surprising.

Despite having conducted additional tests to try and elucidate the functional changes in the *dnaQ* protein, we were unable to develop a conclusive causal mechanism for how this specific genotype may lead to a L-tyrosine overproduction phenotype. This inherent biological complexity, though sometime bewildering and frustrating, once again serves to highlight one of the main advantages of transposon mutagenesis and more generally, combinatorial metabolic engineering: because of the completely global and unbiased nature of this process, one is able to uncover genetic targets that, in all likelihood, would not have even been *considered* under a strictly rational approach.

6.4.2 Phenotype does not require complete Inactivation of gene

Our investigations into the nature of the *dnaQ* mutation led to an interesting revelation regarding the use of transposon mutagenesis for gene deletion. From this example, we discovered that there are unique cases in which *complete* inactivation of a gene is not only unnecessary but also undesirable. As previous studies have verified, deletion of the entire *dnaQ* locus would have negatively altered pol III activity and, as a result, led to inviable

colonies. In marked contrast, transposon integration into just the C-terminal end of this protein had no negative effect on cell viability and instead resulted in significant gains in L-tyrosine production. To add yet another layer of complexity, the desired phenotype was shown to be dependent not just on truncation of the protein but also on the specific residues that were introduced into its C-terminal sequence (Figure 6.3). It is clear from this example that transposon mutagenesis may grant access to an even wider range of phenotypes than previously thought, simply through its ability to integrate at specific locations within a single gene. Although the number of loci exhibiting this degree of sensitivity to the site of integration may be small, this case study still showcases the advantages of harnessing this unique property. In this way, transposon mutagenesis is able to distinguish itself from more traditional and standardized approaches for knockout library generation (Baba, Ara et al. 2006; Zhou, Minami et al. 2010).

6.4.3 Difficulties of inverse metabolic engineering

Despite having formulated a logical connection between acetate production, cell growth, and L-tyrosine production, our attempts to engineer L-tyrosine overproduction by rational means were largely unsuccessful. In this chapter, we therefore see our first introduction into the inherent difficulties of an inverse metabolic engineering approach. Why might have these well-intentioned attempts failed? One possible explanation may be that crude operon deletions could not sufficiently mimic the cellular conditions of our recovered knockout strain, which likely possessed more subtle changes in its cellular makeup. Alternatively, one could also argue the exact opposite: microbes are extremely adaptable organisms and have been known to alter

their own regulation and metabolism to minimize or negate sudden shifts in cellular function. It is quite possible then that the deletion of three biosynthetic genes was not sufficient for altering the cell's metabolism and reallocating resources towards L-tyrosine production. A third and final possibility is that despite a perceived biochemical connection between acetate production and L-tyrosine synthesis, no significant interactions are actually realized within an active cellular network. Regardless of the actual cause for this apparent discrepancy, it is clear that more sophisticated tools must be utilized to gain a better understanding of the underlying changes occurring within a cell. Only then will we be able to formulate and execute effective metabolic engineering strategies for improving L-tyrosine production in these strains.

CHAPTER 7 GLOBAL TRANSCRIPTION MACHINERY ENGINEERING SEARCH

7.1 Introduction

In Chapter 6, we observed that deletion of iHH alone was not sufficient for recovering a high L-tyrosine production phenotype, thus introducing the possibility that an additional unidentified chromosomal mutation may be required. This example gives us an early indication of the potential benefits of introducing genetic modifications concurrently, particularly for eliciting complex phenotypes such as metabolite overproduction. Although greedy search algorithms (sequential target identification) can be used to identify mutations with positive additive effects on cellular phenotype, they are unable to access cellular properties requiring the *simultaneous* modulation of an entire subset of genes. Fortunately, the recently developed technique of global transcription machinery engineering (gTME) lends itself quite nicely for this purpose.

As briefly introduced in Chapter 2, gTME is a unique tool capable of altering the cellular transcriptome on a global level. Such a task is enabled through the targeted mutagenesis (via error prone PCR) of proteins that are known to influence gene transcription by the RNA polymerase holoenzyme (Alper and Stephanopoulos 2007). In this chapter, we focus on the mutagenesis of two subunits of the RNA polymerase - *rpoD*-encoded σ^{70} and the r*poA*-encoded α subunit – for engineering novel cellular properties.

7.2 Structure of RNA polymerase and determinants of specificity

RNA polymerase is a multi-subunit complex that directs and regulates transcription in prokaryotes. The core enzyme, which is comprised of the subunits $\alpha_2\beta\beta'\omega$, is capable of carrying out all stages of transcription except for the process of initiation. This first stage requires binding to one of *E. coli*'s seven sigma factors, each of which is active and preferred during distinct environmental or cellular conditions (Figure 7.1) (Browning and Busby 2004). Despite this specificity, however, all sigma factors proceed with the same mechanism: through their interactions with the -35 and -10 sequences, sigma factors bind and preferentially recruit the RNA polymerase holoenzyme to the promoters of a unique set of genes. It is because of this direct role in promoter recognition and binding that sigma factors were selected as initial targets for transcriptional mutagenesis (Alper and Stephanopoulos 2007; Klein-Marcuschamer and Stephanopoulos 2008). In this study, we build upon this previous work and explore the use of σ^{70} (the main housekeeping sigma factor in *E. coli)* for improving L-tyrosine production in *E. coli*.

Although sigma factors are often recognized as the primary determinants of promoter specificity, the *rpoA*-encoded α subunit of RNA polymerase has also been implicated in promoter recognition. In particular, its C-terminal domain (denoted αCTD) has been shown to interact not only with upstream promoter (UP) enhancer elements of the DNA, but also with several activator and repressor proteins (Figure 7.1) (Ross, Gosink et al. 1993; Gaal, Ross et al. 1996; Browning and Busby 2004; Dangi, Gronenborn et al. 2004). As an added benefit, because the α subunit is part of the core enzyme, it has the potential to change the RNA polymerase's promoter affinities regardless of which of the seven sigma factors is bound (and hence, under

all environmental conditions) (Ishihama 2000; Jishage, Kvint et al. 2002; Browning and Busby 2004). We therefore hypothesized that the α subunit could also be a useful mutagenesis target for engineering L-tyrosine overproduction in our strains.

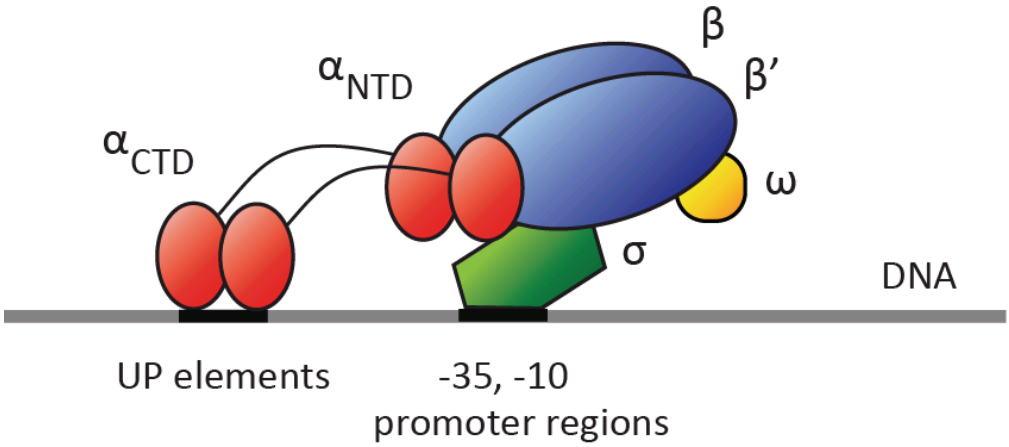


Figure 7.1 Structure and DNA interactions of RNA polymerase

The RNA polymerase holoenzyme is composed of the following subunits: $\alpha_2\beta\beta'\omega$. The C-terminal domain of the α subunit (αCTD) is capable of influencing promoter specificity through its interactions with upstream promoter (UP) enhancer elements and other activator/repressor proteins (not shown). *E. coli*'s seven sigma factors also control transcription through their interactions with the -35 and -10 regions of promoters.

7.3 Materials and Methods

7.3.1 Cultivation conditions

L-tyrosine production experiments were performed at 37°C with 225 rpm orbital shaking in 50 ml MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) cultures supplemented with 5 g/l glucose and an additional 4 g/l NH_4Cl. All liquid cultivations were conducted in at least

triplicates. When appropriate, antibiotics were added in the following concentrations: 34 μg/ml chloramphenicol for maintenance of pHACM-derived plasmids and 100 μg/ml ampicillin for maintenance of pTrc*melA*^{mut1}. IPTG (EMD Chemicals) was added at a concentration of 1 mM for the induction of pTrc*melA*^{mut1}. For L-phenylalanine auxotrophs (Δ*pheA*), L-phenylalanine (Sigma) was supplied at a concentration of 0.35 mM.

7.3.2 Generation and screening of gTME libraries

rpoA and *rpoD* plasmid libraries were generated as described previously (Alper and Stephanopoulos 2007; Klein-Marcuschamer, Santos et al. 2009). Briefly, fragment mutagenesis was performed on wild-type *rpoA* or *rpoD* using the GenemorphII Random Mutagenesis Kit (Stratagene) to induce low (0-4.5 mutations/kb), medium (4.5-9 mutations/kb), and high (9-16 mutations/kb) frequencies of mutation. Following digestion and ligation, preparations of the resulting pHACM-*rpoA/rpoD* plasmid libraries were then transformed into parental strain P2 to generate the *rpoA* and *rpoD* mutant strain libraries. Approximately 7.5 x 10^5 and 3.1 x 10^6 viable colonies (for *rpoA* and *rpoD*, respectively) were screened using the melanin-based high-throughput assay described in Chapter 5 (Santos and Stephanopoulos 2008).

A saturation mutagenesis library for pHACM-*rpoA*14 was constructed using the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene). Primers were designed according to the manufacturer's specifications with degenerate bases substituted into the codon positions corresponding to V257 and L281. This plasmid library was subsequently transformed into a plasmid-cured rpoA14 strain (generated by four rounds of subculturing in LB

medium), and approximately 5000 viable colonies were screened for L-tyrosine production (Santos and Stephanopoulos 2008).

7.3.3 Plasmid curing and retransformation

To investigate the individual contributions of plasmid-based and chromosomal-based mutations, isolated mutant strains were subcultured in LB medium to promote loss of its corresponding pHACM-derived plasmid. Following 3 to 4 rounds of reinoculation, strains were streaked out on LB-agar plates and checked for the loss of chloramphenicol resistance. Routine chemical transformation protocols were utilized in the construction of strains containing a combination of wild-type or mutant backgrounds and plasmids.

7.3.4 Mutational analysis of strains

Mutation frequency analysis protocols were adapted from a previously published report on the use of *rpoB* to analyze the specificity of base substitutions in *E. coli* (Garibyan, Huang et al. 2003). Briefly, cells were inoculated at a starting OD_{600} of 0.01 and grown for 48 hr in 5 ml LB cultures. After appropriate dilution, cultures were plated on LB-agar with and without 100 μg/ml rifampicin (Sigma). (Dilution is necessary to recover distinct colonies as opposed to lawn growth, with an approximate target of less than 200 colonies per plate). The number of colony forming units (CFUs) on each plate was quantified after 16-20 hr, and average mutation frequencies were determined by dividing the average number of rifampicin-resistant colonies by the average number of CFUs growing on unselective media (LB). At least ten culture and plate replicates were used in this calculation for each strain tested.

7.3.5 Analytical methods

For the quantification of L-tyrosine, cell-free culture supernatants were filtered through 0.2 µm PTFE membrane syringe filters (VWR International) and used for HPLC analysis with a Waters 2690 Separations module connected with a Waters 996 Photodiode Array detector set to a wavelength of 278 nm. The samples were separated on a Waters Resolve C18 column with 0.1 % (vol/vol) trifluoroacetic acid (TFA) in water (solvent A) and 0.1 % (vol/vol) TFA in acetonitrile (solvent B) as the mobile phase. The following gradient was used at a flow rate of 1 ml/min: 0 min, 95 % solvent A + 5 % solvent B; 8 min, 20 % solvent A + 80 % solvent B; 10 min, 80 % solvent A + 20 % solvent B; 11 min, 95 % solvent A + 5 % solvent B. Cell densities of cultures for growth rate calculations were determined by measuring their absorbance at 600 nm with an Ultrospec 2100 *pro* UV/Visible spectrophotometer (Amersham Biosciences).

7.4 Results and Discussion

7.4.1 Identification and performance of mutant rpoD3

Because previous studies have demonstrated the effectiveness of modulating σ^{70} for eliciting complex phenotypes (Alper and Stephanopoulos 2007; Klein-Marcuschamer and Stephanopoulos 2008), we began our investigation with the creation of a library of plasmid-expressed *rpoD* variants in the background of parental strain P2. Screening of an *rpoD* mutant library consisting of 3.1×10^6 viable colonies led to the isolation of a strain possessing significant increases (113%) in L-tyrosine production over P2. This mutant, named rpoD3, exhibited final L-titers of 893 mg/l L-tyrosine (Figure 7.2). Final concentrations were measured after 48 hr due

to the slower growth rate of this strain (as alluded to in Chapter 6), although detailed growth and L-tyrosine production curves show that levels remain stable after just 28 hr (Figure 7.3).

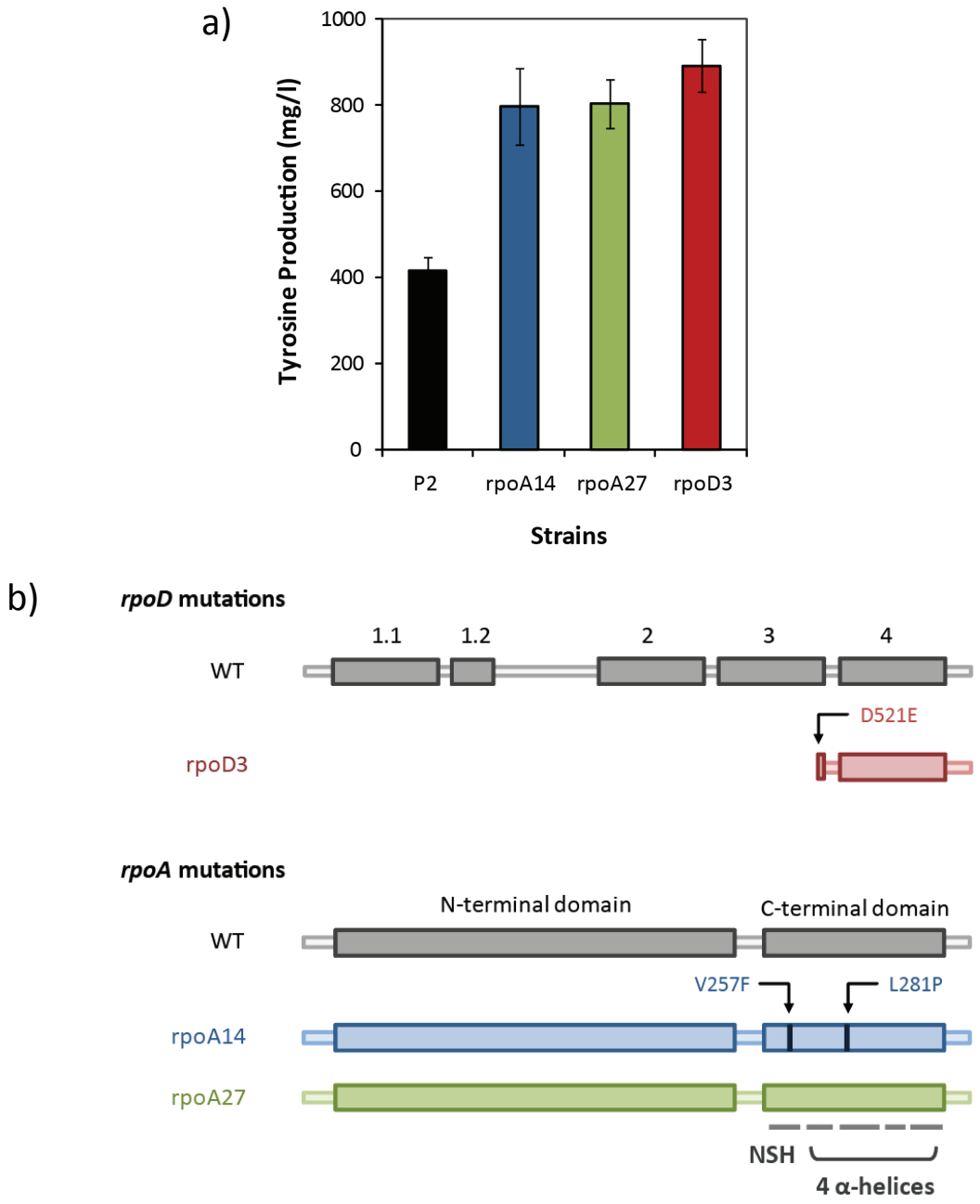


Figure 7.2 Properties of mutants isolated from gTME libraries

a) L-tyrosine titers of isolated mutants after 48 hr b) Sequence analysis of plasmid-encoded mutant *rpoA* and *rpoD*. Numbers above wild-type *rpoD* indicate conserved regions of the protein. NSH stands for the "non-standard helix" portion of the α subunit C-terminal domain.

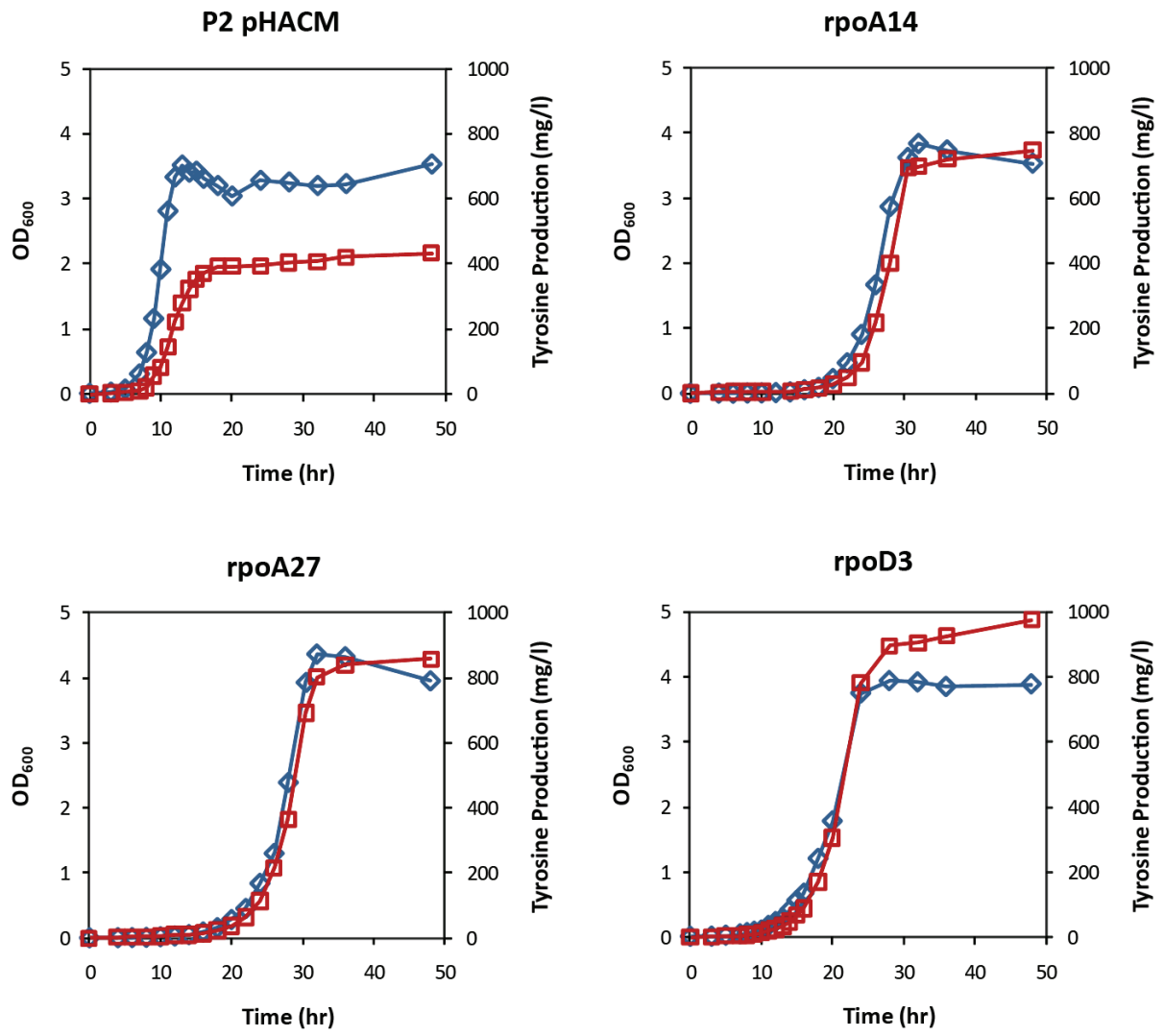


Figure 7.3 Representative growth and L-tyrosine production curves for gTME mutants

OD_{600} measurements (◇) and L-tyrosine production profiles (□) for parental P2 with pHACM (an empty vector) and three gTME mutants – rpoA14, rpoA27, and rpoD3. Specific growth rates for these strains were: 0.617 hr^{-1} (P2 pHACM), 0.296 hr^{-1} (rpoA14), 0.249 hr^{-1} (rpoA27), and 0.290 hr^{-1} (rpoD3).

7.4.2 Sequencing of pHACM-*rpoD*3 and structural analysis of σ^{70}

The σ^{70} protein contains four conserved domains (region 1-4) that play specific roles in mediating binding to RNA polymerase and to DNA. Interestingly, a sequence analysis of the plasmid-encoded *rpoD* from mutant rpoD3 revealed a truncated form of this protein consisting of just region 4 and the tail end of region 3 (Figure 7.2). Because *E. coli*'s RNA polymerase is a well-studied enzyme, much is actually known about the functional role of σ^{70}'s fourth conserved domain. Region 4 is made up of four helices which interact with the -35 region of the promoter and additionally contains a major RNA polymerase binding determinant. More recently, this region has also been implicated in the binding of anti-σ factors such as Rsd, which are expressed during stationary phase (Gruber and Gross 2003). Anti-σ factors play an important role in regulating the metabolic shift from exponential growth to stationary phase by binding to σ^{70} and subsequently blocking its association with the RNA polymerase. It is through this indirect way that anti-σ factors are able to contribute to the preferential use of stationary phase σ factor σ^{S} to mediate the transcription of a novel set of genes (Gruber and Gross 2003; Sharma and Chatterji 2008).

How then might overexpression of a separate Region 4 domain lead to changes in L-tyrosine production within this mutant strain? One effect may be closely related to its efficient ability to bind Rsd. Indeed, truncated sigma factors have been shown to possess higher binding affinities to anti-σ factors relative to the full length protein (Sharma, Ravishankar et al. 1999). We hypothesize then that overexpression of a truncated protein could, in essence, titrate high amounts of Rsd from the cytosol, leaving a smaller number of molecules available to bind the chromosomally-expressed and fully functional wild-type form. Fewer Rsd-$\sigma^{70,WT}$ binding events

would delay the onset of stationary phase and may ultimately lengthen the exponential growth period during which L-tyrosine is produced.

As an alternative hypothesis, the truncated σ^{70} may also affect L-tyrosine production through its association with the RNA polymerase core and the subsequent assembly of nonfunctional enzymes. Even a small fraction of unproductive RNA polymerases could ultimately lead to the slower growth rate observed in these cells. As suggested by previous discussions and data (Figure 6.4, Chapter 6), this reduced growth may ultimately prove to be beneficial for L-tyrosine production. It is important to note that the points we have raised here are simply conjectures; clearly, both theories must be either validated or refuted with additional functional characterizations and tests.

7.4.3 Identification and performance of mutants rpoA14 and rpoA27

Buoyed by these successes, we then turned our attention towards applying the same transcriptional engineering paradigm using a novel genetic target – the RNA polymerase α subunit. In order to properly characterize the potential of this approach, *rpoA* mutagenesis was evaluated with respect to three different cellular phenotypes – butanol tolerance, hyaluronic acid production, and L-tyrosine production. Specific details regarding the engineering of the first two properties can be found in a recent publication from our lab (Klein-Marcuschamer, Santos et al. 2009). In these next sections, we discuss the efficacy of an *rpoA*-mediated transcriptional engineering approach for optimizing the third trait, the overproduction of L-tyrosine. L-tyrosine synthesis in *E. coli* is a natural platform for which to test this novel method,

particularly since transcriptional modifications have already proven to be an effective strategy for eliciting this phenotype.

A plasmid-encoded *rpoA* mutant library was generated by error-prone PCR (Klein-Marcuschamer, Santos et al. 2009) and transformed into parental strain P2. Screening and selection from an initial pool of $7.5x10^5$ mutants led to the identification of two notable strains (denoted rpoA14 and rpoA27) with a 91-93% increase in L-tyrosine titer compared to the parental. Final concentrations of 798 and 806 mg/l L-tyrosine (for rpoA14 and rpoA27, respectively) were reached after 48 hr (Figure 7.2), although as mentioned in Chapter 6, these strains also exhibited a slightly reduced growth rate (Figure 7.3).

7.4.4 Sequence analysis of pHACM-*rpoA*14 and pHACM-*rpoA*27

Sequencing of the mutant plasmid pHACM-*rpoA*14 revealed the presence of two amino acid substitutions within the αCTD in close proximity to residues involved with regulatory factor and UP element interactions (Murakami, Fujita et al. 1996). The first change (V257F) occurred in the so-called "non-standard helix", while the other (L281P) was located in one of the four α-helices of the αCTD (Figure 7.2) (Gaal, Ross et al. 1996). Because both substitutions reside in secondary structural elements of the protein, it is likely that these mutations have altered the overall αCTD structure and, as a result, its interactions with target proteins and sequences. In particular, the amino acid proline has been shown to lead to α-helix destabilization under certain conditions (Li, Goto et al. 1996). A change in helix conformation could change the relative alignment of amino acids that make contacts with promoters, thus altering the affinity of the RNA polymerase for some of its targets.

Surprisingly, no changes were found in the sequence of pHACM-*rpoA*27. The recovery of a wild-type plasmid in rpoA27 was quite unexpected, particularly since the overexpression of *rpoA* by itself does not seem to confer additional advantages in the parental background (Figure 7.4). As with KO-iHH, it is possible that other mutations within the chromosome may be either partial or full contributors to the observed phenotype.

7.4.5 High L-tyrosine titers are dependent on both plasmid and background

Because a wild-type plasmid was isolated from the rpoA27 mutant, it seems likely that at least one additional background mutation may be needed to adequately boost the strain's capacity for L-tyrosine production, a phenomenon that we had also observed for KO-iHH (Chapter 6). To determine whether similar conditions were required for the other two gTME mutants (rpoA14 and rpoD3), we conducted a series of tests to verify the specific genetic components (or combinations thereof) that serve as the main determinants of cellular phenotype.

We first tested whether plasmid transfer into the clean genetic background of parental strain P2 could recover the high L-tyrosine titers observed in the isolated mutants. As we might have suspected, P2 containing pHACM-*rpoA*27 (which has 100% sequence identity as pHACM-rpoA^{WT}) exhibited very similar properties as P2 by itself or with the wild-type *rpoA* control plasmid. Unexpectedly, however, the same trend held true for both pHACM-*rpoA*14 and pHACM-*rpoD*3, plainly indicating that all three recovered strains require additional factors beyond the mutant plasmids in order to achieve high levels of L-tyrosine production (Figure 7.4, 2nd bar within each set). Specifically, alterations located within their respective bacterial chromosomes must offer significant contributions to these dramatic shifts

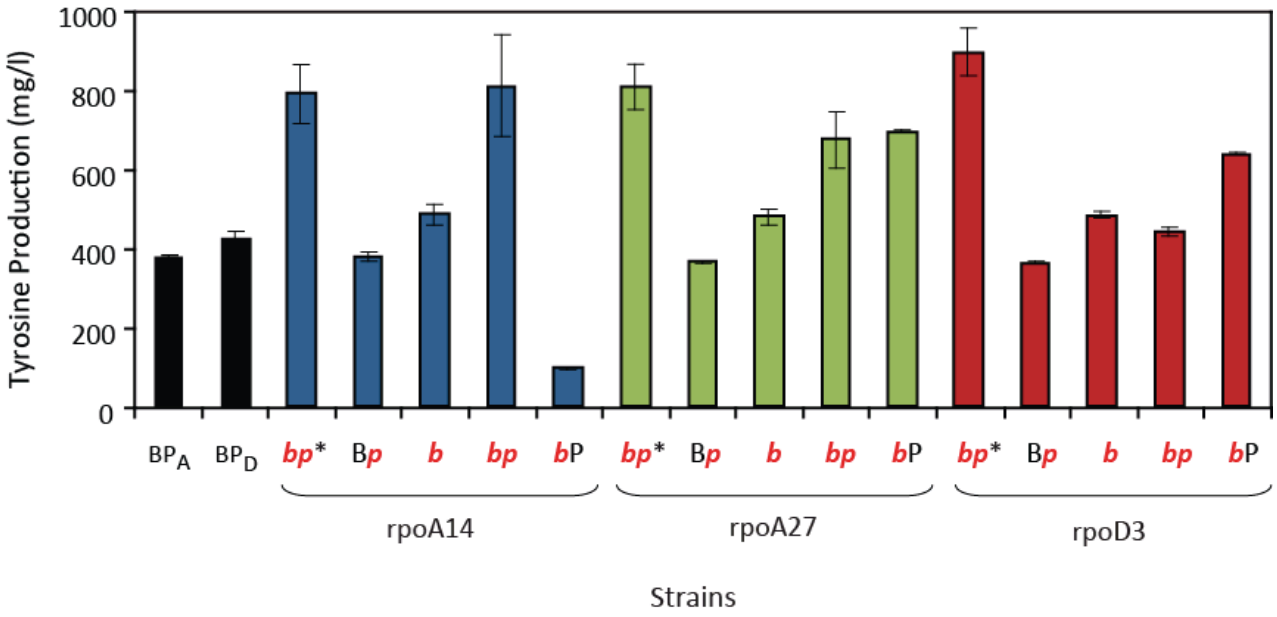


Figure 7.4 Both a mutant *rpoA*/*rpoD* plasmid and a mutant background are required to impart a L-tyrosine overproduction phenotype

B = background, P = plasmid; Wild-type versions of the background or plasmid are denoted by black uppercase letters. Mutant versions are denoted by red lowercase letters. P_A and P_D represent the wild-type *rpoA* and *rpoD* plasmids, respectively. bp* represent the original strains isolated from the mutant libraries.

in cellular states. To determine whether these unknown background mutations could single-handedly elicit the desired phenotypic effects, we cured the three strains of their plasmids to facilitate the recovery of an isolated mutant chromosome. As shown in Figure 7.4 (3rd bar within each set), these plasmid-free strains also possessed a clear deficit in phenotype associated with the loss of their respective mutant plasmids.

Taking these results altogether, it becomes readily apparent that high L-tyrosine titers in these strains are uniquely dependent on *both* a mutated pHACM-*rpoA*/*rpoD* plasmid *and* a mutant background. Indeed, we found that retransforming pHACM-*rpoA*14 and pHACM-*rpoA*27 back into the previously cured strains resulted in complete recovery of the desired

phenotype (Figure 7.4, 4th set of bars within each group). Unusually, this pattern did not hold for rpoD3, with the reconstructed strain exhibiting L-tyrosine titers as low as the parental. This discrepancy may have resulted from additional chromosomal mutational events occurring during the plasmid curing step, which required at least three rounds of subculturing in LB. However, this process was repeated several times with similar results (data not shown), suggesting that rpoD3 may be inherently unstable with regards to the transfer of phenotype. The reason for this instability remains unclear.

As a final control, we also verified that the wild-type *rpoA*/*rpoD* plasmid could not supplant the roles of their mutant counterparts. As expected, strains comprised of a mutant rpoA14 or rpoD3 background with a wild-type plasmid exhibited reduced L-tyrosine production levels compared to the isolated mutants (Figure 7.4, last bar within each group). The effect was particularly striking for rpoA14 which produced L-tyrosine at less than 100 mg/l, indicating a substantial need for the mutant version of the plasmid. The only exception was seen for rpoA27, a result that still makes intuitive sense given that a wild-type sequence was recovered from this particular strain (Figure 7.2).

7.4.6 Overexpression of *rpoA*/*rpoD* does not induce a mutator phenotype

Because an unusually high incidence of chromosomal variation was found within these strains, we decided to measure their mutation frequencies to determine if elevated values could account for the observed discrepancies. To quantify this trait, we utilized the same rifampicin-based mutagenesis assay described in Chapter 6. However, a longer initial incubation period in LB was incorporated (48 hr) to more closely mimic the nutrient-limited conditions that were

likely present during plate selection and screening. For most strains tested, the addition of a wild-type or mutant *rpoA*/*rpoD* plasmid did not result in any increases in mutation frequency when compared to the parental strain P2. The only exception was seen for the rpoA27 mutant, which had a higher observed frequency of about 16 mutants/10^6 cells, a value comparable to that of the *dnaQ*^{mut1} strain introduced in Chapter 6 (Figure 7.5). This slightly elevated may have been necessary to elicit a high-production phenotype for this mutant, particularly since this strain was found to possess only a wild-type allele of *rpoA* on its plasmid.

Altogether, these findings clearly indicate that a high mutator phenotype was largely absent in these isolated strains. As such, it becomes reasonable to assume that a high rate of chromosomal variation must have arisen out of a specific requirement for these background mutations to generate the desired phenotype. Indeed, if these changes are in fact essential for enhanced L-tyrosine synthesis, then such genomic alterations, no matter how "rare" an occurrence within the library, would undoubtedly be detected and selected by the high-throughput screen.

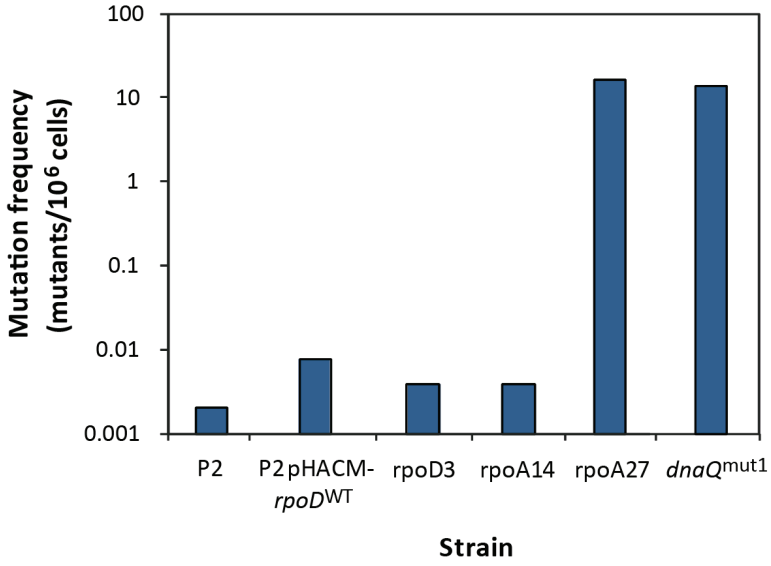


Figure 7.5 Mutation frequencies of select strains

Mutation frequencies were measured after 48 hr growth in 5 ml LB. Values for *dnaQ*^{mut1} were included as a positive control.

7.4.7 Saturation mutagenesis of pHACM-*rpoA*14 for improved productivity

Although significant gains in L-tyrosine titer were already achieved through both *rpoD* and *rpoA* mutagenesis, we were interested in determining whether additional rounds of transcriptional engineering could be further beneficial for this phenotype. However, rather than subjecting the entire locus to error prone PCR as before, we instead restricted our mutagenesis to the two substituted residues found in pHACM-*rpoA*14. Given their locations within secondary structural elements of the αCTD, we hypothesized that such changes may be important determinants of promoter specificity; hence, we were interested in exploring whether other amino acid substitutions within these two sites, V257 and L281, could further enhance L-tyrosine production in these strains.

A saturation mutagenesis library was constructed for pHACM-*rpoA*14 through the introduction of degenerate bases into the codon positions of V257 and L281. The resulting plasmids were then transformed into the cured, plasmid-free version of rpoA14 and screened as before. Our search through a pool of about 5000 viable colonies led to the identification of a mutant strain, denoted rpoA22-sm, that was found to contain a V257R substitution in its α subunit protein sequence. Although this isolate exhibited only a modest increase in L-tyrosine titer compared to mutant rpoA14 (up 20% to 962 mg/l), this was, quite surprisingly, accompanied by a faster rate of L-tyrosine production (Figure 7.6). In fact, the recovered mutant was able to produce 72% (~700 mg/l) of the total amount of L-tyrosine after 24 hr of incubation. By comparison, rpoA14 titers during this time point were so low (~330 mg/l) that they remained largely indistinguishable from those of the parental strain P2. The results were even more striking when pHACM-*rpoA*22-sm was retransformed into a clean rpoA14

background, with this reconstructed strain reaching nearly 90% (~806 mg/l) of its final titer after just 24 hr. These notable differences in productivity are likely a result of a more robust strain constitution, as these strains exhibited a 15% higher growth rate than the library parental, rpoA14 (0.341 hr^{-1} versus 0.296 hr^{-1}). Altogether, the enhanced growth and productivity of mutant rpoA22-sm make it a very interesting candidate for the development of a continuous process for L-tyrosine production.

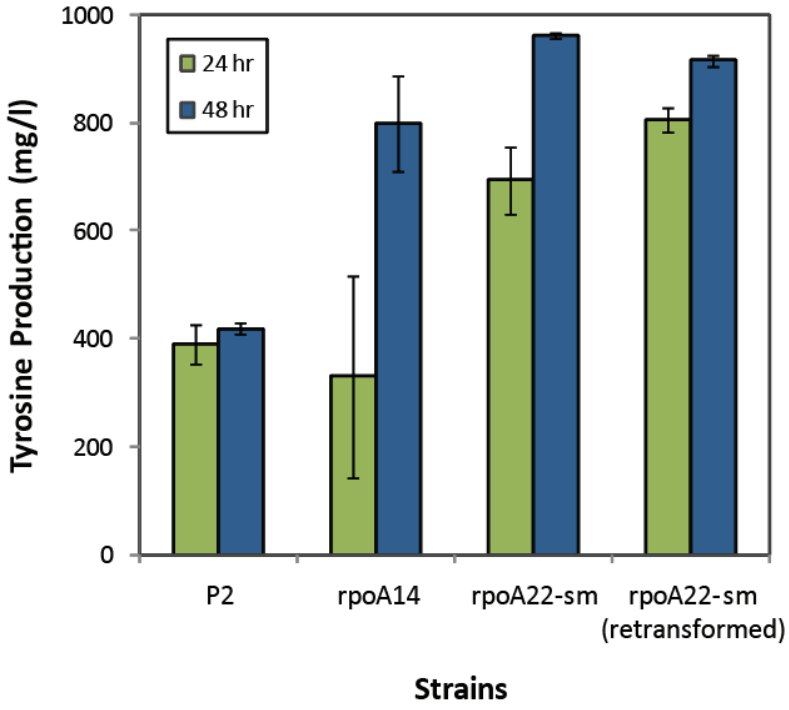


Figure 7.6 Performance of strains isolated from a pHACM-*rpoA*14 saturation mutagenesis library

L-tyrosine titers are shown after 24 hr (green bars) and 48 hr (blue bars). Bars labeled as rpoA22-sm represent the titers for the original isolate from the saturation mutagenesis library. rpoA22-sm (retransformed) refers to a strain in which the pHACM-*rpoA*22-sm plasmid was individually isolated and subsequently transformed into a plasmid-free rpoA14 background.

7.4.8 Mechanistic insights into *rpoA* activity

Although our analysis of the *rpoA* saturation mutagenesis library focused largely on the characterization of rpoA22-sm, it was interesting to note that almost all mutants subjected to the final round of screening (in 50 ml shake flasks) exhibited titers that were either comparable or higher than the library parental, rpoA14 (data not shown). This unusually high percentage of hits was incongruous with our experiences with previous libraries, suggesting that the specific identities of the amino acid substitutions actually play a limited role in determining phenotype. This then implies that a more general, nonspecific mechanism for *rpoA* activity may be responsible for the observed gains in L-tyrosine production. One plausible hypothesis is that helical destabilization in αCTD mediated by substitutions at V257 and/or L281 may lead to a misfolding of the C-terminal domain. This generic, partial loss of function in *rpoA* may be all that is required to impart a L-tyrosine overproduction phenotype.

This notion of *rpoA* nonspecificity has been recently supported by two other studies that have utilized *rpoA* transcriptional engineering for optimizing cellular properties (Klein-Marcuschamer, Santos et al. 2009). In the first, an engineered strain capable of high levels of hyaluronic acid production (rpoA-HA) was recovered from a library and shown to possess a L254Q variant of *rpoA* (Figure 7.7a). As with rpoA14's V257F mutation, this amino acid substitution was mapped to the non-standard helix of αCTD and likely altered or disrupted the local structure of this domain. An even less subtle mutation was found in an *E. coli* strain developed and screened for high butanol tolerance (L33). Surprisingly, sequence analysis revealed that this strain had acquired an *rpoA* variant devoid of almost its entire C-terminal

domain (Figure 7.7a). This severe protein truncation clearly reinforces the idea that a partial or complete loss of αCTD function may be needed to elicit the selected phenotypes.

Given the strong support for a nonspecific *rpoA* mechanism, we then wondered whether all three *rpoA* plasmids could lead to the same transcriptional and metabolic outcomes within a cell. If such were the case, these plasmids should prove to be completely interchangeable; that is, a plasmid that was selected for one cellular property should, in the appropriate strain background, also prove to be effective in eliciting a second, unrelated phenotype. We decided to test this interesting possibility by attempting to recover high L-tyrosine titers using plasmids derived from both the butanol-tolerant strain L33 and the hyaluronic acid producer rpoA-HA (denoted pL33 and pHA, respectively). Transformation of these two plasmids into a cured rpoA14 strain led to some intriguing results. As shown in Figure 7.7b, both plasmids led to L-tyrosine titers that were either comparable or slightly higher than those seen for rpoA14. More impressively, however, both pL33 and pHA were able to elicit significantly faster rates of L-tyrosine production from these strains. Their performance was similar to that seen in rpoA22-sm, with almost all L-tyrosine produced after just 24 hr. These results clearly support the notion that functional shifts or deficiencies in αCTD are sufficient for redirecting substrate flow into the L-tyrosine pathway. However, because slight differences were still seen with regards to L-tyrosine productivity, it appears that specific sequence alterations do still play a subtle role in influencing cellular function.

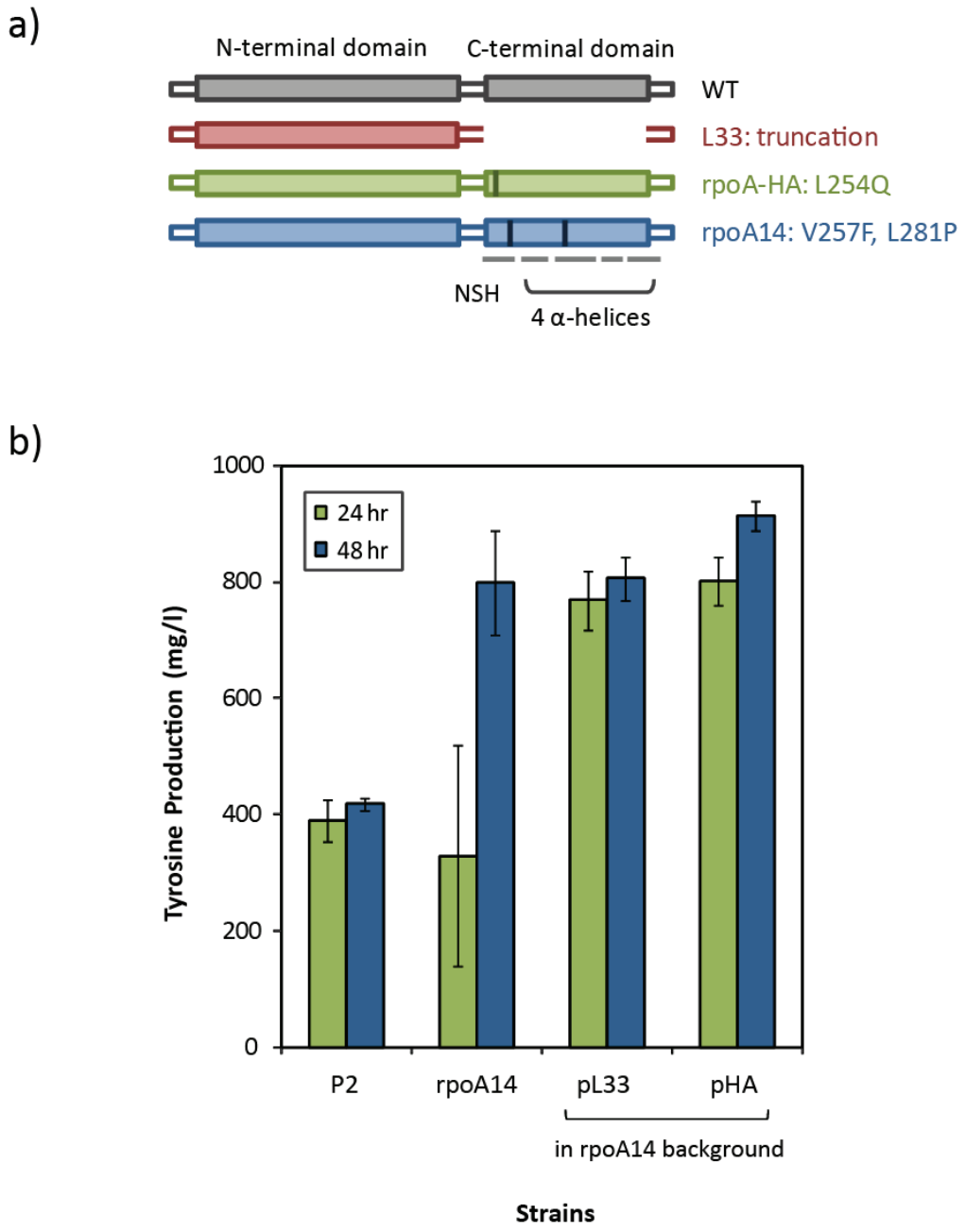


Figure 7.7 Performance of cured rpoA14 with αCTD-mutated *rpoA* variants

a) Schematic of mutant *rpoA* proteins recovered from strains optimized for three separate phenotypes. L33 - butanol tolerance; HA - hyaluronic acid production; rpoA14 - L-tyrosine production. b) Plasmids from the mutant strains L33 and rpoA-HA were isolated and transformed into a cured rpoA14 background. L-tyrosine titers were then measured after 24 hr (green bars) and 48 hr (blue bars). Production of L-tyrosine by parentals P2 and rpoA14 were also included as a reference.

7.4.9 Plasmid and background synergisms are nonspecific

Because the previously decribed plasmid interchangeability experiment was limited to only a single mutant genome (cured rpoA14), we broadened our analysis to examine the effects of these plasmids in other strain backgrounds. Of note, our interests extended beyond just the *rpoA* library, as we also included pHACM-*rpoD*3 and the rpoD3 background in these follow-up studies. Although we are well aware that *rpoA*-encoded α and *rpoD*-encoded σ^{70} possess discrete functions within a cell, we were interested in probing whether this plasmid phenomenon would hold not just between phenotypic groupings but across target/library boundaries as well.

We first constructed a comprehensive set of strains encompassing all possible combinations of the mutant plasmids (pL33, pHA, pHACM-*rpoA*14, pHACM-*rpoA*27, pHACM-*rpoD*3) and cured backgrounds (rpoA14, rpoA27, rpoD3). Strains were then cultivated in MOPS minimal medium and characterized for L-tyrosine production. The results were once again quite startling. As shown in Figure 7.8, *all rpoA* and *rpoD* plasmids were successful in recovering high L-tyrosine titers regardless of the mutant background in which they were present. Although levels were generally lower in a cured rpoA27 strain, all plasmid-background combinations still yielded titers that either closely approached or even exceeded the values seen for our original isolates. Thus, although contrary to our expectations, this experiment reveals that these *rpoA* and *rpoD* plasmids possess the property of interchangeability despite having been selected from completely separate transcriptional libraries and, in some cases, for different phenotypes.

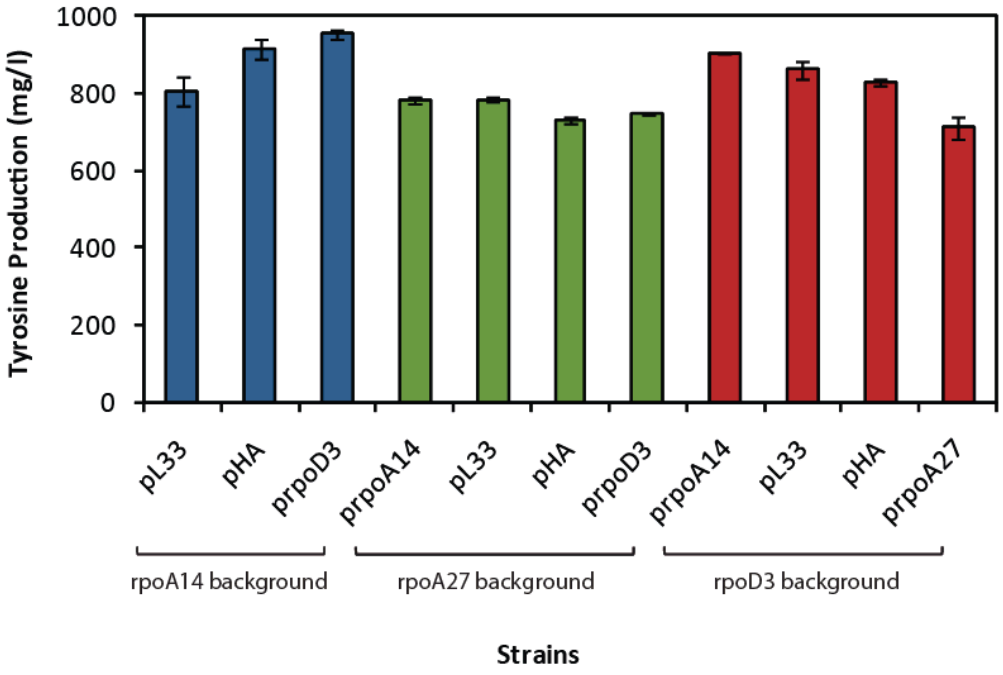


Figure 7.8 Interchangeability of backgrounds and plasmids

L-tyrosine production of various background-plasmid combinations were measured after 48 hr. Blue, green, and red bars represent strains with an rpoA14, rpoA27, or rpoD3 cured background, respectively. Labels preceded by a "p" indicate the strain from which the mutant *rpoA* or *rpoD* plasmid was isolated. Results for p*rpoA*27 (equivalent to pHACM-*rpoA*^{WT}) in an rpoA14 background were not included, as this has previously been shown to be detrimental to phenotype (Figure 7.4).

While this phenomenon remains somewhat of a mystery, one can envision two separate scenarios that could lead to these seemingly unlikely results. The first possibility, as suggested by our analysis of the *rpoA* αCTD, is that all tested plasmids may be working through a very generic mode of action, therefore allowing these constructs to elicit similar phenotypes in all three backgrounds. However, additional support for this hypothesis is clearly still lacking, particularly since interchangeability between two functionally distinct proteins (α subunit, σ^{70}) was also observed. In addition, such a theory does not explain why a wild-type *rpoA* plasmid

(pHACM-*rpoA*27), which contains no deficiencies in its C-terminal domain, could also produce the same results in an rpoD3 background (Figure 7.8). The second possible explanation is a much simpler one which assumes that the same chromosomal aberration occurred in all three backgrounds. However, this scenario seems highly unlikely when one considers the number of possible base pairs substitutions that can occur during any given mutational event (4.6 million). The recovery of the exact same mutation within three distinct strains from two separate libraries then becomes virtually impossible. Given the serious flaws in both conjectures, it is clear that a more in-depth characterization is needed to elucidate the complex interplay between the mutant *rpoA*/*rpoD* plasmids and these strains' altered genomes.

7.5 Conclusions

7.5.1 Success through transcriptional engineering

In this chapter, we have demonstrated the clear potential of using a transcriptional engineering approach for optimizing a L-tyrosine production phenotype. The screening of two libraries constructed through the mutagenesis of *rpoA* and *rpoD* led to significant increases in L-tyrosine production in three strains, which exhibited final titers of 798-893 mg/l L-tyrosine. Remarkably, this more than doubles the previous levels observed in the parental P2, and, to our knowledge, is the highest titer reported for 50 ml *E. coli* cultivations in minimal media. These titers correspond to an overall yield of 0.16-0.18 g L-tyrosine/g glucose, a value that exceeds the best previously reported strain by 44% (Lütke-Eversloh and Stephanopoulos 2007).

It is important to note that the capabilities of these gTME-derived strains far surpass not only those of rationally engineered strain constructions but also those of the random knockout mutants reported in Chapter 6. This ability to reach previously inaccessible phenotypes therefore sets gTME apart, even from other promising and more established combinatorial engineering methods. These advantages likely stem from its unique capacity to reprogram the transcriptome, alter the entire cellular milieu, and as a result, generate an enormous amount of phenotypic diversity. As validated by these experiments, sampling of a wider range of phenotypic space greatly increased our chances of recovering desirable transcriptional mutants.

7.5.2 Limitations of gTME

Despite these successes, however, gTME is unfortunately not without limitations. When we attempted to recover improved mutants through a second round of transcriptional engineering, we found that although productivity was heightened, fewer gains were seen with respect to overall yields and titers. While the absence of a significantly enhanced mutant may have been influenced by our site-restricted saturation mutagenesis approach, we also cannot discount the possibility that an upper limit for this method may have been reached. The best performing mutant from this second round exhibited a titer of 962 mg/l L-tyrosine and a yield of 0.19 g L-tyrosine/g glucose, which represents only 35% of the maximum theoretical value (or 73% when accounting for glucose diverted towards biomass formation). Because there is clearly still additional room for improvement, it may be worthwhile to explore the use of complementary, orthogonal methods for further improving L-tyrosine production in these strains. In addition, process scale-up and optimization may also be required to enhance strain performance as has

been demonstrated in other recent studies (Lütke-Eversloh and Stephanopoulos 2007; Patnaik, Zolandz et al. 2008).

7.5.3 Complications with inverse metabolic engineering

Unfortunately, transcriptional modification by gTME, while certainly a boon for *recovering* desirable mutants, becomes somewhat of a disadvantage when applied towards an inverse metabolic engineering paradigm. Due to the sheer number of changes introduced within the cell, it becomes nearly impossible to pinpoint which specific alterations are responsible for the desired phenotype. This complexity ultimately limits our ability to understand the underlying genotype to phenotype biochemistry, and consequently provides very little insight for the formulation of novel, directed strategies for engineering future generations of strains.

It is important to note that our inability to probe these causal mechanisms does not necessarily limit transferability of the desired phenotype into a novel strain. When recipient and donor strains are sufficiently similar, this task can be easily accomplished simply by transforming the plasmid-encoded transcriptional unit into the new background. Unfortunately, the discovery of chromosomal mutations that appear to be causally linked to L-tyrosine production makes this a much more difficult endeavor. Indeed, this unusual requirement forces us to conduct a full characterization of the isolated strains in order to identify a suitable method for imparting this desired cellular property. In the next chapter, we discuss several approaches for investigating the genetic and biochemical changes that have occurred within these strains and, ambitiously, attempt to utilize this information to generate a novel set of strains capable of L-tyrosine overproduction.

CHAPTER 8 ANALYSIS OF TRANSCRIPTIONAL ENGINEERING MUTANTS

8.1 Introduction

As discussed in the previous chapter, transcriptional engineering of both σ^{70} and the α subunit of RNA polymerase led to the isolation of several strains possessing an enhanced capacity for L-tyrosine production. Unfortunately, however, we discovered early on that this desirable property could not be imparted simply through plasmid transfer but rather, required the contributions of unidentified chromosomal mutations within the strains. In this chapter, we discuss our attempts to fully characterize both the genetics and the underlying biochemistry needed to impart these mutant phenotypes. In doing so, we hope to a) gain a more complete understanding of the factors that influence L-tyrosine production and b) use this information to formulate novel strategies for engineering these properties into future production strains.

8.2 Materials and Methods

8.2.1 Transcriptional Analysis

Strains P2, rpoA14, rpoA27, and rpoD3 were grown in 50 ml MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) to an OD_{600} of approximately 0.4. Triplicates samples of RNA (on three separate days) were then extracted using the QIAGEN RNeasy Mini Kit according to the manufacturer's protocol. We enlisted the microarray services of the David Koch Institute for

Integrative Cancer Research Microarray Technologies Core Facility to run extracted samples on GeneChip *E. coli* Genome 2.0 Arrays (Affymetrix). To allow for statistical confidence in differential gene expression, these arrays were run in triplicates with biological replicates. The resulting microarray data was deposited in the Gene Expression Omnibus database under accession number GSE21652, and a complete MIAME checklist is provided in APPENDIX A. Expression profile deviations between the selected mutants and the parental were identified by a Significance Analysis of Microarrays (SAM) analysis using SAM 3.0 (Tusher, Tibshirani et al. 2001).

8.2.2 Creation of pZE overexpression plasmids

E. coli K12 genomic DNA was prepared using the Wizard Genomic DNA Purification Kit (Promega) and used as a template for the synthesis of genes *gadE*, *ydeO*, *evgA*, and *relA* with Phusion DNA polymerase (New England Biolabs). All PCR fragments were digested with KpnI and BsaI or MluI and subsequently ligated to similarly digested pZE-*gfp* plasmids. To enable testing of both low- (II, AA, Y) and high- (P_L) level expression, two promoters from a synthetic library were used in the creation of each gene construct (Table 8.1) (Alper, Fischer et al. 2005). Following transformation into chemically competent *E. coli* DH5α (Invitrogen), plasmids were isolated and verified by both PCR and sequencing. All enzymes used in the cloning procedure were purchased from New England Biolabs. The names and sequences of all primers used for gene amplification are listed in Table 8.2.

Table 8.1 Relative strengths of four synthetic promoters

Promoter	Average promoter strength metric (0 to 1)[a]
II	0.04
AA	0.22
Y	0.31
P_L	0.87

[a] data provided by C. Fischer

8.2.3 Whole genome sequencing and SNP detection

Whole genome sequencing for strains P2, rpoA14, rpoA27, and rpoD3 was performed using the Illumina/Solexa Genome Analyzer System. Briefly, genomic DNA from all four strains was extracted using the Wizard Genomic DNA Purification Kit (Promega). Samples were then fragmented and prepared for paired-end sequencing using the Paired-End DNA Sample Prep Kit (Illumina). Samples were analyzed and processed at the MIT Biopolymers Laboratory, with duplicate lanes used for each strain. Sequence alignment and analysis, including the detection of insertions, deletions, and SNPs, was performed by the David Koch Institute Bioinformatics Facility. All sequence deviations from P2 were additionally validated by Sanger sequencing. Raw data files are available at the National Center for Biotechnology Information Sequence Read Archive under accession number SRA012672.14.

8.2.4 Reengineering of SNPs into the bacterial chromosome

In order to reintroduce verified SNPs back into a P2 background, we utilized a two-step lambda-red recombination based method (Datsenko and Wanner 2000) to delete the relevant locus and subsequently replace it with a SNP-substituted variant. Two knockout cassettes for the *purF*

and *hisH* genes were generated by amplification of pKD13's *kan*^{FRT} region using Taq DNA polymerase. Primers CS707 pKD13 kan-purF sense and CS708 pKD13 kan-purF anti were used for the generation of *purF*::*kan* integration cassettes, and CS709 pKD13 kan-hisH sense and CS710 pKD13 kan-hisH anti were used to create the *hisH*::*kan* fragment (Table 8.2). All four primers incorporated 50-58 bp of homology with the ends of their respective targets to facilitate integration into the proper locus. These cassettes were transformed into P2 pJM12, and transformants were selected on LB-kanamycin plates. Proper integration was verified by colony PCR and sequencing.

Cassettes for the second round of integration were generated through the amplification of SNP-containing *purF* and *hisH* regions using primers CS699 purF sense, CS700 purF anti, CS701 hisH sense, and CS702 hisH anti (Table 8.2), as well as genomic DNA preparations from mutants rpoA14, rpoA27, and rpoD3 as templates. SNP variants were then transformed into either P2 *purF*::*kan*^{FRT} pJM12 or P2 *hisH*::*kan*^{FRT} pJM12 and grown overnight in 5 ml M9 minimal medium (Sambrook, Fritsch et al. 1989) with 5 g/l glucose. Correct transformants were selected using this alternate method due to the lack of an antibiotic marker for monitoring the second integration event. However, because *purF* and *hisH* deletion strains exhibit much slower growth rates in minimal medium than their SNP-containing counterparts, this selection process resulted in the isolation of correct transformants after just a single round of overnight growth in minimal medium. Individual colonies were tested for the loss of kanamycin resistance (indicating replacement of *kan*^{FRT} with a mutated *purF* or *hisH*) and later validated by colony PCR and sequencing.

Table 8.2 Primers used in this study

Primer Name	Primer Sequence (5'→ 3')
CS555 evgA sense KpnI	CTC GGT ACC ATG AAC GCA ATA ATT ATT GAT GAC CAT CC
CS556 evgA anti MluI	CGA CGC GT T TAG CCG ATT TTG TTA CGT TGT GCG
CS557 ydeO sense KpnI	CTC GGT ACC ATG TCG CTC GTT TGT TCT GTT ATA TTT ATT C
CS558 ydeO anti BsaI	GGT CTC TCT TTT CAA ATA GCT AAA GCA TTC ATC GTG TTG C
CS559 gadE sense KpnI	CTC GGT ACC ATG ATT TTT CTC ATG ACG AAA GAT TCT TTT C
CS560 gadE anti MluI	CGA CGC GTC TAA AAA TAA GAT GTG ATA CCC AGG GTG ACG
CS582 relA sense KpnI	CTC GGT ACC ATG GTT GCG GTA AGA AGT GCA CAT ATC A
CS699 purF sense	GCA GCA ATG GCA GCG AAA ATA TTG
CS700 purF anti	CAG TCT GGT TTA CGG GCT TTG AAG AC
CS701 hisH sense	TCT CAG CAC CGA AAT GAT CGA GCA
CS702 hisH anti	CCG GAA TAA TCA TCA CAT CTC CAG GA
CS707 pKD13 kan-purF sense	TAA CGC ACA TGA CCA ATG CCC ATA TTG CCC TGC AAA CGC TGC ATA TGG CGA GCG TGT AGG CTG GAG CTG CTT C
CS708 pKD13 kan-purF anti	CGG TAC TGT TTA TCG CTA CCC TGA TCG TTG GTG CTA TCG TGA ACT TCG TGA TCC GTC GAC CTG CAG TTC GA
CS709 pKD13 kan-hisH sense	CGT GAC CCG GAC GTC GTG TTG CTG GCC GAT AAA CTG TTT TTA CCC GGC GTT GGC ACT GAT CCG TCG ACC TGC AGT TCG A
CS710 pKD13 kan-hisH anti	CGG CAT TGC GTA GCT GTG AAC AAA GTA AAA GTA CGC GCC GTC TTC AAT CCC CTG TGT AGG CTG GAG CTG CTT C

8.2.5 Cultivation conditions

L-tyrosine production experiments were performed at 37°C with 225 rpm orbital shaking in 50 ml MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) cultures supplemented with 5 g/l glucose and an additional 4 g/l NH_4Cl. All liquid cultivations were conducted in at least triplicates. When appropriate, antibiotics were added in the following concentrations: 34 μg/ml chloramphenicol for maintenance of pHACM-derived plasmids and 20 μg/ml kanamycin

for maintenance of pZE-derived plasmids. L-phenylalanine auxotrophic (Δ*pheA*) cultures were additionally provided with L-phenylalanine (Sigma) at a concentration of 0.35 mM.

8.2.6 Analytical methods

For the quantification of L-tyrosine, cell-free culture supernatants were filtered through 0.2 µm PTFE membrane syringe filters (VWR International) and used for HPLC analysis with a Waters 2690 Separations module connected with a Waters 996 Photodiode Array detector (Waters) set to a wavelength of 278 nm. The samples were separated on a Waters Resolve C18 column with 0.1 % (vol/vol) trifluoroacetic acid (TFA) in water (solvent A) and 0.1 % (vol/vol) TFA in acetonitrile (solvent B) as the mobile phase. The following gradient was used at a flow rate of 1 ml/min: 0 min, 95 % solvent A + 5 % solvent B; 8 min, 20 % solvent A + 80 % solvent B; 10 min, 80 % solvent A + 20 % solvent B; 11 min, 95 % solvent A + 5 % solvent B. Cell densities of cultures for growth rate calculations were determined by measuring their absorbance at 600 nm with an Ultrospec 2100 *pro* UV/Visible spectrophotometer (Amersham Biosciences).

8.2.7 Fed-batch fermentations of rpoA14^{R}

rpoA14^{R} was cultured in 5 ml LB medium until the mid-exponential phase, then transferred into 2-50 ml MOPS minimal or R medium flask cultures at a starting OD_{600} of 0.1. These were cultivated at 37°C with 225 rpm orbital shaking until the mid-exponential phase and were subsequently used as the inoculum for a 1.5-l culture (5% by volume).

Fed-batch fermentations for rpoA14^{R} were performed in 2-l glass vessels using the BioFlo110 modular fermentor system (New Brunswick Scientific). Two separate experiments

were conducted using MOPS minimal medium and R medium (Riesenberg, Schulz et al. 1991) supplemented with 0.5 g/l L-phenylalanine and 68 µg/ml chloramphenicol. Fermentations were performed at 37°C with a pH automatically adjusted to 7.0 using 30% solution of ammonium hydroxide. Dissolved oxygen (pO2) was maintained at ≥25% by agitation speed (100–1,000 rpm) and gas mix control (air/oxygen). Foam formation was controlled by the addition of Tego Antifoam KS911 (Evonik Goldschmidt). Operations were controlled and recorded with the BioCommand Lite Data Logging software (New Brunswick Scientific).

8.3 Results and Discussion

8.3.1 Full transcriptional analysis of three gTME mutants

Mutations introduced during gTME often have global effects within the cell and, as a result, frequently lead to changes in the transcript levels of hundreds if not thousands of genes. The main challenge in strain characterization for inverse metabolic engineering then becomes the identification of the undoubtedly small subset of alterations that possesses a causal link to the phenotype. Thus, to approach this problem, it becomes necessary to conduct a full analysis of the observed transcriptional changes between desirable mutants and their corresponding parental strains. By observing differential patterns in gene expression, it may then be possible to identify specific pathways or mechanisms by which a phenotype is influenced and manipulated.

For this particular study, we were fortunate that three separate mutants were identified to have comparably high capacities for L-tyrosine production. Although these strains may have

acquired these traits by disparate mechanisms, sufficient similarities in gene expression may point us to relevant pathways responsible for influencing strain performance. In this section, we conducted a full transcriptional analysis of four strains – rpoA14, rpoA27, rpoD3, and P2 – using the GeneChip *E. coli* Genome 2.0 Arrays (Affymetrix). Expression profile deviations between mutants and parental were identified by Significance Analysis of Microarrays (SAM) using SAM 3.0 (Tusher, Tibshirani et al. 2001).

Analysis of upregulated genes

Because of the sheer number of differentially expressed loci, we first focused our analysis on the upregulated pathways found in these strains. Quite surprisingly, our transcriptional comparisons returned a very small and manageable number of genes, with only 17 loci identified for rpoD3, 6 for rpoA27, and none for rpoA14. A large percentage of these presumed hits were found to possess functions related to acid stress resistance (Table 8.3).

E. coli possesses three acid resistance pathways, two of which rely on decarboxylase/antiporter systems that utilize external amino acids to protect cells during extreme acid challenges. The genes *gadA*, *gadB*, and *gadC* encode for the most efficient of the three - the glutamate-dependent acid resistance system (AR2) - and were recovered as upregulated targets in both rpoA27 and rpoD3 (Table 8.3). During conditions of acid stress, the GadA and GadB glutamate decarboxylase isozymes actively convert intracellular glutamate and a proton to γ-aminobutyrate. The GadC antiporter then orchestrates the export of this product in exchange for extracellular glutamate. The consumption of internal protons mediated by this system helps the cell maintain a suitable cytosolic pH for survival (Ma, Gong et al. 2003).

Table 8.3 Upregulated genes related to acid resistance

Gene Name	Gene ID	Function	rpoD3		rpoA27[b]	
			Log Fold Change	q-value[a] (%)	Log Fold Change	q-value[a] (%)
b3517	*gadA*	glutamate decarboxylase	2.236	0	1.974	11.05
b1493	*gadB*	glutamate decarboxylase	2.699	0	2.515	0
b1492	*gadC*	glutamate:γ-aminobutyrate antiporter	2.027	0	-	-
b3512	*gadE*	transcriptional activator	2.137	0	2.160	0
b3510	*hdeA*	acid stress chaperone	1.950	0	-	-
b3509	*hdeB*	acid stress chaperone	2.115	0	-	-
b3511	*hdeD*	acid resistance membrane protein	1.707	0	-	-
b3506	*slp*	starvation lipoprotein	1.609	10.17	1.897	5.67

[a] q values represent the lowest false discovery rate at which that gene is considered significant. It functions like the well-known p-value but is adapted to multiple-testing situations.

[b] Dashed lines (-) indicate that a gene was not upregulated.

The bacterial periplasm is known to be quite vulnerable to acid stress due to its permeability to small molecules less than 600 Da in size. Thus, to prevent the acid-induced aggregation of periplasmic proteins, *E. coli* is known to express the *hdeA*- and *hdeB*-encoded acid stress chaperones (Gajiwala and Burley 2000; Kern, Malki et al. 2007). These two proteins, along with the *hdeD*-encoded acid stress membrane protein, were among the list of upregulated components in rpoD3 (Table 8.3). Along with hindering the formation of insoluble periplasmic bodies at low pHs, more recent studies have also demonstrated the roles of HdeA and HdeB in assisting with the resolubilization and renaturation of existing protein aggregates (Malki, Le et al. 2008). Although HdeD is known to be part of the acid resistance transcriptional

network, few studies have investigated its specific role in the cell's response to acid-induced stress (Masuda and Church 2003).

Because several genes involved with acid stress resistance were recovered from our analysis, we decided to investigate the specific transcriptional controls that govern these pathways. Fortunately, a comprehensive analysis of the acid stress transcriptional network has already been performed by other labs, allowing us to simply overlay our data on top of existing transcriptional maps (Figure 8.1) (Masuda and Church 2003). Interestingly, all genes listed in Table 8.3 were found to lie within the same regulon, with the proteins YdeO and GadE (formely known as YhiE) exerting direct transcriptional control over the identified operons. The expression of YdeO is, in turn, influenced by the regulator EvgA which exists one level above in the network. Given this configuration, we hypothesized that both YdeO and GadE could be used as genetic dials for modulating the acid stress response of *E. coli*. Moreover, if acid stress resistance is indeed related to L-tyrosine production as is suggested by our transcriptional data, then altering the expression of these two proteins may have interesting implications for engineering this phenotype.

To explore this possibility, we constructed overexpression plasmids for both *ydeO* and *gadE* using two versions of the pZE vector with a low- (Y) or high-strength (P_L) constitutive promoter. These plasmids were then transformed into the background of parental strain P2 to determine whether overexpression of these regulators could elicit a L-tyrosine overproduction phenotype. Contrary to our expectations, plasmid-encoded *ydeO* and *gadE* yielded only moderate increases in L-tyrosine production over the parental (up 26-32% to 535-552 mg/l L-tyrosine, respectively) (Figure 8.2). It was also interesting to note that such enhancements

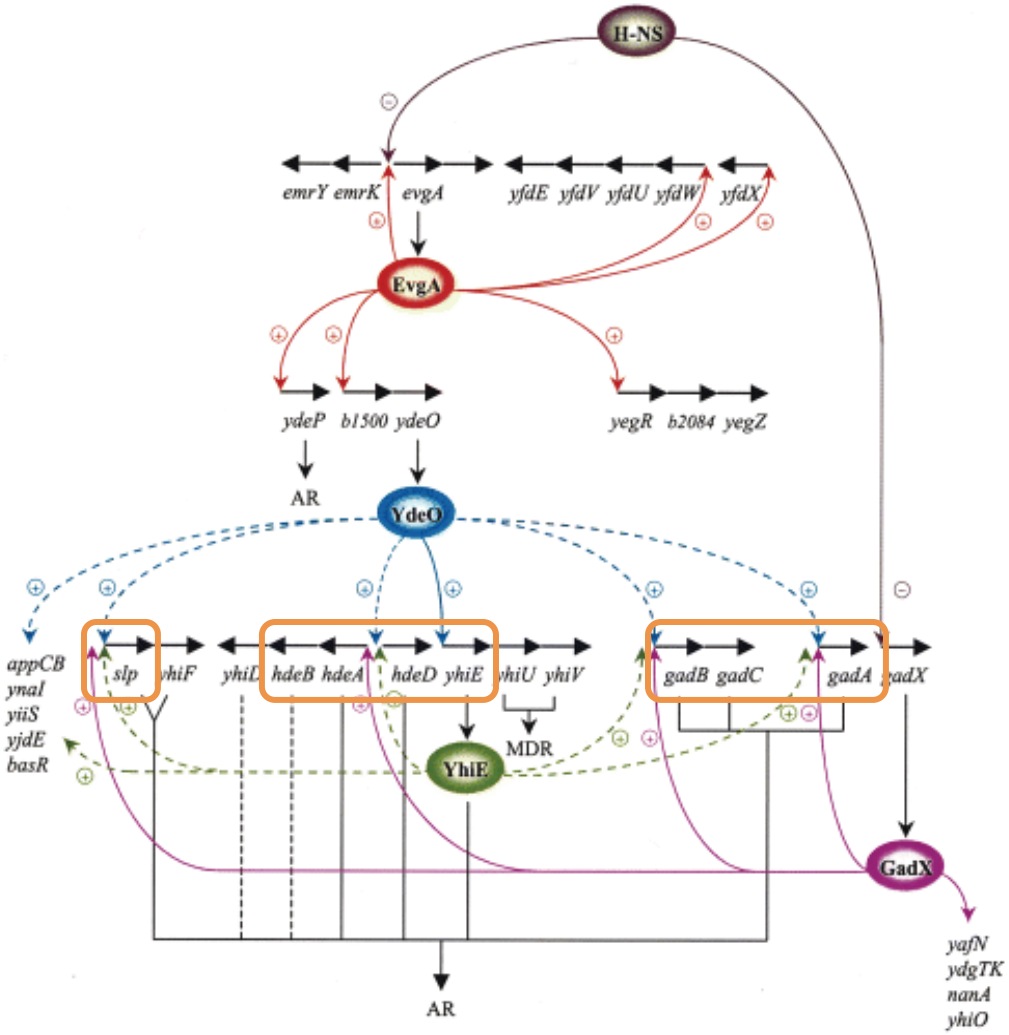


Figure 8.1 Regulatory network of acid resistance genes

Figure taken from (Masuda and Church 2003). Encircled protein names represent transcriptional regulators with "+" representing activation and "-" representing repression. Orange boxes highlight upregulated genes that were recovered through our transcriptional anaylsis. YhiE/*yhiE* and GadE/*gadE* are two different names for the same protein/gene.

were seen only with low levels of *ydeO* and *gadE* overexpression. The use of a stronger P_L promoter resulted in a 17-22% drop in overall titer. Thus, although the acid resistance pathway may still contribute to L-tyrosine overproduction, this data clearly demonstrates that it is not the sole determinant of phenotype.

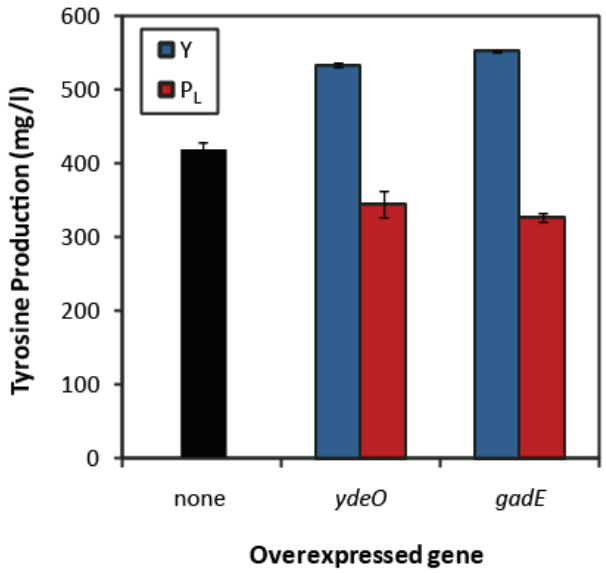


Figure 8.2 Overexpression of *ydeO* and *gadE* in parental strain P2

Two promoter strengths were tested for *ydeO* and *gadE* overexpression: Y (low, blue bars) and P_L (high, red bars). A numerical comparison of relative promoter strengths can be found in Table 8.1. 'None' (black bar) indicates that no plasmid-encoded gene was present.

Analysis of downregulated pathways

In stark contrast to the small number of upregulated genes found in the mutants, several (on the order of hundreds) downregulated genes were recovered by our analysis. Although this large discrepancy seems unusual at first glance, it may also simply reflect the nature of a transcriptional engineering approach. Because this technique relies on the introduction of

mutations into essential components of the transcriptional machinery, it is likely that overall activity along with specificity is altered. Even small decreases in the rate of RNA polymerase function could influence the number of transcripts generated, which would then be reflected as a downregulation in gene expression.

Regardless of the exact mechanism by which this phenomenon occurs, it is clear that the sheer number of genes recovered by our analysis makes it quite difficult to parse out specific patterns of underexpression. Thus, to convert this into a more manageable problem, we made use of Ecocyc's Pathway Tools Omics Viewer (http://ecocyc.org/). Taking transcriptional data as its input, this program is able to overlay over-and under-expression ratios onto a preassembled metabolic network. Simple color mapping then allows one to immediately distinguish pathways that have a significant number of genes exhibiting differential expression. The results of this analysis are summarized in Table 8.4.

Table 8.4 Downregulated pathways in rpoD3, rpoA14, and rpoA27

Pathway	rpoD3	rpoA14	rpoA27
Arginine synthesis	✓	✓	✓
Isoleucine synthesis	✓	✓	✓
Leucine/valine synthesis	✓	✓	✓
Histidine synthesis	✓		✓
Tryptophan synthesis	✓		✓
Lysine synthesis	✓	✓	
Glutamate synthesis	✓	✓	✓
De novo purine/pyrimidine biosynthesis	✓		
DNA replication	✓		✓
Ribosomal proteins and RNA	✓	✓	✓
Fatty acid elongation		✓	

At first glance, the most conspicuous result seems to be the large number of amino acid pathways that were downregulated in these mutant strains. While we cannot offer any theories as to why or how these specific pathways were targeted, it is interesting to note that these findings do appear to be reasonable from a biochemical standpoint. Because many of these amino acids share common precursors with L-tyrosine (and the other aromatic amino acids), the reduced synthesis of these competing compounds could lead to higher overall levels of L-tyrosine within the cell (Figure 8.3). As an example, expression of the leucine and valine biosynthetic pathway was reduced in all three gTME-derived strains. We hypothesize that decreases in the amount of pyruvate consumed by these reactions could ultimately lead to a larger pool of PEP, one of the limiting precursors for L-tyrosine production. We observe that arginine and histidine biosynthesis was also affected, suggesting that higher amounts of the precursor glutamate may be useful to drive the final transamination step of L-tyrosine biosynthesis (Figure 3.2). Finally, tryptophan production was shown to be downregulated in these strains, an unsurprising result given that it shares much of the aromatic amino acid biosynthetic pathway with L-tyrosine and pulls significantly from the supply of chorismate (Figure 3.2). Apart from these examples, however, it remains unclear as to how decreased production of isoleucine, lysine, and glutamate could ultimately benefit L-tyrosine production.

Because the downregulation of several amino acid pathways did not offer any additional insights into the underlying strain biochemistry, we focused our attentions on elucidating the effects of three other identified pathways. In particular, we were intrigued by the reduced production of ribosomal subunit proteins and RNA, a trait that was common among all three mutants. When examined together with reductions in both fatty acid and purine or pyrimidine

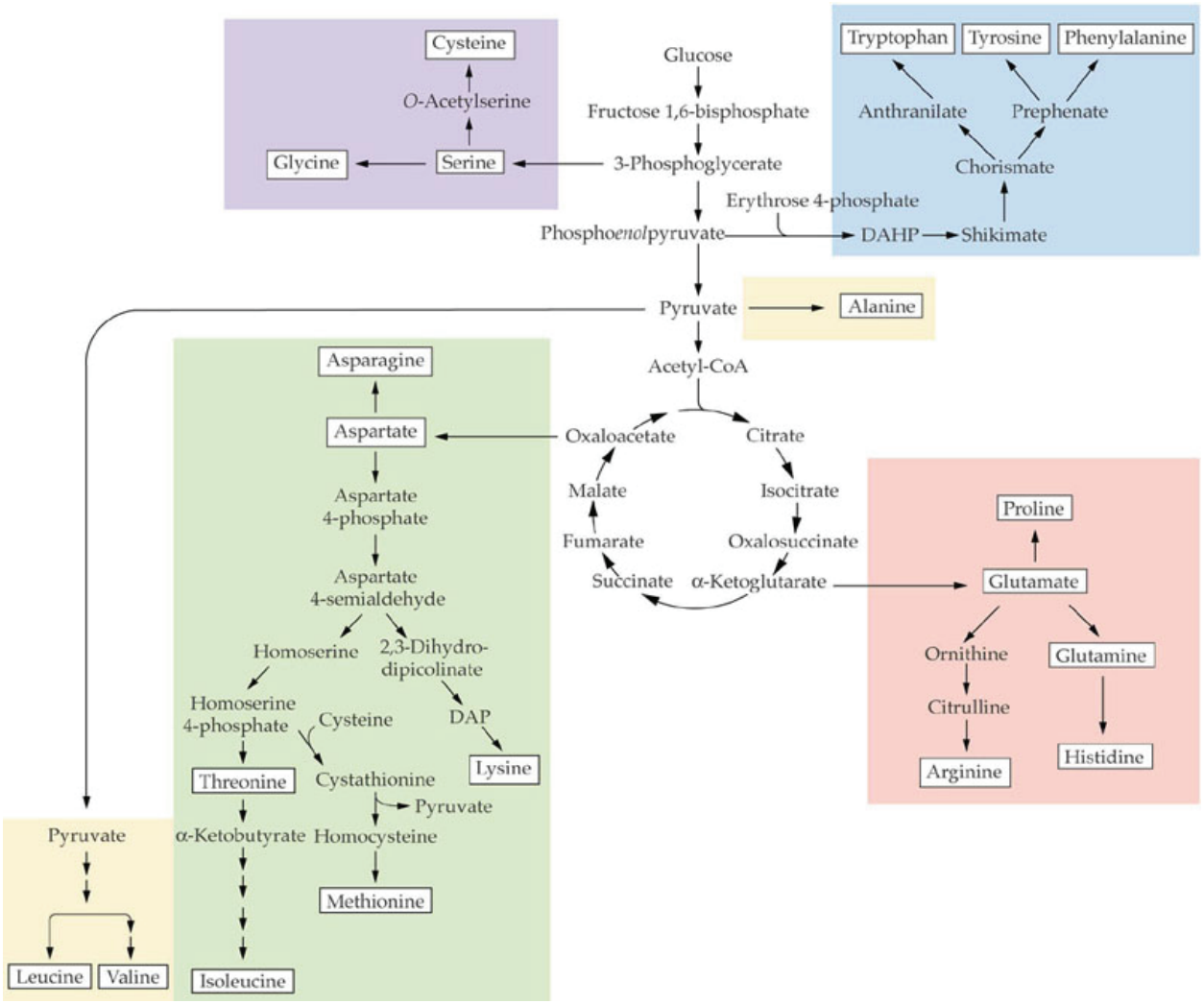


Figure 8.3 Pathways and precursors for amino acid biosynthesis

Source: http://www.uky.edu/~dhild/biochem/24/lect24.html

biosynthesis, a potential role for guanosine tetraphosphate/pentaphosphate (collectively known as (p)ppGpp) and the stringent response began to emerge.

(p)ppGpp is a small nucleotide regulator that is rapidly synthesized during conditions of nutritional stress. Two separate enzymes have been shown to be responsible for (p)ppGpp formation within *E. coli*, each responding to different environmental stressors (Figure 8.4). The first ppGpp synthase, encoded by *relA*, is known to associate with ribosomes and is able to upregulate (p)ppGpp expression during amino acid starvation. Functional characterizations have shown that this reaction is mediated by the binding of uncharged tRNAs to the ribosomal 'A' site, which leads to stalled protein synthesis and subsequently enables this process. Although much less is known about the mechanism for the second *spoT*-encoded ppGpp synthase, it has been shown to be active during other conditions of stress, including phosphorus, iron, carbon source or fatty acid deprivation. Because SpoT also possesses hydrolase activity in addition to its synthase domain, it has the ability to quickly adjust and finely tune the levels of (p)ppGpp in response to changes in the cellular environment (Magnusson, Farewell et al. 2005; Potrykus and Cashel 2008; Srivatsan and Wang 2008).

Although no real scientific consensus has been reached regarding the specific mechanisms for (p)ppGpp action, its upregulation has been proven to have several immediate and widespread effects on cellular function (Figure 8.4). On the whole, resources are often redistributed from cellular activities related to proliferation and growth to those needed for maintenance and survival. In particular, the downregulation of both ribosomal proteins and stable RNA (tRNA and rRNA), as was observed in our gTME mutants, has long been considered the hallmark feature of the stringent response (Table 8.4). Individual strain decreases in other

proliferation-related activities, including DNA replication and fatty acid synthesis are also in accordance with this (p)ppGpp-related hypothesis. Although it was our analysis of the downregulated pathways that led us to expore this avenue, it is also noteworthy to mention that several upregulated targets also corroborate this theory. Indeed, stress survival genes, such as those of the acid resistance pathways, are among the first genetic targets influenced by increases in intracellular (p)pGpp levels (Magnusson, Farewell et al. 2005).

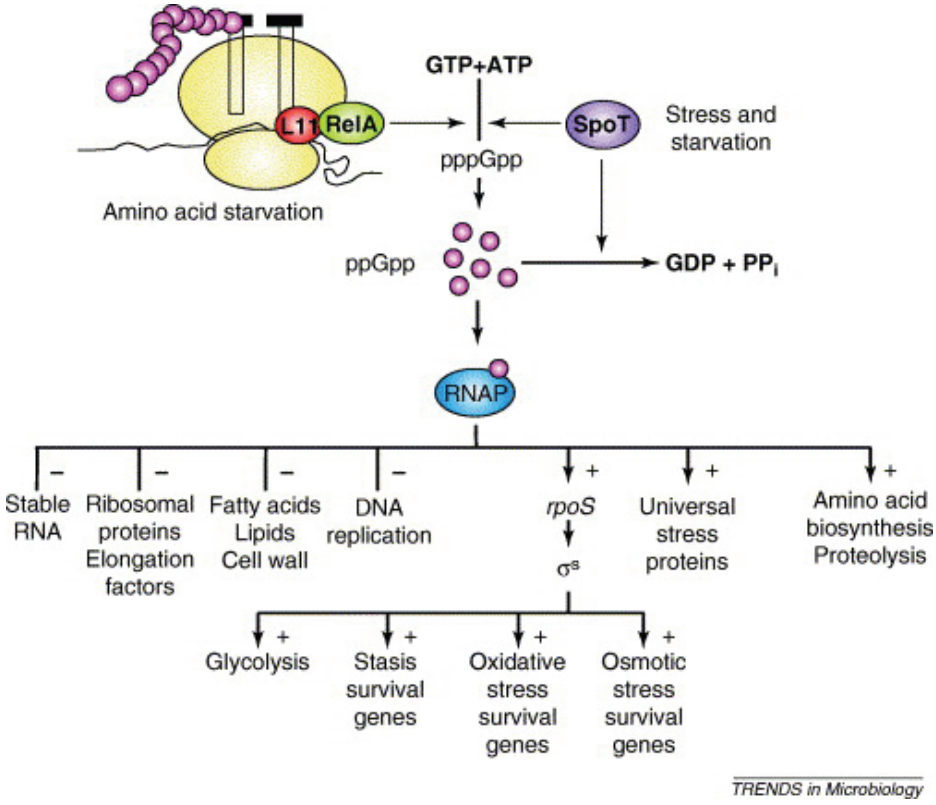


Figure 8.4 The(p)ppGpp-mediated stringent response in *E. coli*

This figure, taken from Magnusson et al, highlights the cellular changes that take place during the stringent response in *E. coli*. Levels of (p)ppGpp are controlled by two enzymes: the RelA ppGpp synthase and SpoT, which possesses both synthase and hydrolase activity (Magnusson, Farewell et al. 2005).

The only major discrepancy found within our data set is the widespread downregulation of several amino acid biosynthetic pathways, a feature which seems to directly contradict past observations on the effects of a stringent response. However, we note that these reports were based on *in vitro* experiments which utilized a fairly narrow set of amino acid promoters and/or operons (histidine, arginine, lysine, phenylalanine, threonine) to generate their data (Stephens, Artz et al. 1975; Paul, Berkmen et al. 2005); thus the notion that (p)ppGpp should upregulate the overproduction of *all* amino acids may be untrue. Indeed, more recent studies have called this general statement into question, with several *in vivo* transcriptional analyses demonstrating that these biosynthetic pathways are rarely upregulated *en masse* (Chang, Smalley et al. 2002; Durfee, Hansen et al. 2008; Traxler, Summers et al. 2008). Of particular interest, Durfee et al. observed that while a serine hydroxamate-induced stringent response in *E. coli* led to activation of seven attenuator-regulated leader sequences, transcript levels of the corresponding *structural genes* were actually found to be lower in these strains (Durfee, Hansen et al. 2008). Similar results were found with a glucose-lactose diauxie model in which most amino acid biosynthetic pathways (with the exception of histidine and arginine) were downregulated (Chang, Smalley et al. 2002).

Taken altogether, our transcriptional analysis points to a potential role for (p)ppGpp in mediating the cellular responses observed in our strains. We therefore wondered whether direct tuning of (p)ppGpp levels could be used to modulate each strain's individual capacities for L-tyrosine overproduction. To explore this possibility, we overexpressed the *relA*-encoded ppGpp synthase in the background of parental P2 to assess whether increasing (p)ppGpp concentrations could have a positive effect on L-tyrosine synthesis. Of note, this strategy was

recently tested for the engineering of both glutamate and lysine production in *E. coli* and successfully led to 21% and 10% increases in yields, respectively (Imaizumi, Kojima et al. 2006).

As seen in Figure 8.5, the overexpression of *relA* in the background of parental P2 led to less than impressive increases in L-tyrosine production, particularly when compared to the titers observed in the original mutant isolates. Gains were limited to only 10% above the control regardless of whether a weak (II) or strong (P_L) promoter was used. These results therefore introduce serious doubts as to whether a causal role for (p)ppGpp actually exists within these strains.

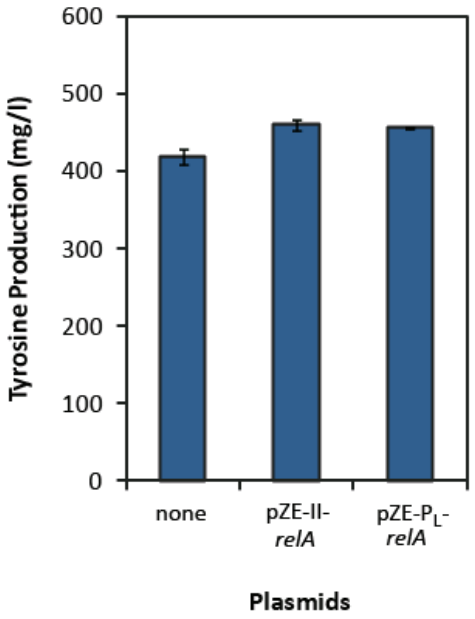


Figure 8.5 Overexpression of *relA* in parental P2

Two promoter strengths for *relA* overexpression were tested: II (low) and PL (high). A numerical comparison of relative promoter strengths can be found in Table 8.1. The bar labeled 'none' represents the titer for strain P2 bearing no additional plasmids.

8.3.2 Overexpression of transcriptional regulators in mutant backgrounds

In our previous experiments, we focused on overexpressing select regulators and enzymes in the parental P2 to ascertain whether such modifications could recover the high L-tyrosine titers observed in our gTME-derived strains. Unfortunately, the answer was a resounding no, with only limited gains made (10-32%) in L-tyrosine production. Indeed, even the highest final titer (achieved with low-level *gadE* overexpression) represented a mere 62% of what had been previously shown to be achievable in the isolated mutant rpoD3. We therefore wondered whether the unidentified background mutations within rpoD3, rpoA14, and rpoA27 were required for L-tyrosine synthesis. In Chapter 7, we saw that these chromosomal variants were absolutely essential for the phenotype, as only minimal increases in L-tyrosine titer were observed if mutated plasmids were provided alone (Figure 7.4). It is therefore feasible that, as with *rpoA* and *rpoD*, the benefits of *ydeO*, *gadE*, and *relA* overexpression may only become manifest when supplied in combination with these background mutations. To investigate this possibility, we transformed the *ydeO*, *gadE*, and *relA* overexpression plasmids into the backgrounds of each of the cured mutant strains, rpoD3, rpoA14, rpoA27. In addition, the transcriptional regulator *evgA* was provided on a pZE backbone and tested; as it exists only one level above *ydeO* in the acid resistance network (Figure 8.1), its overexpression may also have the capacity to influence strain performance.

The results, as seen in Figure 8.6, were wholly unexpected. Although drastic decreases in L-tyrosine production were fairly consistent for *ydeO* and *gadE* overexpression strains, the transfer of *evgA* and *relA* plasmids led to remarkable increases in L-tyrosine production. In fact, the titers seen in the original isolates were not only matched but sometimes even exceeded, as

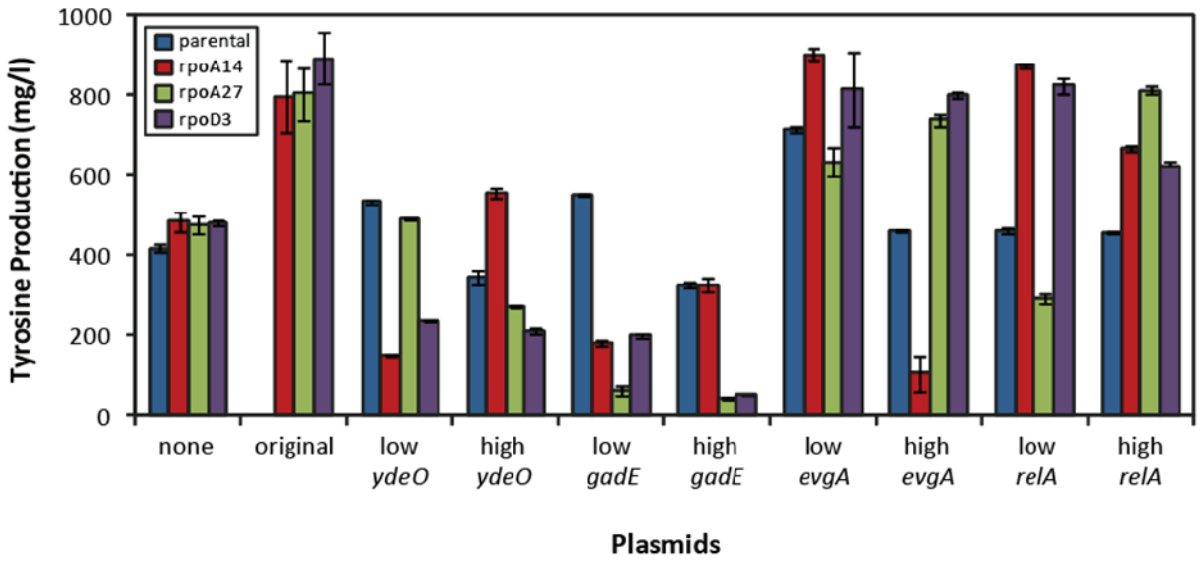


Figure 8.6 Overexpression of *ydeO*, *gadE*, *evgA*, and *relA* in mutant backgrounds

The genes *ydeo*, *gadE*, *evgA,* and *relA* were individually overexpressed in four different backgrounds – the parental P2 (blue bars) and the cured backgrounds of rpoA14 (red bars), rpoA27 (green bars), and rpoD3 (purple bars). Two promoter strengths were tested for each gene in order to ascertain the effects of both low and high level expression on cellular phenotype. All "high" strength promoters were P_L. "Low" strength promoters for *ydeO*, *gadE*, *evgA*, and *relA* were Y, Y, AA, and II, respectively (Table 8.1). L-tyrosine titers were measured after 48 hr cultivation in 50 ml MOPS minimal medium.

seen prominently for the rpoA14 background. The use of two different strength promoters also revealed the presence of expression optima that were dependent on both the transcriptional regulator and the mutant chromosome. As such, fine-tuning the expression of these genes (perhaps through the use of additional synthetic promoters) may lead to even greater titers and yields than those observed here.

It is interesting to note that while *evgA* successfully elicited high L-tyrosine titers, overexpression of either *ydeO* or *gadE* alone did not have the same cellular effects. One

possible explanation for this discrepancy may be the existence of slight inaccuracies in the regulatory scheme mapped out for this complex network. Specifically, recent studies have questioned the proposed linearity of this system (Figure 8.1) and have instead postulated a branched regulatory circuit for EvgA-YdeO-GadE control. In this alternate scheme, EvgA plays a *direct* role in upregulating *gadE* transcription during exponential growth in glucose-supplemented minimal medium (Ma, Masuda et al. 2004); thus, it is possible that e*vgA* overexpression may actually be required to achieve full activation of these pathways. Another likely conjecture stems from our knowledge that EvgA is capable of mediating other responses besides those related to acid resistance, independent of both YdeO and GadE. If these other activated systems play even a limited role in L-tyrosine production, their specific effects on phenotype would only be elicited by the upregulation of the EvgA protein (Nishino and Yamaguchi 2001; Nishino, Inazumi et al. 2003).

The results discussed above clearly have widespread implications for the underlying biochemistry of these strains. Our analysis shows that the mutant *rpoA/rpoD* plasmids can be replaced by overexpression of either *evgA* or *relA* without the loss of phenotype. We therefore maintain that despite the extensive changes induced by these mutant transcriptional components, their effects on L-tyrosine production can be isolated to the upregulation of two specific pathways. Because chromosomal mutations are still required for this phenotype, the exact mechanisms still remain somewhat of a mystery; however, these results clearly point to the important roles of both acid resistance and/or the stringent response in eliciting high L-tyrosine production within these strains.

8.3.3 A whole genome sequencing approach for identifying chromosomal variations

Although we were able to deduce the specific functions of *rpoA*/*rpoD* overexpression in our mutant strains, the identities of the essential background mutations still represent a missing piece of the puzzle. Unfortunately, transfer of this phenotype into a novel strain continues to be difficult without this information and, as a result, future engineering work remains restricted to these specific backgrounds. To overcome these critical limitations, we focused our next efforts on uncovering the precise chromosomal alterations responsible for these strains' unique capacity for L-tyrosine synthesis.

As mentioned in Chapter 2, rapid declines in the cost of whole genome sequencing have made these techniques much more feasible, even for academic explorations. We therefore took advantage of the growing availability and increasing affordability of these technologies to fully evaluate the altered genetic make-ups of the isolated strains. Specifically, the genomes of mutants rpoD3, rpoA14, and rpoA27 were analyzed and subsequently compared to the sequences of both parental P2 and the reference genome, *E. coli* K12 (MG1655).

As demonstrated in previous studies, the number of changes recovered during whole genome sequencing is oftentimes so large that it becomes necessary to restrict one's search to a smaller subset of genes known to be related to phenotype (Ikeda, Ohnishi et al. 2006). We were therefore quite fortunate to have recovered a short and manageable list of possible insertions, deletions, and single nucleotide polymorphisms (SNPs) during this analysis. Even more auspiciously, validation of these presumed changes by Sanger sequencing revealed that the sequences of rpoA14, rpoA27, and rpoD3 each differed from parental P2 by only a *single* mutation. A base pair change in rpoA14 resulted in a L82R substitution in the *hisH*-encoded

imidazole glycerol phosphate (IGP) synthase subunit. Similarly, mutations in rpoD3 led to a V5G shift in the *purF* gene encoding for the enzyme amidophosphoribosyl transferase. For rpoA27, a T→C nucleotide substitution was found 17 bp upstream of the same gene targeted in rpoD3 (*purF*) (Table 8.5, Figure 8.7).

Table 8.5 Validated SNPs in gTME-derived mutants

Position	rpoD3		rpoA14		rpoA27		Annotation
	R	S	R	S	R	S	
2092803			T	G			*hisH*; imidazole glycerol phosphate synthase subunit, glutamine amidotransferase (histidine biosynthesis)
2428247	A	C					*purF*; amidophosphoribosyl transferase (*de novo* purine biosynthesis)
2428277					T	C	

R = reference (P2) sequence; S = substituted base pair

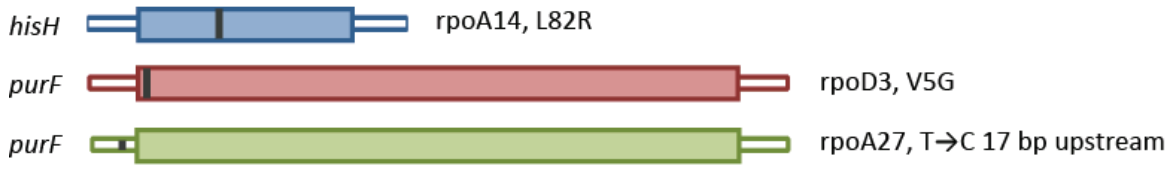


Figure 8.7 Locations of validated SNPs within the *hisH* and *purF* loci

8.3.4 Identified SNPs are necessary for phenotype

To validate the importance of these substitutions, we reintroduced each SNP into the parental P2 background and tested its effect on cellular phenotype. Specifically, we expected strains containing these reconstructed backgrounds and their corresponding *rpoA* or *rpoD* plasmids to possess the ability to produce L-tyrosine at the levels seen with the original isolates. As shown in Figure 8.8, all three chromosomal mutations were successful in eliciting high amounts of L-tyrosine production from strains carrying the *rpoA*/*rpoD* mutant plasmids. Interestingly, measurements for the reconstructed rpoA27 and rpoD3 strains (denoted rpoA27^{R} and rpoD3^{R}) were consistently ~150 mg/l lower than titers of the original isolates, suggesting that additional mutations (such as large-scale insertions or deletions) may have been missed by our sequencing analysis. The exact opposite was found to be true for rpoA14^{R}, which exhibited final levels that were over 100 mg/l higher than the original mutant rpoA14. Despite these slight discrepancies in titers, however, these results clearly demonstrate the importance of these specific *hisH* and *purF* substitutions for eliciting a high production phenotype.

This analysis led us to the construction of a *completely genetically-defined* strain, rpoA14^{R}, which possesses a titer of 902 mg/l L-tyrosine and a yield of 0.18 g L-tyrosine/g glucose. To put these numbers into perspective, this yield on glucose is more than 150% greater than a classically improved L-phenylalanine auxotroph (DPD4195) that is currently being used for the industrial production of L-tyrosine and, when excluding glucose consumed for biomass formation, represents 85% of the maximum theoretical yield (Olson, Templeton et al. 2007).

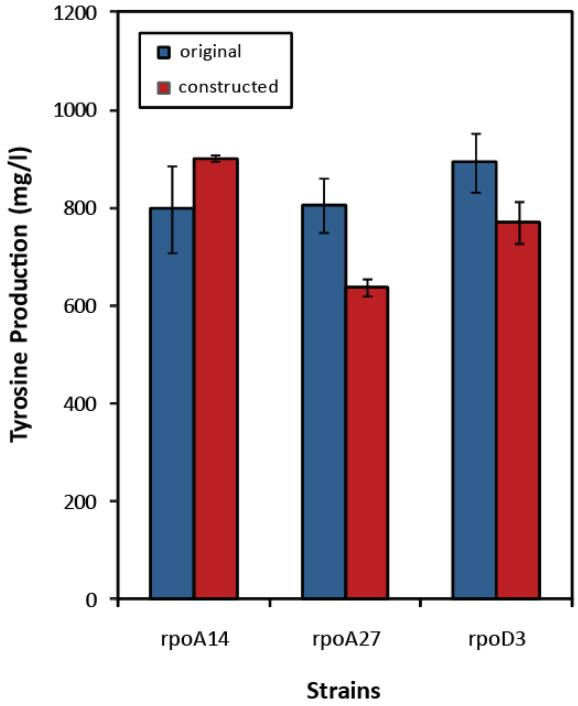


Figure 8.8 Comparison of original and reconstructed gTME strains

L-tyrosine titers for the original isolates are shown in blue. The performance of reconstructed strains containing an introduced SNP and their corresponding *rpoA* or *rpoD* plasmids is shown in red. All L-tyrosine concentrations were measured after 48 hr.

8.3.5 Performance of rpoA14^{R} in large-scale bioreactors

Although the industrial strain DPD4195's yield on glucose was only 0.074 g L-tyrosine/g glucose in 50 ml cultures, the transfer of this strain into a 200-l fermentor led to significant improvements in performance (a yield of 0.28-0.32 g L-tyrosine/g glucose and final titers of 51-57 g/l L-tyrosine) (Olson, Templeton et al. 2007; Patnaik, Zolandz et al. 2008). Because such impressive gains were made through simple scale-up and process optimization, we were naturally curious to see how our genetically-defined strain rpoA14^{R} would perform on the bioreactor scale (2-l fermentors). Two different minimal media formulations were tested for this purpose – MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) and R medium (Riesenberg, Schulz et al. 1991).

When cells were cultivated in MOPS minimal medium, strain characteristics appeared to be quite similar to those previously observed in 50 ml shake flask cultures, with yields on glucose of 0.204 g L-tyrosine/g glucose and a specific growth rate of 0.275 hr^{-1}. Unfortunately, however, cell growth was hampered and peaked at a very low OD_{600} of 10, leading to a titer of just 5.7 g/l L-tyrosine (Figure 8.9). The cells behaved quite differently when cultivated in the second media formulation. Switching to another defined synthetic medium (R medium) led to much higher maximum productivities (2.1 g L-tyrosine/l/hr), titers (13.8 g/), and growth rates (0.405 hr^{-1}) but unfortunately took its toll on overall yield (0.120 g L-tyrosine/g glucose). As has been observed in previous studies, this fermentation also produced ~0.5 g/l L-phenylalanine (in addition to the initial 0.5 g/l supplemented into the medium) during the course of the culture. Scientists have hypothesized this to be due to the nonenzymatic conversion of prephenate to L-phenylpyruvate, which can then be transaminated to form L-phenylalanine (Young, Gibson et al. 1969; Zamir, Jung et al. 1983; Patnaik, Zolandz et al. 2008).

The results of these two crude experiments make it quite clear that additional process optimization can likely have a significant impact on strain performance in these controlled bioreactor environments. Indeed, identifying limiting nutrients in the MOPS minimal medium formulation may extend the growth and production phase of the cell, and modification of R medium may help to increase the yield to previously observed levels. Ultimately, balancing these two media formulations and tweaking other important bioreactor parameters should lead to the development of optimal fermentation protocols that maximize the yield, titers, and productivities of these cultures.

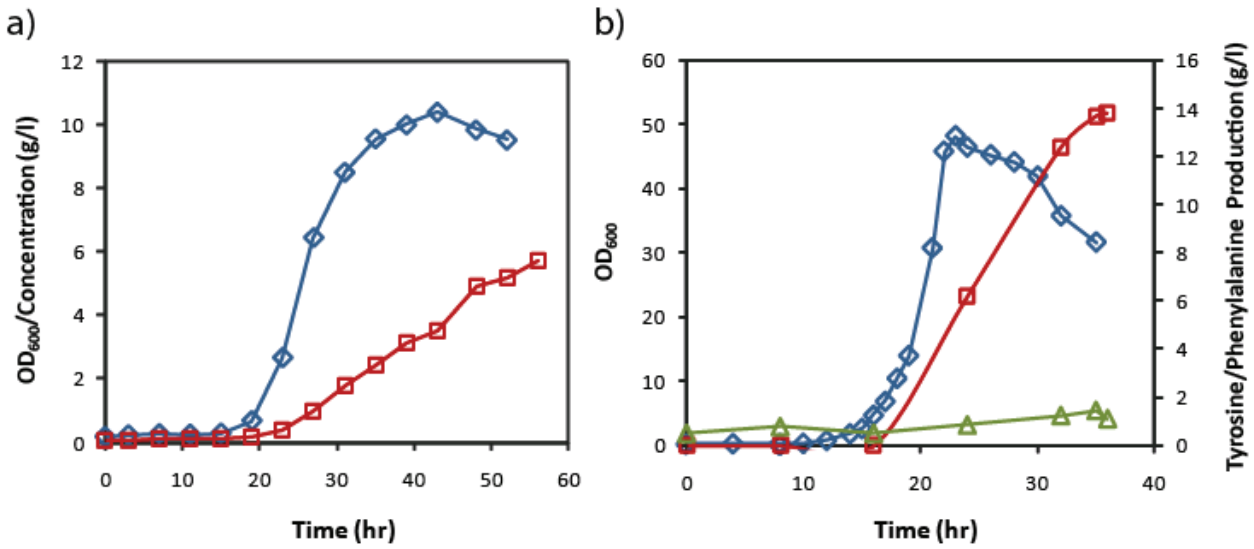


Figure 8.9 Performance of rpoA14^{R} in 2-l reactors

Two different media formulations were tested: a) MOPS minimal medium and b) R medium. Measurements for cell density (OD_{600}) (◇), L-tyrosine concentrations (□), and L-phenylalanine concentrations (△) are reported above.

8.4 Conclusions

In Chapter 7, we saw that while the recovery of gTME mutants rpoD3, rpoA14, and rpoA27 successfully demonstrated the potential of a transcriptional engineering approach for eliciting complex phenotypes, it also highlighted the inherent difficulties associated with applying an inverse metabolic engineering paradigm. This was in large part due to the global nature of this approach, which was additionally complicated by the unusual appearance of supplementary mutations within the bacterial chromosome. Because such properties unfortunately eliminated the benefits of direct plasmid transfer, it became clear that a full characterization would be

necessary to extract the biochemical and genetic information needed to engineer novel strain backgrounds.

In this chapter, we conducted an in-depth analysis of these three gTME-derived strains using two common 'omics techniques – microarray analysis and whole genome sequencing. Although widespread problems with data analysis and integration often limit the overall utility of these methods, we were successful in extracting useful information to guide future metabolic engineering efforts.

8.4.1 Overexpression of acid resistance and stringent response regulators can eliminate the requirement for a mutant *rpoA* or *rpoD*

Although hundreds of genes were recovered by our transcriptional analysis, specific patterns of over- or underexpression led us to investigate the roles of two pathways in *E. coli*. To our astonishment, we found that the individual overexpression of two regulators/enzymes – *evgA* and *relA* – could completely supplant the requirement for either a mutant *rpoA* or *rpoD* for recovering a high L-tyrosine production phenotype. These experiments therefore provide compelling evidence for the specific mechanisms induced by these mutated transcriptional components and clearly implicate both acid resistance and the stringent response in eliciting the desired phenotype. Although two separate proteins were seen to have the same overall effect on L-tyrosine synthesis, it is important to note that both pathways are actually linked through the actions of the small nucleotide regulator (p)ppGpp. Indeed, because (p)ppGpp levels dictate the onset of both the stringent response and more indirectly, the acid stress pathway, it is possible that these varied cellular responses are, in actuality, associated with a

common mode of action. It is also worthwhile to mention that this proposed mechanism is actually corroborated by the strain analysis presented in Chapter 6. Because (p)ppGpp induction shifts cellular resources from growth to cell maintenance, increased levels of this regulator are often accompanied by proportional decreases in growth rate and increases in acid resistance (as reflected by higher culture pHs). These two cellular properties were observed for both the knockout strains generated by transposon mutagenesis and the gTME-derived mutants described here (Figure 6.4, Figure 6.5) (Sarubbi, Rudd et al. 1988).

Owing to the difficulties of analyzing transcriptional data, such success stories are actually quite rare in the literature. We were therefore quite pleased with these results, not only because a feasible mechanism for *rpoA*/*rpoD* activity has been established, but also because this example once again demonstrates the enormous benefits of undertaking an unguided approach. Since our search for differential expression patterns was not restricted to a predetermined set of genes (as has been required in previous microarray studies) (Wahlbom, Cordero Otero et al. 2003; Bro, Knudsen et al. 2005), we were able to identify seemingly unrelated pathways that ultimately proved to be causally linked to the desired cellular phenotype. Indeed, a purely rational approach would not have led us to target global processes, such as the acid stress pathway or the stringent response.

8.4.2 Functional similarities of recovered SNPs

Because *evgA* and *relA* were only effective when expressed in the mutant backgrounds, it became clear that unidentified chromosomal variations still played a key role in determining cellular phenotype. To identify these specific mutations, we analyzed the sequences of each

mutant strain via whole genome sequencing. Although this approach sometimes leads to the identification of more mutations than can be functionally characterized (Ikeda, Ohnishi et al. 2006), we were quite fortunate in that only a *single* mutation was validated to be present within each mutant. This small number, while initially surprising, does however corroborate previous data showing that the mutational frequencies of these strains were not significantly altered (Figure 7.5). One would therefore not expect several genomic modifications to emerge within the short time frame required for library generation, screening, and cultivation.

Our analysis revealed that single base pair substitutions occurred within the *hisH* (IGP synthase subunit, glutamine amidotransferase) and *purF* (amidophosphoribosyl transferase) loci of *E. coli* (L82R in *hisH* for rpoA14 and V5G in *purF* for rpoD3). HisH and HisF, which together form the heterodimeric enzyme IGP synthase, catalyze a key branch point in histidine biosynthesis. During the fifth step of the pathway, nitrogen from glutamine is used for the formation of IGP with concomitant generation of both glutamate and aminoimidazole carboxamide ribonucleotide (AICAR). Through a similar mechanism, PurF catalyzes the first step of the *de novo* purine biosynthetic pathway, during which 5-phosphoribosylpyrophosphate (PRPP) and glutamine are enzymatically converted to 5'-phosphoribosylamine (PRA) and glutamate (Mei and Zalkin 1989; Mei and Zalkin 1990; O'Donoghue, Amaro et al. 2001). It is interesting to note that both altered protein targets reside within biochemically-related systems. Since histidine and purine biosynthesis are actually linked through the intermediate AICAR and the shared precursor PRPP, it is quite possible that both mutations may exert their effects through modulation of a common precursor pathway.

HisH and PurF also possess very similar biochemical functions within the cell. Both proteins belong to the class of enzymes generally referred to as glutamine amidotransferases (GATases), which utilize glutamine as a source of an amide nitrogen with the concomitant production of glutamate. HisH is classified within the group of type-I GATases, which are distinguished by the presence of a catalytic triad internal to the protein at Cys77, His178, and Glu180 (O'Donoghue, Amaro et al. 2001). Though they possess similar functionalities, type-II GATases such as PurF are marked by an N-terminal glutamine amidotransferase domain with catalytic residues at Cys1, Asp29, and His101 (Mei and Zalkin 1989; Mei and Zalkin 1990). Surprisingly, we found that the internal SNPs recovered during our analysis (V5G in rpoD3 and L82R in rpoA14) lie only 4 or 5 amino acids downstream of the catalytic cysteines for both of these enzymes. When compared to a previously constructed *hisH* structural model, these changes were found to reside within a helical structure directly adjacent to these conserved residues (Mei and Zalkin 1989; Mei and Zalkin 1990; O'Donoghue, Amaro et al. 2001). Because both Cys77 (*hisH*) and Cys1 (*purF*) possess important roles in glutamine binding, it seems likely that the introduction of these neighboring SNPs may serve to alter the specific substrate affinities of these enzymes. Such a result would then point to a possible role for glutamine/glutamate levels for influencing cellular phenotype, a theory with plausible grounds given that glutamate is consumed during the L-tyrosine pathway's final transamination step. Interestingly, this idea also ties in with previous transcriptional results involving the acid resistance pathway, as glutamate was found to be the main substrate used in that system.

In rpoA27, a T→C mutation was found in the intergenic region between *purF* and *cvpA,* which comprise a two-gene polycistronic operon. Although *purF* transcript levels were found to

be downregulated by about two-fold in this strain, it is not clear whether such changes were a result of this mutation, as the promoter/operator region for this operon is actually located upstream of *cvpA* (Rolfes and Zalkin 1988; Schumacher, Choi et al. 1994).

8.4.3 Performance of a superior completely genetically-defined strain

Results from both our microarray studies and whole genome sequencing allowed us to apply an inverse metabolic engineering paradigm for the construction of a completely genetically-defined strain for L-tyrosine production. When compared on a 50 ml basis, strain rpoA14^{R} is capable of significantly outperforming industrial strain DPD4195 with respect to both yields and titers (Olson, Templeton et al. 2007). Early bioreactor experiments point to the potential for process scale-up but also illustrate the need for additional process optimization to achieve maximum titers, yields, and productivities.

In Table 8.6, we provide a quick comparison of two experiments conducted with rpoA14^{R}, as well as fermentation parameters for the fully rationally-engineered strain T2 (Lütke-Eversloh and Stephanopoulos 2007) and the industrial producer DPD4195 (Patnaik, Zolandz et al. 2008). As briefly mentioned earlier, two different fermentation media were tested for rpoA14^{R} (MOPS minimal and R media), with each showing a unique ability to maximize either yields or maximum productivities, respectively. Cultivation in MOPS minimal medium resulted in a yield of 0.204 g L-tyrosine/g glucose, exactly twice the value observed with the T2 fermentations. Similarly, maximum productivities in R medium were found to be twofold greater than T2 values as well (188 versus 92.6 mg L-tyrosine/g DCW/hr**)**. Although overall productivities for rpoA14^{R} were low compared to both T2 and DPD4195, this parameter

was skewed by the long lag time seen in our cultures, a feature that can be either partially or completely eliminated by optimizing our inoculation protocols.

It is interesting to note that scale-up and process optimization of DPD4195 cultures led to significant improvements in strain performance, with more than a quadrupling in yield (0.07 to 0.3 g L-tyrosine/g glucose) and an impressive titer of 55 g/l (Young, Gibson et al. 1969; Zamir, Jung et al. 1983; Patnaik, Zolandz et al. 2008). Given this remarkable potential for process-related gains, we are confident that similar improvements can be achieved with our engineered strain in order to transform it into a competitive or superior industrial performer.

Table 8.6 Comparison of bioreactor parameters

	T2 (3-l)	**DPD4195 (200-l)**	**rpoA14^{R} (MOPS)**	**rpoA14^{R} (R)**
Final L-Tyr Titer (g/l)	9.7	55	5.71	13.8
Total Glucose consumed (g/l)	95	183	28	115
Overall yield (g Tyr/g Glc)	0.102	0.3	0.204	0.120
Maximum Productivity (mg Tyr/g DCW/hr)	92.6	-	88.24	188
Maximum Productivity (g Tyr/l/hr)	-	-	0.280	2.06
Overall Productivity (g Tyr/l/hr)	0.626	1.15	0.102	0.391
Growth rate (hr^{-1})	0.26	-	0.275	0.405
Maximum OD_{600}	120	65	10.4	48.1

Dashes lines (-) indicate values that were either not published or unavailable.

CHAPTER 9 FLAVONOID PRODUCTION FROM GLUCOSE

9.1 Introduction

Although L-tyrosine is an industrially-relevant compound in and of itself, it is also often used as a precursor molecule for the synthesis of a variety of polymers, specialty chemicals, and pharmaceutical compounds (Bell and Wheeler 1986; della-Cioppa, Garger et al. 1990; Bonuccelli and Del Dotto 2006; Rajput and Rajput 2006; Leonard, Lim et al. 2007; Qi, Vannelli et al. 2007; Sariaslani 2007; Vannelli, Wei Qi et al. 2007; Leonard, Yan et al. 2008). As a result, several of our engineered L-tyrosine production strains may actually prove to be quite valuable for other industrial processes, and in essence, can be viewed as customizable biological platforms for the synthesis of these other important L-tyrosine derivatives. One potential application for these strains has recently emerged from a renewed interest in flavonoid compounds, which possess pharmicodynamic properties useful for combating a wide range of human health problems from inflammation and cancer to diabetes and heart disease (Knekt, Jarvinen et al. 1996; Hollman and Katan 1998; Fowler and Koffas 2009). Current microbial-based production processes rely on external feeding of expensive aromatic precursors to achieve mg/l quantities of flavonoid production; hence, this application seems particularly well-adapted for exploring the added value of our recombinant L-tyrosine producers.

9.1.1 Flavonoids and the phenylpropanoid pathway

Flavonoids comprise a highly diverse family of plant secondary polyphenols which possess biochemical properties (estrogenic, antioxidant, antiviral, antibacterial, antiobesity, and anticancer) that are useful for the treatment of several human pathologies (Knekt, Jarvinen et al. 1996; Hollman and Katan 1998; Harborne and Williams 2000; Forkmann and Martens 2001; Fowler and Koffas 2009). This wide range of pharmaceutical indications arises from the varied and distinct chemical structures generated in the phenylpropanoid pathway. Indeed, several enzymes exist which are capable of functionalizing or otherwise altering the conformation of the three-ring phenylpropanoid core structure through an assortment of hydroxylation, reduction, alkylation, oxidation, and glucosylation reactions. These modifications, when applied in combination, allow for the generation of over 8000 chemically distinct species that span the six major flavonoid groups: isoflavones, flavanones, flavones, flavonols, catechins, and anthocyanins (Figure 9.1) (Fowler and Koffas 2009).

Although a myriad of enzymes are involved with structural diversification, only four catalytic steps are actually required for the conversion of the aromatic compounds L-phenylalanine and L-tyrosine to the main flavanone precursors, piconembrin and naringenin (Figure 9.1). For the specific case of naringenin production, this process begins with the conversion of L-tyrosine to the phenylpropanoic acid *p*-coumaric acid by the action of a bacterial tyrosine ammonia lyase (TAL). In the plant phenylpropanoid pathway, *p*-coumaric acid may also be formed by an alternate two-step conversion from L-phenylalanine which requires both a phenylalanine ammonia lyase (PAL) and a cytochrome P450 enzyme, *cinnamic acid* 4-hydroxylase (C4H). Once *p*-coumaric acid has been generated, the enzyme 4-coumarate:CoA

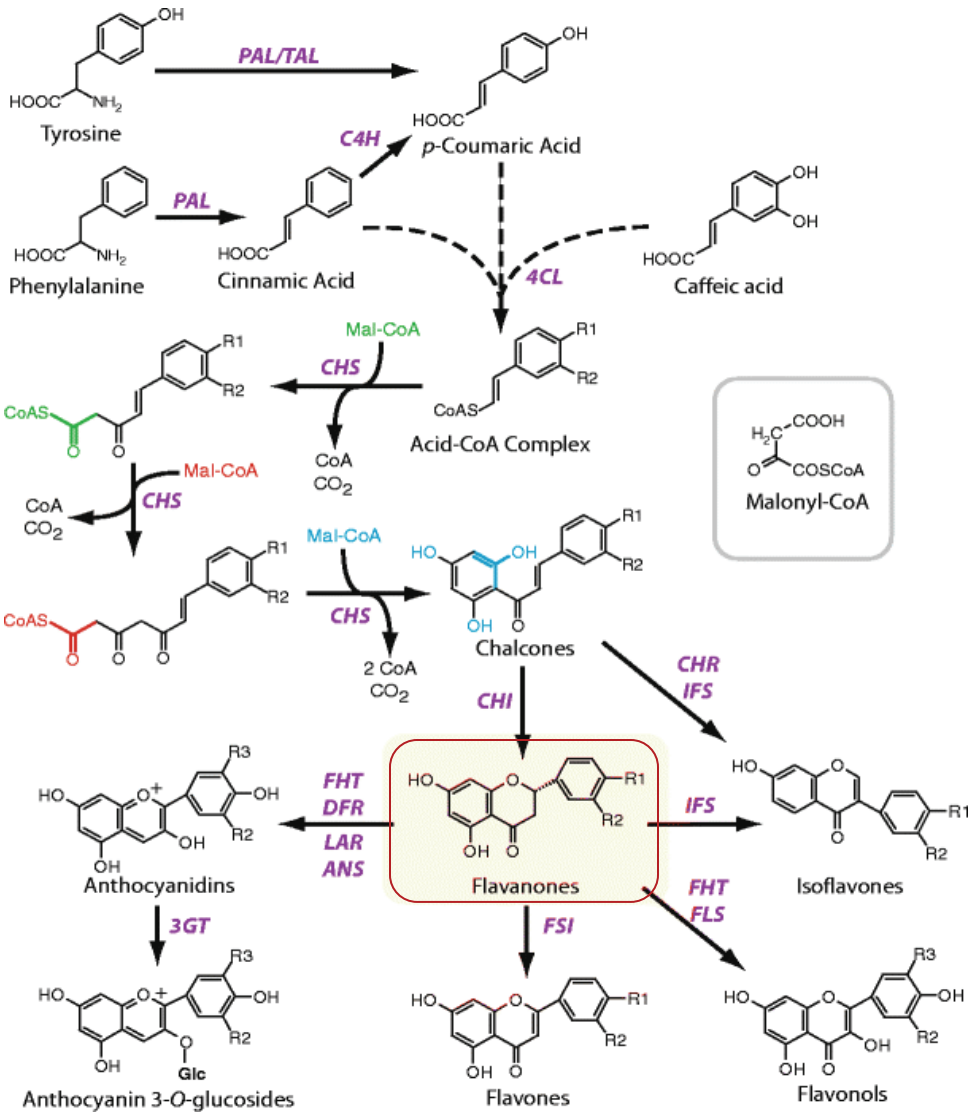


Figure 9.1 Detailed biosynthetic steps for flavanones and the diversification of flavonoids

The three successive carbons added from malonyl-CoA by CHS are shown in green, red, and blue. The R-groups denote the hydroxylation patterns for the natural flavonoid compounds. Enzyme abbreviations are as follows: *DFR* dihydroflavanone reductase, *LAR* leucoanthocynanidin reductase, *ANS* anthocyanidin synthase, *3GT* uridine, *flavanone* 3-glucoside transferase, *FSI* flanone synthase, *CHR* chalcone reductase, *IFS* isoflavanone synthase, *FHT* flavanone hydroxytransferase, *FLS* flavonol synthase. Figure and caption were taken from a recent review (Fowler and Koffas 2009).

ligase (4CL) mediates formation of its corresponding CoA ester, coumaroyl-CoA, which is subsequently condensed with three malonyl-CoA units by the sequential action of the type III polyketide synthase, chalcone synthase (CHS). In the final step, the resulting naringenin chalcone is stereospecifically isomerized by chalcone isomerase (CHI) to form the (*2S*)-flavanone naringenin. This compound provides the phenylpropanoid core structure for a myriad of other flavonoid molecules created through the combined actions of functionalizing enzymes (Kaneko, Hwang et al. 2003; Fowler and Koffas 2009).

9.1.2 Microbial production of flavonoid compounds

Surprisingly, attempts to develop recombinant strains for the microbial production of flavonoids have only emerged within the last seven years from just a handful of academic laboratories. Although these initial investigations explored the use of both *E. coli* and *S. cerevisiae* as production hosts, more recent studies have focused on the engineering of the former, presumably due to the relative ease of its genetic manipulation, as well as its inherent and proven capacity for heterologous protein expression.

The first report of naringenin production in *E. coli* was published in 2003 and described the construction of flavonoid gene clusters containing the plant biosynthetic enzymes *Rhodoturula rubra* PAL (*Rr*PAL), *Streptomyces coelicolor* 4CL (*Sc*4CL), and *Glycyrrhiza echinata* (licorice plant) CHS (*Ge*CHS). Three separate gene configurations were tested in which both the number and location of T7 promoters and ribosome binding sites (RBS) were varied. Although an optimum in naringenin production was found through the addition of a promoter and RBS in front of each gene, production levels remained quite low with a titer of only 0.45 mg/l after 65

hr (Hwang, Kaneko et al. 2003; Kaneko, Hwang et al. 2003). A similar investigation was conducted a year later which explored the use of different genetic sources for the construction of a flavonoid pathway in *E. coli*. This time, a newly characterized TAL enzyme from *Rhodobacter sphaeroides* (*Rs*TAL) (Kyndt, Meyer et al. 2002) was chosen to replace PAL and was expressed in combination with 4CL and CHS from *Arabidopsis thaliana* (*At*4CL, *At*CHS). Titers of up to 20.8 mg/l naringenin were observed after 48 hr, although such numbers were only recovered during cultivation in rich medium (Terrific Broth, TB) and therefore may not be a fair comparison with previous results (Watts, Lee et al. 2004). Interestingly, both studies did not express a CHI enzyme and instead relied on exposure to alkali conditions for the conversion of naringenin chalcone to naringenin; as such, naringenin titers listed above represent those of a racemic mixture rather than the biologically-active (*2S*) enantiomer.

Because of the relatively low levels of naringenin production observed in minimal media, some groups theorized that malonyl-CoA availability may present a rate-limiting component for flavonoid production. Indeed, studies have shown than intracellular concentrations of malonyl-CoA range between 4-20 µM (Takamura and Nomura 1988) and may therefore not be sufficient to support high CHS activity. To tackle this potential problem, researchers focused subsequent efforts on increasing malonyl-CoA pools within *E. coli*. One study found that coexpression of two *Corynebacterium glutamicum* acetyl-CoA carboxylase (ACC) subunits with the previously described *Rr*PAL-*Sc*4CL-*Ge*CHS gene cluster led to the synthesis of 57 mg/l naringenin. Because *Pueraria lobata* CHI (*Pl*CHI) was also provided, these numbers represent production levels for the favored enantiomeric form, (*2S*)-naringenin

(Miyahisa, Kaneko et al. 2005). However, such cultivations were seen to consume a large amount of glucose, which was supplemented at a concentration of 40 g/l.

Although the initial groundwork was set in place by the Horinouchi lab at the University of Tokyo (Hwang, Kaneko et al. 2003; Kaneko, Hwang et al. 2003; Miyahisa, Kaneko et al. 2005), the most recent advances in *E. coli* flavonoid production have come from the Koffas lab at the University at Buffalo. Researchers in this lab began with the construction and evaluation of two base strains for the production of naringenin from supplemented *p*-coumaric acid, thus eliminating the need for a PAL/TAL enzyme. The first strain (E1) was designed to express *Petroselinum crispus* (parlsey) 4CL-2 (*Pc*4CL2) and both CHS and CHI from *Petunia hybrida* (*Ph*CHS, *Ph*CHI) under the control of separate T7 promoters. The second strain (E2) differs from E1 only by the replacement of *Ph*CHI with a *Medicago sativa* variant (*Ms*CHI). Interestingly, this seemingly subtle change resulted in a significant difference in strain performance, with E2 producing 14-fold more naringenin than E1 after 36 hr (42 mg/l versus 3 mg/l) (Leonard, Lim et al. 2007). This discrepancy is presumed to be due to a more robust and efficient *Ms*CHI enzyme and gives us an early indication of the importance of considering various enzyme sources when constructing heterologous pathways.

As seen with other investigations, early successes in the construction of a basal strain led this group to explore additional avenues for increasing the supply of malonyl-CoA within the cell. While the addition of *P. luminescens* ACC (*Pl*ACC) along with (ACC)-biotin ligase (BirA) led to only marginal increases in naringenin production (69 mg/l), the overexpression of *Pl*ACC and the acetate assimilation enzyme, acetyl-CoA synthase (ACS) resulted in a maximum titer of 119 mg/l naringenin (with 2 g/l acetate supplementation) (Leonard, Lim et al. 2007). Following

investigations found that supplementation with malonate and overexpression of *Rhizobium trifolii matB* and *matC* (for its transport and conversion to malonyl-CoA) could increase production levels up to 155 mg/l naringenin (Leonard, Yan et al. 2008). In that same study, authors also explored the use of the fatty acid inhibitor cerulenin to suppress native degradation pathways for malonyl-CoA and reported titers of 186 mg/l naringenin. Although the cost of cerulenin makes it prohibitive for use in actual production processes, such results illustrate the enormous potential for increasing flavonoid production through the targeted manipulation of these biosynthetic pathways. Finally, in their most recent work, a stoichiometric model was employed to find alternate engineering targets for increasing the supply of malonyl-CoA. Deletions of genes *sdhA*, *adhE, brnQ,* and *citE* combined with the overexpression of ACS, ACC, biotin ligase (BPL), and pantothenate kinase (PNK) led to the highest levels of naringenin production reported (270 mg/l). Along with the standard precursors (*p*-coumaric acid), such cultures were supplemented with 6 µM biotin, 50 µM pantothenic acid, and 2 g/l sodium acetate (Fowler, Gikandi et al. 2009).

9.1.3 Feeding and cultivation conditions

Although previous studies have already made significant gains in demonstrating the feasibility of microbial naringenin production in *Escherichia coli*, the established protocols suffer from two disadvantages that could be prohibitive during process scale-up (Miyahisa, Kaneko et al. 2005; Leonard, Lim et al. 2007; Leonard, Yan et al. 2008). The first main shortcoming is that fermentation protocols often require two separate cultivation steps to achieve high flavonoid titers. Typically, strains are first grown in rich media in order to generate biomass and ensure

adequate heterologous protein expression. After reaching a target density, cells are collected and transferred to minimal media for the second stage of the process during which flavonoids are produced from supplemented phenylpropanoic precursors. While the separation of biomass can be performed relatively easily on a laboratory scale, such procedures are significantly more difficult and expensive when translated to large-scale fermentation processes. As such, the development of robust strains that can perform equally well in a single medium formulation is absolutely required for this process.

The second major drawback found in these studies is the heavy reliance on precursor feeding (typically L-tyrosine or *p*-coumaric acid) to achieve high levels of flavonoid production. This requirement is particularly unfavorable for the case of *p*-coumaric acid supplementation, given its high market price, especially in comparison to both L-tyrosine and glucose (Table 9.1). Thus, there is an obvious economic incentive to develop strains capable of converting cheaper feedstocks such as glucose to high value flavonoid compounds.

In this chapter, we outline the construction and evaluation of a series of strains capable of circumventing both of these critical limitations. To mediate the production of naringenin from glucose, a four-enzyme heterologous pathway (consisting of TAL, 4CL, CHS, and CHI) was assembled within two strains which have been previously engineered for high L-tyrosine production (P2 and rpoA14^{R}). Due to the high sensitivity of strain performance on both enzyme source and relative gene expression levels, sequential optimization was required for each step of the pathway. However, once an optimum metabolic balance had been achieved, the resulting strains were found to possess a remarkably robust constitution, exhibiting unfettered

growth and competitive naringenin titers (up to 84 mg/l with the addition of cerulenin), even with a single-stage fermentation in minimal media.

Table 9.1 Cost of naringenin and precursors

Compound	Price ($/g)
Naringenin	6.46
p-Coumaric acid	2.74
L-Tyrosine	0.48
Glucose	0.01

* Calculated from prices of the largest available quantities on Sigma-Aldrich

9.2 Materials and Methods

9.2.1 Construction of (DE3) lysogenic strains for T7 expression

The λDE3 Lysogenization Kit (EMD Chemicals) was used to prepare strains *E. coli* K12, P2, and rpoA14^{R} for the expression of genes cloned in T7 expression vectors. Manufacturer's protocols were followed for lysogenization and strain verification. Strains that have undergone λDE3 lysogenization are indicated by the "(DE3)" notation following their names.

9.2.2 Gene Synthesis

CHI from *Pueraria lobata* (*Pl*CHI) was codon optimized for *E. coli* expression and synthesized using established protocols for gene synthesis (Kodumal, Patel et al. 2004). Oligonucleotides were designed with the software package Gene Morphing System (GeMS), which was

previously available for public use at http://software.kosan.com/GeMS (Jayaraj, Reid et al. 2005). Following assembly, the synthesized *chi* gene was cloned into pTrcHis2B (Invitrogen) using the primers CS420 CHI sense KpnI and CS421 CHI anti HindIII (Table 9.3) and the restriction enzymes KpnI and HindIII. Errors found within the resulting plasmid, pTrc-*Pl*CHI^{syn}, were corrected with the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene) using the manufacturer's protocols. Codon optimization and synthesis of both *Rhodotorula glutinis* tal (*Rg*TAL) and *Petroselinum crispus* (parsley) 4CL-1 (*Pc*4CL) were performed by DNA2.0. In future references, synthetic genes/proteins are denoted by a superscript "syn." The DNA sequences and corresponding amino acids for all synthesized genes are provided in APPENDIX B (Table B.1).

9.2.3 Heterologous pathway construction and assembly

All constructed plasmids described below were verified by colony PCR and sequencing. Routine transformations were performed with chemically competent *E. coli* DH5α cells (Invitrogen) according to the manufacturer's protocols. A list of plasmids and strains used in this study can be found in Table 9.2.

Flavonoid biosynthetic pathway construction in pCS204

The pCS204 flavonoid plasmid contains *Rhodobacter sphaeroides tal* (*Rs*TAL, also known as *hutH*), *Streptomyces coelicolor 4cl-2* (*Sc*4CL), *Arabidopsis thaliana chs* (*At*CHS), and synthetic *P. lobata chi* (*Pl*CHI^{syn}), each under the control of an independent trc promoter. The plasmid was assembled by a three-step cloning process. Briefly, the first three genes were first

independently cloned into pTrcHis2B or pTrcsGFP (pTrcHis2B carrying a codon-optimized superfolder green fluorescent protein (Pedelacq, Cabantous et al. 2006)) (C. Santos, unpublished) using primers CS313-CS318, CS420-CS421, and the restriction enzyme pairs specified in Table 9.3 to form pTrc-*Rs*TAL, pTrc-*Sc*4CL, and pTrc-*At*CHS. *R. sphaeroides* genomic DNA was used as a template for amplification of *Rs*TAL and was obtained from American Type Culture Collection (ATCC 17023). Similarly, *S. coelicolor* genomic DNA was used as a template for *Sc*4CL and was extracted using the Wizard Genomic DNA Kit (Promega). *At*CHS was amplified from an *A. thaliana* cDNA library from American Type Culture Collection (pFL61, ATCC 77500)**.**

In the second round of cloning, the P_{trc}-*Sc*4CL and P_{trc}-*Pl*CHI^{syn} regions were amplified from their respective plasmids with primers CS481-CS485 and cloned into pTrc-*Rs*TAL and pTrc-*At*CHS with the restriction sites HindIII and BstBI, respectively. It is noteworthy to mention that P_{trc}-*Pl*CHI^{syn} was amplified with two rounds of PCR (using the primer pairings CS483-CS484 and CS483-CS485) in order to incorporate a multi-cloning site designed to facilitate the addition of future genes/elements within this plasmid. The resulting plasmids from this second round of cloning were named pTrc-*Rs*TAL-*Sc*4CL and pTrc-*At*CHS-*Pl*CHI^{syn}.

In the third and final round of assembly, P_{trc}-*At*CHS-P_{trc}-*Pl*CHI was amplified from pTrc-*At*CHS-*Pl*CHI with primers CS486 CHS-CHI sense BamHI and CS487 CHS-CHI anti BamHI. This fragment was then cloned into pTrc-*Rs*TAL-*Sc*4CL with restriction enzyme BamHI to form plasmid pCS204.

Gene sequences and orientations were verified by colony PCR and sequencing after each round of cloning.

Table 9.2 Plasmids and strains used in this study

Plasmid or Strain	Relevant characteristics	Source
Plasmids		
pTrcHis2B	trc promoter, pBR322 ori, Amp^R	Invitrogen
pTrcsGFP	pTrcHis2B carrying a superfolder GFP (sGFP) variant (Pedelacq, Cabantous et al. 2006) that was synthesized and codon-optimized for *E. coli*	C. Santos, unpublished
pACYC184	p15A ori, Cm^R	ATCC
pETDuet-1	double T7 promoters, ColE1(pBR322) ori, Amp^R	Novagen
pCDFDuet-1	double T7 promoters, CloDF13 ori, Sp^R	
pRARE2	p15A ori, Cm^R, supplies tRNAs for the rare codons AUA, AGG, AGA, CUA, CCC, GGA, and CGG	Novagen
pCS204	pTrcHis2B carrying *R. sphaeroides* TAL, *S. coelicolor* 4CL, *A. thaliana* CHS, and *P. lobata* CHI^{syn}	This study
pTrc-*Rs*TAL	pTrcHis2B carrying *R. sphaeroides* TAL	This study
pTrc-*Rg*TAL^{syn}	pTrcHis2B carrying codon-optimized *R. glutinis* TAL	This study
pTrc-*Rg*TAL^{syn}-*Sc*4CL	pTrcHis2B carrying codon-optimized *R. glutinis* TAL and *S. coelicolor* 4CL	This study
pTrc-*Rg*TAL^{syn}-*Pc*$4CL^{syn}$	pTrcHis2B carrying codon-optimized *R. glutinis* TAL and codon-optimized *P. crispus* 4CL-1	This study
pACYC-*Sc*4CL	pACYC184 carrying *S. coelicolor* 4CL	This study
pET-*Rg*TAL^{syn}-*Pc*$4CL^{syn}$	pETDuet-1 carrying codon-optimized *R. glutinis* TAL and codon-optimized *P. crispus* 4CL-1	This study
pCDF-*Rg*TAL^{syn}-*Pc*$4CL^{syn}$	pCDFDuet-1 carrying codon-optimized *R. glutinis* TAL and codon-optimized *P. crispus* 4CL-1	This study
pCDF-trc-*Rg*TAL^{syn}-*Pc*$4CL^{syn}$	pCDFDuet-1 carrying codon-optimized *R. glutinis* TAL and codon-optimized *P. crispus* 4CL-1 with trc promoters	This study
pACKm-*At*CHS-*Pl*CHI^{syn}	pACKm carrying *A.thaliana* CHS and codon-optimized *P. lobata* CHI	This study
pCDF-*At*CHS-*Pl*CHI^{syn}	pCDFDuet-1 carrying *A.thaliana* CHS and codon-optimized *P. lobata* CHI	This study

pET-*Ph*CHS-*Ms*CHI	pETDuet-1 carrying *P.hybrida* CHS and *M. sativa* CHI	(Leonard, Lim et al. 2007)
pOM-*Ph*CHS-*Ms*CHI	pOM carrying *P.hybrida* CHS and *M. sativa* CHI with a single GAP (constitutive) promoter	This study
pACYC-MatBC	pACYCDuet-1 carrying *R. trifolii* MatB and MatC	(Leonard, Yan et al. 2008)
Strains		
E. coli K12 (MG1655)	wild-type	ATCC
P2	*E. coli* K12 Δ*pheA* Δ*tyrR lacZ*::$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr} *tyrR*::$P_{LtetO-1}$-*tyrA*^{fbr}*aroG*^{fbr}	Ch 4
rpoA14^{R}	P2 *hisH*(L82R) pHACM-*rpoA*14	Ch 8
E. coli K12 (DE3)	*E. coli* K12 carrying the gene for T7 RNA polymerase	This study
P2 (DE3)	P2 carrying the gene for T7 RNA polymerase	This study
rpoA14^{R} (DE3)	rpoA14^{R} carrying the gene for T7 RNA polymerase	This study

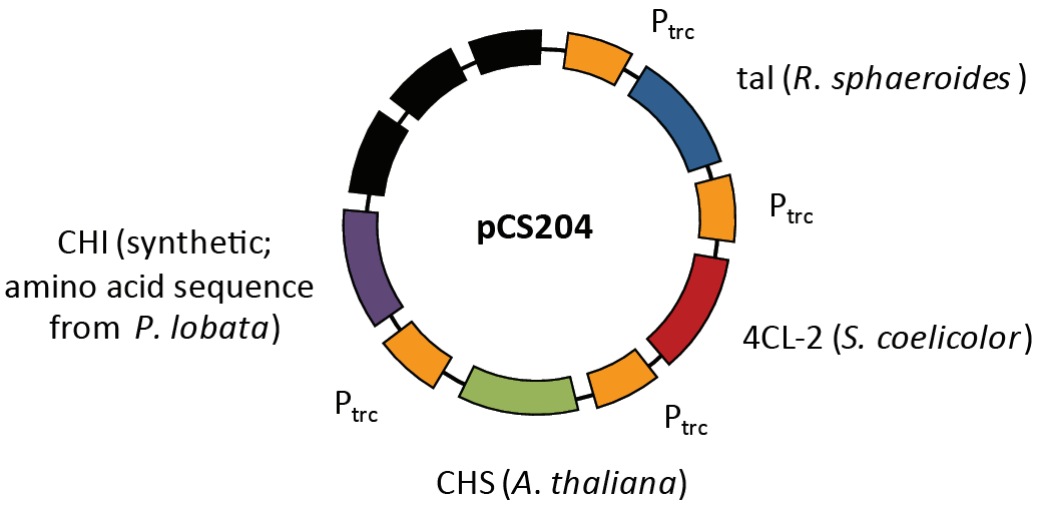


Figure 9.2 Schematic of pCS204 flavonoid plasmid

Table 9.3 Primers used in this study

Primer Name	Primer Sequence (5'→ 3')
CS313 R. sphaeroides hutH sense KpnI	GCTCGGTACC ATGCTCGCCATGAGCCCCC
CS314 R. sphaeroides hutH anti HindIII	ACG AAG CTT TTA GAC GGG AGA TTG CTG CAA GAG G
CS315 S. coelicolor 4CL-2 sense NcoI	TAA ACC ATG GTC CGC AGC GAG TAC GCA G
CS316 S. coelicolor 4CL-2 anti HindIII	ACG AAG CTT TTA TCG CGG CTC CCT GAG CTG T
CS317 A. thaliana CHS sense NcoI	TAA ACC ATG GTG ATG GCT GGT GCT TCT TCT T
CS318 A. thaliana CHS anti KpnI	GCT CGG TAC CTT AGA GAG GAA CGC TGT GCA AGA CG
CS420 CHI sense KpnI	GCT CGG TAC CAT GGC TGC GGC TGC TGC C
CS421 CHI anti HindIII	ACG AAG CTT TTA CAC AAT AAT ACC GTG GCT CAA CAC G
CS481 pTrc 4CL sense[a]	ACG AAG CTT AAT CCT AGG AAC TGA AAT GAG CTG TTG ACA ATT AAT CAT CC
CS482 pTrc 4CL anti[b]	ACG AAG CTT CTT GGA TCC CGA TCC GGA AAT TAT CGC GGC TCC CTG AGC TGT
CS483 pTrc CHI sense[c]	GAG TTC GAA CGA TGT ACA AAC TGA AAT GAG CTG TTG ACA ATT AAT CAT CC
CS484 pTrc CHI anti 1[d]	GCT AGC TTC GTA CGT GCT GAG CAT ATC AAT TGA TTA CAC AAT AAT ACC GTG GCT CAA CAC G
CS485 pTrc CHI anti 2[d,e]	GAG **TTC GAA** CTC GAG ATA CTA GTG TAG ATC TTT GGC CTC GCT GGC CAT GCT AGC TTC GTA CGT GCT GAG CAT ATC
CS486 CHS-CHI sense BamHI	CTT GGA TCC GCC GAC ATC ATA ACG GTT CTG GC
CS487 CHS-CHI anti BamHI	CTT GGA TCC GAG TTC GAA CTC GAG ATA CTA GTG TAG ATC TTT GGC
CS619 tal sense NcoI	TAA ACC ATG GCG CCT CGC C
CS620 tal anti SalI	AAT GTC GAC TTA TGC CAG CAT CTT CAG CAG AAC ATT
CS621 4CL sense NdeI	GCA CTA ACA TAT GGG TGA CTG CGT TGC CCC
CS622 4CL anti AvrII	AAT CCT AGG TTA CTT CGG CAG GTC GCC
CS627 CHS sense NdeI	TAA CAT ATG GTG ATG GCT GGT GC
CS628 CHS anti AvrII	AAT CCT AGG TTA GAG AGG AAC GCT GTG CAA GAC G
CS629 CHI sense NcoI	TAT ACC ATG GCT GCG GCT GCT G
CS630 CHI anti NotI	TAA GCG GCC GCT TAC ACA ATA ATA CCG TGG CTC AAC ACG
CS644 lacI sense AatII	CAT GAC GTC CCG CTT ACA GAC AAG CTG TGA CCG
CS645 CHI anti BsiWI	GCT TCG TAC GTG CTG AGC ATA TCA ATT

CS786 tal sense FseI	TAA CGG CCG GCC CCG ACA TCA TAA CGG TTC TGG CA
CS787 rrnB anti BamHI	TAA GGA TCC CAA CAG ATA AAA CGA AAG GCC CAG TCT
CS792 BsrGI-BglII oligo 1	GTA CGC GCA TGC GC
CS793 BsrGI-BglII oligo 2	GAT CGC GCA TGC GC

[a] Underlined segments indicate the addition of AvrII and HindIII restriction sites

[b] Underlined segments indicate the addition of BspEI, BamHI, and HindIII restriction sites

[c] Underlined segments indicate the addition of BsrGI and BstBI restriction sites

[d] Underlined segments indicate the addition of a multicloning site (MfeI, BlpI, BsiWI, NheI, SfiI, SpeI, XhoI)

[e] Bold segment indicates addition of a BstBI restriction site

TAL/4CL plasmid variants

pJ206-*Rg*TAL^{syn} (from DNA2.0) and pTrcHis2B were digested with restriction enzymes NcoI and HindIII, and the appropriate fragments were ligated to form pTrc-*Rg*TAL^{syn}. pTrc-*Rg*TAL^{syn}-*Sc*4CL was subsequently constructed by amplifying P_{trc}-*Sc*4CL with primers CS481 pTrc 4CL sense and CS482 pTrc 4CL anti (Table 9.3) and cloning the resulting product into the HindIII site of pTrc-*Rg*TAL^{syn}. pTrc-*Rg*TAL^{syn}-*Pc*4CL^{syn} was assembled by digestion of both pJ281-*Pc*4CL^{syn} (from DNA2.0) and pTrc-*Rg*TAL^{syn} with SalI followed by ligation of the appropriate fragments.

pET-*Rg*TAL^{syn} was constructed by amplifying *Rg*TAL^{syn} from pTrc-*Rg*TAL^{syn} using primers CS619 tal sense NcoI and CS620 tal anti SalI (Table 9.3) and cloning the resulting product into the NcoI/SalI sites of pETDuet-1 (Novagen). Because subsequent insertion of *Pc*4CL^{syn} into this plasmid required the restriction site NdeI, the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene) and primer CS657 pETtal (Quikchange) were used to change an internal NdeI sequence (within *Rg*TAL^{syn}) from CATATG to CACATG. *Pc*4CL^{syn} was subsequently cloned into this plasmid through amplification from pJ281-*Pc*4CL^{syn} with primers CS621 4CL sense NdeI and

CS622 4CL anti AvrII followed by insertion into the NdeI/AvrII sites of pET-*Rg*TAL^{syn}. The resulting plasmid was named pET-*Rg*TAL^{syn}-*Pc*4CL^{syn}.

pCDF-*Rg*TAL^{syn}-*Pc*4CL^{syn} was constructed through the digestion of both pET-*Rg*TAL^{syn}-*Pc*4CL^{syn} and pCDFDuet-1 (Novagen) with NcoI and AvrII, followed by ligation of the appropriate fragments. pCDF-trc-*Rg*TAL^{syn}-*Pc*4CL^{syn} was constructed by amplifying P_{trc}-*Rg*TAL^{syn}-P_{trc}-*Pc*4CL^{syn} from pTrc-*Rg*TAL^{syn}-*Pc*4CL^{syn} using primers CS786 tal sense FseI and CS787 rrnB anti BamHI (Table 9.3). FseI/BamHI-digested products were then ligated with a similarly digested pCDFDuet-1 plasmid.

To assemble pACYC-*Sc*4CL, the P_{trc}-*Sc*4CL cassette was first amplified with primers CS481 pTrc 4CL sense and CS482 pTrc 4CL anti (Table 9.3), then cloned into the HindIII restriction site of pACY184.

CHS/CHI plasmid variants

pACKm-*At*CHS-*Pl*CHI^{syn} was constructed by amplifying the *lacI*-P_{trc}-*At*CHS-P_{trc}-*Pl*CHI^{syn} region from pTrc-*At*CHS-*Pl*CHI using primers CS644 lacI sense AatII and CS645 CHI anti BsiWI (Table 9.3). The resulting product was subsequently cloned into the plasmid pACKm-FLP-Trc-MEP (P. Ajikumar, unpublished) using the AatII and BsiWI restriction sites.

pCDF-*At*CHS was constructed by amplifying *At*CHS from pTrc-*At*CHS using primers CS627 CHS sense NdeI and CS628 CHS anti AvrII (Table 9.3) and cloning this PCR product into pCDFDuet-1 with the restriction sites/enzymes NdeI and AvrII. To assemble, pCDF-*At*CHS-*Pl*CHI^{syn}, *Pl*CHI^{syn} was amplified with primers CS629 CHI sense NcoI and CS630 CHI anti NotI using pTrc- *Pl*CHI^{syn} as a template and cloned into pCDF-*At*CHS with the sites NcoI and NotI.

pOM-*Ph*CHS-*Ms*CHI was constructed by digesting pOM-*Ph*CHS-*Ms*CHI-*At*4CL (R. Lim, unpublished) with BsrGI and BglII, followed by ligation with oligos CS792 BsrGI-BglII oligo 1 and CS793 BsrGI-BglII oligo 2 (Table 9.3) (at a ratio of 215 ng oligo per 100 ng digested plasmid).

9.2.4 Cultivation conditions

Two different fermentation protocols were developed for evaluating a strain's potential for flavonoid production. The first approach involved the cultivation of strains in 50 ml medium with 250-300 rpm orbital shaking at a temperature of 30°C. Induction of heterologous pathway expression was performed either at the beginning of the culture or during mid-exponential phase (as indicated for each experiment), and flavonoid production was assayed after 72 hr. In the second fermentation scheme, strains were first cultured in 25 ml medium at 37°C with 250-300 rpm orbital shaking. After a period of 15-24 hr (or after an OD_{600} of 1.0-2.0 had been reached), an additional 25 ml fresh medium was provided, pathway expression was induced, and cultures were subsequently transferred to a lower temperature (30°C) for optimal enzyme synthesis and flavonoid production. Flavonoid concentrations were measured after a total fermentation time of 48 hr.

All liquid cultivations were conducted in MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) cultures supplemented with 5 g/l glucose and an additional 4 g/l NH_4Cl. When appropriate, antibiotics were added in the following concentrations: 100 μg/ml carbenicillin for the maintenance of pTrc- or pET-derived plasmids, 34 μg/ml chloramphenicol for pHACM-derived plasmids, 68 μg/ml chloramphenicol for pACYC-derived plasmids and pRARE2 (Novagen), and 20 μg/ml kanamycin for pACKm-derived plasmids. IPTG (EMD

Chemicals) was provided at a concentration of 1 mM for the induction of expression from both trc and T7 promoters. Cultures for L-phenylalanine auxotrophic (Δ*pheA*) strains were additionally supplemented with L-phenylalanine (Sigma) at a concentration of 0.35 mM. For malonyl CoA availability experiments, cerulenin (Cayman Chemicals) and sodium malonate dibasic (Sigma) were added at a concentration 20 µg/ml and 2 g/l (1 g/l added twice), respectively.

9.2.5 Analytical Methods

For the quantification of L-tyrosine, cell-free culture supernatants were filtered through 0.2 µm PTFE membrane syringe filters (VWR International) and used for HPLC analysis with a Waters 2690 Separations module connected with a Waters 996 Photodiode Array detector (Waters) set to a wavelength of 278 nm. The samples were separated on a Waters Resolve C18 column with 0.1 % (vol/vol) trifluoroacetic acid (TFA) in water (solvent A) and 0.1 % (vol/vol) TFA in acetonitrile (solvent B) as the mobile phase. The following gradient was used at a flow rate of 1 ml/min: 0 min, 95 % solvent A + 5 % solvent B; 8 min, 20 % solvent A + 80 % solvent B; 10 min, 80 % solvent A + 20 % solvent B; 11 min, 95 % solvent A + 5 % solvent B.

To quantify levels of *p*-coumaric acid, cinnamic acid, and naringenin, 1 ml of culture supernatant was first extracted with an equal volume of ethyl acetate (EMD Chemicals). After vortexing and centrifugation, the top organic layer was separated and evaporated to dryness, and the remaining residue was resolubilized with 200 µl methanol (EMD Chemicals). Samples were analyzed using a Shimadzu Prominence HPLC system and a Waters Resolve C18 column using the same buffer system described above. Specifically, flavonoid compounds were

separated with the following acetonitrile/water gradient at a flow rate of 1 ml/min: 0 min, 90 % solvent A + 10 % solvent B; 10 min, 60 % solvent A + 40 % solvent B; 15 min, 60 % solvent A + 40 % solvent B; 17 min, 90 % solvent A + 10 % solvent B. Products were detected by monitoring their absorbance at 250 nm (*p*-coumaric acid) and 312 nm (cinnamic acid, naringenin), and concentrations were determined through the use of the corresponding chemical standards (Sigma).

Cell densities of cultures were determined by measuring their absorbance at 600 nm with an Ultrospec 2100 *pro* UV/Visible spectrophotometer (Amersham Biosciences).

9.3 Results and Discussion

Because L-tyrosine serves as the main precursor for the flavonoid naringenin, strains exhibiting an enhanced capacity for its synthesis provide a natural platform for exploring the potential of microbial flavonoid production from glucose. With a high flux through the aromatic amino acid pathway already in place, the next logical step towards this goal then becomes the assembly and grafting of an appropriate flavonoid biosynthetic gene cluster within these specific strain backgrounds. In this study, our main objective was to engineer a functional pathway consisting of TAL, 4CL, CHS, and CHI in order to mediate the conversion of L-tyrosine to naringenin.

9.3.1 Selection of enzyme sources

The selection of specific enzyme sources for the genes of a pathway remains a challenging task due to the minimal information available regarding the functionality of their products in heterologous hosts. Consequently, to increase our chances of success, we opted to construct a

four-gene assembly comprising solely variants that have previously been shown to be active for other similar applications. For the first step of the phenylpropanoid pathway, we chose a well-characterized TAL variant from *R. sphaeroides* that exhibited a 90 to 160-fold higher catalytic efficiency for L-tyrosine over L-phenylalanine (Watts, Mijts et al. 2006; Schroeder, Kumaran et al. 2008). Since most TAL enzymes exhibit some level of activity on both amino acids, it was important to select a form with a strong preference for L-tyrosine in order to ensure maximal substrate utilization within our strains. For the second catalytic step, 4CL-2 from *S. coelicolor* was selected due to its unique ability to efficiently convert both *p*-coumaric acid and cinnamic acid into their corresponding CoA esters (Kaneko, Hwang et al. 2003; Miyahisa, Kaneko et al. 2005). This dual substrate capacity ensures that the conversion of L-phenylalanine to cinnamic acid by *Rs*TAL does not lead to wasted resources but instead results in the productive synthesis of the flavanone piconembrin. We used *A. thaliana* as the source for the third enzyme, CHS, not only because it had been utilized in previous studies (Watts, Lee et al. 2004) but also because a cDNA library was readily available, thus circumventing the need for both plant cultivations and RNA preparations. Although we also intended to acquire CHI from this library, difficulties during gene amplification ultimately led us to pursue the direct synthesis of the desired locus using oligonucleotide gene assembly (Kodumal, Patel et al. 2004). The sequence of CHI from *P. lobata* was chosen for this purpose due to its demonstrated performance in prior investigations (Miyahisa, Kaneko et al. 2005) and was additionally codon-optimized to ensure adequate expression in *E. coli*.

9.3.2 Construction and evaluation of pCS204 performance

Rather than assembling the genes *Rs*TAL, *Sc*4CL, *At*CHS, and *Pl*CHI^{syn} into a single polycistronic operon, we decided to equip each locus with its own trc promoter to facilitate strong expression within *E. coli*. This four-gene biosynthetic cluster was constructed by sequential cloning into the pTrcHis2B vector to yield the plasmid pCS204 (Figure 9.2). In our initial tests of this plasmid, we monitored the production and accumulation of *p*-coumaric acid and naringenin to determine the functionality of this pathway within *E. coli*. Two distinct strain backgrounds and media formulations were examined during these experiments – a wild-type *E. coli* K12 with 500 mg/l of L-tyrosine supplementation and the previously engineered strain P2 (Chapter 4), which has the capacity for high endogenous L-tyrosine production (~400 mg/l L-tyrosine). Although the results obtained for the latter strain are the most relevant to our target application, we conducted parallel experiments with *E. coli* K12 in order to identify potential discrepancies between the consumption of endogenously-produced and externally supplemented aromatic precursors. Cultivations were conducted in 50 ml MOPS minimal medium at 30°C with IPTG promoter induction performed during culture inoculation.

Contrary to our expectations, strains that were grown for more than 48 hr yielded only small amounts of *p*-coumaric acid (less than 1 mg/l) and non-detectable levels of naringenin. Because other recent studies have relied on supplementation at the level of *p*-coumaric acid (Leonard, Lim et al. 2007; Leonard, Yan et al. 2008; Fowler, Gikandi et al. 2009), we suspected that TAL activity may present a major bottleneck in these strains. We therefore repeated the experiment with supplemented *p*-coumaric acid in the medium to bypass the TAL step and to test the performance of the last three genes of the pathway. Unfortunately, no measurable

levels of naringenin were recovered even for this case, suggesting the presence of severe functional deficiencies within both TAL and at least one other enzyme of this heterologous pathway.

9.3.3 An abundance of rare codons may limit heterologous protein expression

Our initial hypothesis was that the lack of pathway functionality may be due to poor expression of these proteins in *E. coli*. Of the possible reasons for weak expression, codon biases among different organisms are often implicated, particularly during the construction of heterologous pathways (Kane 1995). Indeed, a quick examination of the codon usage of the flavonoid biosynthetic cluster lent support to this theory. As seen in Table 9.4, *Rs*TAL, *Sc*4CL, and *At*CHS require the use of several rare codons in *E. coli*. In particular, the greatest offenders seem to be the amino acid/codon pairs of proline/CCC, glycine/GGA, and arginine/CGG, with some proteins requiring up to 21 instances of the same charged tRNA species for the synthesis of a *single* polypeptide.

This analysis strongly suggested that the presence of rare codons could present a genuine challenge in the translation of these flavonoid enzymes. To address this possibility, we transformed the engineered strains with pRARE2, a commercially-available plasmid which enables the IPTG-inducible overexpression of many of these rare codon tRNAs. If problems with flavonoid production are indeed related to the poor translational capacities of our strains, then supplying this plasmid should result in gains in both protein expression and enzyme activity. However, pRARE2 had no discernible effects on *p*-coumaric acid or naringenin levels, which remained low or undetectable regardless of strain background (*E. coli* K12 or P2) or

precursor supplementation (L-tyrosine and *p*-coumaric acid). Thus, it seems that this inability to produce flavonoids may not be a mere result of poor protein expression but may instead be related to inherent deficiencies in enzyme activity. To elucidate such bottlenecks, we designed an approach for the step-wise validation and optimization of each successive enzyme of the pathway.

Table 9.4 Rare codons found within *Rs*TAL, *Sc*4CL, *At*CHS, and *Pl*CHI^{syn} sequences

Rare codon*	*Rs*TAL	*Sc*4CL	*At*CHS	*Pl*CHI^{syn}
Pro (CCC)	**12 (43%)**	**21 (50%)**	4 (21%)	4 (40%)
Leu (CUA)	1 (1%)	0 (0%)	6 (15%)	0 (0%)
Arg (AGG)	3 (7%)	3 (9%)	4 (22%)	0 (0%)
Gly (GGA)	**8 (16%)**	2 (5%)	**9 (26%)**	1 (5%)
Arg (AGA)	1 (2%)	0 (0%)	3 (17%)	0 (0%)
Ile (AUA)	0 (0%)	0 (0%)	5 (24%)	1 (7%)
Arg (CGG)	**16 (35%)**	7 (21%)	1 (6%)	3 (43%)
TOTAL NUMBER	**41**	**33**	**32**	**9**

* The first value denotes the number of instances that a specified codon appears within a gene/protein. The second number represents the percentage of amino acids encoded by that rare codon. Numbers appearing in red highlight rare codons with a particularly high abundance within these gene/protein sequences.

9.3.4 Comparison of TAL sources

Because very little *p*-coumaric acid (less than 1 mg/l) was detected with pCS204, our first goal was to demonstrate that high levels of *p*-coumaric acid can in fact be recovered from our engineered strains. To this end, we analyzed the performance of TAL in both *E. coli* K12 and P2

in the absence of the other downstream flavonoid enzymes 4CL, CHS, and CHI. Since these strains do not possess any endogenous pathways for *p*-coumaric acid consumption, the levels of *p*-coumaric acid in the culture supernatant should accurately reflect the conversion potential of the TAL enzyme being evaluated.

When *Rs*TAL activity was tested under these experimental conditions, the same problems previously observed with the full biosynthetic gene cluster quickly emerged. As seen in Figure 9.3, *p*-coumaric acid levels were prohibitively low, with values ranging between just 1.5 and 5.5 mg/l. Thus, our initial suspicions were confirmed: the activity of this particular TAL variant was simply not adequate for our intended application. Although we did not provide the rare codon plasmid pRARE2 to assist with translation during these studies, we inferred from previous data that similar results would have likely been obtained.

The poor performance of *Rs*TAL required us to refocus our efforts on the identification of new TAL variants with the requisite expression and catalytic profiles. In particular, our search through the literature led us to explore TAL from the red yeast *R. glutinis* (*Rg*TAL), which exhibited the strongest preference for L-tyrosine over L-phenylalanine and the highest specific activity when evaluated against seven other bacterial and fungal TAL enzymes (Vannelli, Wei Qi et al. 2007). Moreover, a direct comparison of purified *Rs*TAL and *Rg*TAL revealed that the catalytic activity (k_{cat}/K_M) of the latter on L-tyrosine was more than twelve-fold better than what was observed with *Rs*TAL (Xue, McCluskey et al. 2007). Although separate investigations on both variants reported a range of values for these specific kinetic parameters (Watts, Mijts et al. 2006; Vannelli, Wei Qi et al. 2007; Xue, McCluskey et al. 2007; Xue, McCluskey et al. 2007), it is clear by at least a first approximation that *Rg*TAL may be a suitable alternative for

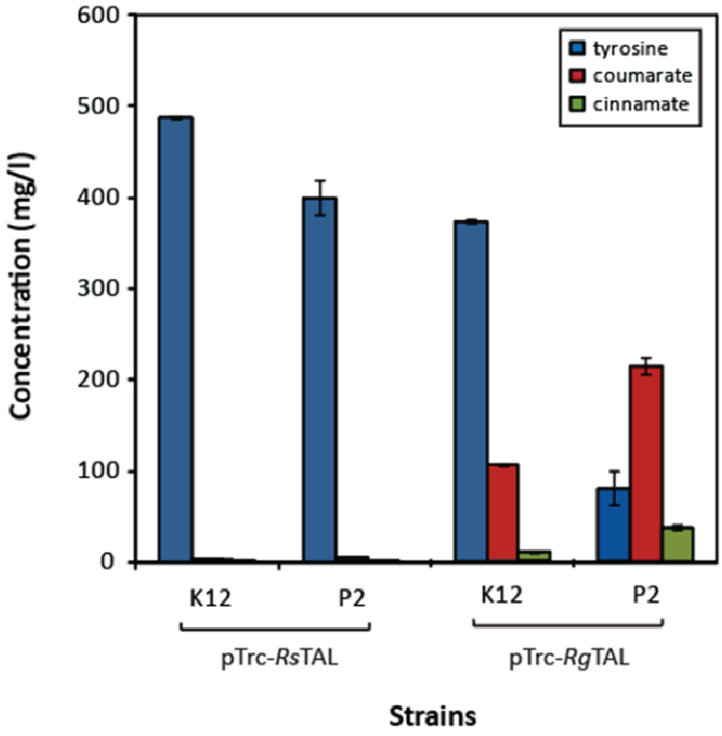


Figure 9.3 Comparison of TAL enzyme activity

Concentration of L-tyrosine (blue), *p*-coumaric acid (red), and cinnamic acid (green) in strains containing plasmid-expressed *R. sphaeroides* TAL (pTrc-*Rs*TAL) or *R. glutinis* TAL (pTrc-*Rg*TAL). K12 strain cultures were supplemented with 500 mg/l L-tyrosine. Values are reported after 72 hr cultivation in MOPS minimal medium.

our application. To bypass any potential issues with protein expression, the *Rg*TAL sequence was codon-optimized for *E. coli* and synthesized for direct cloning into pTrcHis2B.

When tested under the same experimental conditions as before, strains overexpressing *Rg*TAL^{syn} acquired a substantial capacity for *p*-coumaric acid synthesis. As seen in Figure 9.3 and Table 9.5, *E. coli* K12 with *Rg*TAL^{syn} produced more than 104 mg/l *p*-coumaric acid from 500 mg/l of supplemented L-tyrosine. Similarly, P2 with *Rg*TAL^{syn} was successful in generating 213 mg/l *p*-coumaric acid, a titer that is actually quite competitive with the amounts typically added

to the medium for flavonoid production (3 mM, 493 mg/l *p*-coumaric acid) (Leonard, Lim et al. 2007; Leonard, Yan et al. 2008). Since *Rg*TAL does exhibit activity on both L-tyrosine and L-phenylalanine, low levels of cinnamic acid (9-25 mg/l) were also recovered from the culture supernatant.

9.3.5 Addition of *Sc*4CL abolishes *Rg*TAL^{syn} activity

Having verified the functionality of the TAL-catalyzed step, we decided to reintroduce the 4CL enzyme into these *Rg*TAL^{syn}-expressing strains. Surprisingly, adding *Sc*4CL onto pTrc-*Rg*TAL^{syn} completely abolished *p*-coumaric acid accumulation, with measured titers falling to 7 mg/l in *E. coli* K12 and just 0.7 mg/l in P2 (Table 9.5). Although we initially suspected that this drop was related to *p*-coumaric acid *consumption* by 4CL to form coumaroyl-CoA, this notion was immediately refuted by the high concentrations of L-tyrosine still present in the media. Thus, the addition of *Sc*4CL imposed some unknown impediment on TAL activity, leading once again to a nonfunctional biosynthetic cluster.

Although previous studies have shown that *Rg*TAL can be inhibited by fairly low levels of *p*-coumaric acid (Sariaslani 2007), to our knowledge, there have been no reports of either 4CL or its biochemical product exerting any regulation on TAL. We therefore hypothesized that these effects were simply related to peculiarities in transcribing or translating these two genes from the same plasmid. To test this theory, we provided *Sc*4CL on a separate vector (pACYC-*Sc*4CL) to ascertain whether *p*-coumaric acid production could be recovered in these strains. Unfortunately, measured *p*-coumaric acid levels for *E. coli* K12 were comparable to those seen with a single plasmid, indicating that improper tandem gene expression was not the major

problem within this system (Table 9.5). Similarly, only small gains were seen in P2, with final *p*-coumaric acid titers reaching only 9% of the value seen with TAL expression alone.

Table 9.5 Effects of TAL/4CL expression on precursor and intermediate concentrations

Strain	Concentrations after 72 hr (mg/l)		
	L-Tyrosine	*p*-Coumaric acid	Cinnamic acid
***E. coli* K12**			
pTrc-*Rg*TAL^{syn}	374	104	9
pTrc-*Rg*TAL^{syn}-*Sc*4CL	485	7	0.3
pTrc-*Rg*TAL^{syn}, pACYC-*Sc*4CL	569	9	3
pTrc-*Rg*TAL^{syn}-*Pc*4CL^{syn}	461	42	1
P2			
pTrc-*Rg*TAL^{syn}	79	213	35
pTrc-*Rg*TAL^{syn}-*Sc*4CL	503	0.7	0.6
pTrc-*Rg*TAL^{syn}, pACYC-*Sc*4CL	484	19	12
pTrc-*Rg*TAL^{syn}-*Pc*4CL^{syn}	521	18	5

9.3.6 Testing other 4CL enzyme sources

Since the observed effects clearly did not arise from the tandem arrangement of *Rg*TAL^{syn} and *Sc*4CL, we decided to examine whether this phenomenon was common among all 4CL variants. Similar experiments were conducted using a new 4CL enzyme from *P. crispus* (parsley), which was selected based on its proven efficacy for other similar applications (Leonard, Lim et al. 2007; Leonard, Yan et al. 2008). As with *Rg*TAL, the genetic sequence of *Pc*4CL was codon optimized for expression in *E. coli* and synthesized prior to cloning. Once again, however, no increases in *p*-coumaric acid production were observed for either strain background tested (*E.*

coli K12 and P2) (Table 9.5), suggesting that this response may be a generic property common to most, if not all, 4CL enzymes.

9.3.7 Balancing relative gene expression to optimize flux

Given these inconclusive results, we chose to revisit the possibility that 4CL-mediated regulatory mechanisms may be negatively impacting TAL activity. Although we found no previous reports on TAL-4CL interactions in the literature, we postulated that the buildup of the 4CL biochemical product, coumaroyl-CoA, may exert a feedback inhibitory effect on the TAL enzyme. To test this hypothesis, we reintroduced the downstream enzymes CHS and CHI into the synthetic pathway to prevent the accumulation of this intermediate and, consequently, reverse any TAL inhibition within these strains. Rather than cloning all four genes onto the same plasmid as we did with pCS204, *At*CHS and *Pl*CHI^{syn} were provided on a separate vector to minimize potential issues associated with the use of large plasmids (i.e. poor transformability, recombination-mediated modifications). Such an arrangement also facilitated a parallel examination of the effects of varying promoter strengths on flavonoid production. Because all previous flavonoid studies seem to have favored the T7 promoter system over trc (Miyahisa, Kaneko et al. 2005; Leonard, Lim et al. 2007; Leonard, Yan et al. 2008), we suspected that stronger expression of the biosynthetic pathway may be needed to achieve high titers.

Interestingly, as seen in Table 9.6, expressing all four genes under a single strength promoter (either all trc or all T7) had no significant effect on *p*-coumaric acid levels, which ranged between 10-19 mg/l. Although L-tyrosine concentrations were observed to be somewhat lower for the T7 system, the absence of any significant naringenin accumulation

(0.09 mg/l) suggests that this discrepancy may be related to enhanced protein synthesis rather than *p*-coumaric acid production and consumption. Given the stringent cellular demands of expression from four T7 promoters, it would not be wholly unexpected to find lower basal levels of L-tyrosine from this overburdened cell.

Because the addition of *At*CHS and *Pl*CHI^{syn} was unable to restore *Rg*TAL^{syn} activity, we remained convinced that coumaroyl-CoA accumulation may still be a lingering issue within these strains. In particular, we hypothesized that protein translation of *At*CHS may be severely hindered due to the presence of several rare codons within its sequence (Table 9.4). With all other parameters being equal between *Pc*4CL^{syn} and *At*CHS (promoter strength, plasmid copy number), even slight deficiencies in *At*CHS protein synthesis could potentially tip the scale in favor of coumaryl-CoA accumulation. To correct for such a scenario, we overexpressed both *At*CHS and *Pl*CHI^{syn} relative to *Rg*TAL^{syn} and *Pc*4CL^{syn}. We postulated that the expression of *Rg*TAL^{syn}-*Pc*4CL^{syn} from a trc promoter and *At*CHS-*Pl*CHI^{syn} from the much stronger T7 promoter could negate these translational shortcomings.

Although this newly constructed strain behaved exactly like its preceding counterparts when induced at inoculation, we observed an unexpected shift in performance upon delayed IPTG induction. In fact, the addition of IPTG at an OD_{600} of 1.0 led to a complete recovery of *p*-coumaric acid titers, which reached levels that were comparable to those seen with *Rg*TAL^{syn} expression alone (198 mg/l) (Table 9.6). Although the exact regulatory mechanisms behind this 4CL-mediated phenomenon remains unclear, these results suggest that proper pathway balancing is needed to restore *Rg*TAL activity within these strains. For reference, we note that delayed induction could not elicit similar gains in other “unbalanced” strains (data not shown).

Table 9.6 Effects of relative gene expression on precursor and intermediate concentrations

Strain[a]	Concentrations after 72 hr (mg/l)			
	L-Tyrosine	*p*-Coumaric acid	Cinnamic acid	Naringenin
P2[b]				
pTrc-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pACKm-*At*CHS-*Pl*CHI^{syn}	496	10	21	0.09
pET-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pCDF-*At*CHS-*Pl*CHI^{syn}	311	19	5	0.3
pTrc -*Rg*TAL^{syn}-*Pc*4CL^{syn}, pCDF-*At*CHS-*Pl*CHI^{syn}	343	7	4	0.04
pTrc -*Rg*TAL^{syn}-*Pc*4CL^{syn}, pCDF-*At*CHS-*Pl*CHI^{syn}, induced at OD_{600} = 1.0[c]	84	198	48	0.61

[a] Plasmids with a pTrc or pACKm backbone contain trc promoters in front of all genes; plasmids with a pET or pCDF backbone contain T7 promoters in front of all genes. Unless indicated, all cultures were induced with 1 mM IPTG at inoculation.

[b] All T7 promoter plasmids were cultivated in a P2(DE3) background for T7 RNA polymerase expression.

[c] Growth was somewhat hampered for this strain with a maximum OD_{600} of just 2.4 compared to 4-5 for other strains.

9.3.8 Testing alternate sources for CHS and CHI

Despite these promising results, naringenin levels remained quite low in this improved strain (0.61 mg/l). We hypothesized that inherent deficiencies in either *At*CHS or *Pl*CHI^{syn} may have led to additional bottlenecks in the pathway; however, due to technical limitations with the detection of both coumaroyl-CoA and the naringenin chalcone, we had no quick way of determining which of the two enzymes required replacement. Rather than attempting an extensive characterization of each enzyme, we opted to simply swap out both genes in favor of two variants which have been successfully deployed in related applications (Leonard, Lim et al. 2007; Leonard, Yan et al. 2008; Fowler, Gikandi et al. 2009). Specifically, CHS from *P. hybrida*

and CHI from *M. sativa* were expressed under two different promoters to ascertain which cluster configurations could yield the best performers.

As seen in Table 9.7, expressing all four flavonoid genes (*Rg*TAL^{syn}, *Pc*4CL^{syn}, *Ph*CHS, *Ms*CHI) under T7 promoters led to a 10-fold increase in naringenin production (0.6 mg/l up to 6 mg/l), even in the absence of a balanced pathway. Thus, as we saw with *Rs*TAL and *Rg*TAL, these results highlight the importance of selecting an appropriate enzyme source/variant in the construction of heterologous pathways. Additional gains were obtained by driving *Ph*CHS and *Ms*CHI expression with a constitutive promoter (P_{GAP}) to ensure protein expression from the beginning of the culture, with naringenin titers reaching 9 mg/l. However, as expected from the previous analysis, the most significant increases were observed after combining a constitutively-expressed *Ph*CHS and *Ms*CHI gene cluster with trc-driven *Rg*TAL^{syn} and *Pc*4CL^{syn}. As evidence of a properly balanced pathway, *p*-coumaric acid levels increased to 136 mg/l from just 39 mg/l in the previous strain, indicating complete recovery of *Rg*TAL^{syn} activity. More notably, this augmented precursor pool had a direct impact on naringenin production, which increased by three-fold to yield a final titer of 29 mg/l.

To put these results into perspective, we note that the best performing base strain reported in the literature (E2 containing *Pc*4CL-2*, Ph*CHS*,* and *Ms*CHI) had the capacity to produce 42 mg/l naringenin from 3 mM (493 mg/l) of supplemented *p*-coumaric acid (Leonard, Lim et al. 2007). Although our final titers still fall a bit short of this value, this strain is capable of synthesizing naringenin *directly* from glucose, thus eliminating all reliance on expensive phenylpropanoic acid precursors. Furthermore, our constructed strain consistently grew to an OD_{600} of 4.5, signifying a fairly robust constitution that would be amenable to future

engineering efforts in contrast with previous constructions which exhibited impaired growth during naringenin production (Leonard, Lim et al. 2007).

Table 9.7 Evaluation of two novel gene sources – *P. hybrida* CHS and *M. sativa* CHI – for flavonoid production

Strain[a]	Concentrations (mg/l)			
	L-Tyrosine	*p*-Coumaric acid	Cinnamic acid	Naringenin
P2[b]				
pCDF-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pET-*Ph*CHS-*Ms*CHI[c]	543	28	15	6
pCDF-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pOM-*Ph*CHS-*Ms*CHI[d]	251	39	16	9
pCDF-trc-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pOM-*Ph*CHS-*Ms*CHI[d]	397	136	24	29

[a] Plasmids with a pET or pCDF backbone contain individual T7 promoters in front of all genes unless otherwise indicated. Plasmids with a pOM backbone contain a single constitutive promoter (P_{GAP}) to drive expression of both genes.

[b] All T7 promoter plasmids were cultivated in a P2(DE3) background for T7 RNA polymerase expression.

[c] Cultivations were performed in 50 ml MOPS minimal medium at 30°C with 1 mM IPTG induction at OD_{600} = 1.0. Measurements are shown after 72 hr.

[d] Strains were grown in 25 ml MOPS minimal medium at 37°C. After 15-20 hr, 25 ml fresh medium and 1 mM IPTG was added to the culture, and flasks were transferred to 30°C. Measurements are shown after 48 hr (total cultivation time).

[e] Although pCDF-trc-*Rg*TAL^{syn}-*Pc*4CL^{syn} was constructed with a pCDF backbone, it contains a trc promoter in front of both genes.

9.3.9 Engineering malonyl-CoA availability

As demonstrated in numerous reports, the supply of malonyl-CoA often appears as the next major bottleneck of the phenylpropanoid pathway due to both the requirement for *three* malonyl-CoA molecules and the typically low basal levels of this metabolite found within cells

(Takamura and Nomura 1988). Several studies have reported novel strategies for increasing the pool of this important precursor molecule (Miyahisa, Kaneko et al. 2005; Leonard, Lim et al. 2007; Leonard, Yan et al. 2008; Fowler, Gikandi et al. 2009). Rather than repeating all these efforts, we instead focused on the application of two specific techniques for increasing the supply of malonyl-CoA. The first strategy utilizes a recombinant malonate assimilation pathway from *Rhizobium trifolii* (MatB and MatC) for both the transport of supplemented malonate into the cell, as well as its subsequent conversion to malonyl-CoA. The second makes use of the fatty acid pathway inhibitor cerulenin, which represses both *fabB* and *fabF*, thus limiting the amount of malonyl-CoA lost to the synthesis of fatty acids (Figure 9.4) (Leonard, Yan et al. 2008).

As seen in Table 9.8, the addition of both malonate (2 g/l) and *R. trifolii* MatB and MatC (*Rt*MATBC) resulted in a 59% increase in naringenin over the previously constructed strain, with titers reaching 46 mg/l after 48 hr. However, the most significant gains were obtained with cerulenin supplementation, which led to an increase of over 190% and a final titer of 84 mg/l naringenin. Although the high cost of cerulenin prohibits its widespread use in industrial fermentation processes, these results suggest that significant gains in flavonoid production can be engineered by manipulating malonyl-CoA production and utilization within this strain. Thus, we remain confident that other established malonyl-CoA engineering strategies can be successfully implemented for the construction of a superior flavonoid producer.

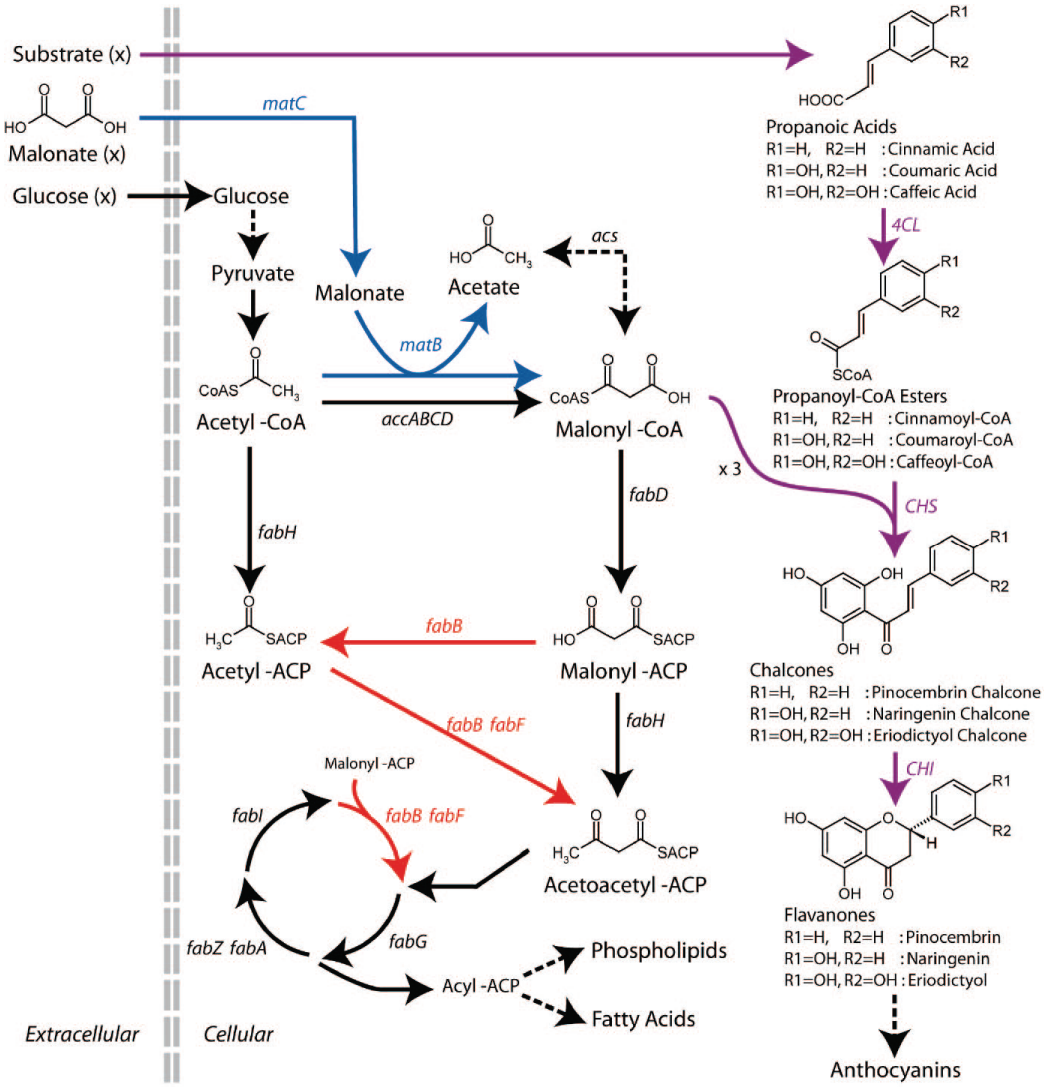


Figure 9.4 Engineering strategies for increasing intracellular concentrations of malonyl-CoA

Figure taken from (Leonard, Yan et al. 2008). The recombinant *R. trifolii matB* and *matC* pathway for malonate assimilation is indicated in blue, and the cerulenin targets *fabB* and *fabF* are indicated in red. Extracellular metabolites are denoted by an (x).

Table 9.8 Engineering malonyl-CoA availability in P2 and rpoA14^{R}

Strain[a]	Concentrations after 48 hr (mg/l)			
	L-Tyrosine	*p*-Coumaric acid	Cinnamic acid	Naringenin
P2				
pCDF-trc-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pOM-*Ph*CHS-*Ms*CHI[b]	397	136	24	29
pCDF-trc-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pOM-*Ph*CHS-*Ms*CHI[b], pACYCMatBC + malonate	42	107	51	46
pCDF-trc-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pOM-*Ph*CHS-*Ms*CHI[b] + cerulenin	439	79	25	84
rpoA14^{R}				
pCDF-trc-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pOM-*Ph*CHS-*Ms*CHI[b]	187	364	107	29
pCDF-trc-*Rg*TAL^{syn}-*Pc*4CL^{syn}, pOM-*Ph*CHS-*Ms*CHI[b] + cerulenin	175	315	101	77

[a] Strains were grown in 25 ml MOPS minimal medium at 37°C. After 15-20 hr, 25 ml fresh medium and 1 mM IPTG was added to the culture, and flasks were transferred to 30°C. Measurements are shown after 48 hr (total cultivation time).

9.3.10 Naringenin production in L-tyrosine overproducing strain rpoA14^{R}

Although P2 was used as the background strain for all our experiments, we recently reported the construction of several other L-tyrosine producers which possessed more than twice the yields and titers observed with P2 (Chapter 8). Given such significant improvements in performance, we were naturally curious to see if these phenotypically superior strains could surpass P2 in the production of flavonoid compounds as well. We explored this potential using the completely genetically-defined strain, rpoA14^{R}, which was reported to produce more than

900 mg/l L-tyrosine in 50 ml cultures. As seen in Table 9.8, final naringenin titers in rpoA14^{R} in both the presence and absence of cerulenin were actually quite comparable to those seen with P2, indicating that malonyl-CoA rather than L-tyrosine truly is the limiting precursor of the pathway. However, it was interesting to note that in contrast to P2, rpoA14^{R} exhibited an unusually enhanced capacity for *p*-coumaric acid synthesis, generating 315-364 mg/l *p*-coumaric acid after 48 hr. Because these *p*-coumaric acid concentrations approach those typically used in supplementation experiments, it is clear that future endeavors for engineering microbial flavonoid production could surely benefit from the use of this superior base strain.

9.4 Conclusions

9.4.1 Lessons learned in heterologous pathway construction

In this study, we have successfully demonstrated the feasibility of utilizing a set of engineered L-tyrosine producers for the synthesis of flavonoid compounds from glucose. Successes in this avenue did not come easily, however, as we experienced several nuances in the construction and assembly of heterologous pathways. During this investigation, we discovered the high sensitivity of this pathway to specific enzyme variants, with certain genetic sources (*Rg*TAL, *Pc*4CL, *Ph*CHS, *Ms*CHI) exhibiting much higher *in vivo* activities than others (*Rs*TAL, *Sc*4CL, *At*CHS, *Pl*CHI). Because these observed discrepancies likely result from an aggregate of factors including protein expression, proper folding, and the enzyme's innate catalytic properties, it becomes quite difficult to make *a priori* predictions on the relative performance of such

variants. We were therefore quite fortunate to have some of this information available to us from prior studies and comparisons conducted by other laboratories.

During the course of these experiments, we also encountered a previously uncharacterized regulatory phenomenon involving 4CL-mediated suppression of TAL enzyme activity. We hypothesized that these effects may be due to the accumulation of and subsequent feedback inhibition by coumaroyl-CoA, a theory that was corroborated by the recovery of TAL activity through adequate pathway balancing. This requirement for gene expression optimization is not an unusual feature of heterologous pathways, particularly for those that may result in the production of potentially toxic intermediates within the cell. As such, several semi-combinatorial tools or approaches have been constructed for the specific purpose of finding these relative expression optima (Alper, Fischer et al. 2005; Pfleger, Pitera et al. 2006; Ajikumar, Xiao et al. *Manuscript submitted*). Although experimenting with these parameters may result in improved production of the precursor *p*-coumaric acid, our most recent results suggest that engineering malonyl-CoA production, at least for the short-term, may have a more significant impact on final naringenin titers.

9.4.2 Demonstrated feasibility of microbial flavonoid production from glucose

Although other laboratories have developed microbial-based processes for flavonoid production, these previous methodologies suffered from two significant shortcomings - the requirement for expensive phenylpropanoic precursors and the need for two separate stages of cultivation for biomass/protein generation and flavonoid production. In this study, we successfully developed a set of strains and protocols capable of circumventing both of these

limitations. The use of previously-engineered L-tyrosine producers enabled us to address the first issue, with the assembly and optimization of a heterologous flavonoid pathway leading to the unique ability to produce naringenin directly from glucose. To our knowledge, this is the first substantiated example of flavonoid synthesis without the presence of expensive phenylpropanoic precursors in the media. Because recovered titers by these engineered strains were comparable to those previously achieved with the addition of *p*-coumaric acid, this corresponds to a several hundred fold decrease in substrate-related costs.

Prior protocols for flavonoid production also required a separate step for biomass buildup using rich media prior to cultivation in minimal media for flavonoid production. Although the rationale for this methodology has not been directly addressed, we presume that this practice is needed to offset the poor growth and protein expression seen in minimal media. In contrast, our engineered strains possessed robust cellular constitutions and exhibited no apparent growth deficiencies in minimal media. As such, these strains could be deployed in a one-medium protocol for the production of flavonoids from glucose, a considerable simplification over processes requiring rich media and cell separation and recycle.

9.4.3 Future attempts to engineer malonyl-CoA availability

The results from our cerulenin supplementation experiments indicate a clear need to engineer malonyl-CoA availability to further improve naringenin yields and titers in these strains. Several investigations have already highlighted potential avenues for introducing such cellular changes through a combination of both rational and model-guided approaches. In one particularly relevant study, researchers found that the simultaneous deletion of genes *sdhA*,

adhE, brnQ, and *citE* and overexpression of the enzymes acetyl-CoA synthase, acetyl-CoA carboxylase, biotin ligase, and pantothenate kinase could increase naringenin levels from 42 mg/l in the base/parental strain to an impressive 270 mg/l in the engineered construct (Fowler, Gikandi et al. 2009). It is very likely that similar gains can be made with our strains, particularly within the rpoA14^{R} background, which already possesses a high capacity for *p*-coumaric acid synthesis. Indeed, such improvements would certainly bring us one step closer to developing an economically viable and scalable process for the microbial production of flavonoid compounds from glucose.

CHAPTER 10 CONCLUSIONS AND RECOMMENDATIONS

As discussed throughout this work, a central goal of this thesis has been to evaluate the utility of combinatorial engineering techniques, such as transposon mutagenesis and gTME, for strain optimization applications. To explore the potential of such strategies, we examined their use within the specific context of L-tyrosine overproduction in *E. coli*. This model system was chosen primarily because this phenotype has been examined in several prior investigations and, as a result, provides a suitable platform for comparing the relative ease and efficacy of a wide range of metabolic engineering techniques.

10.1 Summary of work

In pursuit of this aim, we began by first laying down the foundational elements required to apply various combinatorial strategies for engineering a L-tyrosine overproduction phenotype. In Chapter 4, we detailed the construction of two rationally-engineered parental strains to ensure that the aromatic amino acid production pathway could operate at maximal capacity. Due to the strict transcriptional and allosteric regulation of this pathway in wild-type *E. coli*, it was necessary to delete key regulatory elements, as well as provide feedback-resistant enzymes in order to establish a base line level of production within these strains. In Chapter 5, we described the development of a high-throughput screen capable of sorting through large combinatorial libraries of strains in the search for L-tyrosine overproducers. Although combinatorial search strategies have the potential to unlock a diverse range of cellular

phenotypes, such efforts would be futile in the absence of proper techniques for the identification and isolation of desirable strains.

Once the initial groundwork for these studies was established, we then focused our efforts on applying these newly generated strains and tools towards the evaluation of two different combinatorial techniques - transposon mutagenesis (Chapter 6) and gTME (Chapter 7). Although several improved strains were yielded by these investigations, the best performing mutants were recovered from two gTME libraries. However, due to idiosyncrasies related to phenotypic transferability, it soon became clear that a comprehensive characterization of these mutants would be necessary to reconstruct the desired phenotype in novel strain backgrounds. In Chapter 8, we conducted a full evaluation of these strains using two 'omics approaches – microarray analysis and whole genome sequencing. Fortuitously, whole genome sequencing recovered only a single base pair substitution in each strain, each of which were validated to be related to L-tyrosine overproduction. Similarly, although transcriptomics studies are often plagued by difficulties with data interpretation, we were fortunate that specific patterns of over- and under- expression led us to the discovery of two specific pathways that were directly involved with L-tyrosine overproduction (acid resistance, stringent response). Indeed, the overexpression of two transcriptional regulators of these pathways (*evgA*, *relA*) was found to completely supplant the need for the mutant transcriptional components (*rpoA*, *rpoD*), thus establishing a validated mechanism for these cellular changes. Altogether, these investigations allowed us to implement an inverse metabolic engineering paradigm for the construction of a completely genetically-defined strain for L-tyrosine production (rpoA14^{R}). Of significance, when compared on a 50 ml scale, strain

rpoA14^{R} could outperform DPD4195, an industrial strain currently used for the production of L-tyrosine. Furthermore, this strain proved to be additionally well-suited for the synthesis of other L-tyrosine-derivatives, such as naringenin, the main precursor for all flavonoids compounds (Chapter 9).

10.2 Combinatorial approaches and the advantages/disadvantages of gTME

In light of these results, several key points regarding the use of combinatorial approaches for strain improvement endeavors can be made. First, the isolation of several strains exhibiting significant increases in L-tyrosine production serves to highlight the true potential of these unguided approaches. Recovered mutants from all three libraries (knockout, *rpoD*, *rpoA*) led to the discovery of novel genetic targets that were previously thought to be unrelated to L-tyrosine production. Indeed, some of these mechanisms are so elusive that even full strain characterizations have failed to uncover the exact underpinnings for these complex genotype-phenotype relationships. Although mechanistic insights are difficult to come by through these techniques, the ability to explore phenotypic space in an unbiased and unfettered fashion enables the engineering of significant leaps in strain performance and function.

Secondly, we observed that although combinatorial search strategies as a whole possess enormous potential for strain improvement endeavors, the utility of specific metabolic engineering techniques do vary based on the intended application. In this study, more significant gains were obtained through a transcriptional engineering approach over transposon mutagenesis, presumably because the introduction of *simultaneous* cellular modifications is often needed to engineer complex phenotypes. While we expect this same trend to hold for

other similarly multifaceted cellular properties, we also acknowledge that other applications may benefit more from greedy search algorithms that utilize single gene diversification approaches (random knockouts or genomic complementation). To provide a simplistic example, one could argue that multiple rounds of transposon mutagenesis would still be the most effective way to engineer phenotypes that are affected by as-yet-uncharacterized, competing cellular pathways. Of note, because several of these genetic techniques proceed through somewhat orthogonal approaches, it may also be advantageous to apply two or more different methodologies in a sequential fashion. Although we did not specifically address these possible interactions in this study, we note that such synergisms have already been demonstrated for other phenotypes, such as lycopene production in *E. coli* (Alper and Stephanopoulos 2007).

Ironically, the advantages of sophisticated combinatorial approaches such as gTME also become the biggest impediments when applying an inverse metabolic engineering framework. Because transcriptional engineering introduces an intractable number of changes within the cell, complicated 'omics analyses must be implemented to identify the genetic and/or biochemical basis for the desired phenotype. Although our own efforts culminated in the discovery of two pathways and three nucleotide substitutions that directly contributed to L-tyrosine overproduction, we note that such successes are usually the exception rather than the rule. Thus, one should seriously consider these drawbacks before undertaking such studies, particularly if there is significant interest in using information from recovered mutants to devise new strategies for engineering future strains.

Unusually, we found that all three transcriptional engineering mutants contained an additional chromosomal mutation needed to impart a L-tyrosine overproduction phenotype. Because this phenomenon was not observed in other gTME studies, we assume that this requirement is specific to this cellular property. However, this inability to recover unmutated backgrounds unfortunately also proves the fallibility of a transcriptional engineering approach. As one might have expected, gTME *by itself* cannot always alter metabolism in favor of a desired characteristic and as such, it should not be considered a panacea for all strain optimization studies.

10.3 Rational versus combinatorial approaches

In this work, we witnessed the recovery of combinatorial engineering mutants that produced L-tyrosine at rates and titers greatly exceeding those of their purely rationally-engineered counterparts. Given these considerable improvements in strain performance, would it then be reasonable to advocate for a completely combinatorial methodology for strain optimization? The answer, of course, is a resounding no. It is important to note that despite the sizeable gains afforded by a combinatorial approach, our initial libraries were generated from rationally-engineered parental strains with an already high baseline for L-tyrosine production. The aromatic amino acid biosynthetic pathway is a highly regulated network of genes with multiple checkpoints that prevent the accumulation of these energetically-expensive compounds. As such, deregulation of this pathway was required to ensure that it could operate at maximal capacity and furthermore, be able to benefit from the advantageous shifts in cellular metabolism introduced by combinatorial approaches.

Although rational engineering may be needed to prepare cells for combinatorial studies, the same strategies can and *should* also be revisited once desirable mutants have been isolated. In Chapter 7, we observed that although a second round of transcriptional engineering was able to increase strain productivity, very few gains in final titer were made. Because our yields are still well below the maximum theoretical value, we suspect that additional pathway bottlenecks may have arisen during this work and may now exist as the limiting factors of the system. As such, rational engineering strategies should clearly be implemented once more in order to identify and remove new blocks within the pathway. Since rational and combinatorial methodologies are uniquely suited for disparate applications, these types of collaborative executions provide the best routes for strain improvement endeavors.

10.4 Future recommendations

Despite our fairly comprehensive treatment of microbial L-tyrosine and flavonoid production, additional avenues for future exploration still remain. With regards to our engineered L-tyrosine producers, significant process scale-up and optimization must still be conducted in order to prove the utility of these strains for real-life industrial applications. However, because prior studies have already demonstrated the feasibility of improving strain performance through these routes (Patnaik, Zolandz et al. 2008), it is likely that similar gains can be achieved with our strains to yield a competitive or superior process for L-tyrosine production. For the case of flavonoid synthesis, we outlined the construction of a suitable base strain and remain optimistic that engineering malonyl-CoA availability within this background can result in significant increases in naringenin titers. Such improvements could truly set the stage for an

economically-viable fermentation-based process for the synthesis of these high-value compounds.

Due to the proven successes of a combinatorial approach, we believe that future strain optimization efforts can continue to benefit from the application of these existing methods, as well as the generation of novel, orthogonal methodologies for engineering phenotypic diversity. Unfortunately, the widespread implementation of these strategies is still currently restricted by the limited availability of high-throughput screens for other properties of interest. Although individual assays can be devised and tailored to specific problems (Tyo, Zhou et al. 2006; Pfleger, Pitera et al. 2007; Santos and Stephanopoulos 2008), future advances in this field hinge on the availability of generic or modular screening platforms for accessing a more extensive array of phenotypes. The development of a microfluidics-based detector and cell sorter in our lab has begun to address some of these critical limitations (Benjamin Wang, unpublished); however, future efforts should continue to explore this pervasive metabolic engineering problem. An additional bottleneck lies with the downstream processing of 'omics-generated data, which unfortunately limits the utility of employing an inverse metabolic engineering paradigm. Although we were still able to recover useful information from our microarray and genome sequencing studies, the true benefits of these 'omics technologies can only be realized with the development of improved data mining tools capable of parsing through the enormous amounts of data generated by these approaches. While these are clearly not trivial problems, advances in both these critical areas of metabolic engineering could significantly strengthen the value of these combinatorial techniques.

REFERENCES

Adrio, J. L. and A. L. Demain (2006). "Genetic improvement of processes yielding microbial products." FEMS Microbiol Rev **30**(2): 187-214.

Ajikumar, P. K., W. Xiao, et al. (*Manuscript submitted*). "Isoprenoid pathway optimization by a multivariate-modular approach for Taxol precursor overproduction in Escherichia coli."

Alexeyev, M. F. and I. N. Shokolenko (1995). "Mini-Tn10 transposon derivatives for insertion mutagenesis and gene delivery into the chromosome of gram-negative bacteria." Gene **160**(1): 59-62.

Alper, H., C. Fischer, et al. (2005). "Tuning genetic control through promoter engineering." Proc Natl Acad Sci U S A **102**(36): 12678-83.

Alper, H., Y. S. Jin, et al. (2005). "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli." Metab Eng **7**(3): 155-64.

Alper, H., K. Miyaoku, et al. (2005). "Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets." Nat Biotechnol **23**(5): 612-6.

Alper, H., J. Moxley, et al. (2006). "Engineering yeast transcription machinery for improved ethanol tolerance and production." Science **314**(5805): 1565-8.

Alper, H. and G. Stephanopoulos (2007). "Global transcription machinery engineering: A new approach for improving cellular phenotype." Metab Eng **9**(3): 258-67.

Askenazi, M., E. M. Driggers, et al. (2003). "Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains." Nat Biotechnol **21**(2): 150-6.

Baba, T., T. Ara, et al. (2006). "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection." Mol Syst Biol **2**: 2006 0008.

Backman, K. C. (1992). Method of biosynthesis of phenylalanine. USPTO. **US Patent 5,169,768**.

Badarinarayana, V., P. W. Estep, 3rd, et al. (2001). "Selection analyses of insertional mutants using subgenic-resolution arrays." Nat Biotechnol **19**(11): 1060-5.

Bailey, J. E. (1991). "Toward a science of metabolic engineering." Science **252**(5013): 1668-75.

Bailey, J. E. (1999). "Lessons from metabolic engineering for functional genomics and drug discovery." Nat Biotechnol **17**(7): 616-8.

Bailey, J. E., A. Sburlati, et al. (1996). "Inverse metabolic engineering: A strategy for directed genetic engineering of useful phenotypes." Biotechnol Bioeng **52**(1): 109-21.

Barnett, H. M. (1935). Method of preparing leucine. USPTO. **US Patent 2,009,868**.

Bell, A. A. and M. H. Wheeler (1986). "Biosynthesis and Functions of Fungal Melanins." Annual Review of Phytopathology **24**: 411-451.

Berry, A. (1996). "Improving production of aromatic compounds in Escherichia coli by metabolic engineering." Trends in Biotechnology **14**(7): 250-256.

Bongaerts, J., M. Kramer, et al. (2001). "Metabolic engineering for microbial production of aromatic amino acids and derived compounds." Metab Eng **3**(4): 289-300.

Bonuccelli, U. and P. Del Dotto (2006). "New pharmacologic horizons in the treatment of Parkinson disease." Neurology **67**(7 Suppl 2): S30-8.

Bravo, A. and J. Mora (1988). "Ammonium Assimilation in Rhizobium-Phaseoli by the Glutamine Synthetase-Glutamate Synthase Pathway." Journal of Bacteriology **170**(2): 980-984.

Breuer, M., K. Ditrich, et al. (2004). "Industrial methods for the production of optically active intermediates." Angew Chem Int Ed Engl **43**(7): 788-824.

Bro, C., S. Knudsen, et al. (2005). "Improvement of galactose uptake in Saccharomyces cerevisiae through overexpression of phosphoglucomutase: example of transcript analysis as a tool in inverse metabolic engineering." Appl Environ Microbiol **71**(11): 6465-72.

Bro, C. and J. Nielsen (2004). "Impact of 'ome' analyses on inverse metabolic engineering." Metab Eng **6**(3): 204-11.

Browning, D. F. and S. J. Busby (2004). "The regulation of bacterial transcription initiation." Nat Rev Microbiol **2**(1): 57-65.

Cabrera-Valladares, N., A. Martinez, et al. (2006). "Expression of the melA gene from Rhizobium etli CFN42 in Escherichia coli and characterization of the encoded tyrosinase." Enzyme and Microbial Technology **38**(6): 772-779.

Cardinal, E. V. (1953). Separation of tyrosine and cystine. USPTO. US Patent 2,650,242.

Chang, D. E., D. J. Smalley, et al. (2002). "Gene expression profiling of Escherichia coli growth transitions: an expanded stringent response model." Mol Microbiol **45**(2): 289-306.

Chavez-Bejar, M. I., A. R. Lara, et al. (2008). "Metabolic engineering of Escherichia coli for L-tyrosine production by expression of genes coding for the chorismate mutase domain of

the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from Zymomonas mobilis." Appl Environ Microbiol **74**(10): 3284-90.

Chen, S., S. Vincent, et al. (2003). "Mapping of chorismate mutase and prephenate dehydrogenase domains in the Escherichia coli T-protein." Eur J Biochem **270**(4): 757-63.

Claus, H. and H. Decker (2006). "Bacterial tyrosinases." Syst Appl Microbiol **29**(1): 3-14.

Dai, M. and S. D. Copley (2004). "Genome shuffling improves degradation of the anthropogenic pesticide pentachlorophenol by Sphingobium chlorophenolicum ATCC 39723." Appl Environ Microbiol **70**(4): 2391-7.

Dangi, B., A. M. Gronenborn, et al. (2004). "Versatility of the carboxy-terminal domain of the alpha subunit of RNA polymerase in transcriptional activation: use of the DNA contact site as a protein contact site for MarA." Mol Microbiol **54**(1): 45-59.

Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." Proc Natl Acad Sci U S A **97**(12): 6640-5.

De Backer, M. D., B. Nelissen, et al. (2001). "An antisense-based functional genomics approach for identification of genes critical for growth of Candida albicans." Nat Biotechnol **19**(3): 235-41.

Deijen, J. B. and J. F. Orlebeke (1994). "Effect of tyrosine on cognitive function and blood pressure under stress." Brain Res Bull **33**(3): 319-23.

della-Cioppa, G., S. J. Garger, et al. (1990). "Melanin production in Escherichia coli from a cloned tyrosinase gene." Biotechnology (N Y) **8**(7): 634-8.

Demain, A. L. (1981). "Industrial microbiology." Science **214**(4524): 987-95.

Demain, A. L. (2001). "Molecular genetics and industrial microbiology--30 years of marriage." J Ind Microbiol Biotechnol **27**(6): 352-6.

Dittrich, C. R., G. N. Bennett, et al. (2005). "Characterization of the acetate-producing pathways in Escherichia coli." Biotechnol Prog **21**(4): 1062-7.

Dueber, J. E., G. C. Wu, et al. (2009). "Synthetic protein scaffolds provide modular control over metabolic flux." Nat Biotechnol **27**(8): 753-9.

Durfee, T., A. M. Hansen, et al. (2008). "Transcription profiling of the stringent response in Escherichia coli." J Bacteriol **190**(3): 1084-96.

Enei, H., H. Matsui, et al. (1972). "Distribution of Tyrosine Phenol Lyase in Microorganisms." Agricultural and Biological Chemistry **36**(11): 1861-1868.

Enei, H., H. Nakazawa, et al. (1972). "Enzymatic preparation of L-tyrosine or 3,4-dihydroxyphenyl-L-alanine from pyruvate, ammonia and phenol or pyrocatechol." FEBS Lett **21**(1): 39-41.

Enei, H., H. Nakazawa, et al. (1973). "Microbiological Synthesis of L-Tyrosine and 3,4-Dihydroxyphenyl-L-Alanine .5. Synthesis of L-Tyrosine or 3,4-Dihydroxyphenyl-L-Alanine from Pyruvic Acid, Ammonia and Phenol or Pyrocatechol." Agricultural and Biological Chemistry **37**(4): 725-735.

Flores, N., J. Xiao, et al. (1996). "Pathway engineering for the production of aromatic compounds in Escherichia coli." Nature Biotechnology **14**(5): 620-623.

Forkmann, G. and S. Martens (2001). "Metabolic engineering and applications of flavonoids." Curr Opin Biotechnol **12**(2): 155-60.

Fowler, Z. L., W. W. Gikandi, et al. (2009). "Increased malonyl coenzyme A biosynthesis by tuning the Escherichia coli metabolic network and its application to flavanone production." Appl Environ Microbiol **75**(18): 5831-9.

Fowler, Z. L. and M. A. Koffas (2009). "Biosynthesis and biotechnological production of flavanones: current state and perspectives." Appl Microbiol Biotechnol **83**(5): 799-808.

Fukui, S., S. Ikeda, et al. (1975). "Comparative Studies on Properties of Tryptophanase and Tyrosine Phenol-Lyase Immobilized Directly on Sepharose or by Use of Sepharose-Bound Pyridoxal 5'-Phosphate." European Journal of Biochemistry **51**(1): 155-164.

Fukui, S., S. I. Ikeda, et al. (1975). "Production of L-Tryptophan, L-Tyrosine and Their Analogs by Use of Immobilized Tryptophanase and Immobilized Beta-Tyrosinase." European Journal of Applied Microbiology **1**(1): 25-39.

Gaal, T., W. Ross, et al. (1996). "DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture." Genes Dev **10**(1): 16-26.

Gajiwala, K. S. and S. K. Burley (2000). "HDEA, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria." J Mol Biol **295**(3): 605-12.

Garibyan, L., T. Huang, et al. (2003). "Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome." DNA Repair (Amst) **2**(5): 593-608.

Garner, C. C. and K. M. Herrmann (1985). "Operator mutations of the Escherichia coli aroF gene." J Biol Chem **260**(6): 3820-5.

Gerth, T. D., R. W. Mann, et al. (1999). Dietary supplement composition. USPTO. US Patent 5,925,377. **5,925,377**.

Gill, R. T. (2003). "Enabling inverse metabolic engineering through genomics." Curr Opin Biotechnol **14**(5): 484-90.

Gill, R. T., S. Wildt, et al. (2002). "Genome-wide screening for trait conferring genes using DNA microarrays." Proc Natl Acad Sci U S A **99**(10): 7033-8.

Giovannini, M., E. Verduci, et al. (2007). "Phenylketonuria: dietary and therapeutic challenges." J Inherit Metab Dis **30**(2): 145-52.

Gong, G. L., X. Sun, et al. (2007). "Mutation and a high-throughput screening method for improving the production of Epothilones of Sorangium." J Ind Microbiol Biotechnol **34**(9): 615-23.

Gonzalez, V., P. Bustos, et al. (2003). "The mosaic structure of the symbiotic plasmid of Rhizobium etli CFN42 and its relation to other symbiotic genome compartments." Genome Biol **4**(6): R36.

Gosset, G., J. Yong-Xiao, et al. (1996). "A direct comparison of approaches for increasing carbon flow to aromatic biosynthesis in Escherichia coli." J Ind Microbiol **17**(1): 47-52.

Gruber, T. M. and C. A. Gross (2003). "Multiple sigma subunits and the partitioning of bacterial transcription space." Annu Rev Microbiol **57**: 441-66.

Harborne, J. B. and C. A. Williams (2000). "Advances in flavonoid research since 1992." Phytochemistry **55**(6): 481-504.

Herring, C. D., A. Raghunathan, et al. (2006). "Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale." Nat Genet **38**(12): 1406-12.

Hida, H., T. Yamada, et al. (2007). "Genome shuffling of Streptomyces sp. U121 for improved production of hydroxycitric acid." Appl Microbiol Biotechnol **73**(6): 1387-93.

Hollman, P. C. and M. B. Katan (1998). "Bioavailability and health effects of dietary flavonols in man." Arch Toxicol Suppl **20**: 237-48.

http://ecocyc.org/ Ecocyc: Encyclopedia of Escherichia coli K-12 Genes and Metabolism.

Hwang, E. I., M. Kaneko, et al. (2003). "Production of plant-specific flavanones by Escherichia coli containing an artificial gene cluster." Appl Environ Microbiol **69**(5): 2699-706.

Ikeda, M. (2006). "Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering." Appl Microbiol Biotechnol **69**(6): 615-26.

Ikeda, M., J. Ohnishi, et al. (2006). "A genome-based approach to create a minimally mutated Corynebacterium glutamicum strain for efficient L-lysine production." J Ind Microbiol Biotechnol **33**(7): 610-5.

Imaizumi, A., H. Kojima, et al. (2006). "The effect of intracellular ppGpp levels on glutamate and lysine overproduction in Escherichia coli." J Biotechnol **125**(3): 328-37.

Ishihama, A. (2000). "Functional modulation of Escherichia coli RNA polymerase." Annu Rev Microbiol **54**: 499-518.

Jayaraj, S., R. Reid, et al. (2005). "GeMS: an advanced software package for designing synthetic genes." Nucleic Acids Res **33**(9): 3011-6.

Jensen, E. B. and S. Carlsen (1990). "Production of recombinant human growth hormone in Escherichia coli: expression of different precursors and physiological effects of glucose, acetate, and salts." Biotechnol Bioeng **36**(1): 1-11.

Jin, Y., R. M. Watt, et al. (2009). "Small noncoding RNA GcvB is a novel regulator of acid resistance in Escherichia coli." BMC Genomics **10**: 165.

Jin, Y. S., H. Alper, et al. (2005). "Improvement of xylose uptake and ethanol production in recombinant Saccharomyces cerevisiae through an inverse metabolic engineering approach." Appl Environ Microbiol **71**(12): 8249-56.

Jin, Y. S. and G. Stephanopoulos (2007). "Multi-dimensional gene target search for improving lycopene biosynthesis in Escherichia coli." Metab Eng **9**(4): 337-47.

Jishage, M., K. Kvint, et al. (2002). "Regulation of sigma factor competition by the alarmone ppGpp." Genes Dev **16**(10): 1260-70.

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

Kaneko, M., E. I. Hwang, et al. (2003). "Heterologous production of flavanones in Escherichia coli: potential for combinatorial biosynthesis of flavonoids in bacteria." J Ind Microbiol Biotechnol **30**(8): 456-61.

Karp, P. D., I. M. Keseler, et al. (2007). "Multidimensional annotation of the Escherichia coli K-12 genome." Nucleic Acids Res **35**(22): 7577-90.

Kelman, Z. and M. O'Donnell (1995). "DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine." Annu Rev Biochem **64**: 171-200.

Kern, R., A. Malki, et al. (2007). "Escherichia coli HdeB is an acid stress chaperone." J Bacteriol **189**(2): 603-10.

Kim, D. Y., E. Rha, et al. (2007). "Development of bioreactor system for L-tyrosine synthesis using thermostable tyrosine phenol-lyase." Journal of Microbiology and Biotechnology **17**(1): 116-122.

Kim, T. Y., H. U. Kim, et al. (2010). "Data integration and analysis of biological networks." Curr Opin Biotechnol **21**(1): 78-84.

Klein-Marcuschamer, D., C. N. Santos, et al. (2009). "Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes." Appl Environ Microbiol **75**(9): 2705-11.

Klein-Marcuschamer, D. and G. Stephanopoulos (2008). "Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains." Proc Natl Acad Sci U S A **105**(7): 2319-24.

Knekt, P., R. Jarvinen, et al. (1996). "Flavonoid intake and coronary mortality in Finland: a cohort study." BMJ **312**(7029): 478-81.

Kodumal, S. J., K. G. Patel, et al. (2004). "Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster." Proc Natl Acad Sci U S A **101**(44): 15573-8.

Kramer, M., J. Bongaerts, et al. (2003). "Metabolic engineering for microbial production of shikimic acid." Metab Eng **5**(4): 277-83.

Kyndt, J. A., T. E. Meyer, et al. (2002). "Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein." FEBS Lett **512**(1-3): 240-4.

Lagunas-Munoz, V. H., N. Cabrera-Valladares, et al. (2006). "Optimum melanin production using recombinant Escherichia coli." J Appl Microbiol **101**(5): 1002-8.

Lease, R. A., D. Smith, et al. (2004). "The small noncoding DsrA RNA is an acid resistance regulator in Escherichia coli." J Bacteriol **186**(18): 6179-85.

Leonard, E., K. H. Lim, et al. (2007). "Engineering central metabolic pathways for high-level flavonoid production in Escherichia coli." Appl Environ Microbiol **73**(12): 3877-86.

Leonard, E., Y. Yan, et al. (2008). "Strain improvement of recombinant Escherichia coli for efficient production of plant flavonoids." Mol Pharm **5**(2): 257-65.

Li, S. C., N. K. Goto, et al. (1996). "Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment." Proc Natl Acad Sci U S A **93**(13): 6676-81.

Liao, J. C., S. Y. Hou, et al. (1996). "Pathway analysis, engineering, and physiological considerations for redirecting central metabolism." Biotechnol Bioeng **52**(1): 129-40.

Lloyd-George, I. and T. M. S. Chang (1993). "Free and microencapsulated Erwinia herbicola for the production of tyrosine." Biomater Artif Cells Immobilization Biotechnol **21**(3): 323-33.

Lloyd-George, I. and T. M. S. Chang (1995). "Characterization of Free and Alginate-Polylysine-Alginate Microencapsulated Erwinia-Herbicola for the Conversion of Ammonia, Pyruvate, and Phenol into L-Tyrosine." Biotechnology and Bioengineering **48**(6): 706-714.

Lu, J. L. and J. C. Liao (1997). "Metabolic engineering and control analysis for production of aromatics: Role of transaldolase." Biotechnol Bioeng **53**(2): 132-8.

Lutke-Eversloh, T., C. N. S. Santos, et al. (2007). "Perspectives of biotechnological production of L-tyrosine and its applications." Appl Microbiol Biotechnol **77**(4): 751-62.

Lütke-Eversloh, T., C. N. S. Santos, et al. (2007). "Perspectives of biotechnological production of L-tyrosine and its applications." Appl Microbiol Biotechnol **77**(4): 751-62.

Lutke-Eversloh, T. and G. Stephanopoulos (2007). "L-Tyrosine production by deregulated strains of Escherichia coli." Appl Microbiol Biotechnol **75**(1): 103-10.

Lütke-Eversloh, T. and G. Stephanopoulos (2005). "Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of Escherichia coli: generation and characterization of tyrosine-insensitive mutants." Appl Environ Microbiol **71**(11): 7224-8.

Lütke-Eversloh, T. and G. Stephanopoulos (2007). "L-Tyrosine production by deregulated strains of Escherichia coli." Appl Microbiol Biotechnol **75**(1): 103-10.

Lütke-Eversloh, T. and G. Stephanopoulos (2007). "A semi-quantitative high-throughput screening method for microbial L: -tyrosine production in microtiter plates." J Ind Microbiol Biotechnol **34**(12): 807-811.

Lütke-Eversloh, T. and G. Stephanopoulos (2008). "Combinatorial pathway analysis for improved L-tyrosine production in Escherichia coli: identification of enzymatic bottlenecks by systematic gene overexpression." Metab Eng **10**(2): 69-77.

Lynch, M. D., R. T. Gill, et al. (2004). "Mapping phenotypic landscapes using DNA micro-arrays." Metab Eng **6**(3): 177-85.

Lynch, M. D., T. Warnecke, et al. (2007). "SCALEs: multiscale analysis of library enrichment." Nat Methods **4**(1): 87-93.

Ma, Z., S. Gong, et al. (2003). "GadE (YhiE) activates glutamate decarboxylase-dependent acid resistance in Escherichia coli K-12." Mol Microbiol **49**(5): 1309-20.

Ma, Z., N. Masuda, et al. (2004). "Characterization of EvgAS-YdeO-GadE branched regulatory circuit governing glutamate-dependent acid resistance in Escherichia coli." J Bacteriol **186**(21): 7378-89.

Magnusson, L. U., A. Farewell, et al. (2005). "ppGpp: a global regulator in Escherichia coli." Trends Microbiol **13**(5): 236-42.

Maki, H. and A. Kornberg (1987). "Proofreading by DNA polymerase III of Escherichia coli depends on cooperative interaction of the polymerase and exonuclease subunits." Proc Natl Acad Sci U S A **84**(13): 4389-92.

Malki, A., H. T. Le, et al. (2008). "Solubilization of protein aggregates by the acid stress chaperones HdeA and HdeB." J Biol Chem **283**(20): 13679-87.

March, J. C., M. A. Eiteman, et al. (2002). "Expression of an anaplerotic enzyme, pyruvate carboxylase, improves recombinant protein production in Escherichia coli." Appl Environ Microbiol **68**(11): 5620-4.

Mark, A. M. (1939). Isolation of leucine and tyrosine from corn gluten. USPTO. **US Patent 2,178,210**.

Mascarenhas, D., D. J. Ashworth, et al. (1991). "Deletion of pgi alters tryptophan biosynthesis in a genetically engineered strain of Escherichia coli." Appl Environ Microbiol **57**(10): 2995-9.

Masuda, N. and G. M. Church (2003). "Regulatory network of acid resistance genes in Escherichia coli." Mol Microbiol **48**(3): 699-712.

Mei, B. and H. Zalkin (1989). "A cysteine-histidine-aspartate catalytic triad is involved in glutamine amide transfer function in purF-type glutamine amidotransferases." J Biol Chem **264**(28): 16613-9.

Mei, B. G. and H. Zalkin (1990). "Amino-terminal deletions define a glutamine amide transfer domain in glutamine phosphoribosylpyrophosphate amidotransferase and other PurF-type amidotransferases." J Bacteriol **172**(6): 3512-4.

Miller, J. H. (1992). A short course in bacterial genetics : a laboratory manual and handbook for Escherichia coli and related bacteria. Plainview, N.Y., Cold Spring Harbor Laboratory Press.

Miyahisa, I., M. Kaneko, et al. (2005). "Efficient production of (2S)-flavanones by Escherichia coli containing an artificial biosynthetic gene cluster." Appl Microbiol Biotechnol **68**(4): 498-504.

Murakami, K., N. Fujita, et al. (1996). "Transcription factor recognition surface on the RNA polymerase alpha subunit is involved in contact with the DNA enhancer element." Embo J **15**(16): 4358-67.

Nagasawa, T., T. Utagawa, et al. (1981). "Syntheses of L-tyrosine-related amino acids by tyrosine phenol-lyase of Citrobacter intermedius." Eur J Biochem **117**(1): 33-40.

Nakano, K., M. Rischke, et al. (1997). "Influence of acetic acid on the growth of Escherichia coli K12 during high-cell-density cultivation in a dialysis reactor." Appl Microbiol Biotechnol **48**(5): 597-601.

Neidhardt, F. C., P. L. Bloch, et al. (1974). "Culture medium for enterobacteria." J Bacteriol **119**(3): 736-47.

Neri, D. F., D. Wiegmann, et al. (1995). "The effects of tyrosine on cognitive performance during extended wakefulness." Aviat Space Environ Med **66**(4): 313-9.

Nishino, K., Y. Inazumi, et al. (2003). "Global analysis of genes regulated by EvgA of the two-component regulatory system in Escherichia coli." J Bacteriol **185**(8): 2667-72.

Nishino, K. and A. Yamaguchi (2001). "Overexpression of the response regulator evgA of the two-component signal transduction system modulates multidrug resistance conferred by multidrug resistance transporters." J Bacteriol **183**(4): 1455-8.

Nosanchuk, J. D. and A. Casadevall (2003). "The contribution of melanin to microbial pathogenesis." Cell Microbiol **5**(4): 203-23.

O'Brien, C., C. Mahoney, et al. (2007). "Dietary tyrosine benefits cognitive and psychomotor performance during body cooling." Physiol Behav **90**(2-3): 301-7.

O'Donnell, M. (2006). "Replisome architecture and dynamics in Escherichia coli." J Biol Chem **281**(16): 10653-6.

O'Donoghue, P., R. E. Amaro, et al. (2001). "On the structure of hisH: protein structure prediction in the context of structural and functional genomics." J Struct Biol **134**(2-3): 257-68.

Ochi, K. (2007). "From microbial differentiation to ribosome engineering." Biosci Biotechnol Biochem **71**(6): 1373-86.

Ochi, K., S. Okamoto, et al. (2004). "Ribosome engineering and secondary metabolite production." Adv Appl Microbiol **56**: 155-84.

Olson, M. M., L. J. Templeton, et al. (2007). "Production of tyrosine from sucrose or glucose achieved by rapid genetic changes to phenylalanine-producing Escherichia coli strains." Appl Microbiol Biotechnol **74**(5): 1031-40.

Para, G., P. Lucciardi, et al. (1985). "Synthesis of L-Tyrosine by Immobilized Escherichia-Intermedia Cells." Applied Microbiology and Biotechnology **21**(5): 273-279.

Parekh, S., V. A. Vinci, et al. (2000). "Improvement of microbial strains and fermentation processes." Appl Microbiol Biotechnol **54**(3): 287-301.

Park, K. S., Y. S. Jang, et al. (2005). "Phenotypic alteration and target gene identification using combinatorial libraries of zinc finger proteins in prokaryotic cells." J Bacteriol **187**(15): 5496-9.

Park, K. S., D. K. Lee, et al. (2003). "Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors." Nat Biotechnol **21**(10): 1208-14.

Patnaik, R. and J. C. Liao (1994). "Engineering of Escherichia coli central metabolism for aromatic metabolite production with near theoretical yield." Appl Environ Microbiol **60**(11): 3903-8.

Patnaik, R., S. Louie, et al. (2002). "Genome shuffling of Lactobacillus for improved acid tolerance." Nat Biotechnol **20**(7): 707-12.

Patnaik, R., R. R. Zolandz, et al. (2008). "L-tyrosine production by recombinant Escherichia coli: fermentation optimization and recovery." Biotechnol Bioeng **99**(4): 741-52.

Paul, B. J., M. B. Berkmen, et al. (2005). "DksA potentiates direct activation of amino acid promoters by ppGpp." Proc Natl Acad Sci U S A **102**(22): 7823-8.

Pedelacq, J. D., S. Cabantous, et al. (2006). "Engineering and characterization of a superfolder green fluorescent protein." Nat Biotechnol **24**(1): 79-88.

Perrino, F. W., S. Harvey, et al. (1999). "Two functional domains of the epsilon subunit of DNA polymerase III." Biochemistry **38**(48): 16001-9.

Pfleger, B. F., D. J. Pitera, et al. (2007). "Microbial sensors for small molecules: development of a mevalonate biosensor." Metab Eng **9**(1): 30-8.

Pfleger, B. F., D. J. Pitera, et al. (2006). "Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes." Nat Biotechnol **24**(8): 1027-32.

Pittard, J., H. Camakaris, et al. (2005). "The TyrR regulon." Mol Microbiol **55**(1): 16-26.

Pohnert, G., S. Zhang, et al. (1999). "Regulation of phenylalanine biosynthesis. Studies on the mechanism of phenylalanine binding and feedback inhibition in the Escherichia coli P-protein." Biochemistry **38**(38): 12212-7.

Polen, T., M. Kramer, et al. (2005). "The global gene expression response of Escherichia coli to L-phenylalanine." J Biotechnol **115**(3): 221-37.

Potrykus, K. and M. Cashel (2008). "(p)ppGpp: still magical?" Annu Rev Microbiol **62**: 35-51.

Qi, W. W., T. Vannelli, et al. (2007). "Functional expression of prokaryotic and eukaryotic genes in Escherichia coli for conversion of glucose to p-hydroxystyrene." Metab Eng **9**(3): 268-76.

Rajput, A. and A. H. Rajput (2006). "Parkinson's disease management strategies." Expert Rev Neurother **6**(1): 91-9.

Riesenberg, D., V. Schulz, et al. (1991). "High cell density cultivation of Escherichia coli at controlled specific growth rate." J Biotechnol **20**(1): 17-27.

Rohr, F. J., D. Lobbregt, et al. (1998). "Tyrosine supplementation in the treatment of maternal phenylketonuria." Am J Clin Nutr **67**(3): 473-6.

Rolfes, R. J. and H. Zalkin (1988). "Regulation of Escherichia coli purF. Mutations that define the promoter, operator, and purine repressor gene." J Biol Chem **263**(36): 19649-52.

Ross, W., K. K. Gosink, et al. (1993). "A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase." Science **262**(5138): 1407-13.

Ruan, L., W. He, et al. (2005). "Cloning and expression of mel gene from Bacillus thuringiensis in Escherichia coli." Antonie Van Leeuwenhoek **87**(4): 283-8.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Santos, C. N. and G. Stephanopoulos (2008). "Melanin-based high-throughput screen for L-tyrosine production in Escherichia coli." Appl Environ Microbiol **74**(4): 1190-7.

Santos, C. N. S. and G. Stephanopoulos (2008). "Melanin-based high-throughput screen for L-tyrosine production in Escherichia coli." Appl Environ Microbiol **74**(4): 1190-7.

Sariaslani, F. S. (2007). "Development of a combined biological and chemical process for production of industrial aromatics from renewable resources." Annu Rev Microbiol **61**: 51-69.

Sarubbi, E., K. E. Rudd, et al. (1988). "Basal ppGpp level adjustment shown by new spoT mutants affect steady state growth rates and rrnA ribosomal promoter regulation in Escherichia coli." Mol Gen Genet **213**(2-3): 214-22.

Sauer, U. (2001). "Evolutionary engineering of industrially important microbial phenotypes." Adv Biochem Eng Biotechnol **73**: 129-69.

Schroeder, A. C., S. Kumaran, et al. (2008). "Contributions of conserved serine and tyrosine residues to catalysis, ligand binding, and cofactor processing in the active site of tyrosine ammonia lyase." Phytochemistry **69**(7): 1496-506.

Schumacher, M. A., K. Y. Choi, et al. (1994). "Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices." Science **266**(5186): 763-70.

Segre, D., D. Vitkup, et al. (2002). "Analysis of optimality in natural and perturbed metabolic networks." Proc Natl Acad Sci U S A **99**(23): 15112-7.

Selifonova, O., F. Valle, et al. (2001). "Rapid evolution of novel traits in microorganisms." Appl Environ Microbiol **67**(8): 3645-9.

Sharma, U. K. and D. Chatterji (2008). "Differential mechanisms of binding of anti-sigma factors Escherichia coli Rsd and bacteriophage T4 AsiA to E. coli RNA polymerase lead to diverse physiological consequences." J Bacteriol **190**(10): 3434-43.

Sharma, U. K., S. Ravishankar, et al. (1999). "Study of the interaction between bacteriophage T4 asiA and Escherichia coli sigma(70), using the yeast two-hybrid system: neutralization of asiA toxicity to E. coli cells by coexpression of a truncated sigma(70) fragment." J Bacteriol **181**(18): 5855-9.

Smith, D. R., A. R. Quinlan, et al. (2008). "Rapid whole-genome mutational profiling using next-generation sequencing technologies." Genome Res **18**(10): 1638-42.

Sprenger, G. A. (2007). "From scratch to value: engineering Escherichia coli wild type cells to the production of L: -phenylalanine and other fine chemicals derived from chorismate." Appl Microbiol Biotechnol **75**(4): 739-49.

Srivatsan, A. and J. D. Wang (2008). "Control of bacterial transcription, translation and replication by (p)ppGpp." Curr Opin Microbiol **11**(2): 100-5.

Stefan, A., L. Reggiani, et al. (2003). "Silencing of the gene coding for the epsilon subunit of DNA polymerase III slows down the growth rate of Escherichia coli populations." FEBS Lett **546**(2-3): 295-9.

Steinmetzer, W. (1983). Process for recovering amino acids from protein hydrolysates. USPTO. US Patent 4,384,136.

Stephanopoulos, G. (1999). "Metabolic fluxes and metabolic engineering." Metab Eng **1**(1): 1-11.

Stephanopoulos, G., H. Alper, et al. (2004). "Exploiting biological complexity for strain improvement through systems biology." Nat Biotechnol **22**(10): 1261-7.

Stephens, J. C., S. W. Artz, et al. (1975). "Guanosine 5'-diphosphate 3'-diphosphate (ppGpp): positive effector for histidine operon transcription and general signal for amino-acid deficiency." Proc Natl Acad Sci U S A **72**(11): 4389-93.

Storz, G., J. A. Opdyke, et al. (2004). "Controlling mRNA stability and translation with small, noncoding RNAs." Curr Opin Microbiol **7**(2): 140-4.

Studwell, P. S. and M. O'Donnell (1990). "Processive replication is contingent on the exonuclease subunit of DNA polymerase III holoenzyme." J Biol Chem **265**(2): 1171-8.

Taft-Benz, S. A. and R. M. Schaaper (1998). "Mutational analysis of the 3'-->5' proofreading exonuclease of Escherichia coli DNA polymerase III." Nucleic Acids Res **26**(17): 4005-11.

Taft-Benz, S. A. and R. M. Schaaper (1999). "The C-terminal domain of dnaQ contains the polymerase binding site." J Bacteriol **181**(9): 2963-5.

Takai, A., R. Nishi, et al. (2009). L-Tyrosine producing bacterium
and a method for producing L-tyrosine. USPTO. **US Patent 7,482,140 B2**.

Takamura, Y. and G. Nomura (1988). "Changes in the intracellular concentration of acetyl-CoA and malonyl-CoA in relation to the carbon and energy metabolism of Escherichia coli K12." J Gen Microbiol **134**(8): 2249-53.

Tatarko, M. and T. Romeo (2001). "Disruption of a global regulatory gene to enhance central carbon flux into phenylalanine biosynthesis in Escherichia coli." Curr Microbiol **43**(1): 26-32.

Tomar, A., M. A. Eiteman, et al. (2003). "The effect of acetate pathway mutations on the production of pyruvate in Escherichia coli." Appl Microbiol Biotechnol **62**(1): 76-82.

Traxler, M. F., S. M. Summers, et al. (2008). "The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli." Mol Microbiol **68**(5): 1128-48.

Tribe, D. E. (1987). Novel microorganism and method. USPTO. **US Patent 4,681,852**.

Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A **98**(9): 5116-21.

Tyo, K. E., H. S. Alper, et al. (2007). "Expanding the metabolic engineering toolbox: more options to engineer cells." Trends Biotechnol **25**(3): 132-7.

Tyo, K. E., H. Zhou, et al. (2006). "High-throughput screen for poly-3-hydroxybutyrate in Escherichia coli and Synechocystis sp. strain PCC6803." Appl Environ Microbiol **72**(5): 3412-7.

Tysyachnaya, I. V., V. I. Yakovleva, et al. (1979). "Kinetic study of tyrosine synthesis in tyrosine phenol lyase reaction catalyzed by Citrobacter freundii cells." Biochemistry-Moscow **44**(12): 1739-1744.

van Spronsen, F. J., M. van Rijn, et al. (2001). "Phenylketonuria: tyrosine supplementation in phenylalanine-restricted diets." Am J Clin Nutr **73**(2): 153-7.

Vannelli, T., W. Wei Qi, et al. (2007). "Production of p-hydroxycinnamic acid from glucose in Saccharomyces cerevisiae and Escherichia coli by expression of heterologous genes from plants and fungi." Metab Eng **9**(2): 142-51.

Vassel, B. (1956). Separation of tyrosine. USPTO. US Patent 2,738,366.

Waalkes, T. P. and S. Udenfriend (1957). "A fluorometric method for the estimation of tyrosine in plasma and tissues." J Lab Clin Med **50**(5): 733-6.

Wahlbom, C. F., R. R. Cordero Otero, et al. (2003). "Molecular analysis of a Saccharomyces cerevisiae mutant with improved ability to utilize xylose shows enhanced expression of proteins involved in transport, initial xylose metabolism, and the pentose phosphate pathway." Appl Environ Microbiol **69**(2): 740-6.

Wang, G., A. Aazaz, et al. (2000). "Cloning and overexpression of a tyrosinase gene mel from Pseudomonas maltophila." FEMS Microbiol Lett **185**(1): 23-7.

Wang, H. H., F. J. Isaacs, et al. (2009). "Programming cells by multiplex genome engineering and accelerated evolution." Nature **460**(7257): 894-8.

Watts, K. T., P. C. Lee, et al. (2004). "Exploring recombinant flavonoid biosynthesis in metabolically engineered Escherichia coli." Chembiochem **5**(4): 500-7.

Watts, K. T., B. N. Mijts, et al. (2006). "Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family." Chem Biol **13**(12): 1317-26.

Williams, M. (2005). "Dietary Supplements and Sports Performance: Amino Acids." Journal of the International Society of Sports Nutrition **2**(2): 63-67.

Wisselink, H. W., M. J. Toirkens, et al. (2007). "Engineering of Saccharomyces cerevisiae for efficient anaerobic alcoholic fermentation of L-arabinose." Appl Environ Microbiol **73**(15): 4881-91.

Xue, Z., M. McCluskey, et al. (2007). "Improved production of p-hydroxycinnamic acid from tyrosine using a novel thermostable phenylalanine/tyrosine ammonia lyase enzyme." Enzyme and Microbial Technology **42**(1): 58-64.

Xue, Z., M. McCluskey, et al. (2007). "Identification, characterization and functional expression of a tyrosine ammonia-lyase and its mutants from the photosynthetic bacterium Rhodobacter sphaeroides." J Ind Microbiol Biotechnol **34**(9): 599-604.

Yakandawala, N., T. Romeo, et al. (2008). "Metabolic engineering of Escherichia coli to enhance phenylalanine production." Appl Microbiol Biotechnol **78**(2): 283-91.

Yamada, H., H. Kumagai, et al. (1972). "Synthesis of L-tyrosine from pyruvate, ammonia and phenol by crystalline tyrosine phenol lyase." Biochem Biophys Res Commun **46**(2): 370-4.

Yomano, L. P., S. W. York, et al. (1998). "Isolation and characterization of ethanol-tolerant mutants of Escherichia coli KO11 for fuel ethanol production." J Ind Microbiol Biotechnol **20**(2): 132-8.

Young, I. G., F. Gibson, et al. (1969). "Enzymic and nonenzymic transformations of chorismic acid and related cyclohexadienes." Biochim Biophys Acta **192**(1): 62-72.

Zamir, L. O., E. Jung, et al. (1983). "Co-accumulation of prephenate, L-arogenate, and spiro-arogenate in a mutant of Neurospora." J Biol Chem **258**(10): 6492-6.

Zhang, Y. X., K. Perry, et al. (2002). "Genome shuffling leads to rapid phenotypic improvement in bacteria." Nature **415**(6872): 644-6.

Zhou, Y., T. Minami, et al. (2010). "Systematic screening of Escherichia coli single-gene knockout mutants for improving recombinant whole-cell biocatalysts." Appl Microbiol Biotechnol: DOI 10.1007/s00253-010-2505-7.

APPENDIX A MIAME CHECKLIST

Brief Description of the Experiment: We measured transcriptional changes in four strains – P2, rpoD3, rpoA14, and rpoA27 - in an effort to understand mechanisms by which L-tyrosine production is positively influenced by mutant *rpoA*- and *rpoD*-encoded transcriptional components.

Experimental factors: We explored the effects of genetic variation (in the form of plasmid-encoded mutated transcriptional machinery) on the cellular transcriptome.

Experimental design: Strains P2, rpoA14, rpoA27, and rpoD3 were grown in 50 ml MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) to an OD_{600} of approximately 0.4. Triplicates samples of RNA (on three separate days) were then extracted using the QIAGEN RNeasy Mini Kit according to the manufacturer's protocol.

Quality control steps taken: Arrays were run in triplicate with biological replicates to allow for statistical confidence in differential gene expression.

Links to the publication, any Supplementary websites or database accession numbers: GEO accession number GSE21652.

SAMPLES USED, EXTRACT PREPARATION AND HANDLING

The origin of each biological sample: *E. coli* K12-derived strains P2, rpoD3, rpoA14, and rpoA27

Manipulation of biological samples and protocols used/Technical protocols for preparing the hybridization extract (e.g., the RNA or DNA extraction and purification protocol), and labeling: Strains P2, rpoA14, rpoA27, and rpoD3 were grown in 50 ml MOPS minimal medium (Teknova) (Neidhardt, Bloch et al. 1974) at 37C with 225 rpm orbital shaking. The cultures were harvested at an OD_{600} of approximately 0.4, and triplicates samples of RNA (on three separate days) were then extracted using the QIAGEN RNeasy Mini Kit according to the manufacturer's protocol. The extracted RNA was processed by the David Koch Institute for Integrative Cancer Research Microarray Technologies Core Facility.

HYBRIDIZATION PROCEDURES AND PARAMETERS

The protocol and conditions used for hybridization, blocking and washing, including any post-processing steps such as staining: Hybridization was conducted by the David Koch Institute for Integrative Cancer Research Microarray Technologies Core Facility using standard Affymetrix protocols.

MEASUREMENT DATA AND SPECIFICATIONS

Data: GEO accession number GSE21652.

Data extraction and processing protocols/Image scanning hardware and software, and processing procedures and parameters: Arrays were scanned and images were processed by Affymetrix hardware and software by the David Koch Institute for Integrative Cancer Research Microarray Technologies Core Facility.

Normalization, transformation and data selection procedures and parameters: Normalization and batch analysis was performed with the Affymetrix Gene Chip Operating Software, v1.4 using default settings. Expression profile deviations between mutants and parental were identified by Significance Analysis of Microarrays (SAM) using SAM 3.0 (Tusher, Tibshirani et al. 2001).

Array design: Affymetrix GeneChip E. Coli Genome 2.0 Array, Part 900551

APPENDIX B SEQUENCES OF SYNTHESIZED GENES AND PROTEINS

Table B.1 DNA and protein sequences of synthesized genes and proteins

Gene	DNA/Protein Sequence
P. lobata chalcone isomerase ($PlCHI^{syn}$) (DNA)	ATGGCTGCGGCTGCTGCCGTGGCGACCATTAGCGCGGTGCAAGTGGAGTTTCT GGAATTTCCAGCGGTAGTGACCAGCCCGGCATCAGGCCGTACCTATTTTCTTGG TGGCGCTGGGGAGCGTGGCCTGACGATTGAGGGCAAGTTTATCAAGTTCACCG GCATTGGCGTGTATTTGGAAGATAAGGCGGTTAGCTCCCTGGCGGCGAAATGG AAAGGCAAACCGAGCGAAGAACTGGTGGAGACCCTGGACTTCTACCGGGATA TCATAAGCGGTCCCTTCGAGAAACTGATCCGTGGCAGCAAAATTCTGCCACTGT CGGGCGTCGAATACAGCAAGAAAGTGATGGAAAACTGCGTGGCGCATATGAA AAGCGTCGGAACCTATGGCGATGCGGAAGCCGCTGCCATCGAGAAGTTCGCG GAGGCCTTCAAAAACGTGAATTTTCAACCTGGCGCGACCGTGTTTTATCGGCAA AGCCCAGATGGCGTTCTGGGCCTGAGTTTCAGCGAGGATGTGACCATTCCCGA TAATGAAGCGGCGGTGATTGAAAACAAAGCCGTCTCCGCTGCGGTGTTAGAAA CCATGATTGGCGAACATGCAGTAAGCCCCGATCTGAAACGTAGCTTGGCGAGC CGGTTACCCGCCGTGTTGAGCCACGGTATTATTGTGTAA
P. lobata chalcone isomerase ($PlCHI^{syn}$) (Protein)	MAAAAAVATISAVQVEFLEFPAVVTSPASGRTYFLGGAGERGLTIEGKFIKFTGIGV YLEDKAVSSLAAKWKGKPSEELVETLDFYRDIISGPFEKLIRGSKILPLSGVEYSKKVM ENCVAHMKSVGTYGDAEAAAIEKFAEAFKNVNFQPGATVFYRQSPDGVLGLSFSE DVTIPDNEAAVIENKAVSAAVLETMIGEHAVSPDLKRSLASRLPAVLSHGIIV

R. glutinis tyrosine ammonia lyase ($Rgtal^{syn}$) (DNA)	ATGGCGCCTCGCCCGACTTCGCAAAGCCAGGCCCGCACTTGCCCGACGACGCA GGTTACCCAAGTTGATATCGTTGAGAAAATGTTGGCGGCTCCTACTGATAGCAC GCTGGAGCTGGACGGTTATAGCCTGAATCTGGGTGATGTCGTGAGCGCTGCGC GTAAGGGTCGTCCTGTCCGTGTCAAAGATAGCGATGAAATCCGCAGCAAAATC GACAAGAGCGTTGAATTCCTGCGCAGCCAACTGAGCATGTCGGTTTACGGTGT GACGACCGGCTTTGGCGGCTCCGCGGACACGCGCACGGAGGACGCAATTAGC CTGCAAAAGGCGTTGCTGGAACACCAGCTGTGTGGTGTGTTGCCGAGCAGCTT CGACAGCTTTCGCTTGGGTCGTGGTCTGGAGAATAGCCTGCCGTTGGAAGTCG TTCGCGGTGCAATGACCATTCGTGTGAATTCGCTGACCCGTGGCCATAGCGCTG TTCGTCTGGTTGTTCTGGAAGCACTGACGAACTTTCTGAACCACGGTATTACCC CGATTGTTCCGCTGCGCGGTACGATCTCCGCGAGCGGCGATCTGTCTCCACTGT CGTACATTGCAGCGGCGATTAGCGGTCACCCGGATAGCAAAGTTCACGTGGTC CATGAAGGCAAAGAGAAGATCCTGTACGCGCGCGAAGCGATGGCGCTGTTTA ACCTGGAGCCGGTGGTTTTGGGTCCGAAGGAGGGCCTGGGTCTGGTGAATGG TACGGCAGTCTCCGCGAGCATGGCAACGCTGGCACTGCACGACGCGCATATGT TGAGCCTGTTGAGCCAATCGCTGACCGCGATGACCGTGGAGGCGATGGTCGGT CACGCGGGCAGCTTCCATCCATTCCTGCACGATGTTACGCGTCCGCACCCGACG CAAATCGAGGTCGCGGGTAACATTCGCAAACTGCTGGAGGGCTCGCGCTTCGC GGTCCACCACGAGGAAGAGGTTAAGGTCAAGGATGATGAAGGCATTTTGCGT CAGGATCGTTATCCGTTGCGCACGAGCCCGCAATGGTTGGGTCCGCTGGTGTC CGACCTGATTCACGCTCATGCCGTCTTGACGATCGAAGCGGGTCAAAGCACCA CCGATAACCCACTGATCGATGTTGAGAATAAGACCAGCCATCACGGTGGCAAC TTTCAAGCGGCAGCGGTTGCCAACACGATGGAAAAGACCCGTCTGGGCTTGGC CCAAATCGGTAAACTGAATTTCACCCAGCTGACGGAGATGCTGAACGCGGGCA TGAATCGTGGCTTGCCGAGCTGCCTGGCGGCTGAAGACCCATCCCTGAGCTAT CATTGCAAAGGTCTGGACATTGCGGCGGCTGCATATACGAGCGAACTGGGCCA CCTGGCTAACCCGGTCACCACCCACGTCCAACCGGCTGAAATGGCAAACCAGG CGGTGAATAGCTTGGCGTTGATTAGCGCACGTCGTACCACGGAATCTAACGAC GTTCTGTCCCTGCTGCTGGCAACGCACCTGTACTGCGTGCTGCAGGCGATCGAC CTGCGTGCGATTGAGTTCGAGTTCAAGAAACAGTTTGGTCCTGCCATTGTTAGC CTGATCGACCAACACTTTGGTAGCGCGATGACGGGTAGCAATCTGCGTGATGA GCTGGTTGAAAAGGTCAATAAGACTCTGGCCAAGCGTTTGGAGCAAACCAATA GCTACGATCTGGTTCCGCGCTGGCACGACGCTTTTAGCTTCGCTGCAGGCACTG TTGTCGAGGTTCTGTCCAGCACGAGCCTGAGCTTGGCGGCCGTGAACGCATGG AAGGTTGCGGCAGCCGAGAGCGCGATCTCCTTGACGCGCCAGGTCCGTGAAAC GTTTTGGTCCGCTGCAAGCACCTCCAGCCCGGCGTTGTCTTACTTGAGCCCGCG CACGCAGATCCTGTACGCATTTGTGCGTGAGGAACTGGGTGTCAAAGCCCGCC GTGGTGACGTCTTCTTGGGTAAACAAGAAGTTACCATCGGCAGCAACGTTAGC AAGATTTACGAAGCCATCAAGAGCGGCCGTATCAACAATGTTCTGCTGAAGAT GCTGGCATAA

R. glutinis tyrosine ammonia lyase (Rgtal^{syn}) (Protein)	MAPRPTSQSQARTCPTTQVTQVDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARK GRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTTGFGGSADTRTEDAISLQKALLE HQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTN FLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVHVVHEGKEKILYAREAMAL FNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQSLTAMTVEAMVG HAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDDEGILRQDR YPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQAAAV ANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLY CVLQAIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLE QTNSYDLVPRWHDAFSFAAGTVVEVLSSTSLSLAAVNAWKVAAAESAISLTRQVRE TFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYE AIKSGRINNVLLKMLA
P. crispus 4-coumarate:coA ligase (Pc4CL^{syn}) (DNA)	ATGGGTGACTGCGTTGCCCCGAAAGAGGATCTGATCTTCCGCAGCAAACTGCC GGACATTTACATTCCAAAGCATCTGCCGCTGCATACGTATTGTTTTGAGAATATC AGCAAGGTTGGCGACAAGAGCTGTCTGATCAACGGCGCAACCGGCGAAACGT TTACCTACAGCCAGGTCGAGCTGCTGTCCCGTAAAGTTGCCAGCGGCCTGAAC AAGCTGGGCATTCAACAAGGTGATACCATTATGCTGTTGCTGCCGAATTCCCCG GAGTACTTTTTCGCTTTCCTGGGTGCGAGCTATCGCGGTGCAATCAGCACCATG GCGAATCCATTCTTTACCAGCGCAGAAGTGATCAAGCAACTGAAAGCGAGCCA AGCGAAGCTGATTATCACCCAGGCATGCTATGTTGACAAGGTCAAGGACTACG CAGCGGAGAAAAACATCCAGATCATTTGTATTGACGATGCACCGCAGGATTGC CTGCACTTTAGCAAGCTGATGGAAGCGGATGAGAGCGAAATGCCGGAAGTGG TCATTAACAGCGATGATGTGGTGGCATTGCCGTACAGCTCTGGCACCACCGGC CTGCCGAAAGGCGTTATGCTGACCCACAAGGGTCTGGTTACGAGCGTTGCACA ACAGGTGGATGGTGATAACCCGAACCTGTATATGCACTCCGAGGATGTCATGA TCTGCATCCTGCCACTGTTCCATATCTATAGCCTGAACGCTGTTCTGTGTTGTGG TCTGCGTGCGGGCGTCACCATTCTGATCATGCAAAAGTTCGACATTGTGCCGTT TCTGGAGCTGATTCAGAAGTATAAGGTTACGATTGGTCCGTTTGTCCCGCCGAT CGTGCTGGCCATCGCGAAAAGCCCGGTCGTTGACAAGTACGACTTGTCTAGCG TGCGCACCGTCATGAGCGGTGCAGCGCCGCTGGGTAAAGAGTTGGAGGACGC TGTCCGTGCGAAATTCCCGAACGCGAAGCTGGGTCAAGGCTATGGCATGACCG AAGCCGGTCCGGTCCTGGCGATGTGTCTGGCGTTCGCCAAAGAGCCGTATGAG ATTAAGTCTGGCGCATGCGGTACCGTTGTGCGTAATGCCGAGATGAAAATCGT TGACCCAGAAACGAATGCGTCTCTGCCGCGTAATCAGCGTGGTGAGATTTGCA TCCGTGGTGATCAGATTATGAAAGGTTACCTGAATGACCCGGAAAGCACCCGC ACCACGATCGACGAAGAGGGTTGGTTGCACACGGGTGACATTGGTTTCATCGA CGATGACGATGAACTGTTCATTGTCGATCGTTTGAAAGAAATCATTAAGTACAA AGGTTTTCAAGTTGCTCCGGCGGAGTTGGAAGCACTGCTGCTGACGCACCCGA CGATCAGCGATGCCGCGGTGGTTCCGATGATTGACGAGAAAGCGGGTGAAGT GCCAGTGGCGTTTGTCGTGCGTACCAATGGTTTTACCACGACCGAAGAAGAAA TCAAACAATTTGTGAGCAAACAGGTCGTGTTCTACAAACGTATCTTCCGCGTCT TCTTCGTTGACGCTATTCCGAAATCCCCGAGCGGCAAGATTTTGCGTAAGGATC TGCGCGCTCGTATTGCGAGCGGCGACCTGCCGAAGTAA

P. crispus 4-coumarate:coA ligase ($Pc4CL^{syn}$) (Protein)	MGDCVAPKEDLIFRSKLPDIYIPKHLPLHTYCFENISKVGDKSCLINGATGETFTYSQV ELLSRKVASGLNKLGIQQGDTIMLLLPNSPEYFFAFLGASYRGAISTMANPFFTSAEV IKQLKASQAKLIITQACYVDKVKDYAAEKNIQIICIDDAPQDCLHFSKLMEADESEMP EVVINSDDVVALPYSSGTTGLPKGVMLTHKGLVTSVAQQVDGDNPNLYMHSEDV MICILPLFHIYSLNAVLCCGLRAGVTILIMQKFDIVPFLELIQKYKVTIGPFVPPIVLAIA KSPVVDKYDLSSVRTVMSGAAPLGKELEDAVRAKFPNAKLGQGYGMTEAGPVLA MCLAFAKEPYEIKSGACGTVVRNAEMKIVDPETNASLPRNQRGEICIRGDQIMKGY LNDPESTRTTIDEEGWLHTGDIGFIDDDDELFIVDRLKEIIKYKGFQVAPAELEALLLT HPTISDAAVVPMIDEKAGEVPVAFVVRTNGFTTTEEEIKQFVSKQVVFYKRIFRVFF VDAIPKSPSGKILRKDLRARIASGDLPK